

# Sistemas de DW

## (Data Warehouse)

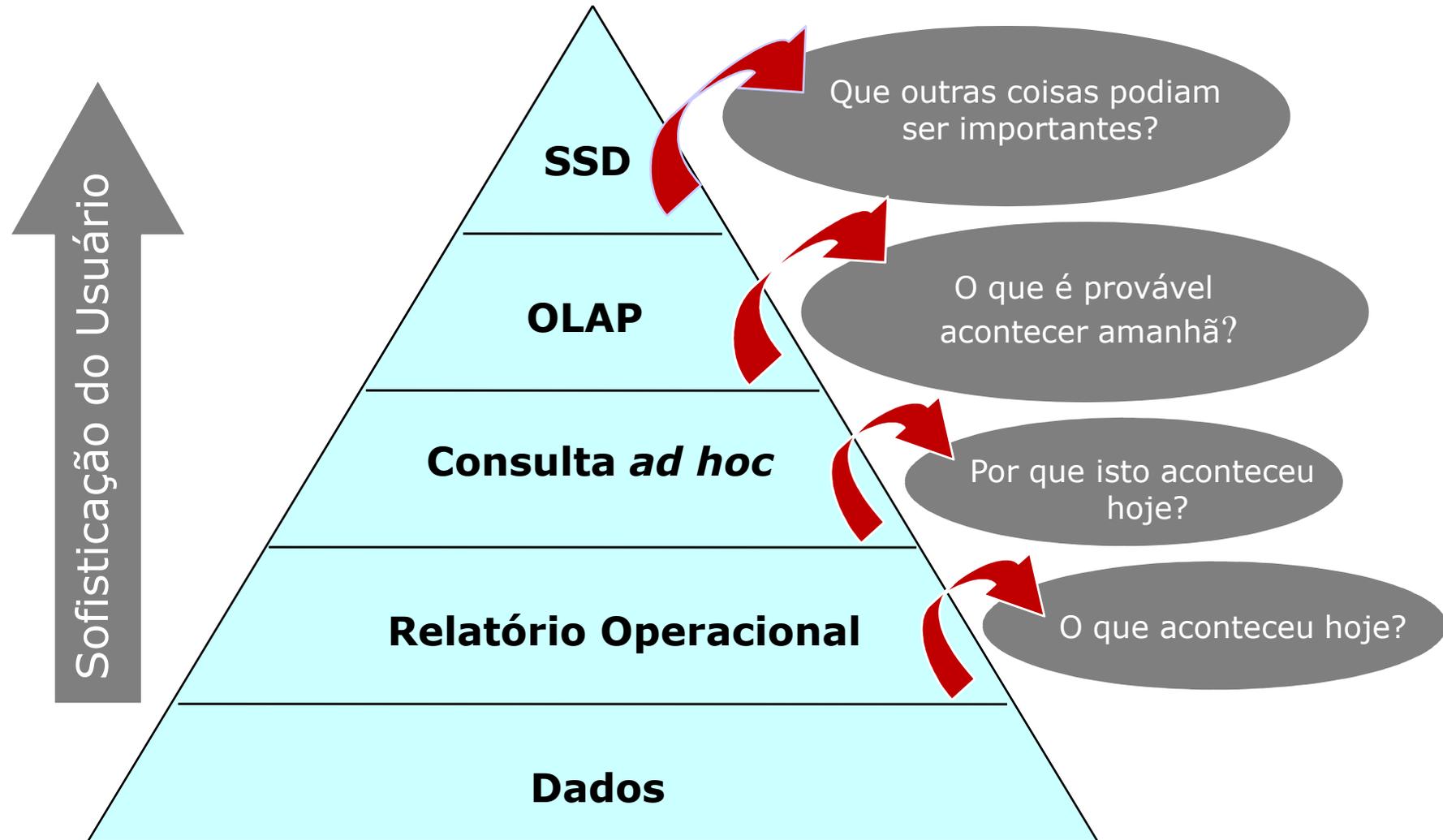
Valéria Cesário Times  
vct@cin.ufpe.br

- Introdução
  - Hierarquia da Informação
  - Sistemas de Suporte à Decisão
  - Evolução de Sistemas de DW
  - Comparativo entre SSD e SSO
- Conceitos Básicos
  - Banco de Dados Multidimensional
  - Data Warehouse (DW)
  - Data Mart (DM)
  - ETL e ODS

- Conceitos Básicos (Cont.)
  - Data Warehousing(DWing)
  - OLAP
- Modelos Dimensionais
  - Estrela
  - Flocos de neve
- Modelagem Dimensional
- Arquitetura de DW
- Processamento OLAP x OLTP

- Cubo Multidimensional
  - Definição
  - Principais Operações
- Arquiteturas OLAP
  - ROLAP
  - MOLAP
  - HOLAP
- Exemplos de Relatórios OLAP
- Bibliografia

# Hierarquia da Informação



# Sistemas de Suporte à Decisão

- O que são Sistemas de Suporte a Decisão(SSD)?
  - Permitem **armazenar** e **analisar** grandes BD para extrair informações que auxiliam a compreensão do comportamento dos dados sobre o negócio
    - Exemplo: Sistemas de Data Warehouse (SDW)
  - Fornecem apoio ao suporte estratégico de decisão
    - Vantagem competitiva e aumento da produtividade na tomada de decisão
  - Consistem em sistemas computacionais de suporte à decisão que integram dados oriundos de diversas fontes de dados
  - São usados por gerentes de negócio e gestores

# Evolução de Sistemas de DW

- Por quê surgiram Sistemas de DW?
  - Necessidade de gerenciamento das atividades de negócio das organizações
  - Organizações acumularam uma grande quantidade de dados ao longo do tempo
    - Necessidade de se fazer o cruzamento dos dados
  - Organizações sentiram a necessidade de avaliar tendências e o comportamento dos dados
    - Necessidade de manter não apenas dados atuais, mas também dados históricos
  - Necessidade de uma ferramenta de BD para SD

# Comparativo entre SDW e SSO

- SDW diferem de SSO (ou transacional)
  - Operacional → orientado por transação
    - Ex: o produto vendido, o artigo publicado, o imposto predial pago
    - Muitas das consultas feitas são previstas a priori
  - Decisão → orientado por assunto
    - Ex: Vendas, Publicações de Artigos, Gerenciamento Urbano
    - Consultas mudam continuamente
- Dividir os dados em dois ambientes ortogonais
  - Operacional (normalmente já existente)
  - Estratégico (construir um DW)

# Comparativo entre SDW e SSO

	Transacional	Data Warehousing
<b>Frequência de atualização</b>	Tempo real	Periodicamente
<b>Estruturado para</b>	Integridade de dados	Facilidade de consulta
<b>Otimizado para</b>	Processamento de transações	Performance de consultas

# Comparativo entre SDW e SSO

	Transacional	Data Warehousing
<b>Frequência de atualização</b>	Tempo real	Periodicamente
<b>Estruturado para</b>	Integridade de dados	Facilidade de consulta
<b>Otimizado para</b>	Processamento de transações	Performance de consultas

# Comparativo entre SDW e SSO

	Transacional	Data Warehousing
<b>Frequência de atualização</b>	Tempo real	Periodicamente
<b>Estruturado para</b>	Integridade de dados	Facilidade de consulta
<b>Otimizado para</b>	Processamento de transações	Performance de consultas

# Comparativo entre SDW e SSO

	Transacional	Data Warehousing
<b>Frequência de atualização</b>	Tempo real	Periodicamente
<b>Estruturado para</b>	Integridade de dados	Facilidade de consulta
<b>Otimizado para</b>	Processamento de transações	Performance de consultas

# Comparativo entre SDW e SSO

- Diferenças entre os ambientes

<b>Características</b>	<b>Operacional</b>	<b>Estratégico</b>
Objetivo	Op. diárias do negócio	Análise do negócio
Visão dos dados	Relacional ou Obi-Rel.	Dimensional
Op. com os dados	Incl., Alt, Excl e Cons.	Carga e Consulta
Atualização	Contínua (tempo real)	Temporal (lote)
Nº de usuários	Milhares	Dezenas
Tipo de usuário	Operacional	Gerencial
Interação c/ usuário	Pré-definida(predom.)	Pré-definida e ad-hoc
Granularidade dados	Detalhados	Detalhados e Resumos
Redundância dados	Ocorrência minimizada	Ocorrência maximizada
Volume dos dados	Megabytes-Gigabytes	Gigabytes-Terabytes
Histórico dos dados	Até a última atualização	5 a 10 anos
Acesso a registro	Dezenas	Milhares

# Comparativo entre SDW e SSO

- Diferenças entre os ambientes

<b>Características</b>	<b>Operacional</b>	<b>Estratégico</b>
Objetivo	Op. diárias do negócio	Análise do negócio
Visão dos dados	Relacional ou Obi-Rel.	Dimensional
Op. com os dados	Incl., Alt, Excl e Cons.	Carga e Consulta
Atualização	Contínua (tempo real)	Temporal (lote)
Nº de usuários	Milhares	Dezenas
Tipo de usuário	Operacional	Gerencial
Interação c/ usuário	Pré-definida(predom.)	Pré-definida e ad-hoc
Granularidade dados	Detalhados	Detalhados e Resumos
Redundância dados	Ocorrência minimizada	Ocorrência maximizada
Volume dos dados	Megabytes-Gigabytes	Gigabytes-Terabytes
Histórico dos dados	Até a última atualização	5 a 10 anos
Acesso a registro	Dezenas	Milhares

# Comparativo entre SDW e SSO

- Diferenças entre os ambientes

<b>Características</b>	<b>Operacional</b>	<b>Estratégico</b>
Objetivo	Op. diárias do negócio	Análise do negócio
Visão dos dados	Relacional ou Obi-Rel.	Dimensional
Op. com os dados	Incl., Alt, Excl e Cons.	Carga e Consulta
Atualização	Contínua (tempo real)	Temporal (lote)
Nº de usuários	Milhares	Dezenas
Tipo de usuário	Operacional	Gerencial
Interação c/ usuário	Pré-definida(predom.)	Pré-definida e ad-hoc
Granularidade dados	Detalhados	Detalhados e Resumos
Redundância dados	Ocorrência minimizada	Ocorrência maximizada
Volume dos dados	Megabytes-Gigabytes	Gigabytes-Terabytes
Histórico dos dados	Até a última atualização	5 a 10 anos
Acesso a registro	Dezenas	Milhares

# Comparativo entre SDW e SSO

- Diferenças entre os ambientes

<b>Características</b>	<b>Operacional</b>	<b>Estratégico</b>
Objetivo	Op. diárias do negócio	Análise do negócio
Visão dos dados	Relacional ou Obi-Rel.	Dimensional
Op. com os dados	Incl., Alt, Excl e Cons.	Carga e Consulta
Atualização	Contínua (tempo real)	Temporal (lote)
Nº de usuários	Milhares	Dezenas
Tipo de usuário	Operacional	Gerencial
Interação c/ usuário	Pré-definida(predom.)	Pré-definida e ad-hoc
Granularidade dados	Detalhados	Detalhados e Resumos
Redundância dados	Ocorrência minimizada	Ocorrência maximizada
Volume dos dados	Megabytes-Gigabytes	Gigabytes-Terabytes
Histórico dos dados	Até a última atualização	5 a 10 anos
Acesso a registro	Dezenas	Milhares

# Comparativo entre SDW e SSO

- Diferenças entre os ambientes

<b>Características</b>	<b>Operacional</b>	<b>Estratégico</b>
Objetivo	Op. diárias do negócio	Análise do negócio
Visão dos dados	Relacional ou Obi-Rel.	Dimensional
Op. com os dados	Incl., Alt, Excl e Cons.	Carga e Consulta
Atualização	Contínua (tempo real)	Temporal (lote)
Nº de usuários	Milhares	Dezenas
Tipo de usuário	Operacional	Gerencial
Interação c/ usuário	Pré-definida(predom.)	Pré-definida e ad-hoc
Granularidade dados	Detalhados	Detalhados e Resumos
Redundância dados	Ocorrência minimizada	Ocorrência maximizada
Volume dos dados	Megabytes-Gigabytes	Gigabytes-Terabytes
Histórico dos dados	Até a última atualização	5 a 10 anos
Acesso a registro	Dezenas	Milhares

# Comparativo entre SDW e SSO

- Diferenças entre os ambientes

<b>Características</b>	<b>Operacional</b>	<b>Estratégico</b>
Objetivo	Op. diárias do negócio	Análise do negócio
Visão dos dados	Relacional ou Obi-Rel.	Dimensional
Op. com os dados	Incl., Alt, Excl e Cons.	Carga e Consulta
Atualização	Contínua (tempo real)	Temporal (lote)
<b>Nº de usuários</b>	Milhares	Dezenas
Tipo de usuário	Operacional	Gerencial
Interação c/ usuário	Pré-definida(predom.)	Pré-definida e ad-hoc
Granularidade dados	Detalhados	Detalhados e Resumos
Redundância dados	Ocorrência minimizada	Ocorrência maximizada
Volume dos dados	Megabytes-Gigabytes	Gigabytes-Terabytes
Histórico dos dados	Até a última atualização	5 a 10 anos
Acesso a registro	Dezenas	Milhares

# Comparativo entre SDW e SSO

- Diferenças entre os ambientes

<b>Características</b>	<b>Operacional</b>	<b>Estratégico</b>
Objetivo	Op. diárias do negócio	Análise do negócio
Visão dos dados	Relacional ou Obi-Rel.	Dimensional
Op. com os dados	Incl., Alt, Excl e Cons.	Carga e Consulta
Atualização	Contínua (tempo real)	Temporal (lote)
Nº de usuários	Milhares	Dezenas
<b>Tipo de usuário</b>	Operacional	Gerencial
Interação c/ usuário	Pré-definida(predom.)	Pré-definida e ad-hoc
Granularidade dados	Detalhados	Detalhados e Resumos
Redundância dados	Ocorrência minimizada	Ocorrência maximizada
Volume dos dados	Megabytes-Gigabytes	Gigabytes-Terabytes
Histórico dos dados	Até a última atualização	5 a 10 anos
Acesso a registro	Dezenas	Milhares

# Comparativo entre SDW e SSO

- Diferenças entre os ambientes

<b>Características</b>	<b>Operacional</b>	<b>Estratégico</b>
Objetivo	Op. diárias do negócio	Análise do negócio
Visão dos dados	Relacional ou Obi-Rel.	Dimensional
Op. com os dados	Incl., Alt, Excl e Cons.	Carga e Consulta
Atualização	Contínua (tempo real)	Temporal (lote)
Nº de usuários	Milhares	Dezenas
Tipo de usuário	Operacional	Gerencial
Interação c/ usuário	Pré-definida(predom.)	Pré-definida e ad-hoc
Granularidade dados	Detalhados	Detalhados e Resumos
Redundância dados	Ocorrência minimizada	Ocorrência maximizada
Volume dos dados	Megabytes-Gigabytes	Gigabytes-Terabytes
Histórico dos dados	Até a última atualização	5 a 10 anos
Acesso a registro	Dezenas	Milhares

# Comparativo entre SDW e SSO

- Diferenças entre os ambientes

<b>Características</b>	<b>Operacional</b>	<b>Estratégico</b>
Objetivo	Op. diárias do negócio	Análise do negócio
Visão dos dados	Relacional ou Obi-Rel.	Dimensional
Op. com os dados	Incl., Alt, Excl e Cons.	Carga e Consulta
Atualização	Contínua (tempo real)	Temporal (lote)
Nº de usuários	Milhares	Dezenas
Tipo de usuário	Operacional	Gerencial
Interação c/ usuário	Pré-definida(predom.)	Pré-definida e ad-hoc
<b>Granularidade dados</b>	Detalhados	Detalhados e Resumos
Redundância dados	Ocorrência minimizada	Ocorrência maximizada
Volume dos dados	Megabytes-Gigabytes	Gigabytes-Terabytes
Histórico dos dados	Até a última atualização	5 a 10 anos
Acesso a registro	Dezenas	Milhares

# Comparativo entre SDW e SSO

- Diferenças entre os ambientes

<b>Características</b>	<b>Operacional</b>	<b>Estratégico</b>
Objetivo	Op. diárias do negócio	Análise do negócio
Visão dos dados	Relacional ou Obi-Rel.	Dimensional
Op. com os dados	Incl., Alt, Excl e Cons.	Carga e Consulta
Atualização	Contínua (tempo real)	Temporal (lote)
Nº de usuários	Milhares	Dezenas
Tipo de usuário	Operacional	Gerencial
Interação c/ usuário	Pré-definida(predom.)	Pré-definida e ad-hoc
Granularidade dados	Detalhados	Detalhados e Resumos
Redundância dados	Ocorrência minimizada	Ocorrência maximizada
Volume dos dados	Megabytes-Gigabytes	Gigabytes-Terabytes
Histórico dos dados	Até a última atualização	5 a 10 anos
Acesso a registro	Dezenas	Milhares

# Comparativo entre SDW e SSO

- Diferenças entre os ambientes

<b>Características</b>	<b>Operacional</b>	<b>Estratégico</b>
Objetivo	Op. diárias do negócio	Análise do negócio
Visão dos dados	Relacional ou Obi-Rel.	Dimensional
Op. com os dados	Incl., Alt, Excl e Cons.	Carga e Consulta
Atualização	Contínua (tempo real)	Temporal (lote)
Nº de usuários	Milhares	Dezenas
Tipo de usuário	Operacional	Gerencial
Interação c/ usuário	Pré-definida(predom.)	Pré-definida e ad-hoc
Granularidade dados	Detalhados	Detalhados e Resumos
Redundância dados	Ocorrência minimizada	Ocorrência maximizada
<b>Volume dos dados</b>	Megabytes-Gigabytes	Gigabytes-Terabytes
Histórico dos dados	Até a última atualização	5 a 10 anos
Acesso a registro	Dezenas	Milhares

# Comparativo entre SDW e SSO

- Diferenças entre os ambientes

<b>Características</b>	<b>Operacional</b>	<b>Estratégico</b>
Objetivo	Op. diárias do negócio	Análise do negócio
Visão dos dados	Relacional ou Obi-Rel.	Dimensional
Op. com os dados	Incl., Alt, Excl e Cons.	Carga e Consulta
Atualização	Contínua (tempo real)	Temporal (lote)
Nº de usuários	Milhares	Dezenas
Tipo de usuário	Operacional	Gerencial
Interação c/ usuário	Pré-definida(predom.)	Pré-definida e ad-hoc
Granularidade dados	Detalhados	Detalhados e Resumos
Redundância dados	Ocorrência minimizada	Ocorrência maximizada
Volume dos dados	Megabytes-Gigabytes	Gigabytes-Terabytes
<b>Histórico dos dados</b>	Até a última atualização	5 a 10 anos
Acesso a registro	Dezenas	Milhares

# Comparativo entre SDW e SSO

- Diferenças entre os ambientes

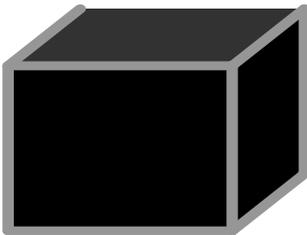
<b>Características</b>	<b>Operacional</b>	<b>Estratégico</b>
Objetivo	Op. diárias do negócio	Análise do negócio
Visão dos dados	Relacional ou Obi-Rel.	Dimensional
Op. com os dados	Incl., Alt, Excl e Cons.	Carga e Consulta
Atualização	Contínua (tempo real)	Temporal (lote)
Nº de usuários	Milhares	Dezenas
Tipo de usuário	Operacional	Gerencial
Interação c/ usuário	Pré-definida(predom.)	Pré-definida e ad-hoc
Granularidade dados	Detalhados	Detalhados e Resumos
Redundância dados	Ocorrência minimizada	Ocorrência maximizada
Volume dos dados	Megabytes-Gigabytes	Gigabytes-Terabytes
Histórico dos dados	Até a última atualização	5 a 10 anos
<b>Acesso a registro</b>	Dezenas	Milhares

- Conceitos e Terminologia de SDW
  - O que é Banco de Dados Multidimensional?
  - O que é Data Warehouse (DW)?
  - O que é Data Mart (DM)?
  - O que são ETL e ODS?
  - O que é Data Warehousing(DWing)?
  - O que são Sistemas de DW e OLAP?

- Conceitos e Terminologia de SDW

- O que é Banco de Dados Multidimensional?
- O que é Data Warehouse (DW)?
- O que é Data Mart (DM)?
- O que são ETL e ODS?
- O que é Data Warehousing(DWing)?
- O que são Sistemas de DW e OLAP?

## Banco de Dados MultiDimensional



**Pense (conceitualmente) em um Banco de Dados Multidimensional (BDMD) como um ARRAY gigantesco**

**BDMD(*i,j,l,m,n,o,p,r,s,t, ... ,*)**

**Tamanho Máximo do BDMD =  $i_{\max} * j_{\max} * k_{\max} * l_{\max} * \dots$**

## BIDIMENSIONAL

	Leste	Oeste	Sul
Produto1	50	80	100
Produto2	40	70	80
Produto3	90	120	140
Produto4	20	10	30

- Conceitos e Terminologia de SDW
  - O que é Banco de Dados Multidimensional?
  - O que é Data Warehouse (DW)?
  - O que é Data Mart (DM)?
  - O que são ETL e ODS?
  - O que é Data Warehousing(DWing)?
  - O que são Sistemas de DW e OLAP?

“Coleção de dados orientada a **assunto, integrada, não-volátil e variante no tempo**, utilizada para tomada de decisões”.

W. H. Inmon

"a copy of transaction data specifically structured for query and analysis"

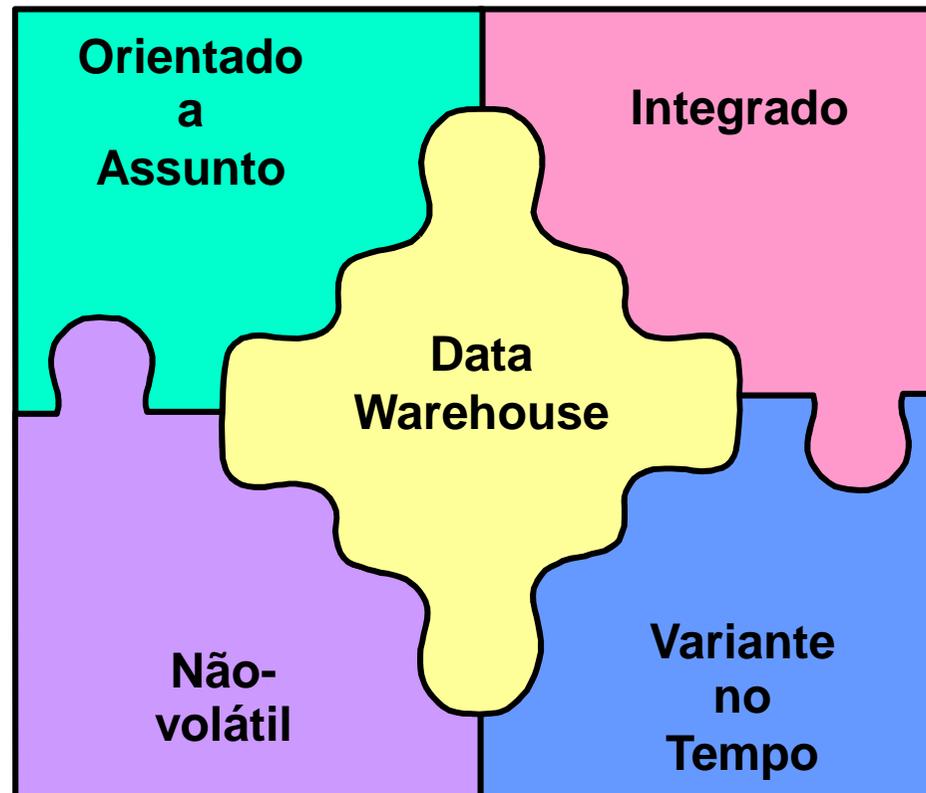
R. Kimball

“Repositório estruturado e corporativo de dados orientados a **assunto, variantes no tempo e históricos**, usados para recuperação de informações e **suporte à decisão**. O DW armazena dados atômicos e sumarizados”.

Definição de DW da Oracle

# Data Warehouse - Definição

- O que é um DW?
  - DW é uma base de dados que facilita a execução de consultas de apoio à decisão
- Propriedades de um DW

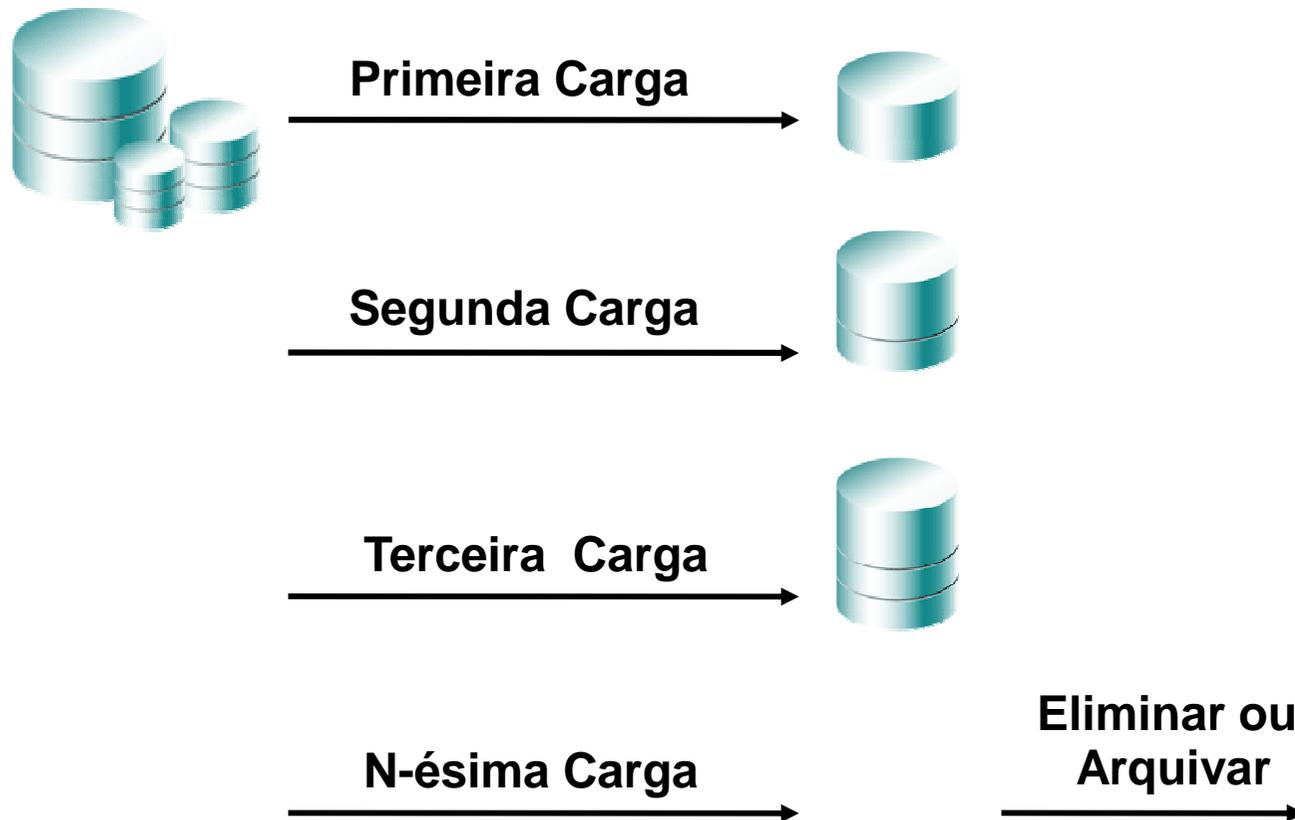


# Data Warehouse

- Alterando Dados em um DW

Bancos de Dados Operacionais

Banco de Data Warehouse



- Características dos dados de um DW

Característica	Descrição
Consolidado	Centralizado em toda a empresa
Consistente	Dentro do data warehouse
Orientado ao assunto	Organizado à perspectiva do usuário
Histórico	Snapshots no tempo
Somente-leitura	Não pode atualizar
Resumido	Ao nível apropriado de detalhe

- Características dos dados de um DW

Característica	Descrição
Consolidado	Centralizado em toda a empresa
Consistente	Dentro do data warehouse
Orientado ao assunto	Organizado à perspectiva do usuário
Histórico	Snapshots no tempo
Somente-leitura	Não pode atualizar
Resumido	Ao nível apropriado de detalhe

- Características dos dados de um DW

Característica	Descrição
Consolidado	Centralizado em toda a empresa
Consistente	Dentro do data warehouse
Orientado ao assunto	Organizado à perspectiva do usuário
Histórico	Snapshots no tempo
Somente-leitura	Não pode atualizar
Resumido	Ao nível apropriado de detalhe

- Características dos dados de um DW

Característica	Descrição
Consolidado	Centralizado em toda a empresa
Consistente	Dentro do data warehouse
Orientado ao assunto	Organizado à perspectiva do usuário
Histórico	Snapshots no tempo
Somente-leitura	Não pode atualizar
Resumido	Ao nível apropriado de detalhe

- Características dos dados de um DW

Característica	Descrição
Consolidado	Centralizado em toda a empresa
Consistente	Dentro do data warehouse
Orientado ao assunto	Organizado à perspectiva do usuário
Histórico	Snapshots no tempo
Somente-leitura	Não pode atualizar
Resumido	Ao nível apropriado de detalhe

- Características dos dados de um DW

Característica	Descrição
Consolidado	Centralizado em toda a empresa
Consistente	Dentro do data warehouse
Orientado ao assunto	Organizado à perspectiva do usuário
Histórico	Snapshots no tempo
Somente-leitura	Não pode atualizar
Resumido	Ao nível apropriado de detalhe

- Características dos dados de um DW

Característica	Descrição
Consolidado	Centralizado em toda a empresa
Consistente	Dentro do data warehouse
Orientado ao assunto	Organizado à perspectiva do usuário
Histórico	Snapshots no tempo
Somente-leitura	Não pode atualizar
Resumido	Ao nível apropriado de detalhe

- Conceitos e Terminologia de SDW
  - O que é Banco de Dados Multidimensional?
  - O que é Data Warehouse (DW)?
  - O que é Data Mart (DM)?
  - O que são ETL e ODS?
  - O que é Data Warehousing(DWing)?
  - O que são Sistemas de DW e OLAP?

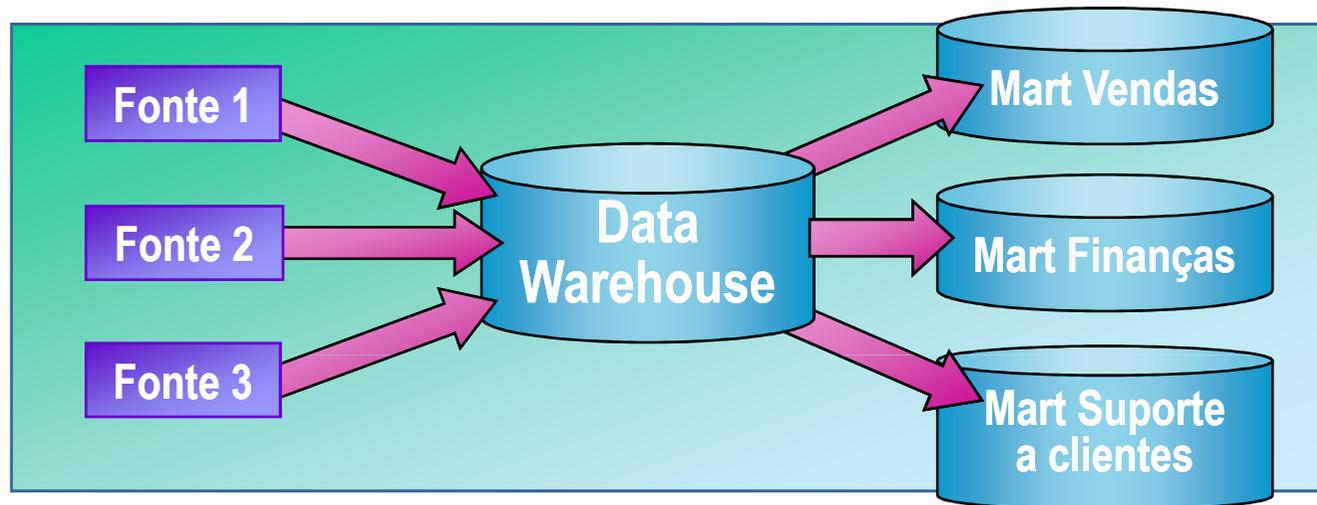
# Data Mart

- Definição
  - DM é um subconjunto de um DW
    - Subconjunto do DW que satisfaz os requisitos de um certo tema ou atividade de negócio
    - Projetado para um dado grupo de usuários
- Específico a um assunto particular ou atividade de negócio
- Pode ser visto como uma solução tática
- Pode ser construído antes ou depois do DW
  - Antes  $\Rightarrow$  pode representar fragmentos de dados, mas reduz a complexidade de desenvolvimento
  - Depois  $\Rightarrow$  produz uma visão integrada dos dados, mas aumenta a complexidade de desenvolvimento

Abordagem atual consiste em construir primeiro os DM, mas garantindo a consistência dos dados!

# Data Mart

- Movendo dados de um DW para DM



- **Vantagens**

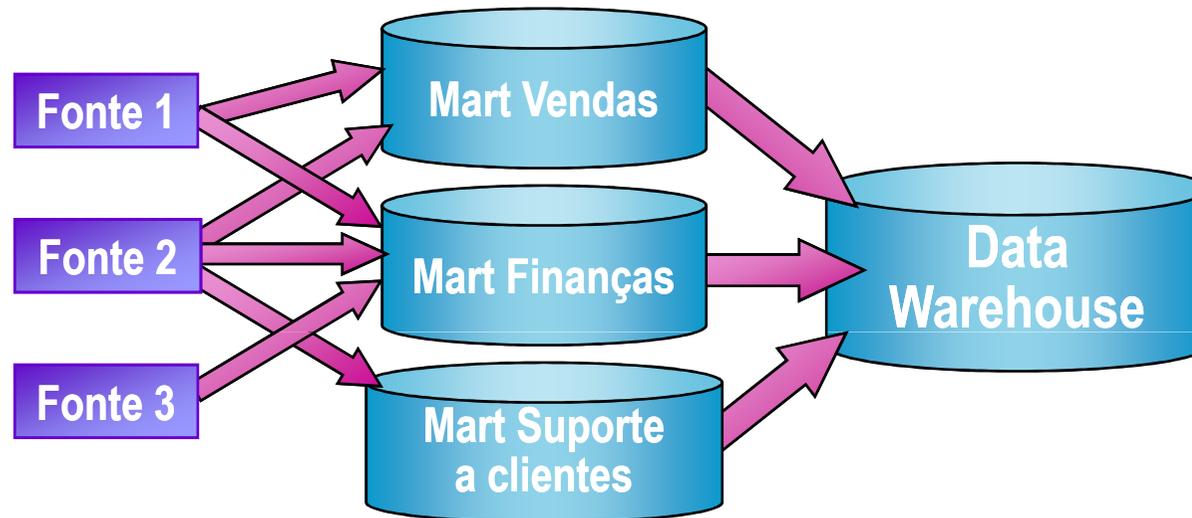
- Campos compartilhados
- Fonte comum
- Processamento distribuído

- **Desvantagens**

- Tempo mais longo de desenvolvimento
- Mais complexo, custoso e demorado

# Data Mart

- Movendo dados de DM para um DW



- **Vantagens**

- Mais simples e rápido
- Dados específicos de cada departamento
- Menos custoso

- **Desvantagens**

- Duplicação de dados
- Data Marts incompatíveis

- Benefícios
  - Tempo e dificuldade de implantação minimizados
  - São mais facilmente entendidos
  - Consultas mais rápidas e com menor número de usuários
- Restrição
  - Requer planejamento prévio para evitar **ilhas de dados (DM independentes)**
    - Garantir a consistência dos dados
    - Geração de DM integrados
- Atualização
  - Carga incorreta de DM deve ser corrigida
  - Mudanças podem gerar necessidade de atualização de DM

# Data Mart

- Razões Adicionais para Criação de DM
  - Permitir aos usuários, acesso aos dados que eles analisam mais frequentemente
  - Permitir o uso de dados em um formato que se assemelha com a visão coletiva de um grupo de usuários
  - Aumentar o tempo de resposta pela redução do volume de dados acessados
  - DM são mais facilmente extraídos, transformados, carregados e integrados
    - Implementação de DM é mais simples do que de um DW inteiro
    - Custo de implementação de DM é também menor
  - Potenciais usuários de DM são mais facilmente identificados em uma dada empresa
    - Facilita o levantamento de requisitos das atividades de negócio

- Comparativo entre DW e DM

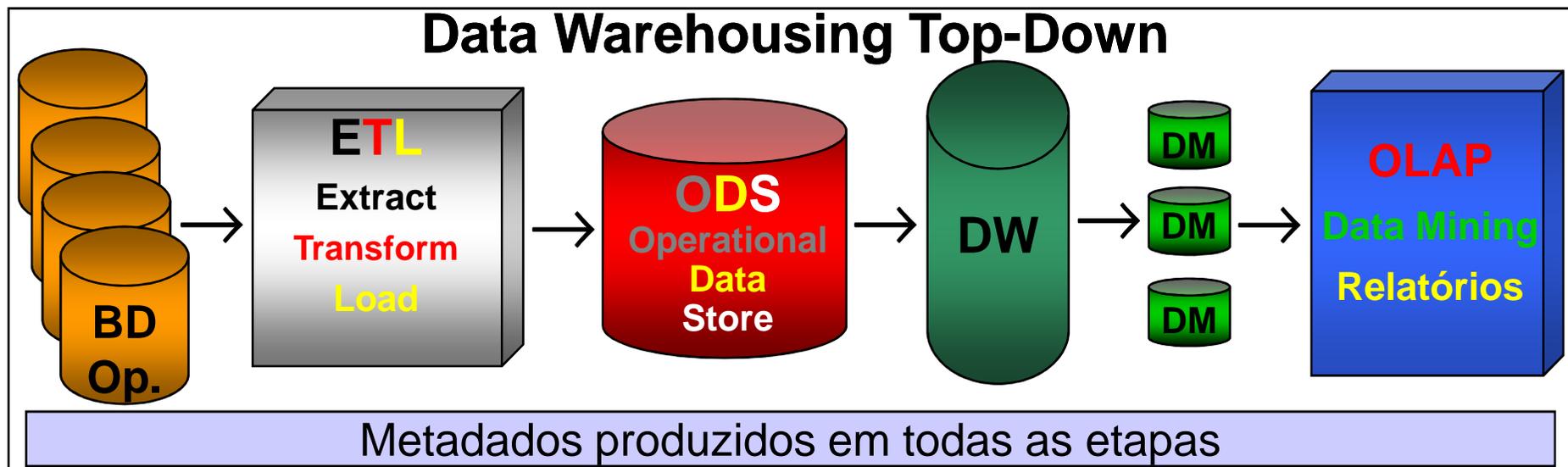
<b>Propriedade</b>	<b>Data Warehouse</b>	<b>Data Mart</b>
Escopo	Empresa	Departamento
Assuntos	Vários	Um único assunto
Fonte de Dados	Várias	Poucas
Tempo de Implementação	Meses a anos	Meses

- Conceitos e Terminologia de SDW
  - O que é Banco de Dados Multidimensional?
  - O que é Data Warehouse (DW)?
  - O que é Data Mart (DM)?
  - O que são ETL e ODS?
  - O que é Data Warehousing(DWing)?
  - O que são Sistemas de DW e OLAP?

- Ferramentas de ETL
  - Responsáveis pela conversão dos dados do ambiente operacional para o de suporte à decisão
  - Realizam Acesso, Extração, Transformação, Validação e Carga dos dados
- Operational Datastore (ODS)
  - Repositório de dados operacionais integrados
  - Benefícios
    - Otimiza a criação do DW
    - Possibilita a realização de consultas relacionais sobre dados históricos
    - Permite interações de tempo real (e.g. gerenciamento de relacionamentos de cliente)

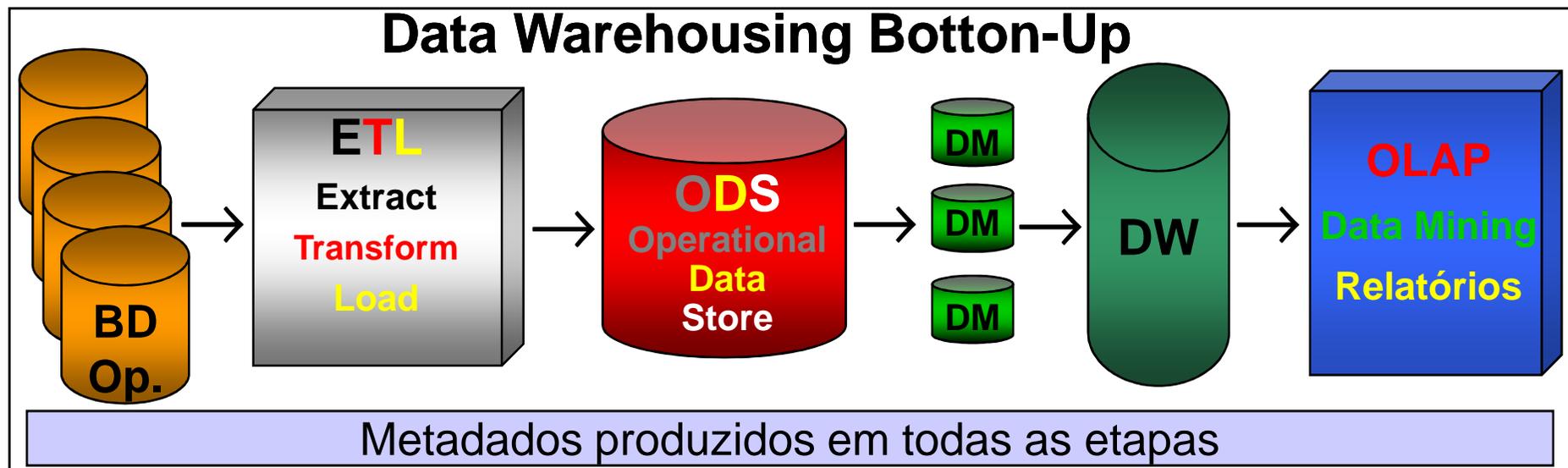
# ETL e ODS

- Abordagem Top-Down
  - Visão DW corporativo (DW  $\Rightarrow$  DM)
    - Grande abrangência
    - Mais Complexo, Custoso e Demorado
    - Alta probabilidade de insucesso
      - Só pode ser avaliado quando terminado



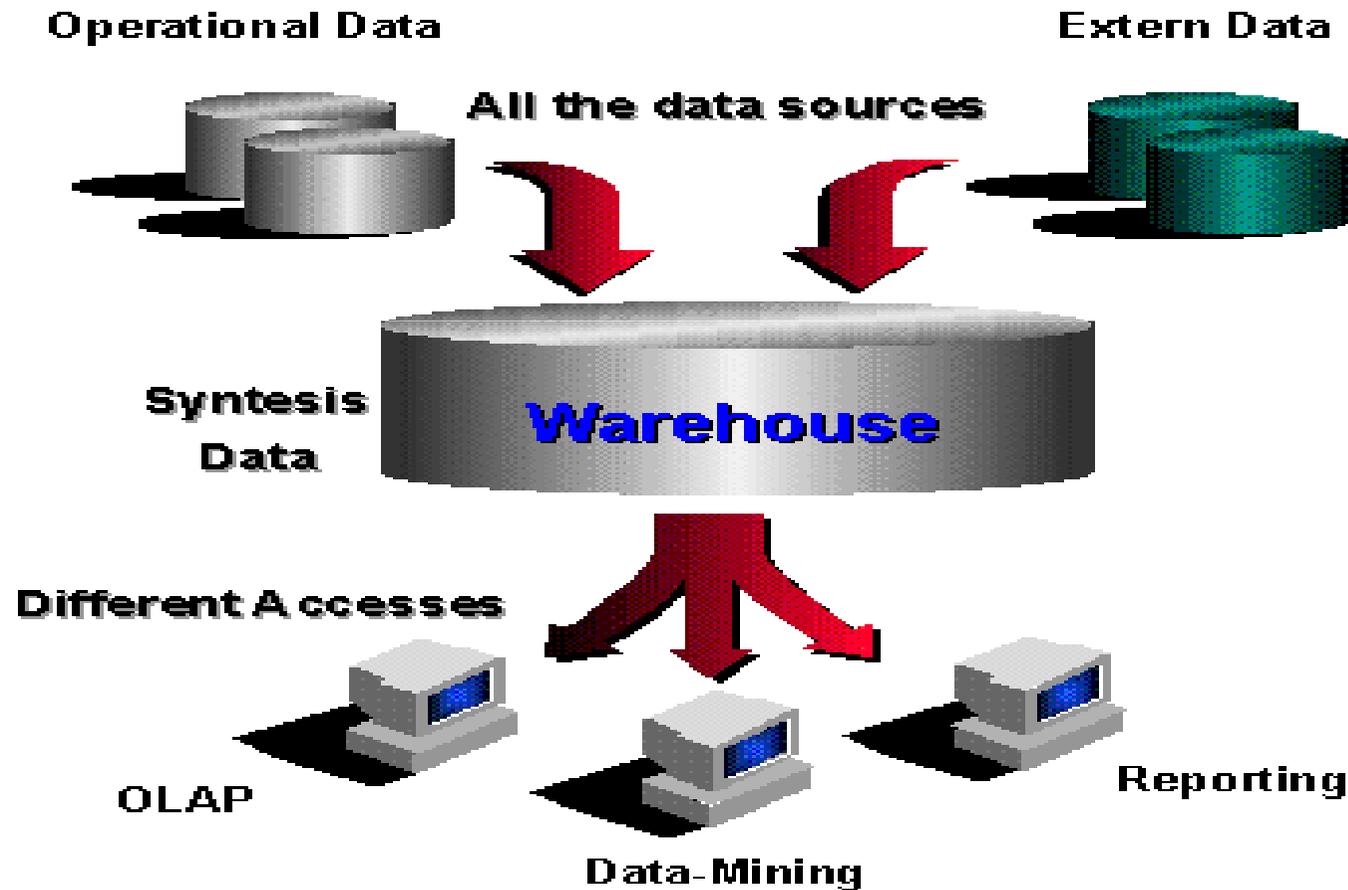
# ETL e ODS

- Abordagem Botton-Up
  - Visão DW departamental (DM  $\Rightarrow$  DW)
    - Foco específico nos aspectos mais críticos
    - Menos Complexo, Custoso, Demorado
    - Probabilidade de ilhas de dados
      - Pode acontecer dos DM não se integrarem



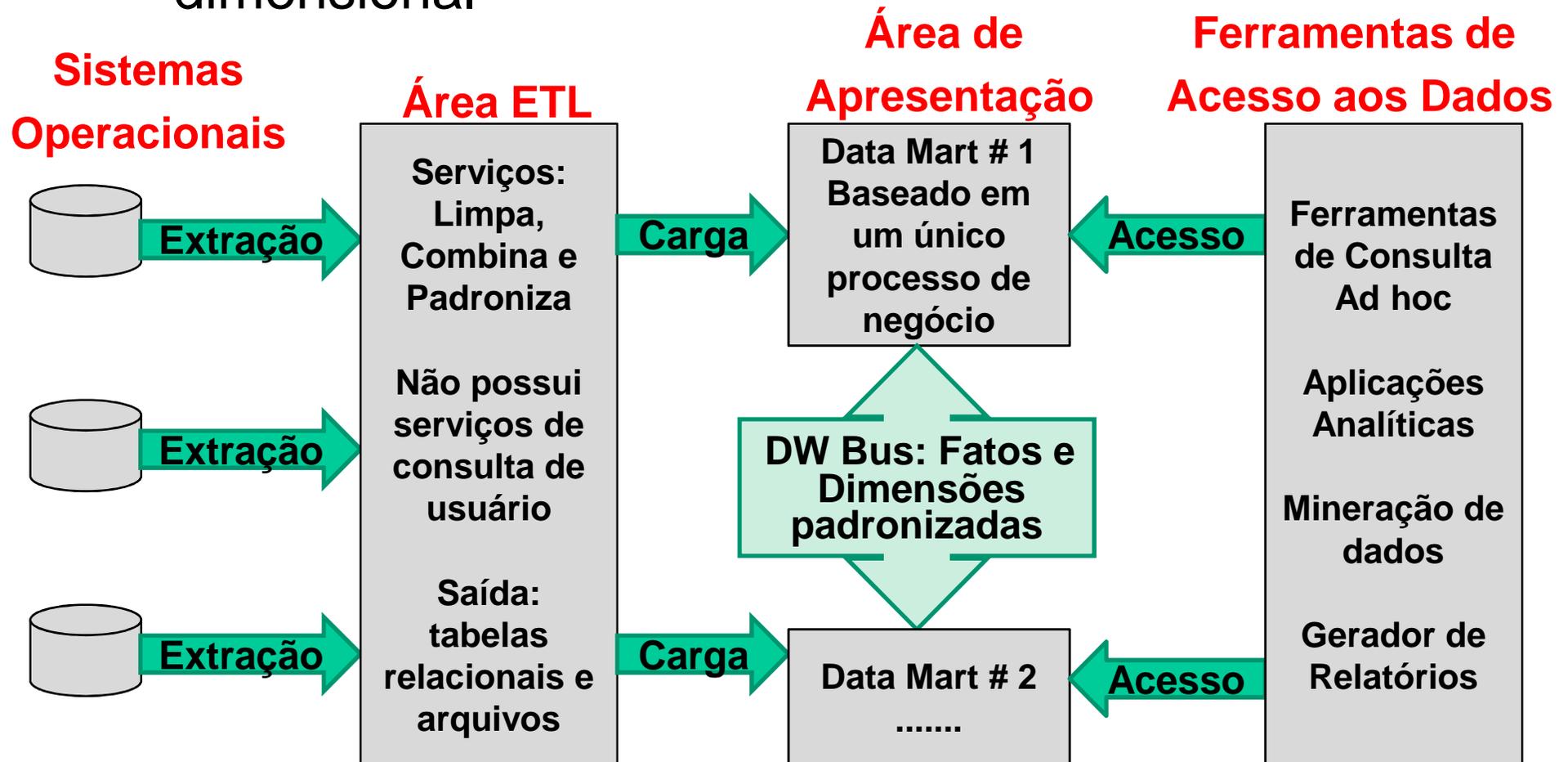
- Conceitos e Terminologia de SDW
  - O que é Banco de Dados Multidimensional?
  - O que é Data Warehouse (DW)?
  - O que é Data Mart (DM)?
  - O que são ETL e ODS?
  - O que é Data Warehousing(DWing)?
  - O que são Sistemas de DW e OLAP?

# Data Warehousing (DWing)



# Data Warehousing

- Definição
  - Processo de construção do **DW = base** de dados dimensional



- Conceitos e Terminologia de SDW
  - O que é Banco de Dados Multidimensional?
  - O que é Data Warehouse (DW)?
  - O que é Data Mart (DM)?
  - O que é ETL e ODS?
  - O que é Data Warehousing(DWing)?
  - O que é Sistemas de DW e OLAP?

# Sistemas de DW

- Provêm diferentes níveis de análise
  - São chamados de sistemas analíticos
- Permitem que usuários naveguem nos diferentes níveis de detalhes sobre os dados
  - Dados são organizados por meio de modelos dimensionais
  - Resultados de consultas são interpretados em uma variedade de visões multidimensionais
    - Consultas são providas pelas ferramentas OLAP (*On-Line Analytical Processing*)

# Sistemas de DW

E  
X  
E  
M  
P  
L  
O

## Relatório mensal de vendas do nordeste – Maio/1999

<u>Estado</u>	<u>Cidade</u>	<u>Un. Vendidas</u>	<u>Vendas \$</u>
PE	Olinda	2.500	\$12.850
PE	Recife	<u>2.750</u>	<u>\$14.135</u>
PE Totais			
CE	Fortaleza	3.200	\$16.800
CE	Crato	<u>1.725</u>	<u>\$ 9.143</u>
CE Totais			
AL	Maceió	<u>1.900</u>	<u>\$ 9.595</u>
AL Totais			
Região Nordeste Total		12.075	\$62.523

- Principais objetivos dos SDW
  - Tornar a informação da organização facilmente acessível
  - Exibir a informação de forma consistente
  - Ser adaptativo e acomodar facilmente mudanças
  - Servir de base para a tomada de decisão
  - Satisfazer as necessidades de análise dos usuários
  - Garantir a segurança das informações mantidas no DW

- Responsabilidades do administrador do SDW
  - Identificar que decisões seus usuários desejam tomar
  - Escolher o subconjunto de dados a ser mantido pelo DM
  - Monitorar a precisão dos dados e dos relatórios gerados
  - Garantir a confiabilidade dos dados, rotulando-o com consistência
  - Garantir interfaces e aplicações sejam simples de usar

# Modelos Dimensionais

- Objetivos da Modelagem Dimensional
  - Fornecer uma imagem global da realidade do negócio
  - Exibir informações em níveis apropriados de detalhes (**resumido** ou **detalhado**)
  - Otimizar o processamento de consultas complexas (**Modelo Estrela** ou **Flocos de Neve**)
  - Integrar dados de diversas fontes em uma única BD para facilitar a geração de relatórios

- Modelo Relacional
  - Usado para identificar relacionamentos entre tipos de relações
  - Visa remover a redundância de dados
  - Processamento de Transações On-Line (OLTP)
  
- Modelo Dimensional
  - Apresenta dados em uma estrutura intuitiva permitindo alta performance de acesso
  - Independe da representação física dos dados
  - Organiza dados em tabelas de  **fatos**  e  **dimensões**
  - Processamento Analítico On-Line (OLAP)

# Modelos Dimensionais

- Comparando Modelos de Dados

## Operacional: OLTP

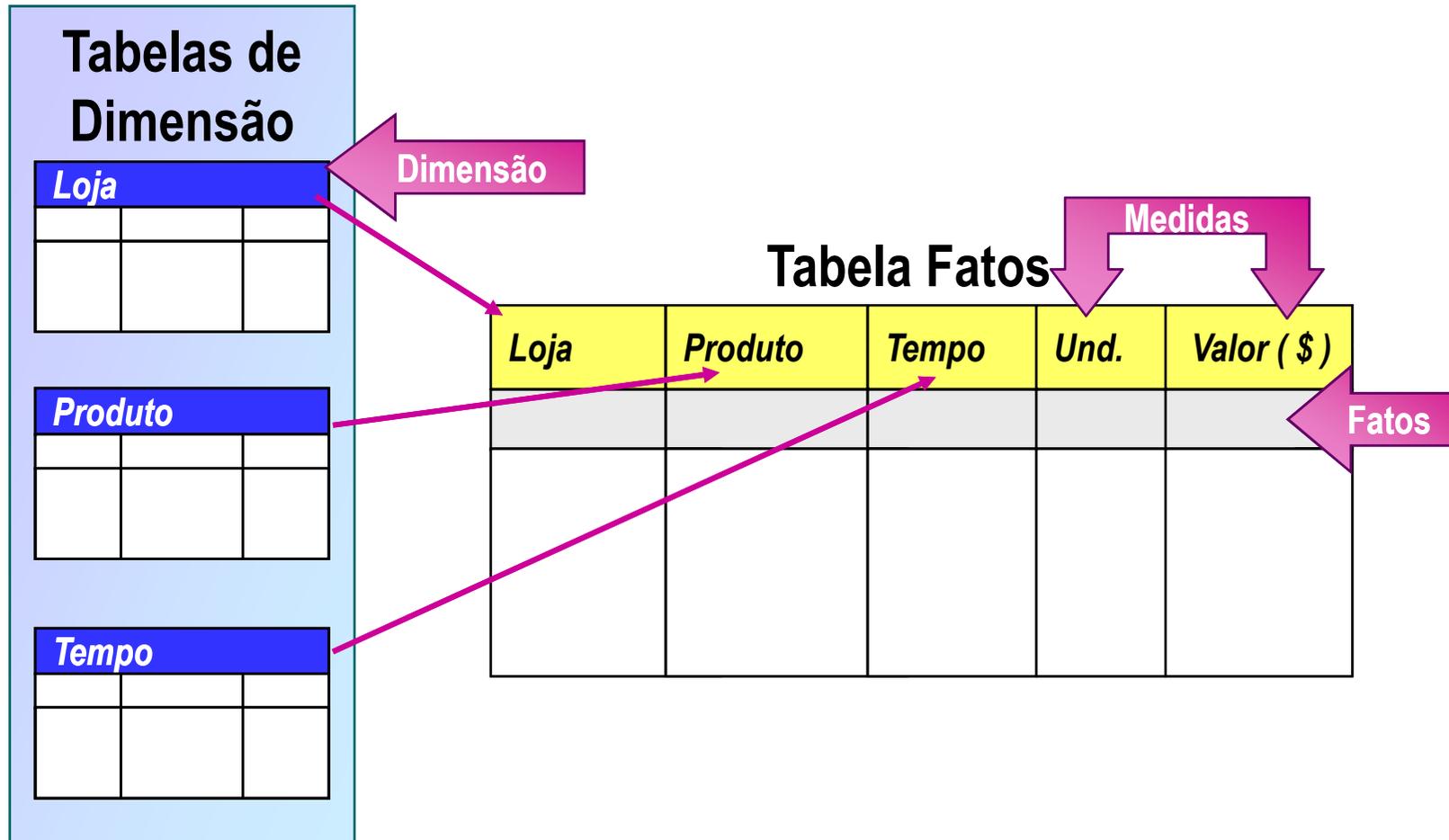
- Entidades normalizadas
- Segue terceira forma normal ou maior
- Produz um design complexo de BD
- Armazena dados no nível transacional mais baixo
- Aumenta o nível de JOIN de tabelas em consultas
- Estrutura tipicamente estática

## Analítico: Data Warehouse

- Entidades desnormalizadas
- Produz um único design de BD mais facilmente compreensível pelos usuários
- Armazena dados
  - Nível de transação
  - Nível de sumário
- Diminui o número de JOIN de tabelas em consultas
- Estrutura dinâmica

# Modelos Dimensionais

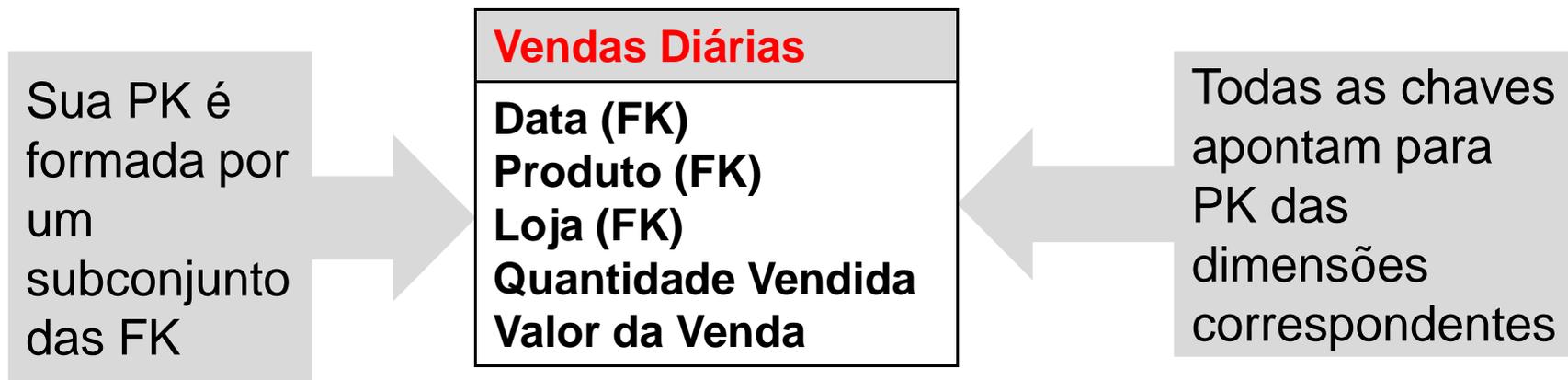
- Componentes do Modelo Dimensional



- Tabela de Fatos
  - Principal tabela do modelo dimensional na qual medidas numéricas sobre o desempenho da atividade de negócio são mantidas
    - Maioria dos fatos são numéricos e aditivos
    - Existem fatos semi-aditivos que podem ser adicionados por meio de algumas dimensões apenas
    - Existem fatos não aditivos que não podem ser adicionados
      - Sumarização é obtida por contagem ou cálculo da média
      - Exemplos incluem proporções e taxas
  - *Fato* representa uma medida de negócio
  - Uma linha da tabela de fato corresponde ao valor de uma medida

- Tabela de Fatos
  - Todas as medidas da tabela de fatos devem ter a mesma granularidade
  - Fatos possuem variação contínua
  - Na teoria, um fato pode ser textual
    - Uma medida textual é uma descrição de algo, i.e. temperatura
    - É obtida de uma lista discreta de valores
      - Por exemplo: 20 – 40C (quente)
  - Pode conter fatos pré-calculados que melhoram o desempenho mas aumentam o tamanho da tabela

- Tabela de Fatos
  - Exemplo



Se não existirem vendas em um dado dia, em uma dada loja e para um dado produto?

- Tabela de Fatos
  - Tabela central com grande volume de dados
  - Armazena as medidas numéricas do negócio e chaves das dimensões (**ID** das dimensões)
  - Cada fato é a interseção entre todas dimensões
  - Idealmente medidas são numéricas e aditivas
    - Vendas(R\$), Valor\_unitário(R\$), Despesas(R\$), QtdVendida
  - Exemplo de medida não numérica: **Temperatura**
  - Possui uma grande quantidade de linhas mas pouca quantidade de colunas
  - Existem três tipos: transacional, periódica e acumulativa

- Tabela de Fatos
  - Tabelas de fatos são **esparsas** e **únicas** por esquema
  - Requer determinação do nível de detalhe
  - Expressam relacionamentos N:M entre dimensões
  - Quando um dado numérico é Medida ou Atributo?
    - Medida  $\Rightarrow$  varia continuamente a cada amostragem
      - **Quantidade vendida de um produto**
    - Atributo  $\Rightarrow$  praticamente constante
      - **Peso de um produto**

# Modelos Dimensionais

- Tabelas de Dimensões
  - Exemplo

Dimensão Produto
Produto (PK)
Descrição
Código (Natural Key)
Marca
Categoria Departamento
Tipo da Embalagem
Tamanho da Embalagem
Peso
.....

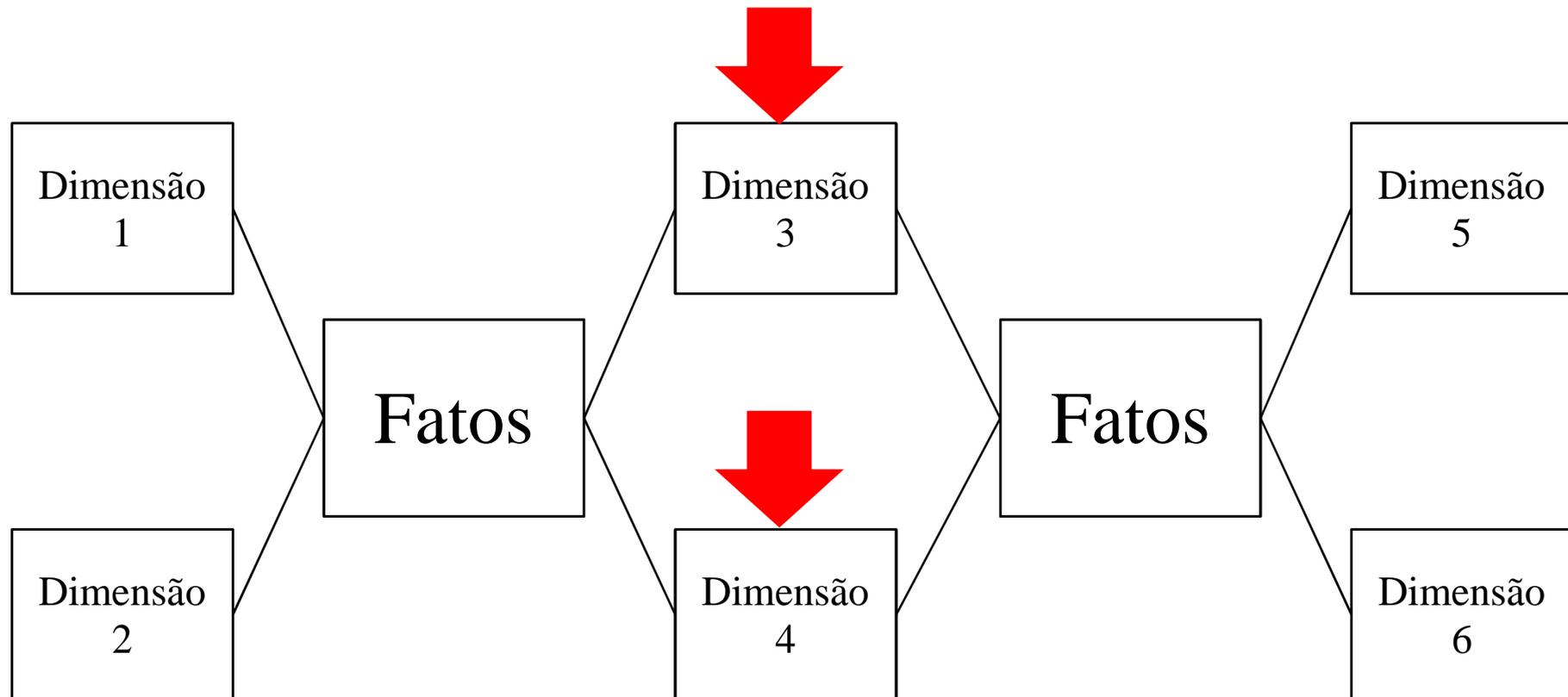
Possui PK para garantir integridade referencial com qualquer tabela de fato

- Embora **tamanho** e **peso** sejam numéricos, eles são atributos descritivos (não são medidas!)
  - Eles são descritores constantes e discretos de um dado produto

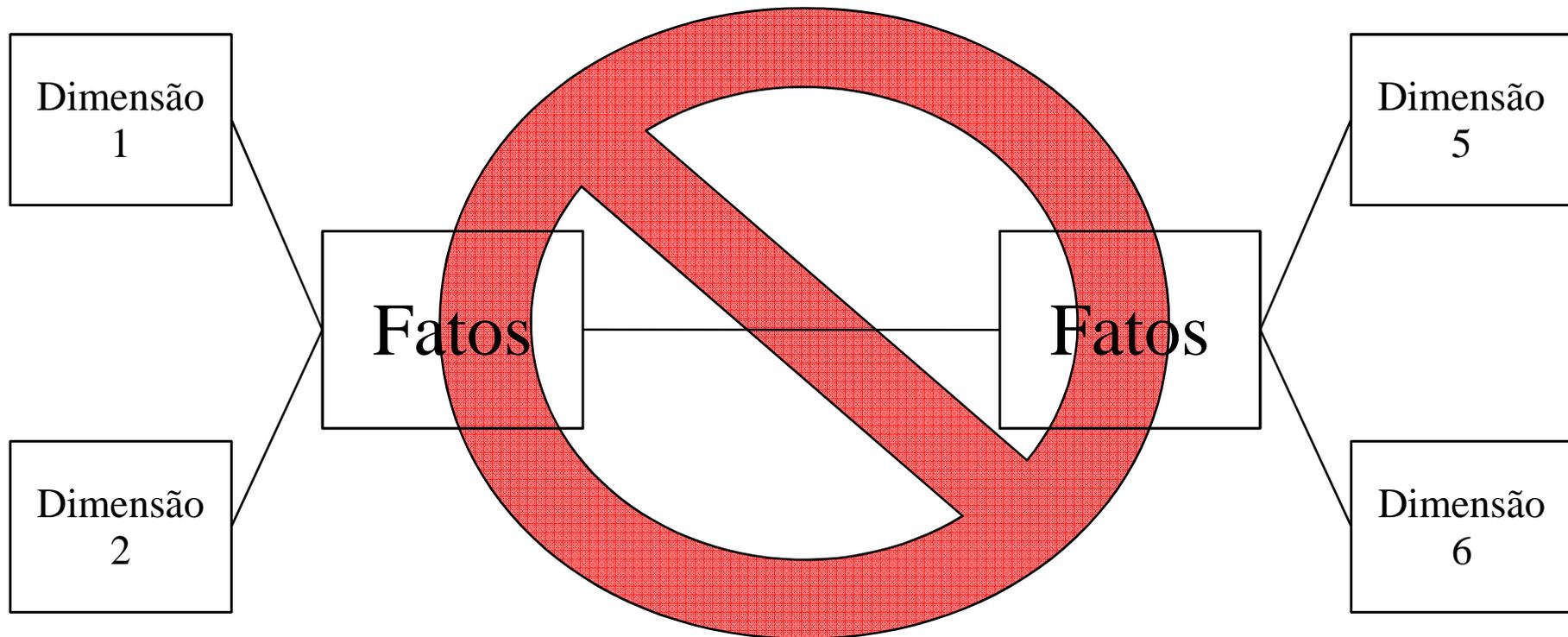
- Tabelas de Dimensões
  - Tabelas periféricas com pouco volume de dados
  - Armazenam as descrições do negócio
    - São usadas como filtros, agrupamentos e rótulos
    - Não devem ser abreviadas para aumentar legibilidade
  - Possuem poucas linhas mas muitos atributos
  - São normalmente desnormalizadas (**esquema estrela**)
  - Atributos  $\Rightarrow$  organizados em hierarquias
    - Produto (Categoria  $\rightarrow$  Marca  $\rightarrow$  Descrição)
    - Loja (Tipo  $\rightarrow$  Endereço  $\rightarrow$  Nome\_Loja)
    - Tempo (Ano  $\rightarrow$  Mês  $\rightarrow$  Dia\_Do\_Mês)
  - Podem ser compartilhadas ou privadas

# Modelos Dimensionais

- Dimensões Compartilhadas



- Fatos  $\leftrightarrow$  Fatos



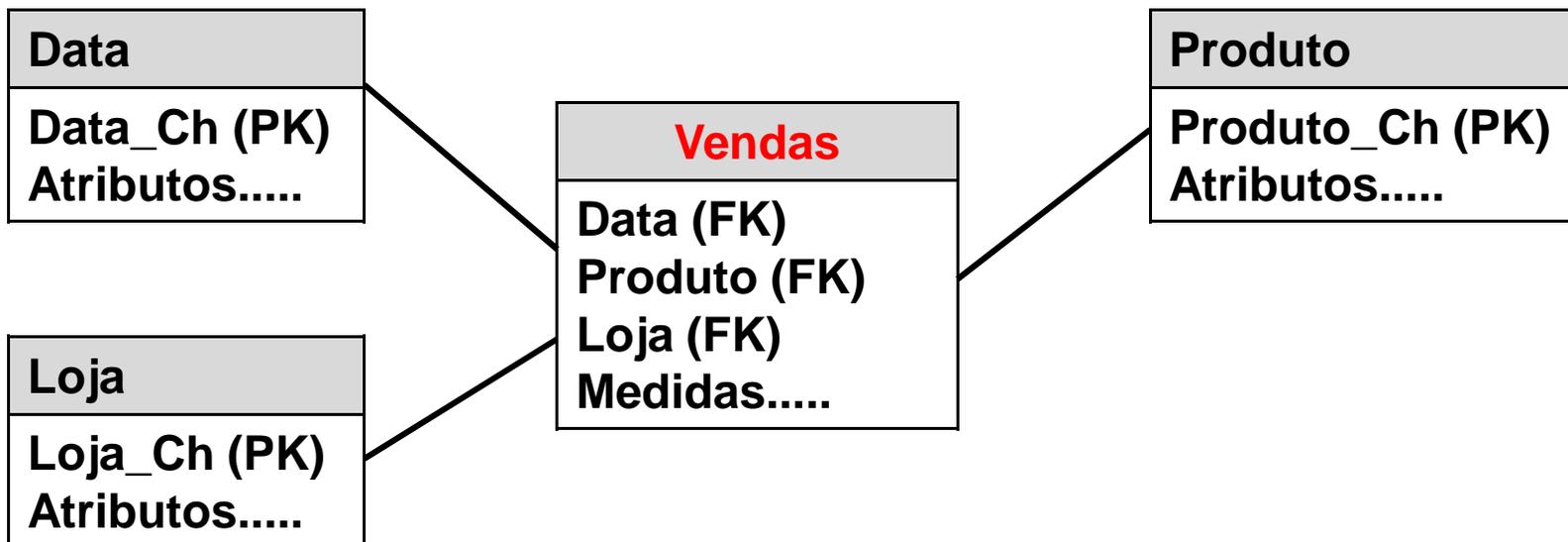
# Modelos Dimensionais

- Tabelas de Dimensões
  - Geralmente representam relacionamentos hierárquicos (1:N) entre seus atributos
    - Informação descritiva hierárquica é armazenada redundantemente
      - Por exemplo, para cada linha da dimensão produto, as descrições de **marca** e **categoria** são repetidas para cada uma de suas **subcategorias**
    - Elas são pequenas e representam menos de 10% do custo total de armazenamento do BD dimensional
    - Desnormalização favorece o desempenho de consultas e facilita o uso do sistema

Dimensões Desnormalizadas  $\Rightarrow$  modelo estrela  
Dimensões Normalizadas  $\Rightarrow$  modelo snowflake

# Modelos Dimensionais

- Tabelas de Dimensões
  - Códigos operacionais são quebrados em atributos separados das tabelas de dimensões
    - Permite que eles possam ser diretamente filtrados, agrupados e usados em relatórios facilmente



# Modelos Dimensionais

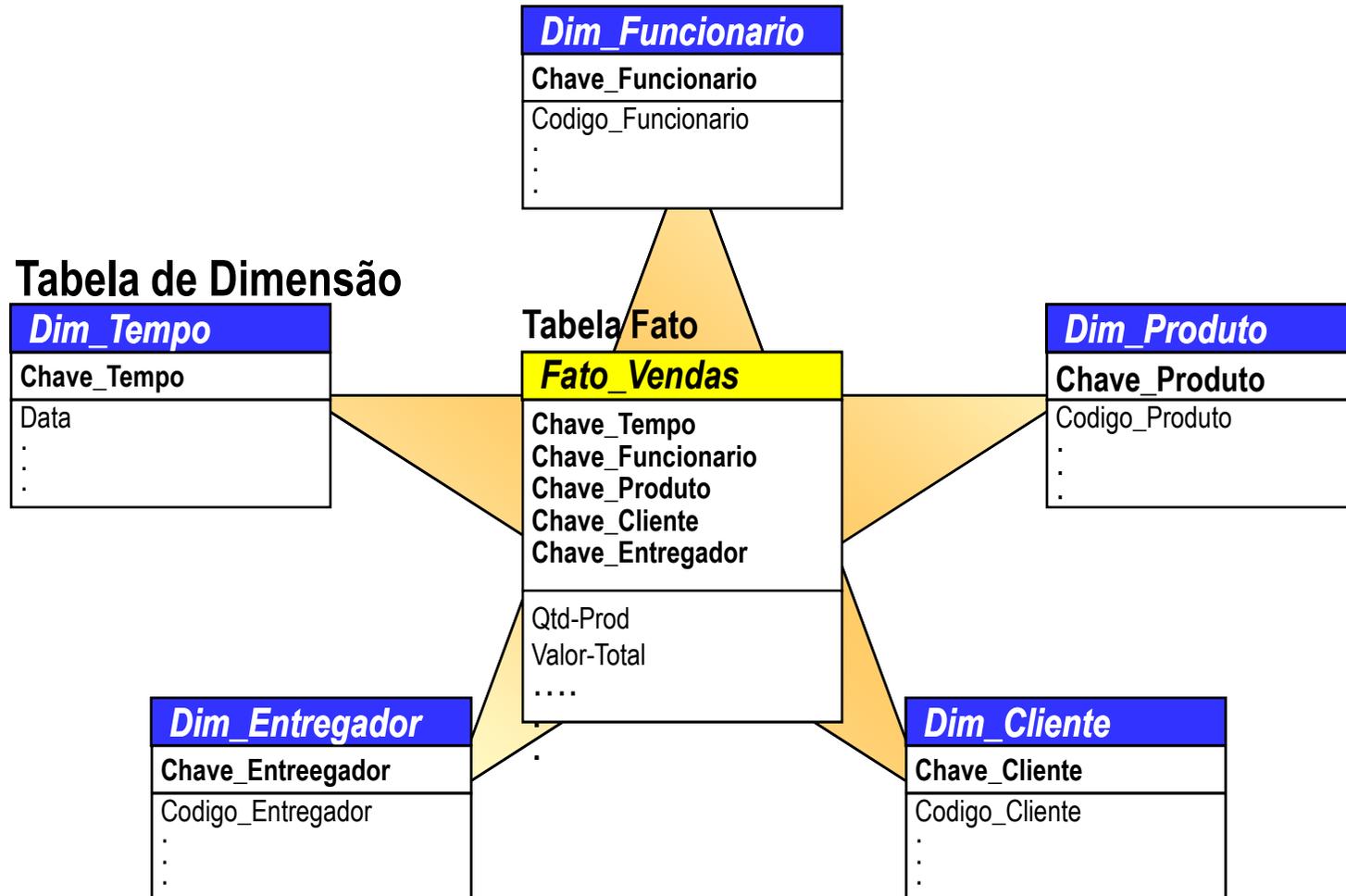
- Comparativo entre Tabela de Fatos e de Dimensão
- Fatos
  - Atributos **quantitativos** sobre o desempenho do negócio em um determinado ramo
  - Exemplo: sobre o fato vendas, a quantidade vendida, o preço da venda, a margem de lucro, etc
- Dimensões
  - Atributos **qualitativos** sobre os ramos do negócio envolvidos na medida de desempenho de um determinado fato
  - Exemplo: sobre a dimensão produto, a descrição, o código, o preço, etc

# Modelos Dimensionais

- Identificadores Artificiais
  - Cada dimensão deve possuir um número inteiro como identificador
    - Facilita as operações de junção-estrela
    - Códigos de ambientes operacionais mudam com o tempo
      - Códigos de produtos obsoletos
      - Números de contas inativas
    - Permite diferenciar instâncias de um mesmo elemento do ambiente operacional
    - Permite representar situações inexistentes no ambiente operacional (data a ser definida ou n/a, sem promoção)
    - Identificadores compostos não devem ser usados

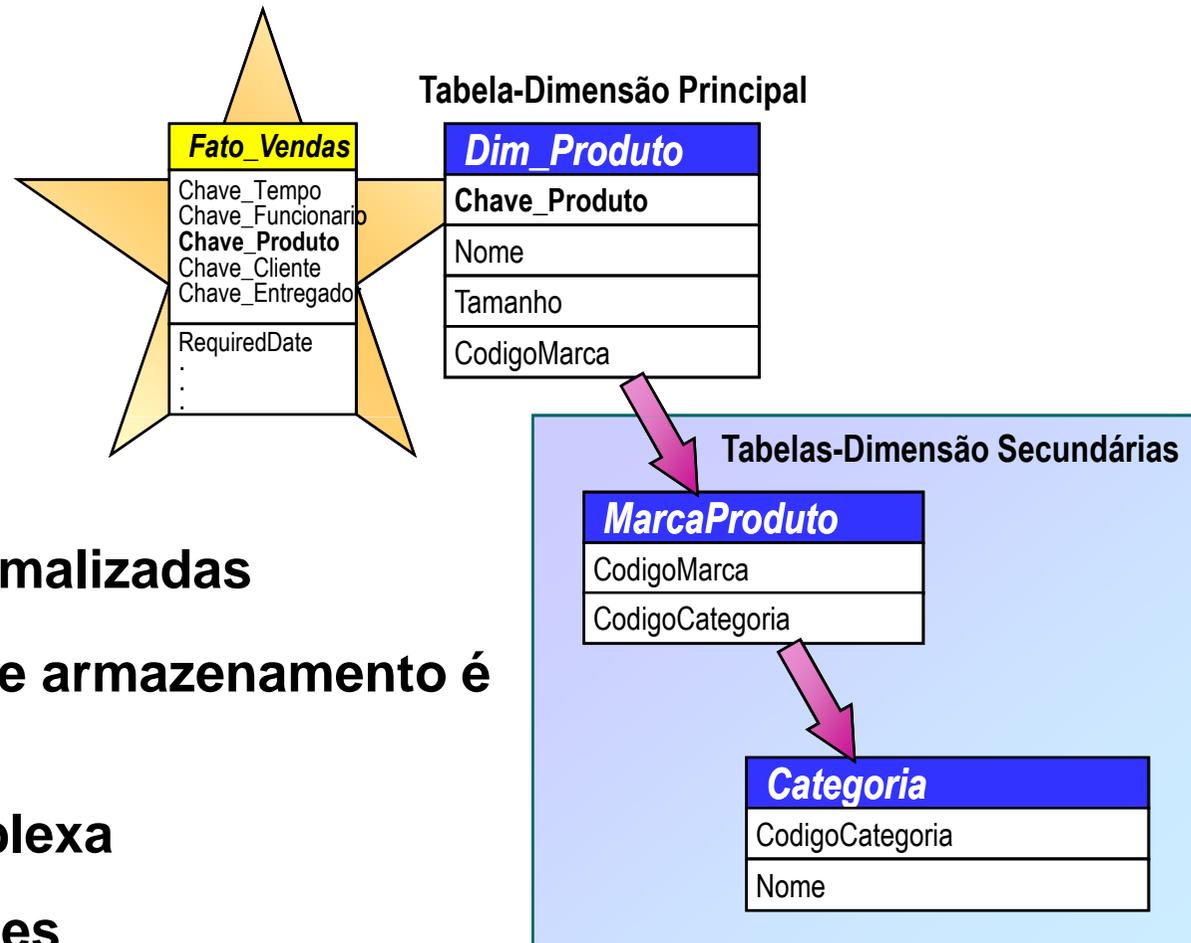
# Modelos Dimensionais

- Outro Exemplo de Esquema Estrela (*Star Schema*)



# Modelos Dimensionais

- Exemplo de Esquema Flocos de Neve (*Snowflake Schema*)



- Dimensões são normalizadas
- Ganho de espaço de armazenamento é pouco relevante
  - Estrutura complexa
  - Custo de junções

# Modelos Dimensionais

- Comparação entre Modelos de Dados Dimensionais

	Star	Snowflake
Clareza	<i>+ fácil</i>	<i>+ difícil</i>
Número de tabelas	<	>
Complexidade de consultas	<i>+ simples</i>	<i>+ complexo</i>
Performance de consulta	<i>+ rápido</i>	<i>+ lento</i>

Economia de espaço de disco obtida pelo *Snowflake* é geralmente menos de 1% do espaço de disco total para manter o BD multidimensional!

# Modelagem Dimensional

- Construção do Modelo Dimensional
  - Análise dos dados do ambiente operacional
  - Levantamento de requisitos das atividades de negócio dos usuários do DM
- A partir do esquema relacional, vários modelos dimensionais podem ser gerados
  - Identificar as atividades de negócio e modelá-las separadamente
  - Relacionamentos N:M com propriedades numéricas e aditivas são mapeados em tabelas fato
  - Desnormalizar todas as tabelas restantes e mapeá-las em tabelas de dimensão

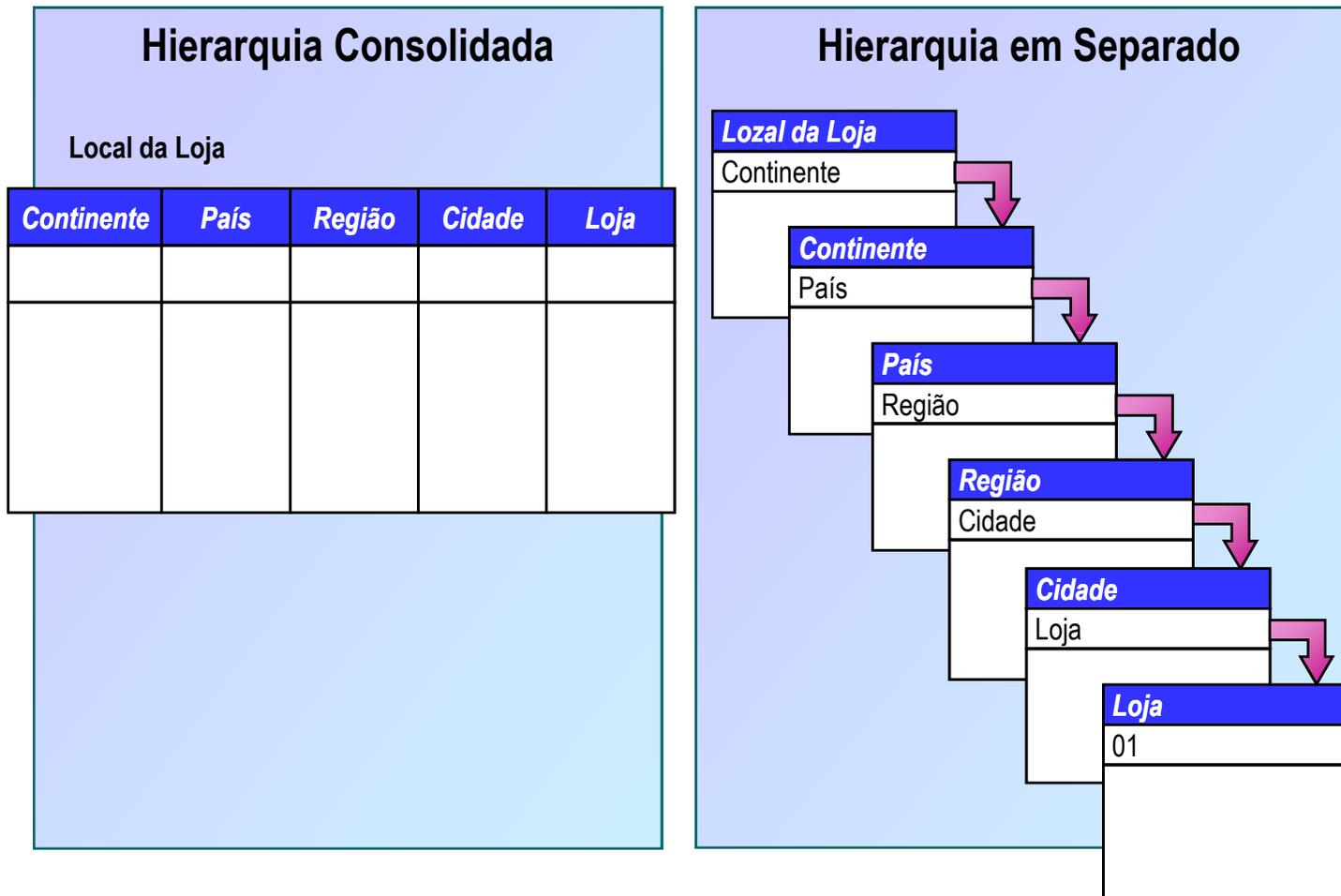
- Escolhendo a granularidade
  - Determinar requisitos dos dados
  - Escolher o nível mais baixo de detalhe
    - Requer espaço em disco
    - Envolve maior tempo de processamento
    - Permite capacidade de análise de dados detalhada
  - Adaptar medidas à granularidade estabelecida
  - Considerações de projeto de medidas
    - Usar medidas aditivas e numéricas

# Modelagem Dimensional

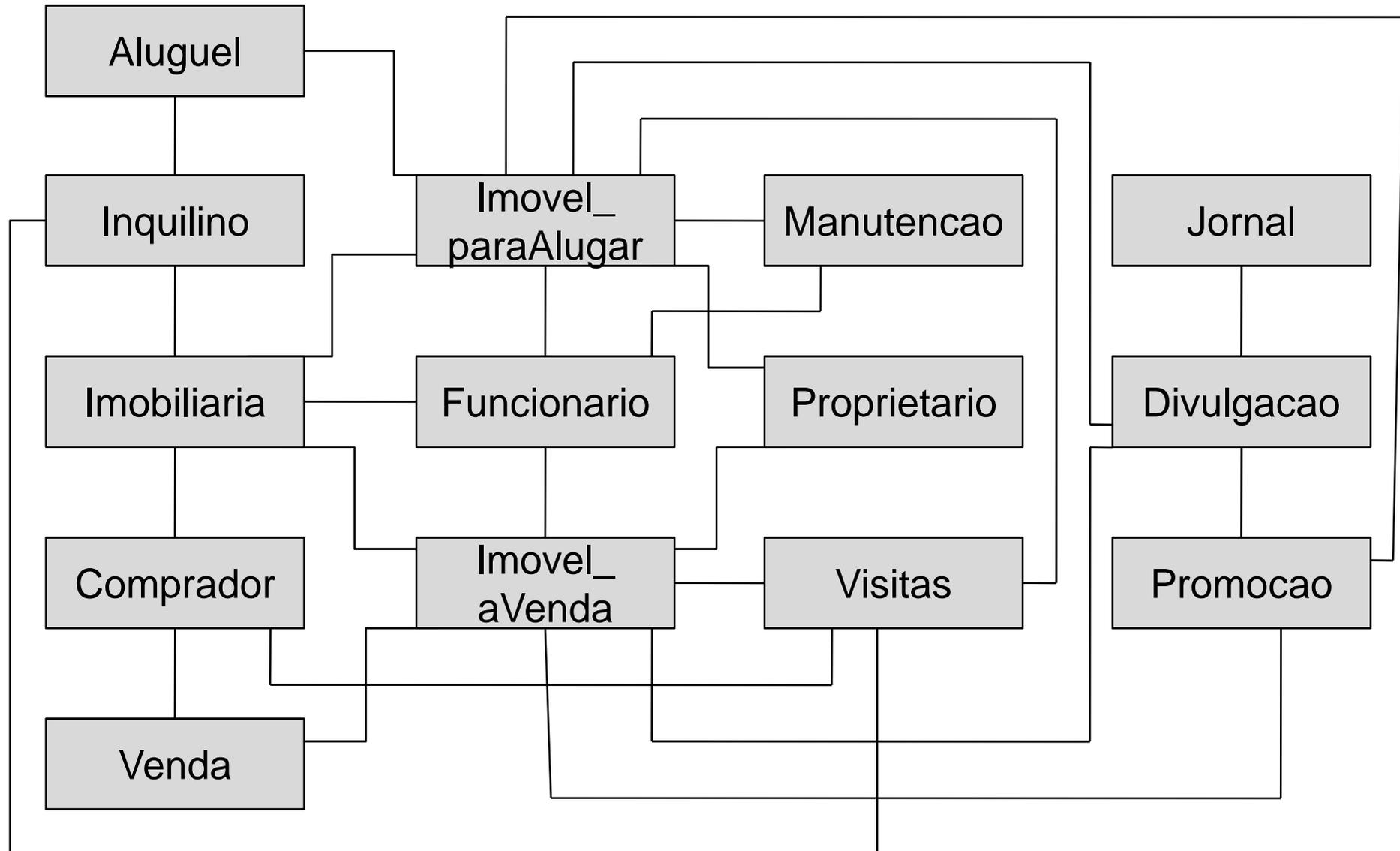
- Definindo características das dimensões
  - Aplicando características a tabelas de dimensão
    - Definir PRIMARY KEY
    - Incluir colunas altamente correlacionadas e descritivas
  - Projetando para Usabilidade e Extensibilidade
    - Minimizar ou evitar uso de códigos e abreviações
    - Criar colunas úteis para níveis de agregação
    - Evitar valores nulos ou inexistentes
    - Minimizar o número de registros que mudam ao longo do tempo

# Modelagem Dimensional

- Identificando hierarquias nas dimensões



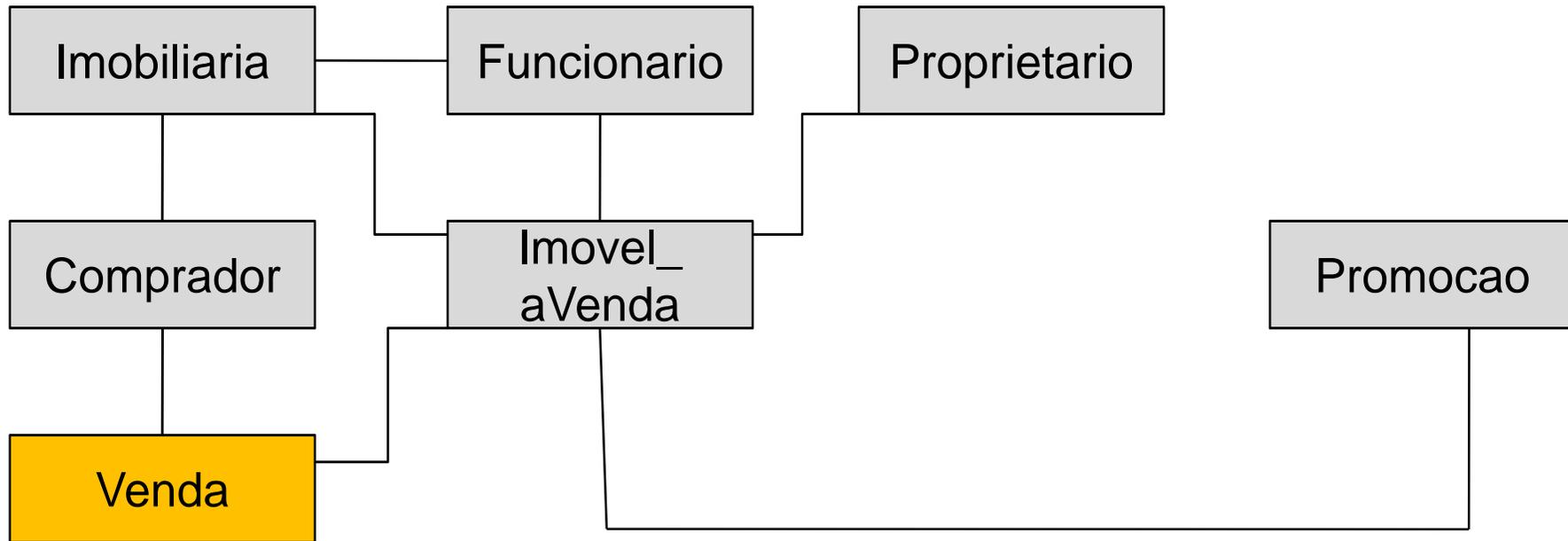
# Modelagem Dimensional





# Modelagem Dimensional

- Entidades da atividade de negócio *Vendas de Imóvel*



- Escolher a granularidade
  - Vendas de Imóveis Individuais
- Escolher as dimensões
  - Entidades restantes + dimensão temporal

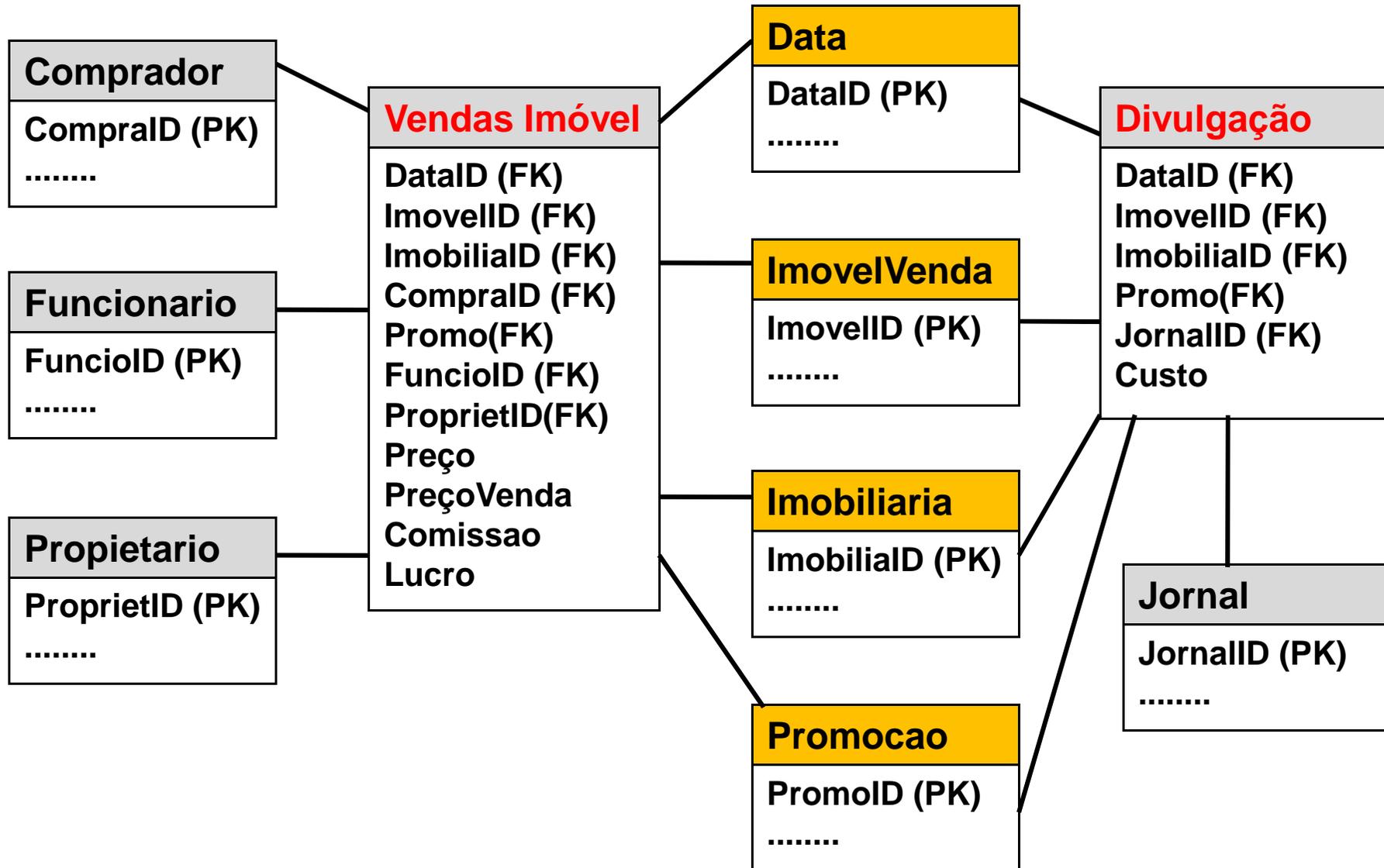
# Modelagem Dimensional

- Identificando e padronizando dimensões
  - Dimensões definem o contexto das consultas sobre os fatos analisados
  - O projeto bem elaborado de um conjunto de dimensões facilita o entendimento e uso do DM
    - Um conjunto incompleto de dimensões limita o universo de análises que podem ser feitas
  - Se duas dimensões existem em dois DM, então:
    - Elas devem ser exatamente as mesmas dimensões ou
    - Uma deve ser um subconjunto da outra



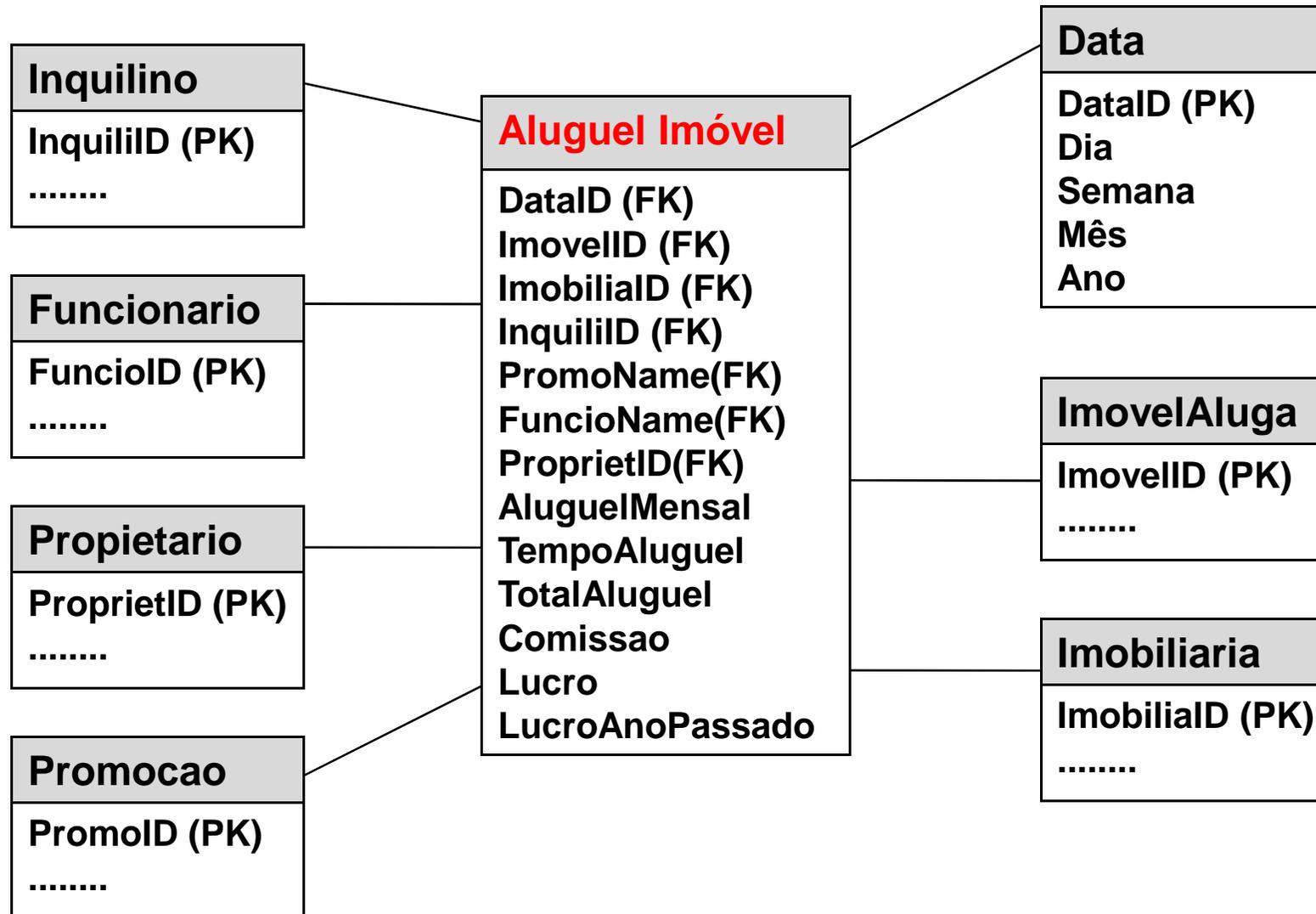
Dois DM podem compartilhar 2 ou mais dimensões em uma mesma aplicação.

# Modelagem Dimensional



# Modelagem Dimensional

- Esquema estrela da atividade *Aluguel de Imóvel*



# Modelagem Dimensional

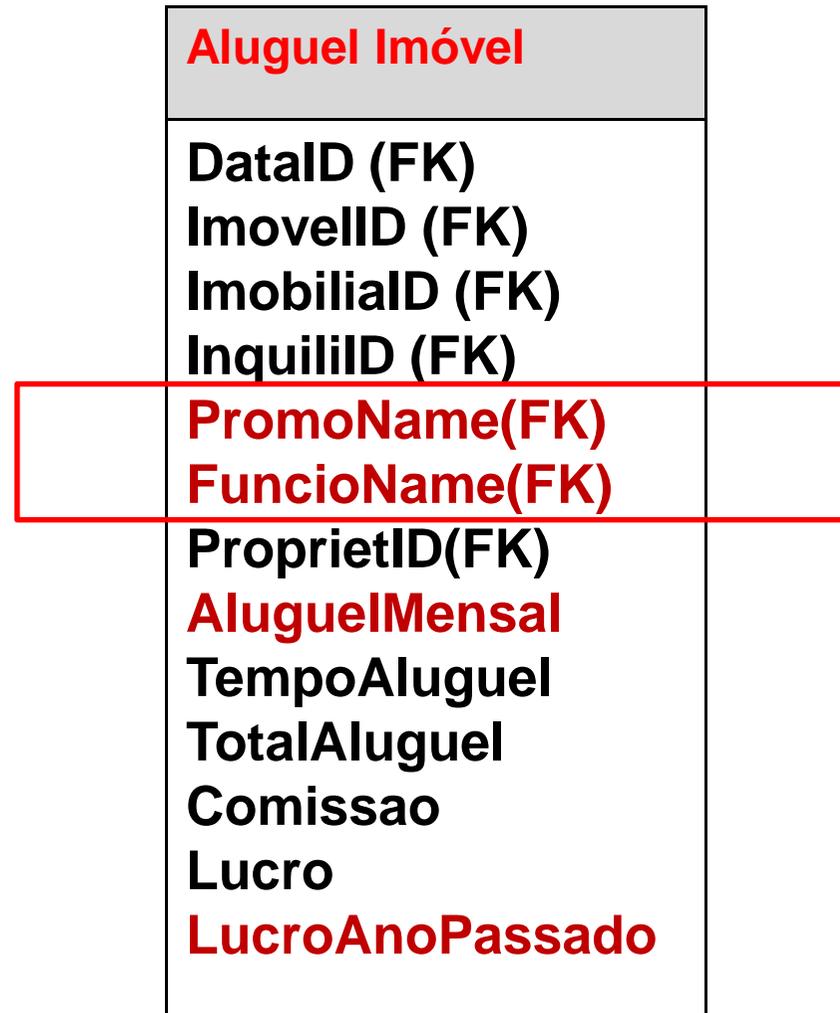
- Esquema estrela da atividade *Aluguel de Imóvel*

## **Aluguel Imóvel**

**DataID (FK)**  
**ImovelID (FK)**  
**ImobiliarID (FK)**  
**Inquilid (FK)**  
**PromoName(FK)**  
**FuncioName(FK)**  
**ProprietID(FK)**  
**AluguelMensal**  
**TempoAluguel**  
**TotalAluguel**  
**Comissao**  
**Lucro**  
**LucroAnoPassado**

# Modelagem Dimensional

- Esquema estrela da atividade *Aluguel de Imóvel*



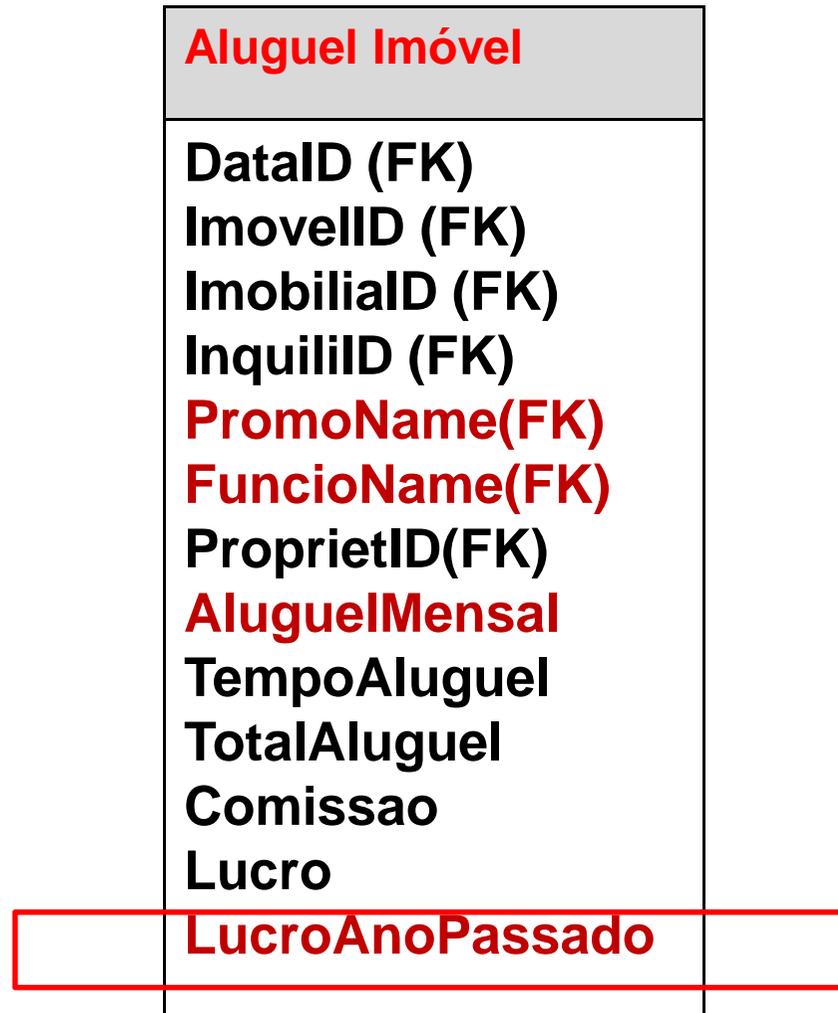
# Modelagem Dimensional

- Esquema estrela da atividade *Aluguel de Imóvel*



# Modelagem Dimensional

- Esquema estrela da atividade *Aluguel de Imóvel*



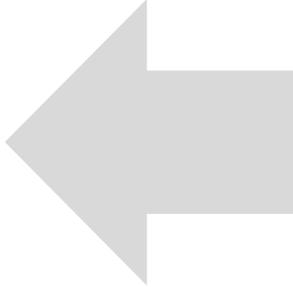
# Modelagem Dimensional

- Medidas Não Aditivas (Kimball):
  - Dependendo do contexto da aplicação e das necessidades do usuário
    - A soma das medidas pode fazer menos sentido do que a aplicação de uma outra função de agregação
      - Média
      - Mínimo ou Máximo
      - Desvio Padrão
    - Exemplos
      - Temperatura
      - Preços Unitários
      - Taxas

# Modelagem Dimensional

- Medidas Não Aditivas (Kimball):
  - Temperatura
    - » Não faz sentido dizer que a temperatura de 2008 foi 200 graus
    - » Faz sentido obter a temperatura média do ano
  - Preço Unitário

Preço de Venda do Milho	R\$
Janeiro	10,00
Fevereiro	11,00
Março	12,00
Abril	10,00
Mai	8,00
Junho	7,00

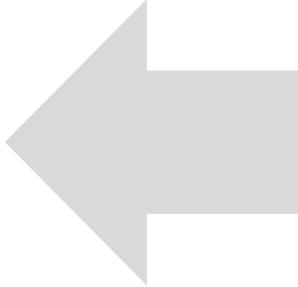


Não faz sentido dizer que o preço de venda da mão de milho no primeiro semestre foi R\$ 58,00

# Modelagem Dimensional

- Medidas Não Aditivas (Kimball):
  - Temperatura
    - » Não faz sentido dizer que a temperatura de 2008 foi 200 graus
    - » Faz sentido obter a temperatura média do ano
  - Preço Unitário

Preço de Venda do Milho	R\$
Janeiro	10,00
Fevereiro	11,00
Março	12,00
Abril	10,00
Mai	8,00
Junho	7,00



Mas faz sentido:

- Preço médio do semestre = R\$ 9,66
- ou
- Preço mínimo = R\$ 7,00

# Modelagem Dimensional

- Metodologia de Projeto de BD Dimensional (Kimball):
  1. Selecione a atividade de negócio a ser modelada
    - a) Não se refere a uma função ou departamento da organização
    - b) Vários fluxos de dados direcionados para modelos de dados separados geram inconsistências
    - c) Primeiro modelo dimensional deve ser o mais simples
  2. Defina a granularidade da atividade de negócio
    - a) Definir exatamente o que uma linha da tabela de fatos representa.
    - b) Modelos dimensionais devem representar informação atômica
  3. Escolha as dimensões que se aplicam a cada linha da tabela de fatos
  4. Identifique os fatos numéricos que popularão cada linha da tabela de fatos

# Modelagem Dimensional

- DW sempre têm uma dimensão temporal
  - Existem várias funcionalidades de cálculo de datas não previstas por SQL
    - Períodos fiscais
    - Estações do ano
    - Feriados
    - Finais de semana
    - Eventos especiais
      - Carnaval
      - Páscoa
      - São João
      - Natal

Dimensão Data
DataKey (PK)
Data
Dia da Semana
IndicadorUltimoDiaSemana
IndicadorUltimoDiaMes
Mês
Ano
SemanaFiscal
NumSemanaFiscalAno
MesFiscal
NumMesFiscalAno
TrimestreFiscal
AnoFiscal
IndicadorFeriado
IndicadorFinalSemana

# Modelagem Dimensional

- Exemplo de uma Tabela de Dimensão Temporal

Key	Data	Descrição	Dia da Semana	Mes	Ano	Ano Fiscal	Feriado	Dia da Semana
1	1/1/02	1 de Jan/02	Terça	Jan	2002	F02/01	Feriado	Semana
2	2/1/02	2 de Jan/02	Quarta	Jan	2002	F02/01	Não	Semana
3	3/1/02	3 de Jan/02	Quinta	Jan	2002	F02/01	Não	Semana
4	4/1/02	4 de Jan/02	Sexta	Jan	2002	F02/01	Não	Semana
5	5/1/02	5 de Jan/02	Sábado	Jan	2002	F02/01	Não	Final
6	6/1/02	6 de Jan/02	Domingo	Jan	2002	F02/01	Não	Final
7	7/1/02	7 de Jan/02	Segunda	Jan	2002	F02/01	Não	Semana
8	8/1/02	8 de Jan/02	Terça	Jan	2002	F02/01	Não	Semana

# Modelagem Dimensional

- Exemplo de uma Tabela de Dimensão Temporal

Key	Data	Descrição	Dia da Semana	Mes	Ano	Ano Fiscal	Feriado	Dia da Semana
1	1/1/02	1 de Jan/02	Terça	Jan	2002	F02/01	Feriado	Semana
2	2/1/02	2 de Jan/02	Quarta	Jan	2002	F02/01	Não	Semana
3	3/1/02	3 de Jan/02	Quinta	Jan	2002	F02/01	Não	Semana
4	4/1/02	4 de Jan/02	Sexta	Jan	2002	F02/01	Não	Semana
5	5/1/02	5 de Jan/02	Sábado	Jan	2002	F02/01	Não	Final
6	6/1/02	6 de Jan/02	Domingo	Jan	2002	F02/01	Não	Final
7	7/1/02	7 de Jan/02	Segunda	Jan	2002	F02/01	Não	Semana
8	8/1/02	8 de Jan/02	Terça	Jan	2002	F02/01	Não	Semana

# Modelagem Dimensional

- Dimensão Causal
  - Descreve fatos que causam mudanças em outros fatos

Dimensão Promoção
PromocaoID (PK)
NomePromocao
TipoReducaoPreço
TipoPropaganda
TipoDisplay
TipoCupom
Custo
DataInicial
DataFinal

Valores de chave nulos não são usados na **tabela de fatos!**

Uma tupla na dimensão deve indicar que ela não é aplicável a uma dada medida.

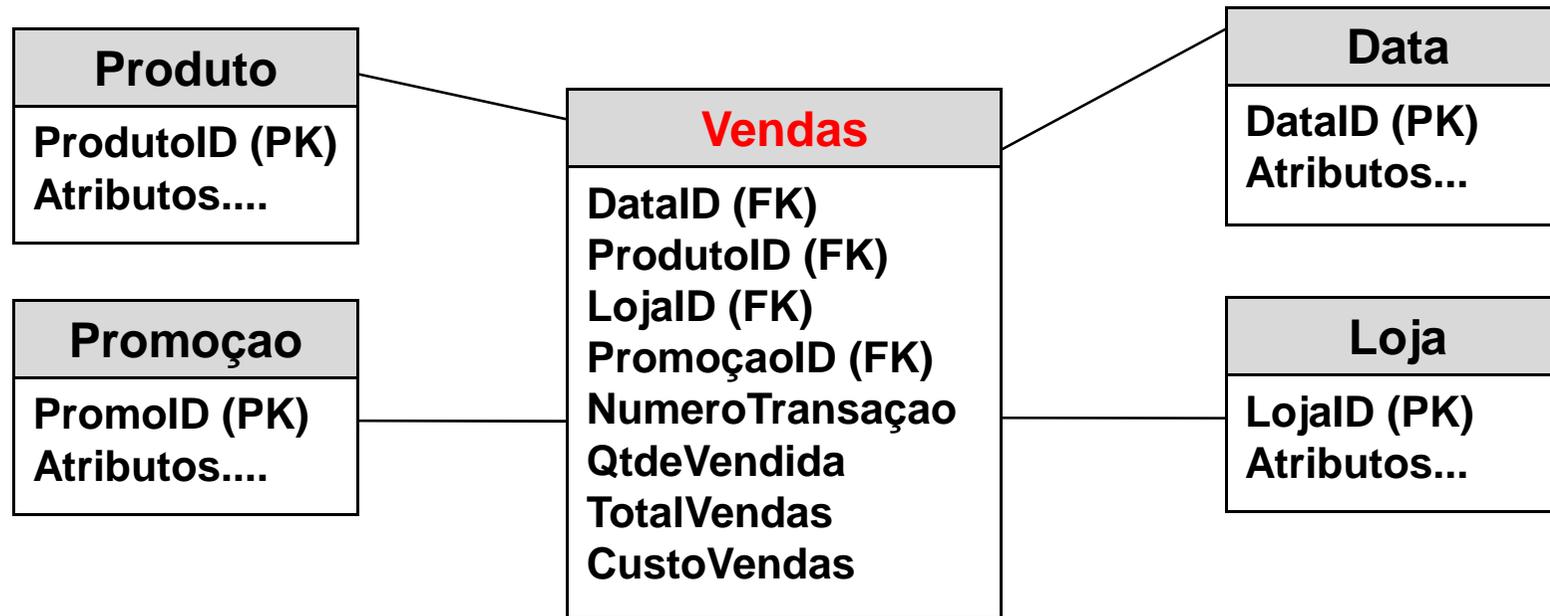
Como existem 4 tipos distintos de promoção, então seria melhor ter 4 dimensões distintas?

# Modelagem Dimensional

- Dimensão Causal
  - Vantagens do uso de uma única dimensão
    - Tamanho da dimensão unificada não é muito maior do que cada uma das visões separadas
    - Dimensão unificada facilita a busca sobre os vários tipos de causas
      - Informação sobre quais dados foram afetados pelos vários tipos de causa não é obtida
  - Vantagens do uso de várias dimensões
    - Se usuários pensam sobre as causas separadamente, a visão separada pode ter maior legibilidade
    - Administração de visões separadas pode ser facilitada

# Modelagem Dimensional

- Esquema estrela da atividade *Vendas*

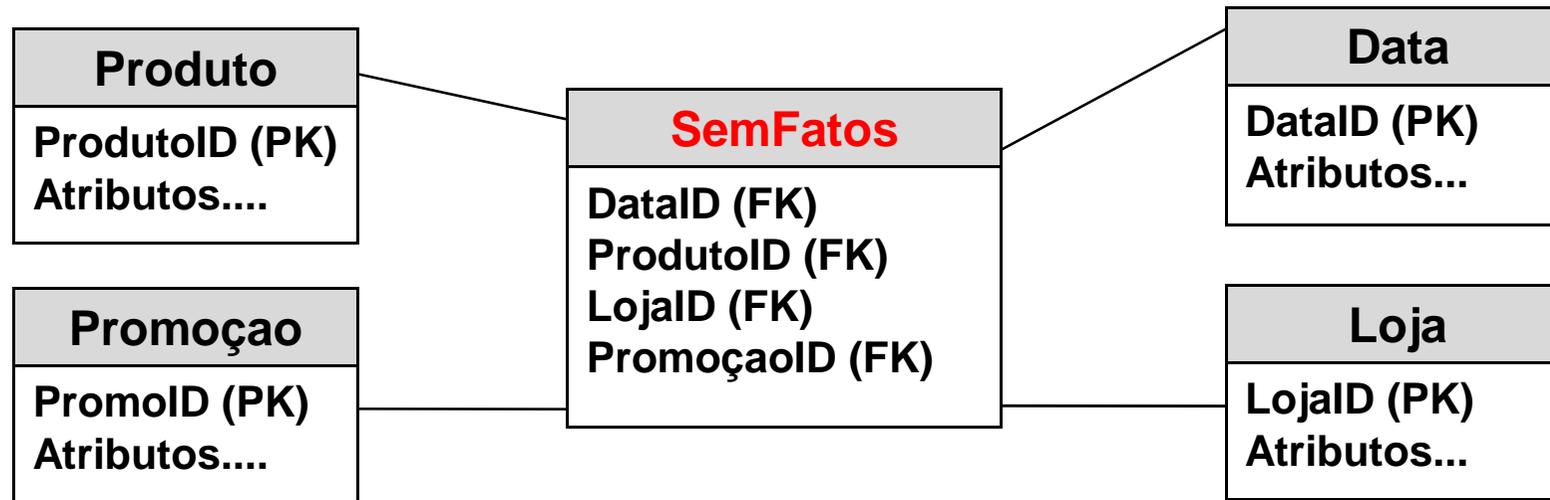


Quais produtos da promoção que não foram vendidos?

Tabela de Fatos Sem Fatos

# Modelagem Dimensional

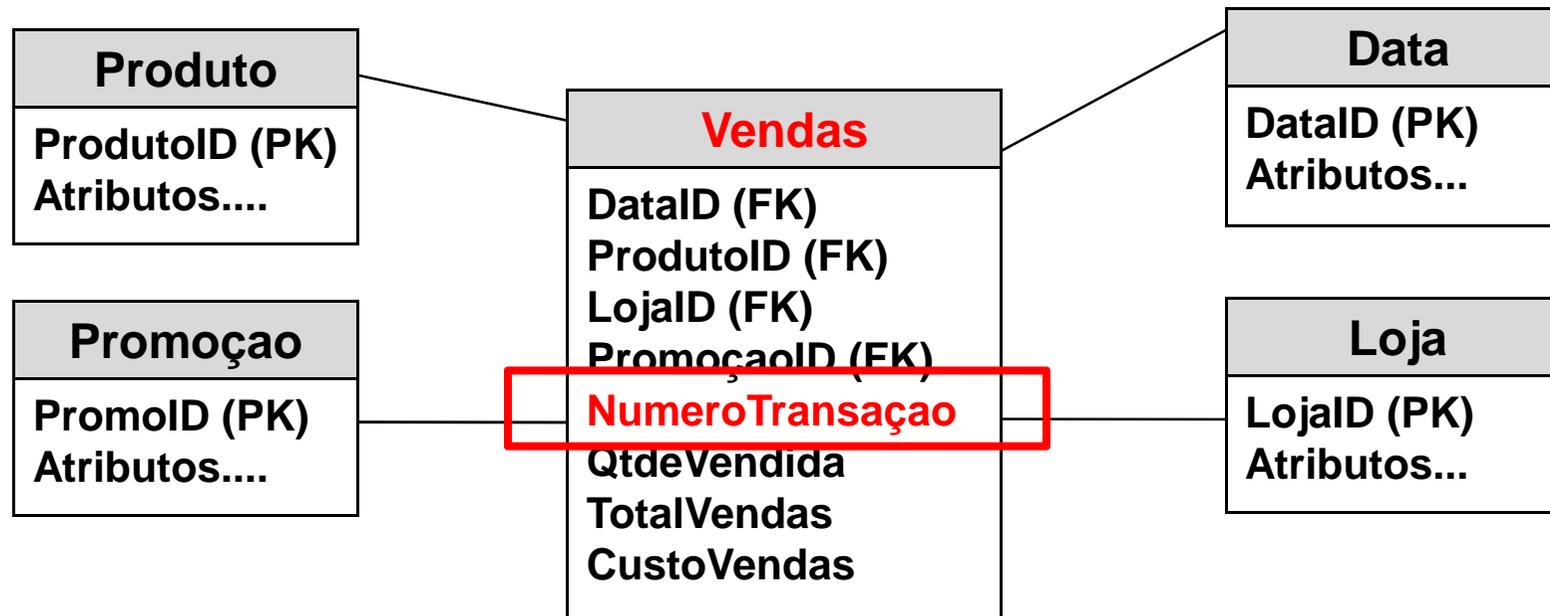
- Esquema estrela da atividade *Promoção sem Vendas*



- Para determinar quais produtos da promoção não venderam
  - Consulta à tabela **SemFatos** para determinar quais produtos da promoção em um dado dia
  - Consulta à tabela **Vendas** para determinar quais produtos foram vendidos no mesmo dia
  - Obter a diferença entre as duas listas de produtos

# Modelagem Dimensional

- Dimensão Degenerada



- É útil para agrupar todos os produtos vendidos em uma única transação
- É útil para identificar cada tupla da tabela de fatos

Chave = NumeroTransação + ProdutoID

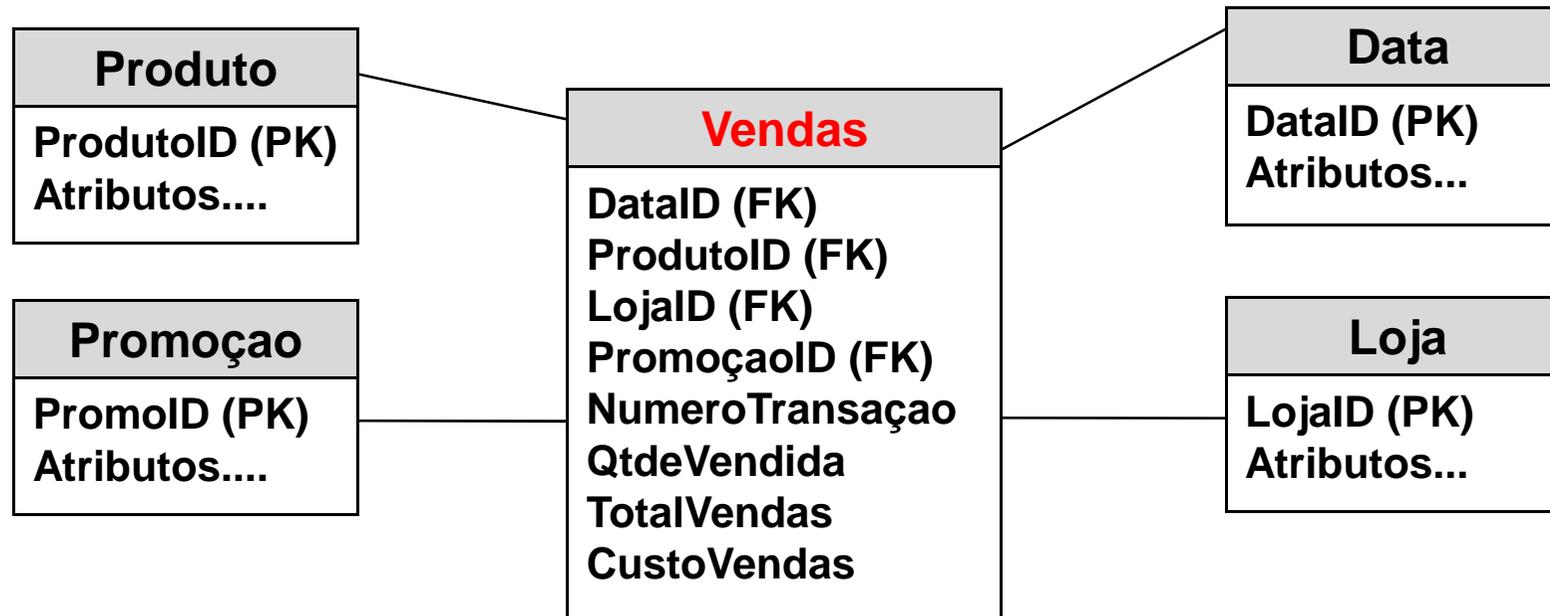
- Dimensão Degenerada
  - Consiste em uma dimensão vazia
    - Contém apenas uma chave que é incluída na tabela de fatos
  - Números de controle do ambiente operacional são geralmente exemplos deste tipo de dimensão
    - Número de Pedido
    - Número de Fatura
  - Se atributos adicionais forem identificados no futuro
    - E eles forem relacionados com a dimensão degenerada
    - Uma nova dimensão normal pode ser criada e a dimensão degenerada removida

# Modelagem Dimensional

- Modelos Dimensionais
  - Acomodam mudanças mais facilmente
    - Novas dimensões podem ser adicionadas
      - Para cada linha existente na tabela de fatos, um valor único da nova dimensão é atribuído
    - Novas medidas podem ser adicionadas à tabela de fatos
      - Assumindo que nível de detalhes seja consistente
    - Dimensões podem:
      - Receber novos atributos
      - Aumentar o nível de detalhes da hierarquia, quebrando os atributos existentes

# Modelagem Dimensional

- Estendendo o esquema da atividade *Vendas*



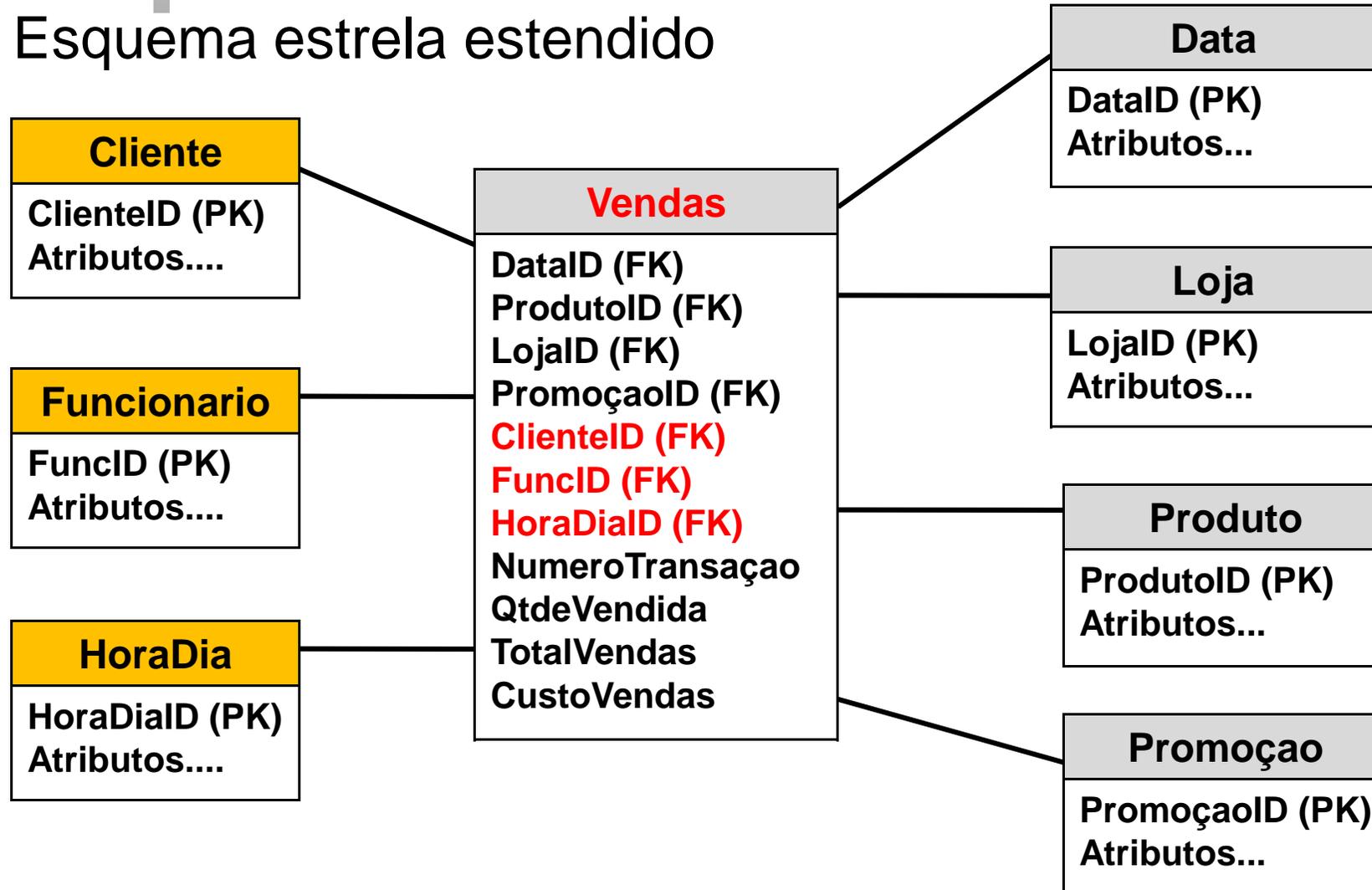
Como estender este esquema para permitir análises sobre padrões de compra de clientes cadastrados?

# Modelagem Dimensional

- Estendendo o esquema com dados sobre Clientes
  - Criar a dimensão cliente
  - Adicionar uma nova chave estrangeira na tabela de fatos
  - Nem toda compra estará associada a um cliente cadastrado
    - Criar um cliente do tipo **não cadastrado** na nova dimensão
  - Criar um cliente do tipo **anterior ao cadastramento** de clientes para associá-los aos fatos já existentes
  - Criar as dimensões Funcionário e Hora do Dia e as respectivas chaves estrangeiras na tabela de fatos
- Dimensões e fatos existentes permanecem inalterados
  - Esquema original foi modelado no maior nível de detalhe

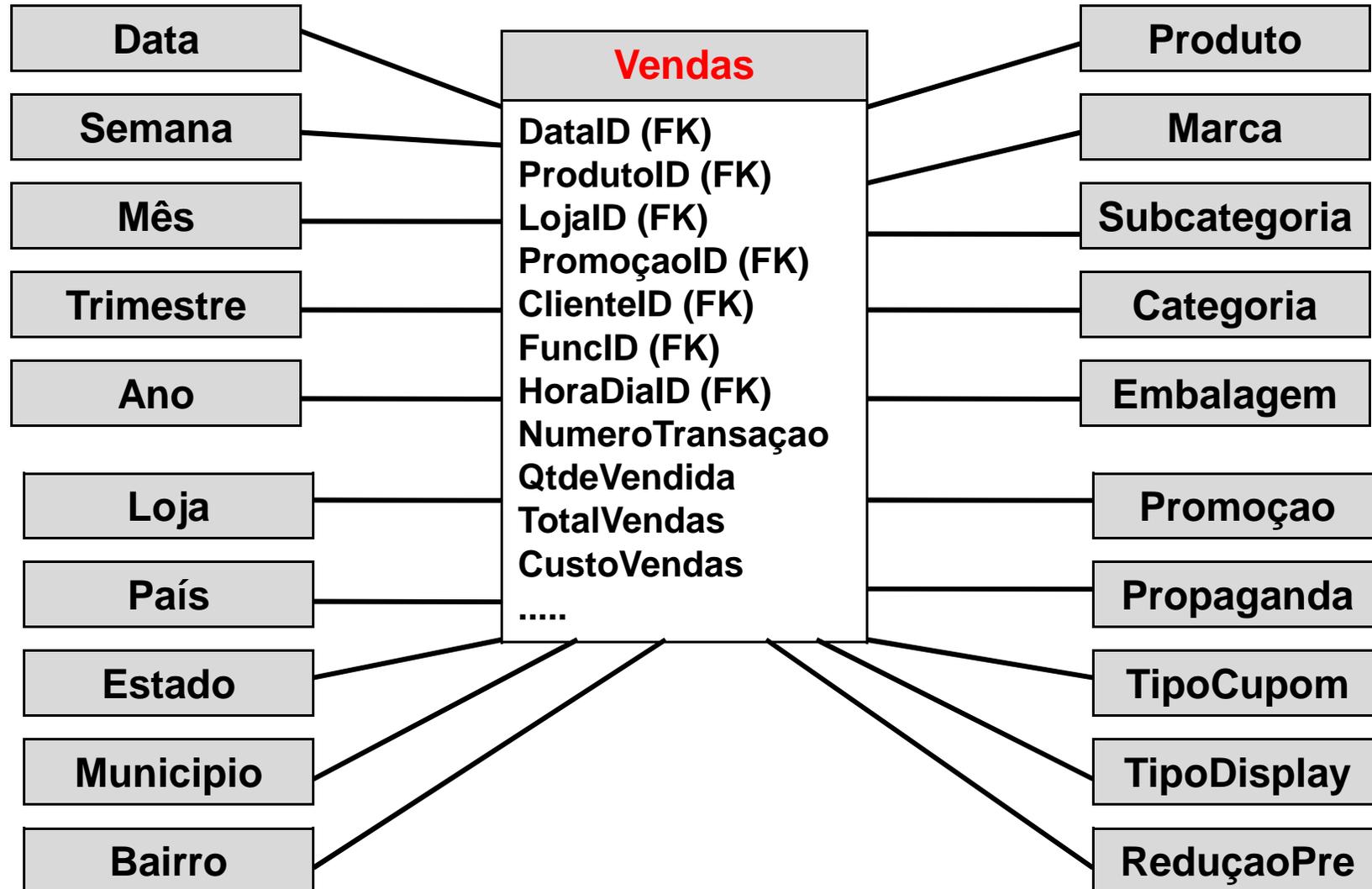
# Modelagem Dimensional

- Esquema estrela estendido



# Modelagem Dimensional

- Muitas Dimensões



# Modelagem Dimensional

- Muitas Dimensões
  - Um grande número de dimensões no esquema é uma indicação que:
    - Várias dimensões não são completamente independentes
    - Elas devem ser combinadas em uma única dimensão
  - Consiste em um erro de modelagem representar elementos de uma hierarquia como dimensões separadas
  - Dimensões correlacionadas de um esquema com várias dimensões devem ser combinadas

# Modelagem Dimensional

- Mudanças em Dimensões
  - Pensava-se que dimensões eram independentes do tempo
  - Mudanças ocorrem no mundo real
    - Linhas de produto são reestruturadas, causando mudanças em hierarquias de produto
    - Clientes se mudam, causando mudanças em seus formatos de endereços
  - Não dá para registrar todas mudanças como fatos
  - Existem quatro abordagens:
    - Sobrescrita do valor
    - Adição de uma tupla
    - Adição de uma coluna
    - Abordagem híbrida

# Modelagem Dimensional

- Sobrescrita do Valor

Chave Produto	Descrição	Departamento	Código Produto
12345	BrushKids	Higiene Pessoal	ABC922-Z



Chave Produto	Descrição	Departamento	Código Produto
12345	BrushKids	Infantil	ABC922-Z

- Não há mudanças em chaves de qualquer tabela
- É a abordagem mais simples, tendo como vantagens a rapidez e facilidade de implementação
- Não mantém o histórico dos valores anteriores
- É adequado para realização de correções em carga de dados

# Modelagem Dimensional

- Adição de uma tupla

Chave Produto	Descrição	Departamento	Código Produto
12345	BrushKids	Higiene Pessoal	ABC922-Z
25984	BrushKids	Infantil	ABC922-Z

- Uma nova tupla na tabela de dimensão é inserida para refletir a mudança no valor do atributo.
- Tabela de fatos não é modificada
- Histórico das modificações é mantido
  - Fatos ocorridos antes da mudança estão associados ao valor do atributo anterior
- Aumenta o tamanho da tabela de dimensões
- Não permite associações entre o novo valor de atributo e fatos antigos e vice-versa

# Modelagem Dimensional

- Adição de uma coluna

Chave Produto	Descrição	Novo Depto	Antigo Depto	Codigo
12345	BrushKids	Infantil	Higiene Pessoal	ABC922-Z

- Uma nova coluna na tabela de dimensão é inserida para refletir a mudança no valor do atributo.
- Fatos antigos e novos podem ser sumarizados em termos dos antigos e novos valores de atributo
- Difere da abordagem anterior porque tanto valores antigos como novos do atributo podem ser considerados
- Não é frequentemente usada
- Não permite análise sobre o impacto das inúmeras mudanças feitas ao valor do atributo

# Modelagem Dimensional

- Abordagem híbrida

Chave	Descrição	Atual Depto	Historico Depto	Codigo
12345	BrushKids	Higiene Pessoal	Higiene Pessoal	ABC922-Z

# Modelagem Dimensional

- Abordagem híbrida

Chave	Descrição	Atual Depto	Historico Depto	Codigo
12345	BrushKids	Infantil	Higiene Pessoal	ABC922-Z
25984	BrushKids	Infantil	Infantil	ABC922-Z

# Modelagem Dimensional

- Abordagem híbrida

Chave	Descrição	Atual Depto	Historico Depto	Codigo
12345	BrushKids	Educacional	Higiene Pessoal	ABC922-Z
25984	BrushKids	Educacional	Infantil	ABC922-Z
31726	BrushKids	Educacional	Educacional	ABC922-Z

- Permite registrar o histórico das mudanças
- Permite sumarizar fatos de acordo com qualquer valor do atributo modificado
- As abordagens vistas anteriormente são apropriadas para *slowly changing dimensions* apenas!
  - Isto pode ser uma indicação que o levantamento de requisitos não foi bem realizado
- Existem outras abordagens na literatura de DW

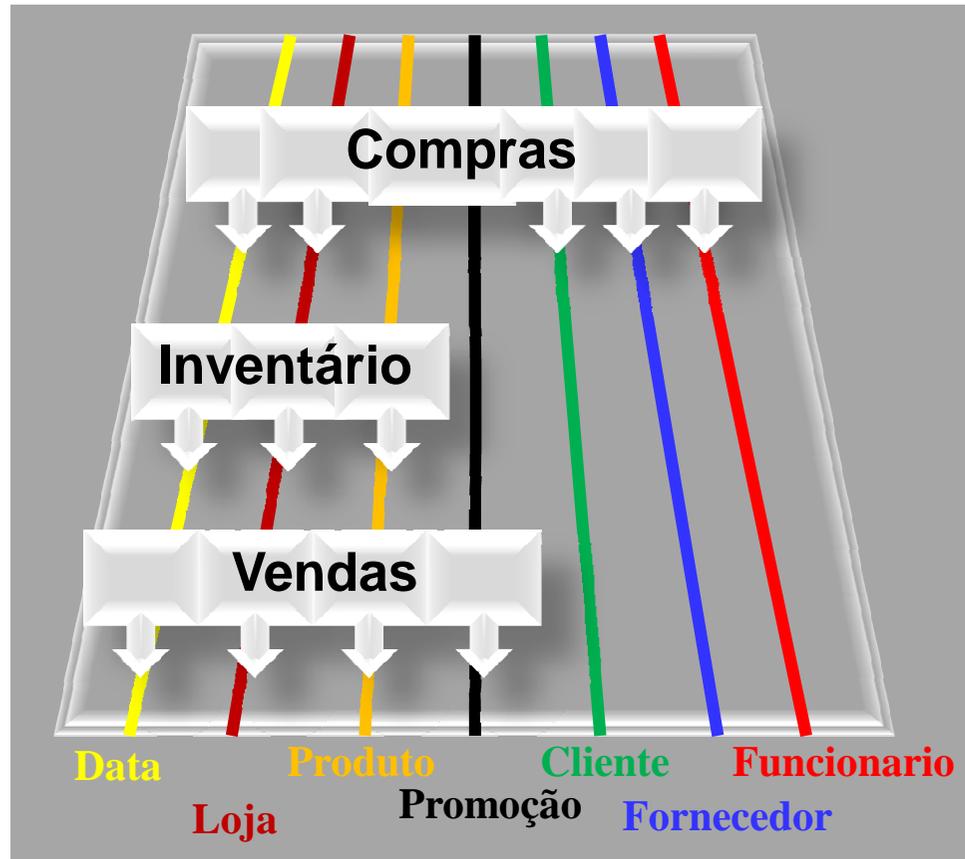
# Modelagem Dimensional

- Mitos da Modelagem Dimensional (Kimball):
  1. Modelos dimensionais e DM são para dados sumarizados apenas
  2. Modelos dimensionais e DM são soluções departamentais, não organizacionais
  3. Modelos dimensionais e DM não são escaláveis
  4. Modelos dimensionais e DM são apenas adequados quando existe um padrão de uso previsível
  5. Modelos dimensionais e DM não podem ser integrados

# Arquitetura de DW (DW bus)

- Integração de dados é essencial
  - Geração de relatórios consistentes a partir de diferentes DM
  - Aumento de desempenho no desenvolvimento do sistema a partir do uso de componentes reusáveis
- Tabelas de fatos mantidas em DM separados representam várias atividades de negócio
- Uso de dimensões compartilhadas permite a integração dos DM
  - Medidas de desempenho de várias atividades de negócio podem ser exibidas em um único relatório
  - *Drill Across* permite a combinação de resultados de consultas feitas à DM distintos por meio do atributo comum da dimensão compartilhada

# Arquitetura de DW (DW bus)



- Consiste em um conjunto de modelos dimensionais que compartilham um conjunto de dimensões padronizadas
- Possui várias atividades de negócio acopladas

# Arquitetura de DW (DW bus)

- Matriz DW Bus
  - Usada para criar, documentar e divulgar a arquitetura de DW

Atividades de Negócio	Dimensões Padronizadas					
	Data	Produto	Loja	Promoção	Vendedor	Fornecedor
Vendas	X	X	X	X		
Inventário	X	X	X			
Entrega	X	X	X			
Compras	X	X			X	X

- Colunas são dimensões compartilhadas e linhas são DM
- Possibilita a visualização de quais dimensões merecem atenção especial por participarem de vários DM

# Arquitetura de DW (DW bus)

- Atividades de projeto da arquitetura de DW
  - Listar todas as dimensões usadas pela organização
  - Listar DM de **primeiro nível** que possuem menos dimensões e muitas delas não são compartilhadas
  - Listar DM **consolidados** que possuem um número maior de dimensões e muitas são compartilhadas
    - Lucro é um exemplo de DM consolidado o qual pode compartilhar dimensões com outros DM (rendimentos e custos)
    - Mantêm dados de vários DM de primeiro nível, mas podem também conter informações específicas

# Arquitetura de DW (DW bus)

- Dimensões Padronizadas
  - São dimensões idênticas ou uma é um subconjunto de uma outra dimensão mais detalhada
  - Possuem propriedades consistentes que incluem:
    - Chaves de dimensão
    - Definições de atributos (i.e. valores e nomes iguais)
      - Agrupamentos consistentes
      - Mesmos rótulos usados nos relatórios
  - Uma dimensão padronizada pode ser a mesma tabela física do BD
    - Dado o ambiente complexo de DW com muitas plataformas de BD, elas podem ser duplicadas e sincronizadas em cada DM

# Arquitetura de DW (DW bus)

- Dimensões padronizadas possuem:
  - Conteúdo de dados
  - Interpretação de dados
  - Formatos de exibição } Consistentes
- Dimensões padronizadas:
  - São definidas no maior nível de detalhe possível
  - Podem pertencer a esquemas de atividades de negócio definidas em granularidades diferentes
    - Basta que uma dimensão seja um subconjunto da outra
    - Antes de seus esquemas serem comparados, uma delas é sumarizada para o mesmo nível de detalhe da outra

# Arquitetura de DW (DW bus)

- Exemplo:

Vendas ⇒ captura dados no nível de produto atômico

Previsão ⇒ gera dados no nível marca de produto

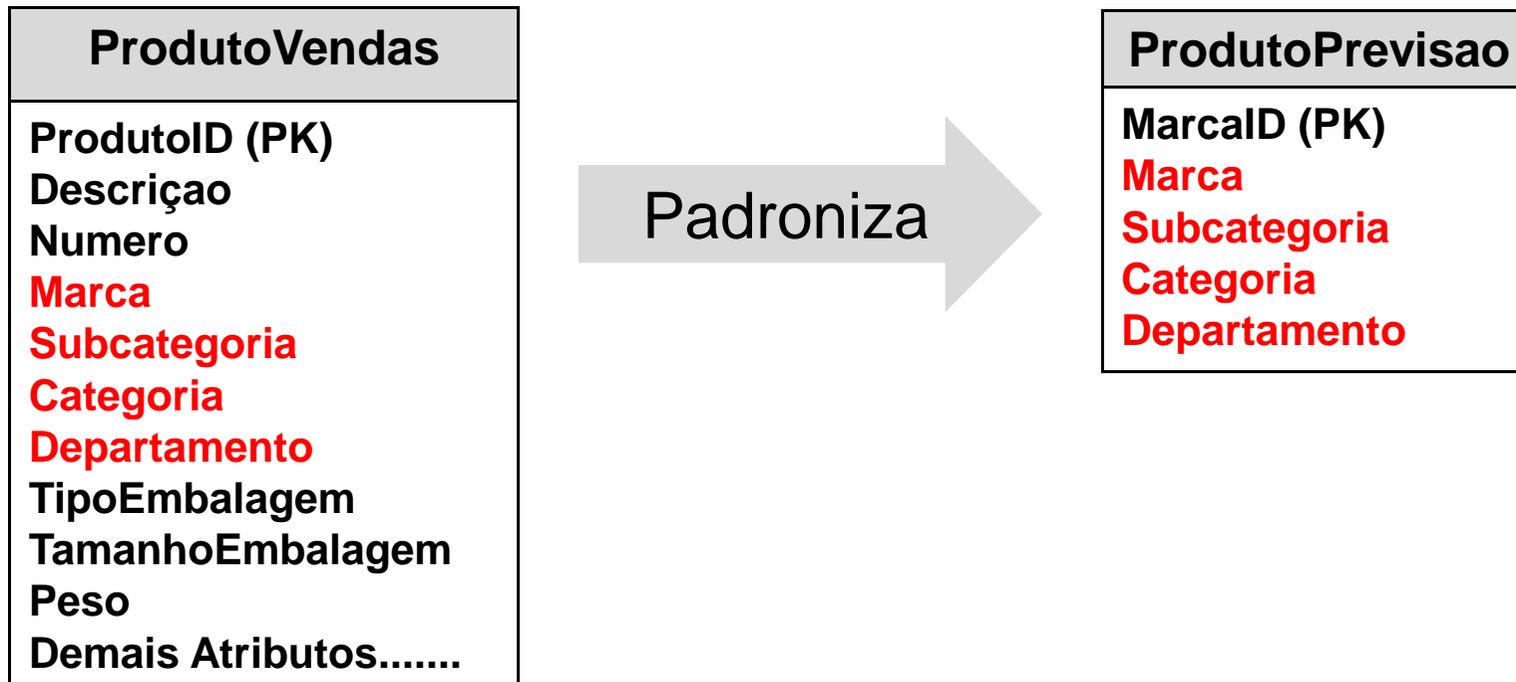
A tabela de dimensão Produto só pode ser compartilhada entre os esquemas destas 2 atividades de negócio se:

A tabela Produto do esquema Previsão for um subconjunto da tabela Produto do esquema Vendas!

Atributos comuns de ambas tabelas (i.e. marca e categoria) devem ser rotulados, definidos e instanciados identicamente!

# Arquitetura de DW (DW bus)

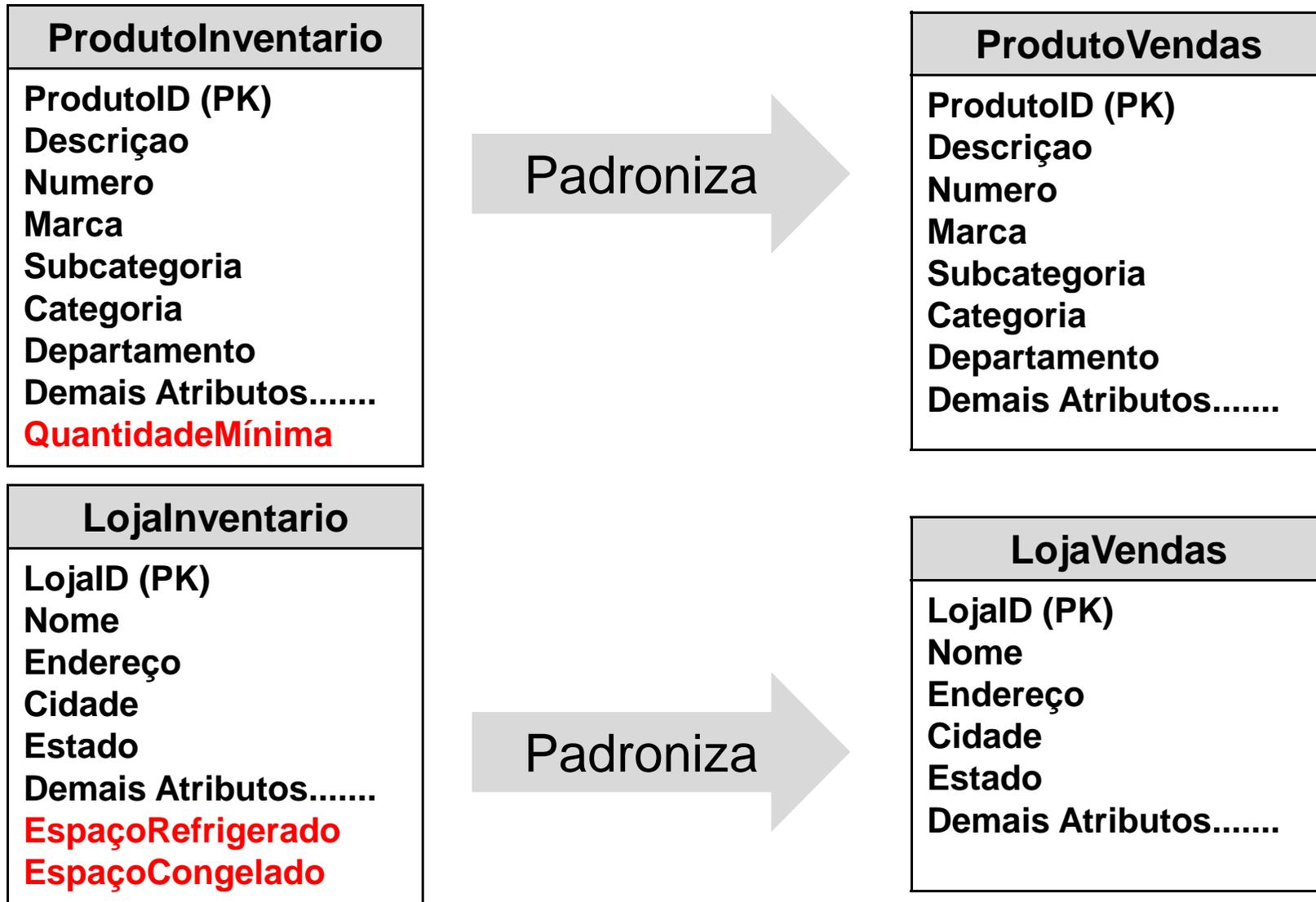
- Exemplo de dimensões padronizadas



- Operações DrillAcross só podem ser realizadas por meio de:
  - atributos comuns das tabelas de dimensões padronizadas
  - tuplas comuns das tabelas para não violar a restrição de integridade referencial

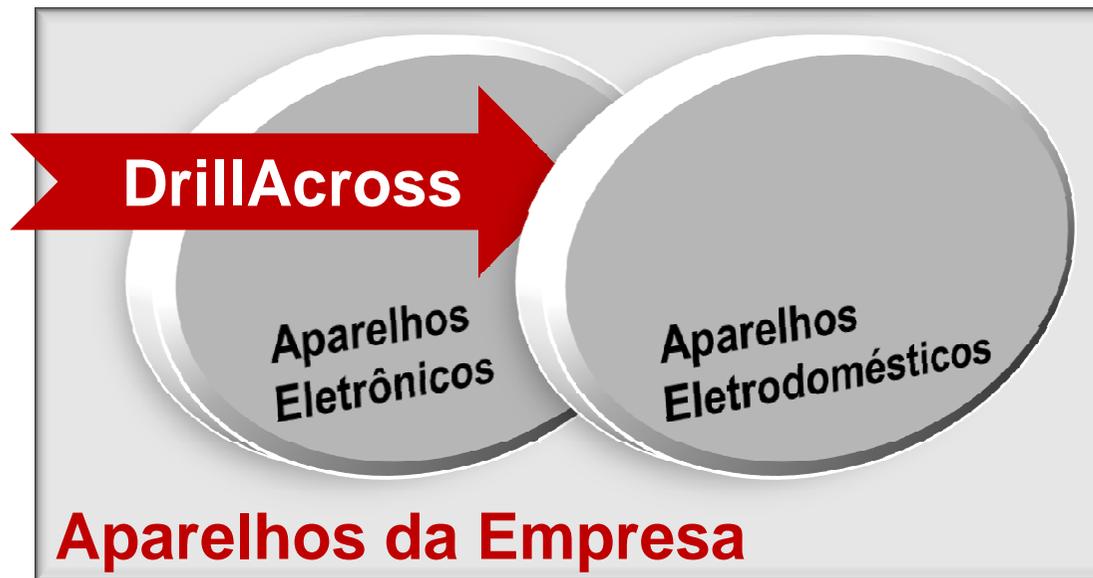
# Arquitetura de DW (DW bus)

- Outro exemplo de dimensões padronizadas



# Arquitetura de DW (DW bus)

- Outro exemplo de dimensões padronizadas ocorre:
  - Elas estão no mesmo nível de detalhes
  - Mas uma possui um subconjunto de tuplas da outra



- Dimensões padronizadas podem ser replicadas logicamente ou fisicamente dentro da empresa
  - Elas são construídas apenas uma vez em cada ETL

# Arquitetura de DW (DW bus)

- Uso de dimensões padronizadas requer
  - Coordenação de implementação para que adição ou modificação de atributos seja revista
  - Disciplina na nomeação do dados
  - Estratégia de liberação da dimensão para área de apresentação
    - Dimensões idênticas devem ser replicadas de forma síncrona para todos DM associados
  - Não é sempre o caso que o uso da arquitetura DW bus é aplicável
    - Pode não haver necessidade de relacionar dados de uma atividade de negócio com dados de outra
    - DW separados podem ser construídos

# Arquitetura de DW (DW bus)

- Fatos padronizados
  - Possuem as mesmas definições e equações
    - Lucro, rendimentos, preços, custos, medidas de qualidade e medidas de satisfação do cliente
  - Em geral, a tabela de fatos não é duplicada explicitamente em vários DM
    - Se eles possuem o mesmo identificador então eles necessitam ter sido definidos
      - No mesmo contexto dimensional
      - Com base nas mesmas unidades de medida
    - Se não for possível padronizar os fatos então eles devem ser nomeados diferentemente para que possam ter diferentes interpretações

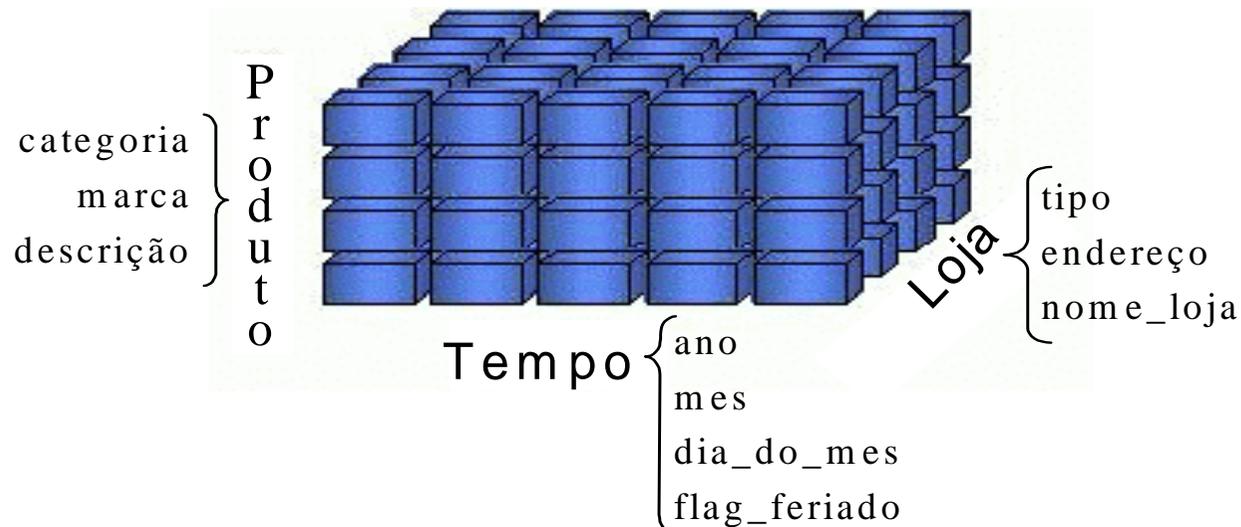
# Processamento OLAP

- O que é OLAP (*On-Line Analytical Processing*)?
  - Conjunto de tecnologias projetadas para analisar e acessar dados típicos de suporte a decisão que estão no DW
  - Fornece dados em alto nível (totais, médias, min..)
  - Acessa vários registros
  - Tem alto desempenho e consultas fáceis e interativas
  - Lida com dados históricos (dimensão temporal)
  - Oferece visões multidimensionais (perspectivas)

- O que é OLAP ? (Cont.)
  - Exemplos de consultas típicas de OLAP:
    - Quais os produtos mais bem vendidos no mês passado?
    - Quais os 10 piores vendedores dos departamentos da filial Recife?
    - Qual a média salarial dos funcionários de informática na região sul nos últimos 5 anos?

# Cubo Multidimensional

- Cubo Multidimensional (Abstração do DW)
  - Metáfora de uma abordagem multidimensional para visualização e organização dos dados
  - Várias dimensões podem ser usadas simultaneamente
  - Dados são manipulados mais rapidamente e facilmente (agregação em níveis de hierarquia)

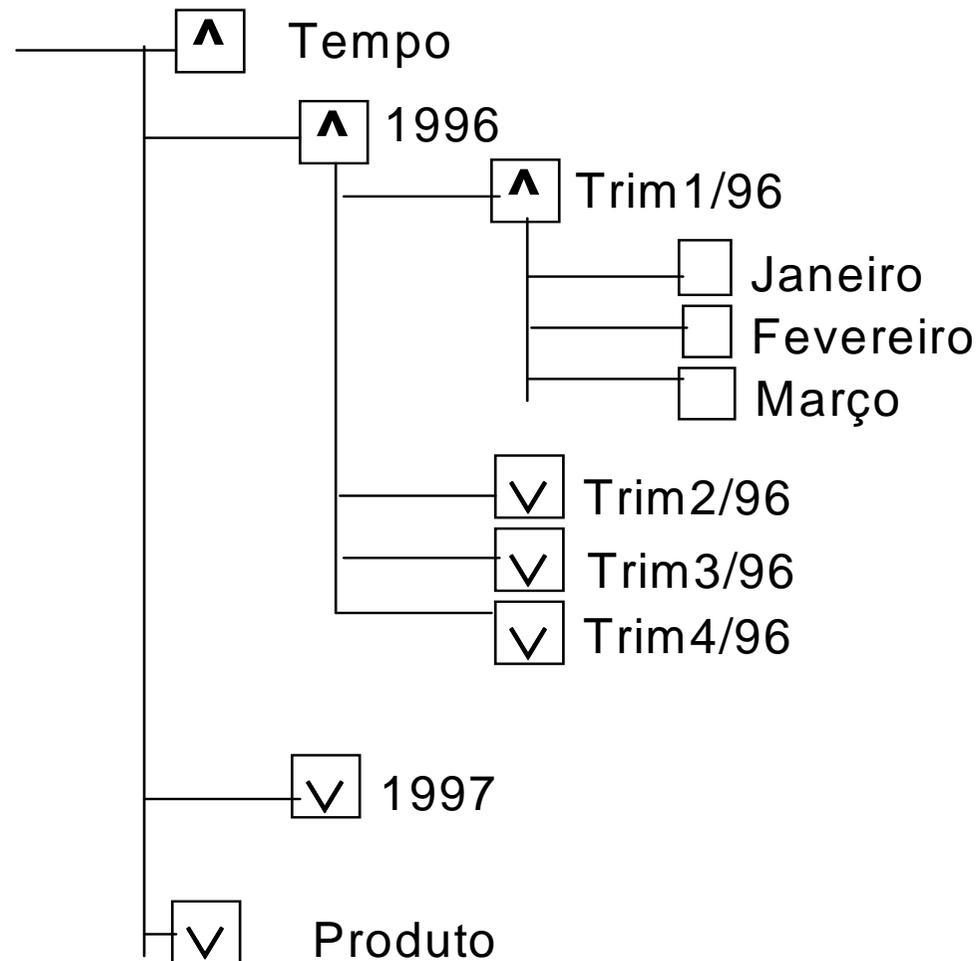


# Cubo Multidimensional

- Agregação em Níveis de Hierarquias

Dimensão Tempo

Chave_Tempo
Mes
Trimestre
Ano

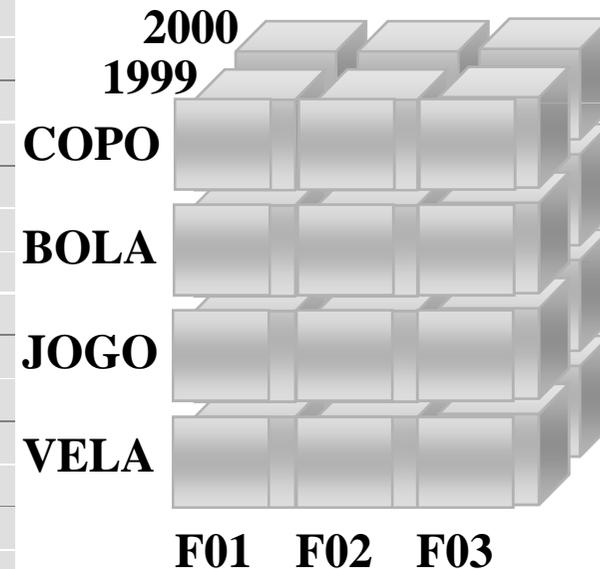


# Cubo Multidimensional

- Como representar as dimensões no cubo?

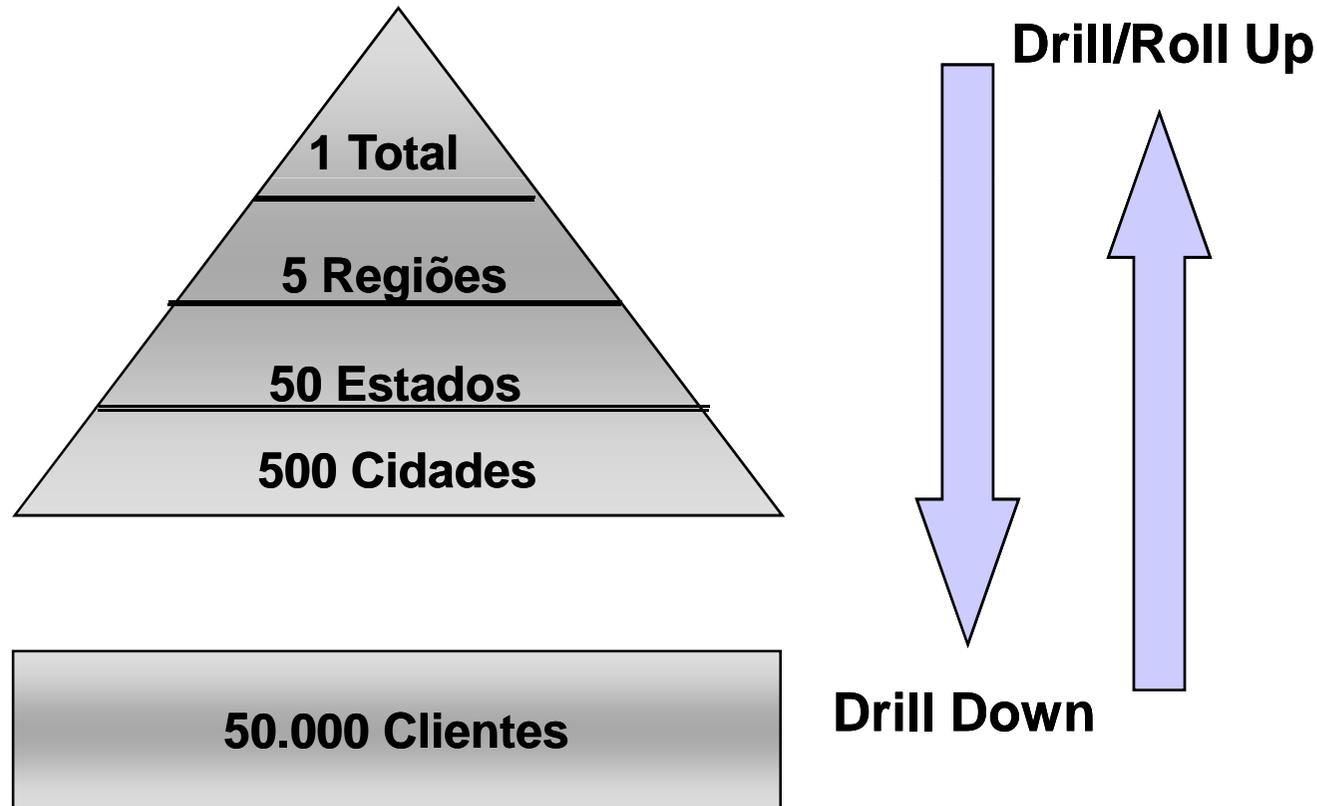
Produto	Loja	Vendas
COPO	F01	50
COPO	F02	60
COPO	F03	100
BOLA	F01	40
BOLA	F02	70
BOLA	F03	80
JOGO	F01	90
JOGO	F02	120
JOGO	F03	140
VELA	F01	20
VELA	F02	10
VELA	F03	30

Produto	Loja	Tempo	Vendas
COPO	F01	1999	50
COPO	F02	1999	60
COPO	F03	1999	100
BOLA	F01	1999	40
BOLA	F02	1999	70
BOLA	F03	1999	80
JOGO	F01	1999	90
JOGO	F02	1999	120
JOGO	F03	1999	140
VELA	F01	1999	20
VELA	F02	1999	10
VELA	F03	1999	30
COPO	F01	2000	50
COPO	F02	2000	60
COPO	F03	2000	100
BOLA	F01	2000	40
BOLA	F02	2000	70
BOLA	F03	2000	80
JOGO	F01	2000	90
JOGO	F02	2000	120
JOGO	F03	2000	140
VELA	F01	2000	20
VELA	F02	2000	10
VELA	F03	2000	30



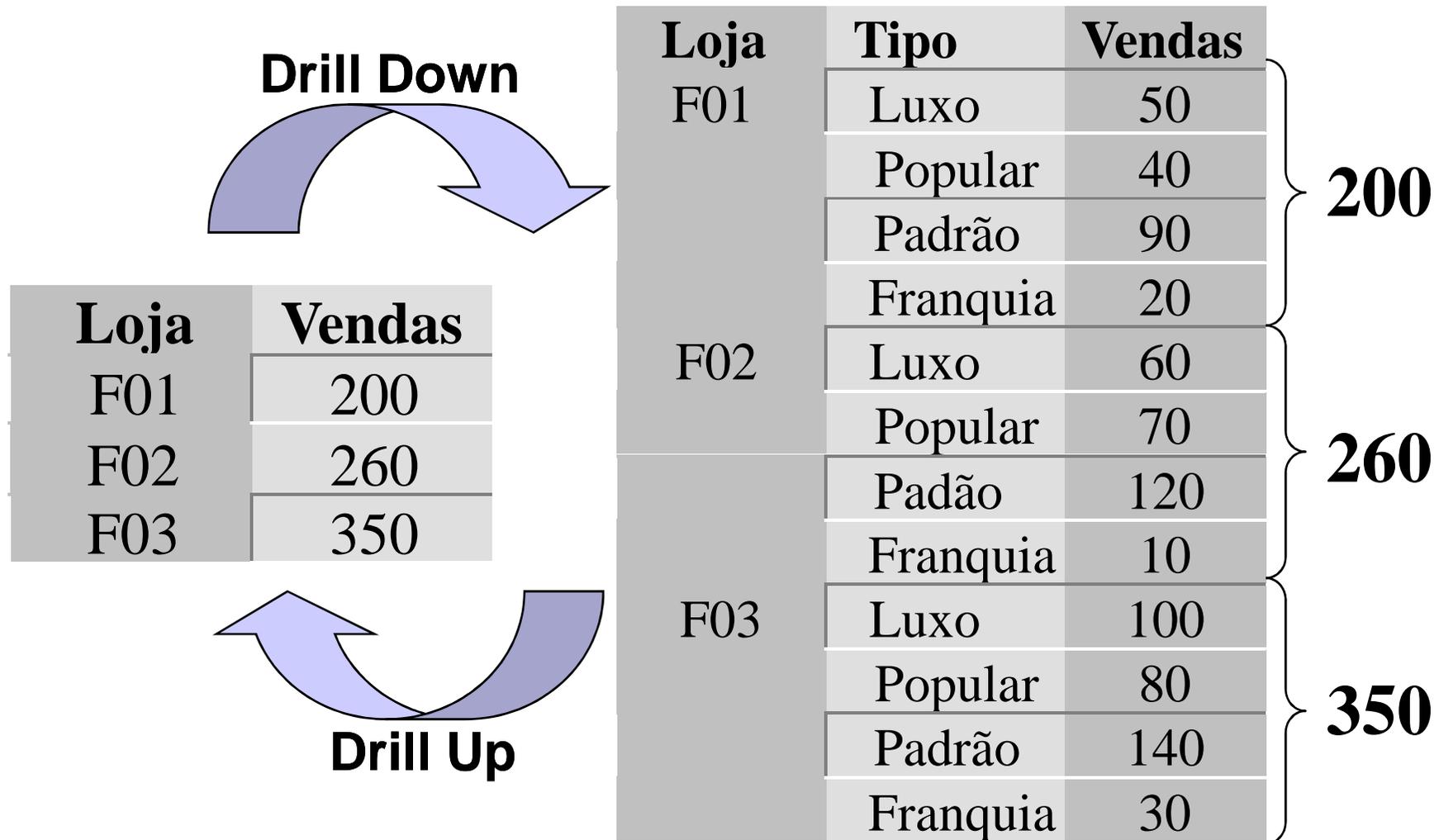
# Cubo Multidimensional

- Drill Down e Roll Up (ou Drill Up)
  - Técnica que possibilita caminhar pela estrutura multidimensional (hierarquias), permitindo ver diferentes níveis de dados



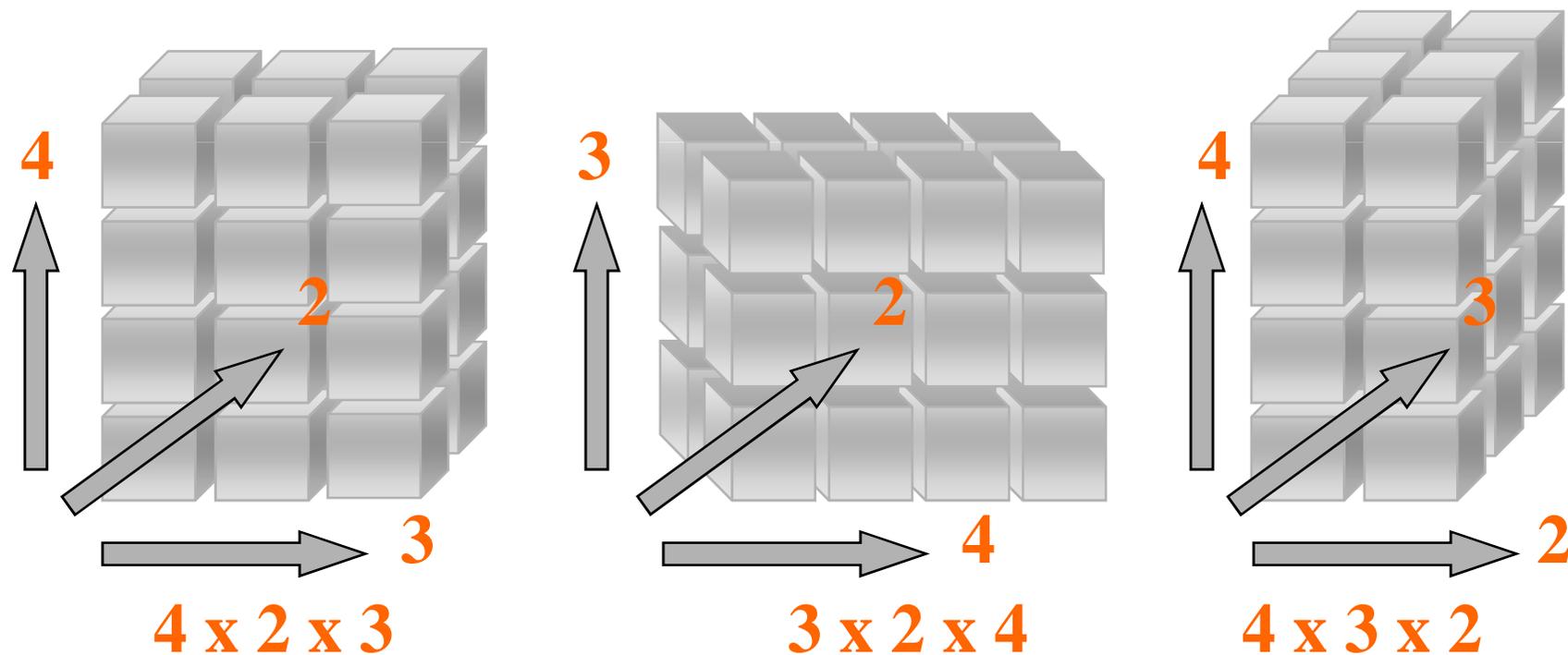
# Cubo Multidimensional

- Drill Down e Roll Up (ou Drill Up)



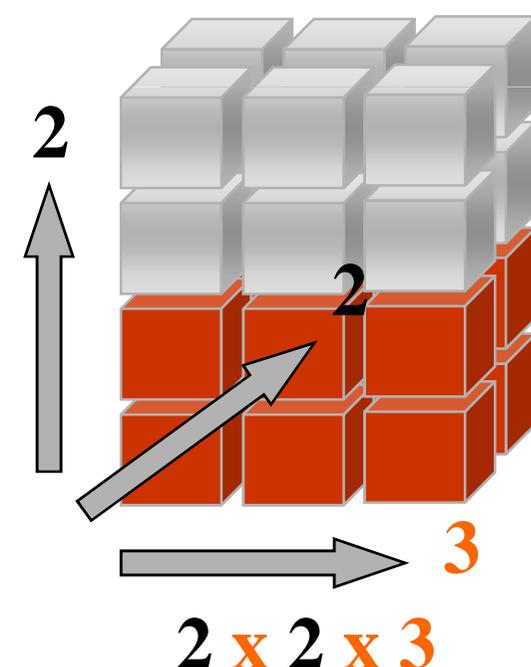
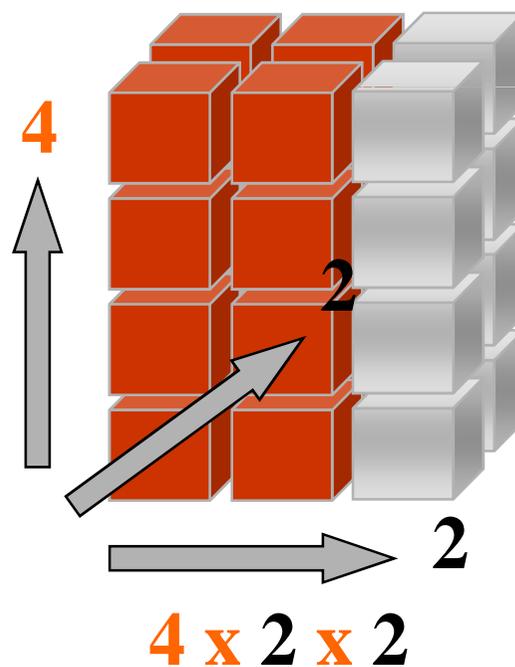
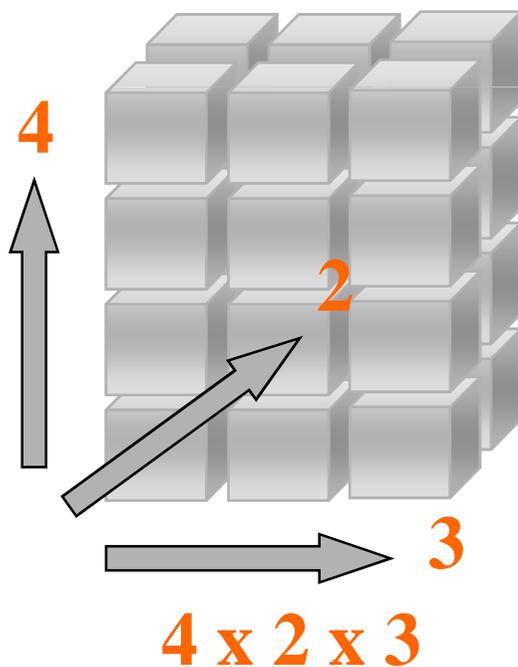
# Cubo Multidimensional

- Rotação
  - Técnica que gira o cubo, permitindo ter diferentes visões dos dados

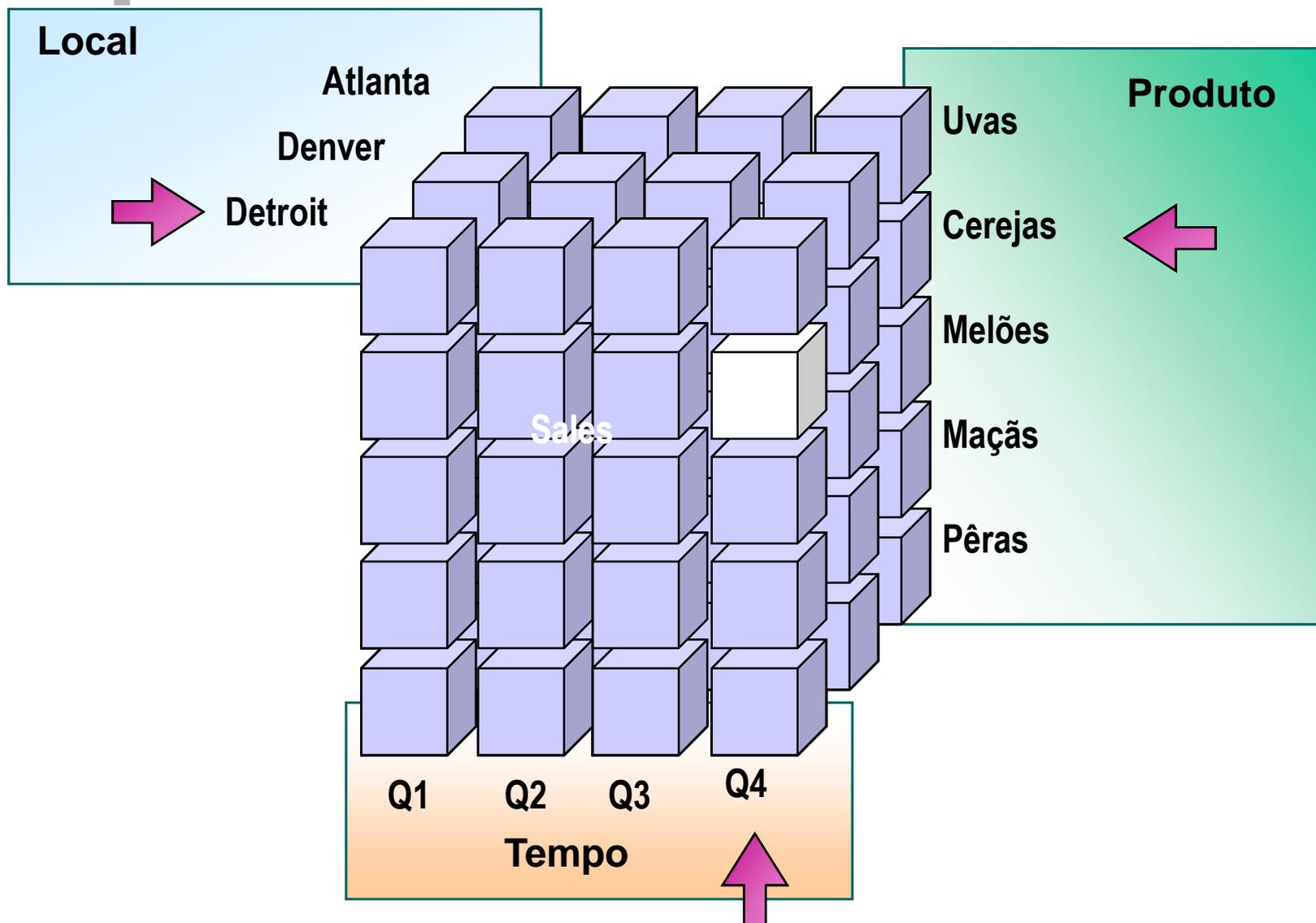


# Cubo Multidimensional

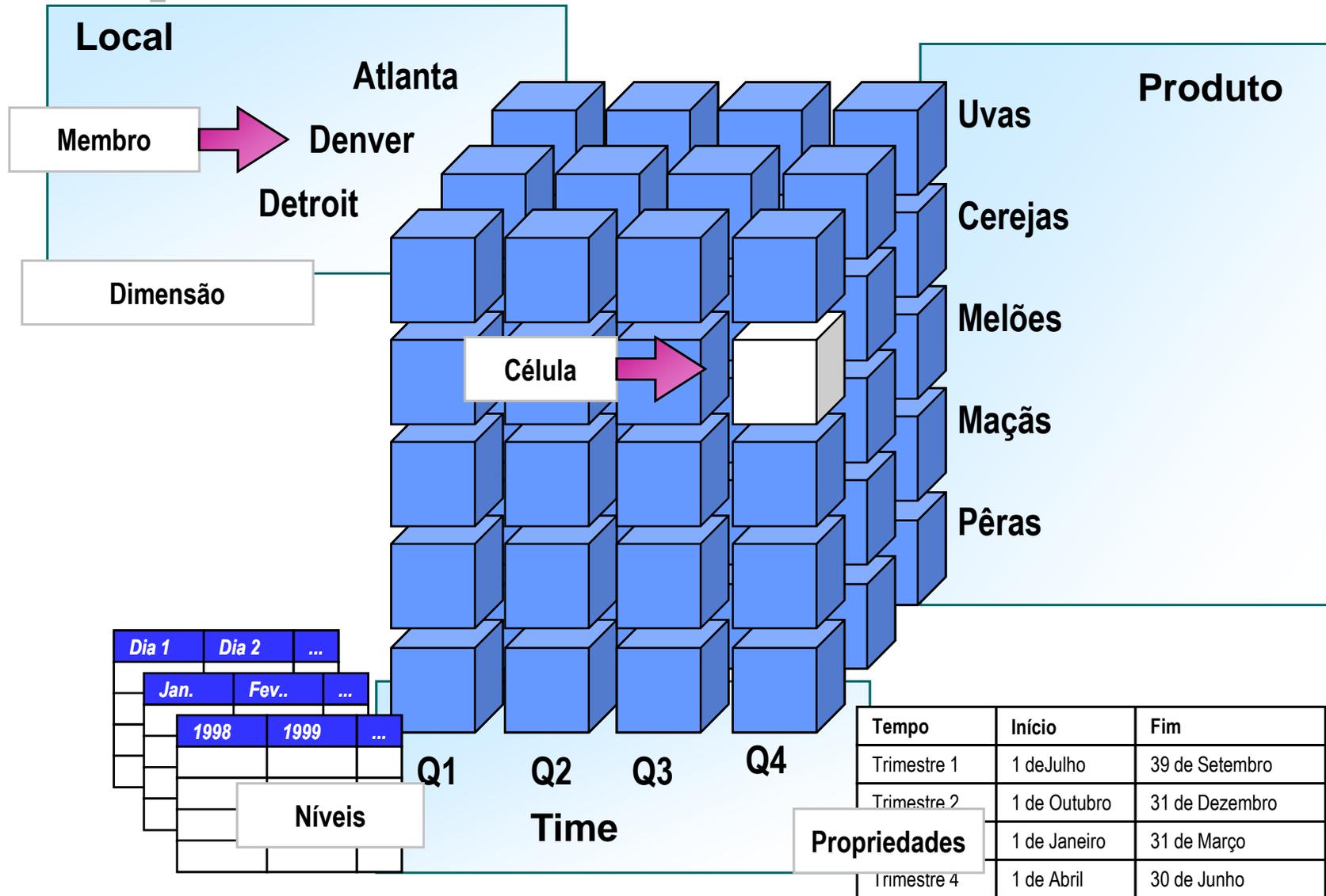
- Slice and Dice
  - Técnica que fatia o cubo, permitindo restringir a análise aos dados, **sem inversão** de eixos
  - “Semelhante” a cláusula WHERE de SQL



# Cubo Multidimensional



# Cubo Multidimensional



# Arquiteturas OLAP

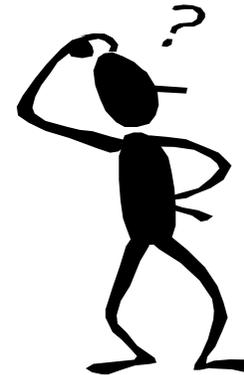
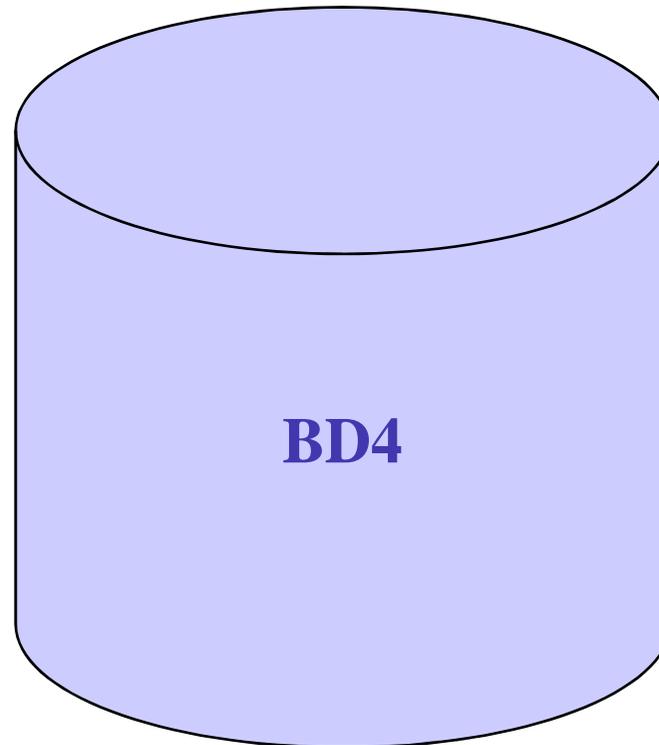
- Formas de armazenamento do cubo multidimensional
  - Relacional OLAP (ROLAP)
  - Multidimensional OLAP (MOLAP)
  - Híbrido OLAP (HOLAP)



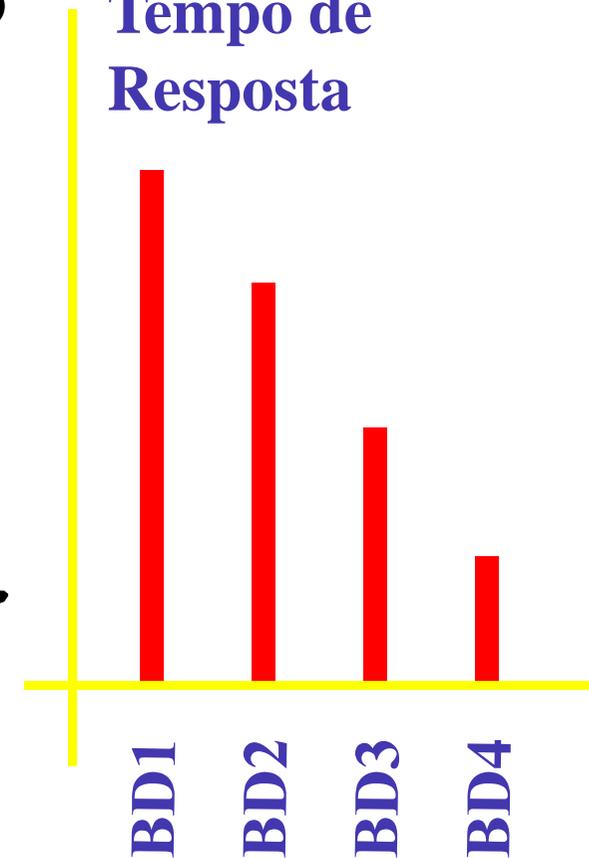
Definem as categorias das ferramentas OLAP!

# Desempenho x Armazenamento

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



**Tempo de Resposta**



**Quanto maior o número de agregados, melhor o desempenho e mais dados para serem armazenados e gerenciados!**

Produto	Região	Vendas no Mês	Comparação com o Mês Anterior
Pasta Colgate	Sul	110	**12%
Pasta Colgate	Sudeste	179	-3%
Pasta Colgate	Nordeste	55	5%
Total		344	**6%

Produto	Região	Tamanho	Vendas no Mês	Comp. Com o Mês Anterior
Colgate	Sul	A	34	**10%
Colgate	Sul	B	36	**13%
Colgate	Sul	C	40	**11%
Colgate	Total		110	**12%
Colgate	Sudeste	A	63	<b>-2.8%</b>
Colgate	Sudeste	B	60	<b>-3.1%</b>
Colgate	Sudeste	C	56	<b>-2.9%</b>
Colgate	Total		179	<b>-3%</b>
Colgate	Nordeste	A	19	5%
Colgate	Nordeste	B	17	4%
Colgate	Nordeste	C	19	6%
Colgate	Total		55	5%
Total			344	6%

Produto	Região	Equipe de Vendas	Vendas no Mês	Comp. Com o Mês Anterior
Colgate	Sul	P. Alegre	52	**21%
Colgate	Sul	Curitiba	28	5%
Colgate	Sul	Florianóp.	30	6%
Colgate			110	**12%
Colgate	Sudeste	São Paulo	93	4%
Colgate	Sudeste	Rio	75	5%
Colgate	Sudeste	Belo Horiz.	11	-15%
Colgate			179	-3%
Colgate	Nordeste	Salvador	21	5%
Colgate	Nordeste	Fortaleza	18	4%
Colgate	Nordeste	Recife	16	6%
Colgate			55	5%
Total			344	6%

- DW e OLAP
  - Data Warehousing, Data Mining & OLAP  
Alex Berson, Stephen J. Smith. McGraw-Hill
  - The Data Warehouse Toolkit.  
Ralph Kimball, Margy Ross. John Wiley & Sons, Inc.
  - Data Warehouse Brasil  
(<http://www.dwbrasil.com.br/>)
  - Artigos de Kimball  
(<http://www.ralphkimball.com/html/articles.html>)
  - Data Warehousing Institute  
(<http://www.dw-institute.com/>)
  - OLAP Report - (<http://www.olapreport.com/>)

cin.ufpe.br



# Centro de Informática

U · F · P · E



UNIVERSIDADE FEDERAL DE PERNAMBUCO