

SISTEMAS DE APOIO À INTELIGÊNCIA DE NEGÓCIOS

Asterio K. Tanaka

<http://www.uniriotec.br/~tanaka/SAIN>
tanaka@uniriotec.br



OLAP e Modelagem Dimensional – Conceitos Básicos

Material baseado em originais de Maria Luiza Campos (<http://dataware.nce.ufrj.br/>)
Complementado com referências atuais de Ralph Kimball (<http://www.kimballgroup.com/>)
Agosto de 2007

- **Visão multidimensional de dados**
- **Agregados e hierarquias de dimensões**
- **Ferramentas OLAP**
 - As doze regras de Codd
 - Operações dimensionais para OLAP
 - Tipos de ferramentas: MOLAP, ROLAP, HOLAP
- **Conceitos de modelagem dimensional**
 - Esquema estrela: Fatos e Dimensões
- **Tabelas de Fatos**
 - Fatos aditivos, semi-aditivos, não-aditivos
- **Tabelas de Dimensões**
 - Hierarquias, Normalização/Desnormalização
 - Esquema Snow Flake
- **Modelagem dimensional e projeto de DW**
 - Data Warehouse Bus Architecture & Matrix
 - Mitos, Passos, Dicas, Armadilhas
 - Processo de projeto

Modelagem de DW para OLAP

- Requisitos diferentes das aplicações do ambiente transacional:
 - flexibilidade quanto às análises a suportar
 - medidas a analisar precisam ser vistas sob diferentes perspectivas (dimensões)
- Enfoque diferente da modelagem no ambiente operacional
- Abordagem utilizada:

▶▶ **MODELAGEM DIMENSIONAL**

Visão multidimensional

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

A visão lógica é multidimensional, embora a estrutura física possa ter a mesma visão tabular do modelo relacional.

Estrutura Relacional

Volume de vendas (do revendedor GLEASON)

MODEL	COLOR	SALES VOLUME
MINI VAN	BLUE	6
MINI VAN	RED	5
MINI VAN	WHITE	4
SPORTS COUPE	BLUE	3
SPORTS COUPE	RED	5
SPORTS COUPE	WHITE	5
SEDAN	BLUE	4
SEDAN	RED	3
SEDAN	WHITE	2

Visão matricial ou multidimensional

Volume de Vendas (do revendedor Gleason)

MODEL	Mini Van	6	5	4
	Coupe	3	5	5
	Sedan	4	3	2
		Blue	Red	White

COLOR

- ✓ Um array multidimensional tem um número fixo de dimensões e os valores são armazenados nas células
- ✓ Cada dimensão consiste de um número de elementos

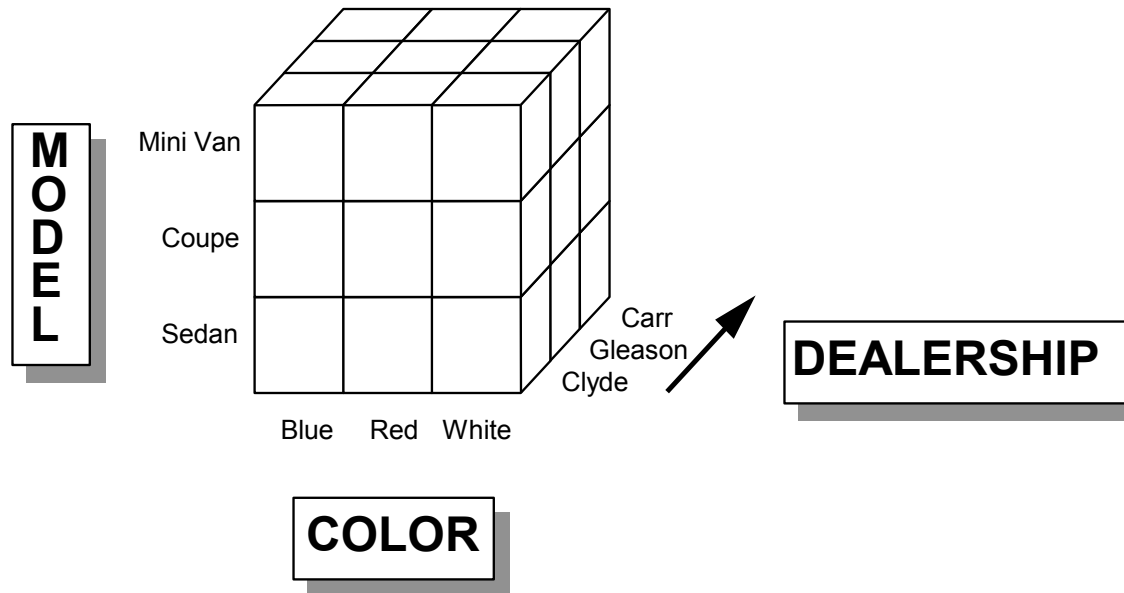
Acrescentando mais uma coluna...

Volume de Vendas
de todos os
revendedores

MODEL	COLOR	DEALERSHIP	VOLUME
MINI VAN	BLUE	CLYDE	6
MINI VAN	BLUE	GLEASON	6
MINI VAN	BLUE	CARR	2
MINI VAN	RED	CLYDE	3
MINI VAN	RED	GLEASON	5
MINI VAN	RED	CARR	5
MINI VAN	WHITE	CLYDE	2
MINI VAN	WHITE	GLEASON	4
MINI VAN	WHITE	CARR	3
SPORTS COUPE	BLUE	CLYDE	2
SPORTS COUPE	BLUE	GLEASON	3
SPORTS COUPE	BLUE	CARR	2
SPORTS COUPE	RED	CLYDE	7
SPORTS COUPE	RED	GLEASON	5
SPORTS COUPE	RED	CARR	2
SPORTS COUPE	WHITE	CLYDE	4
SPORTS COUPE	WHITE	GLEASON	5
SPORTS COUPE	WHITE	CARR	1
SEDAN	BLUE	CLYDE	6
SEDAN	BLUE	GLEASON	4
SEDAN	BLUE	CARR	2
SEDAN	RED	CLYDE	1
SEDAN	RED	GLEASON	3
SEDAN	RED	CARR	4
SEDAN	WHITE	CLYDE	2
SEDAN	WHITE	GLEASON	2
SEDAN	WHITE	CARR	3

Visão multidimensional

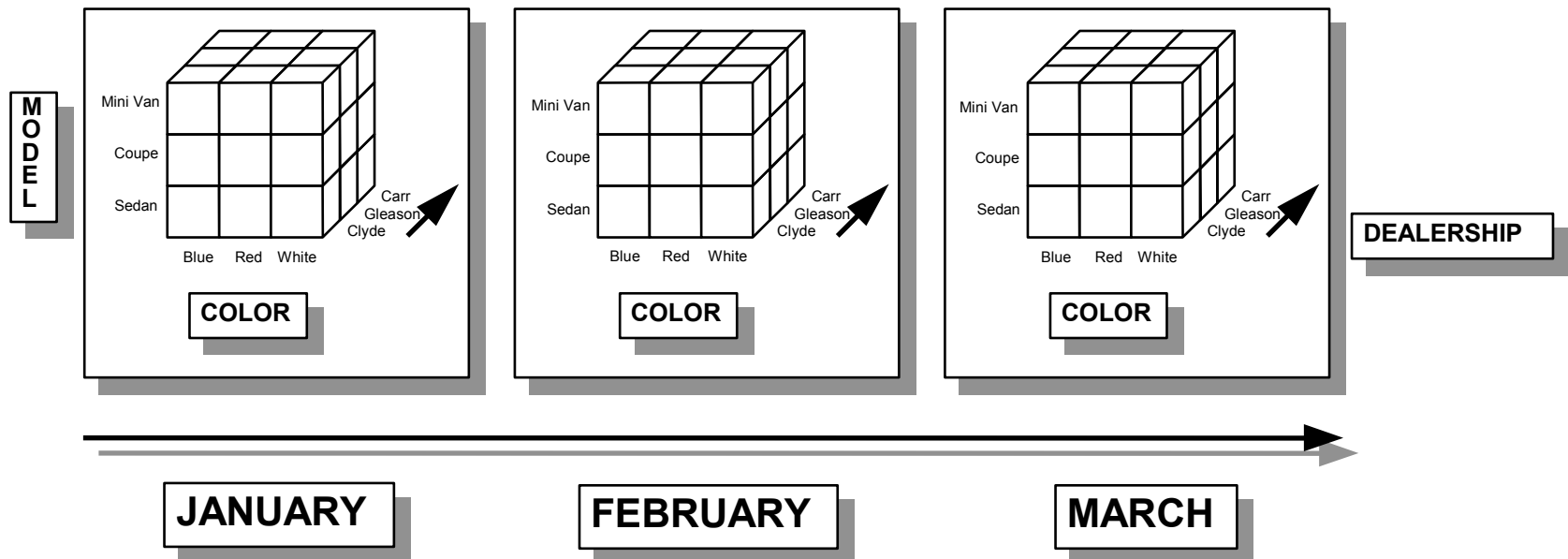
Volume de Vendas



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

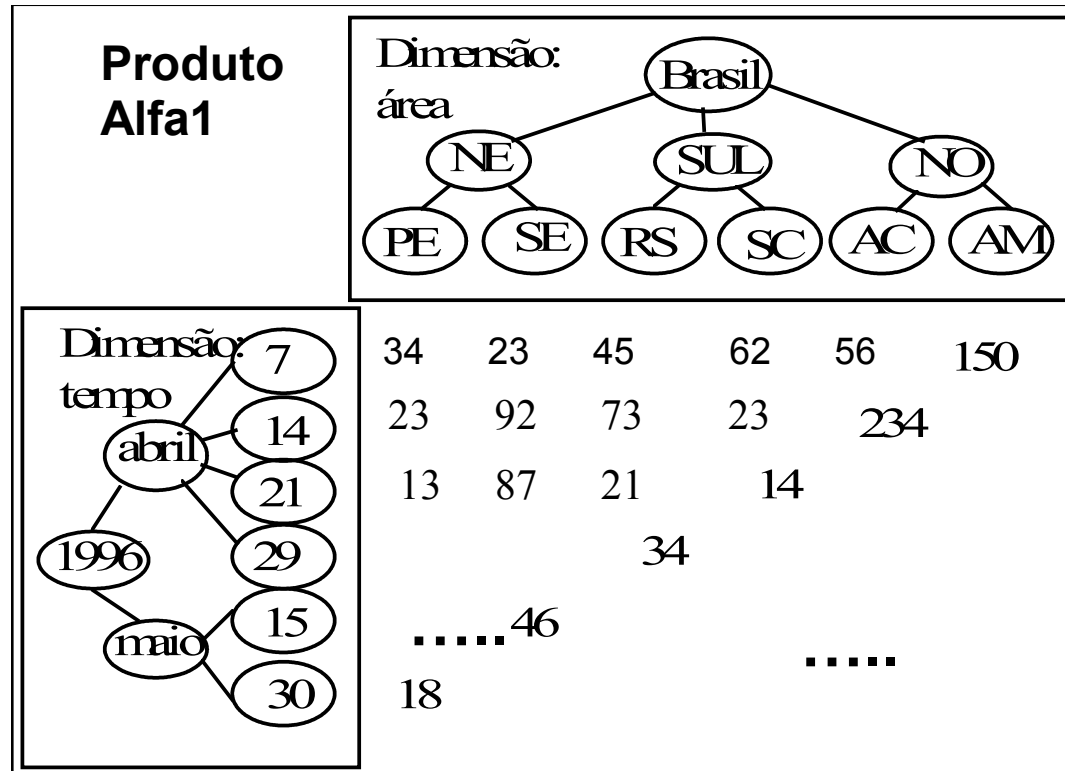
Adicionando Dimensões - Hipercubos

Volume de Vendas



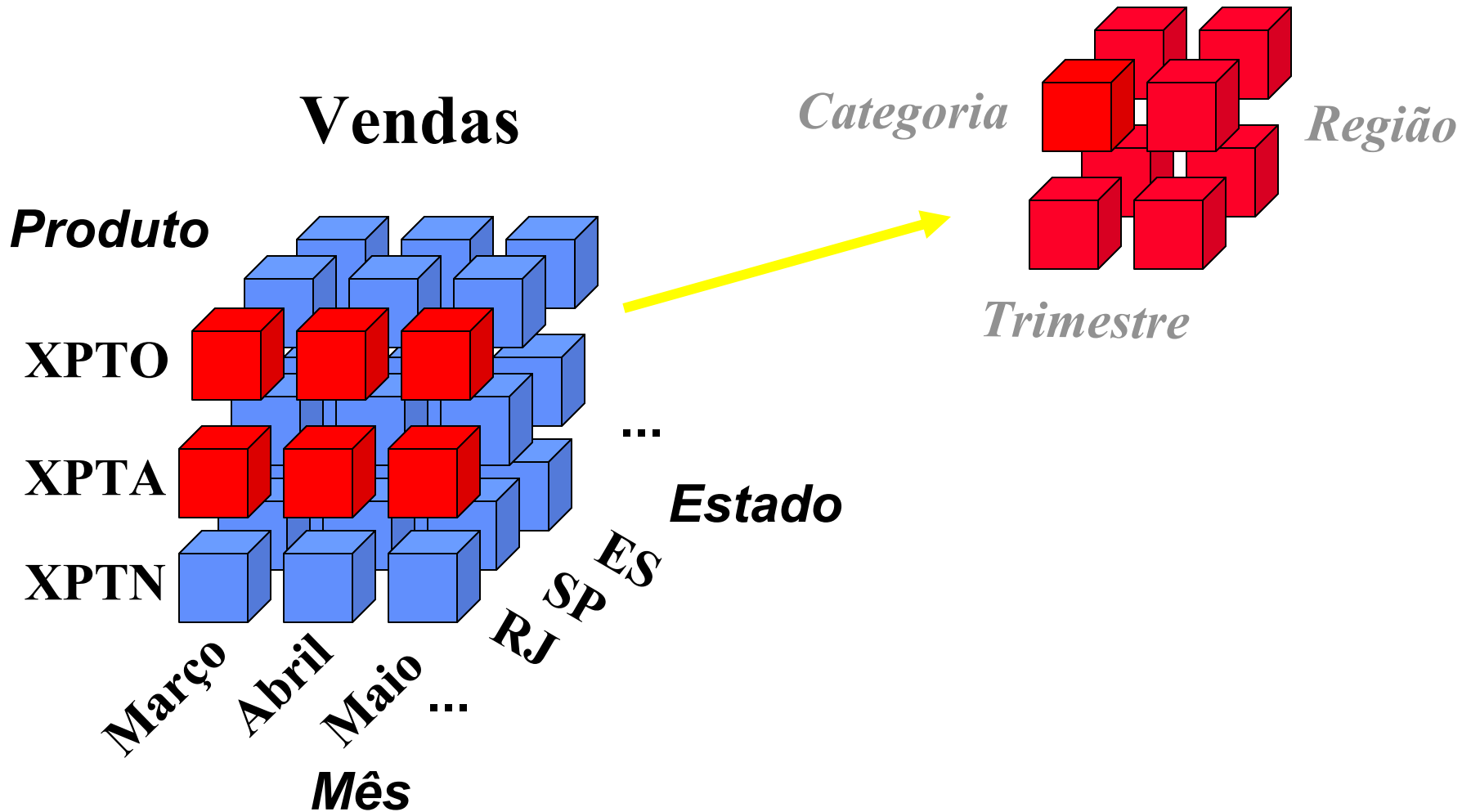
Níveis nas dimensões ou Hierarquias

Total de vendas



- Hierarquias são a base das agregações

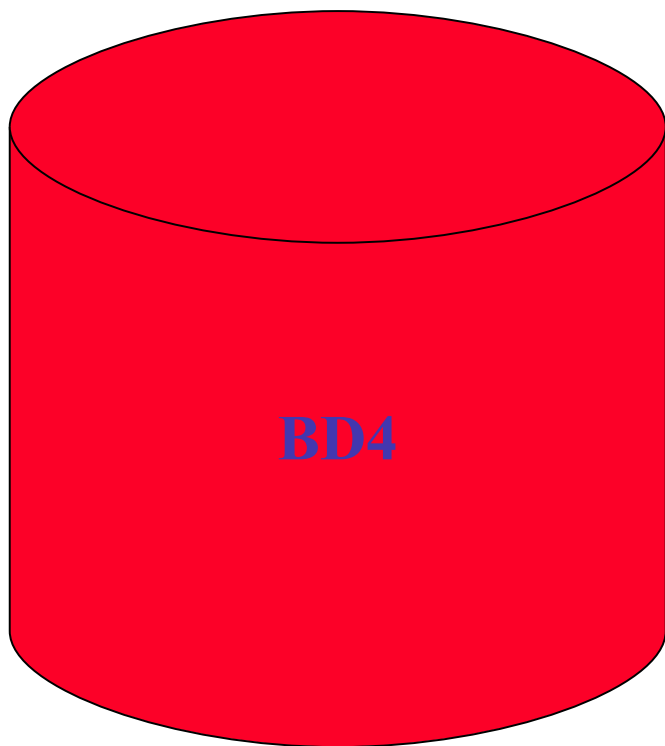
Agregados



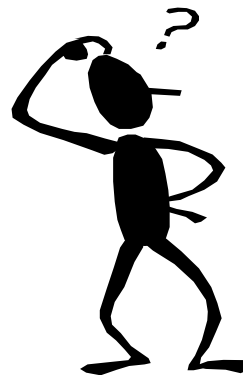
Problemas

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

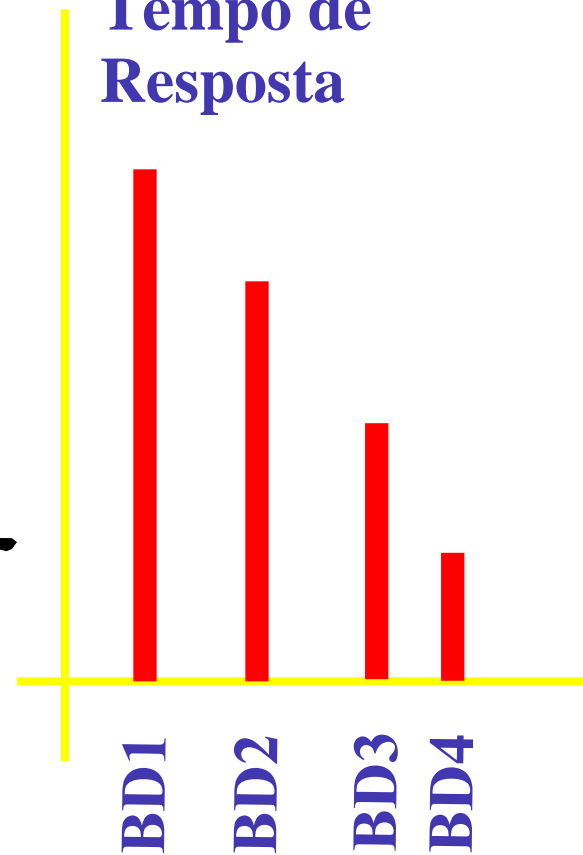
Armazenamento



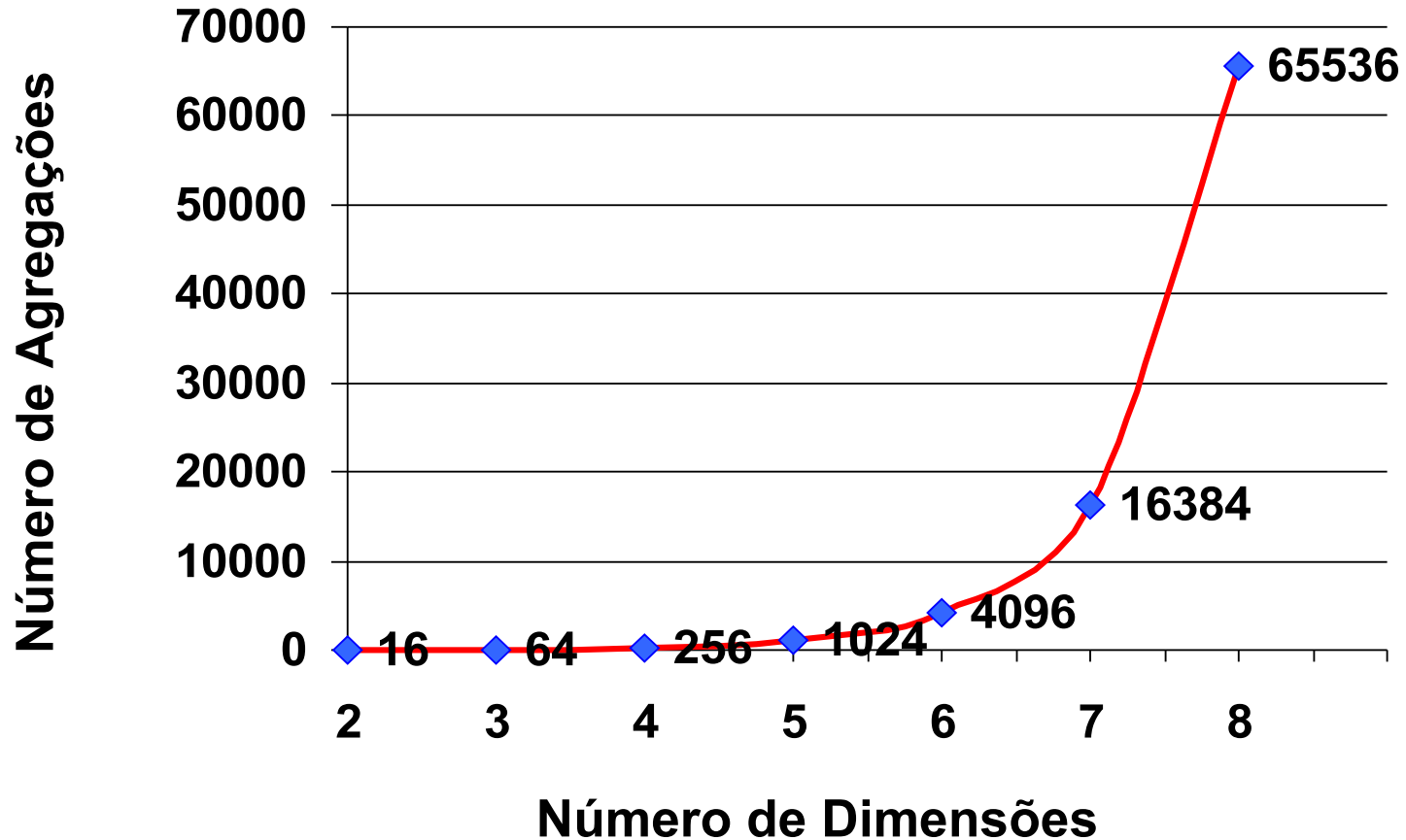
X



Tempo de Resposta



A Síndrome da Explosão no Volume de Dados

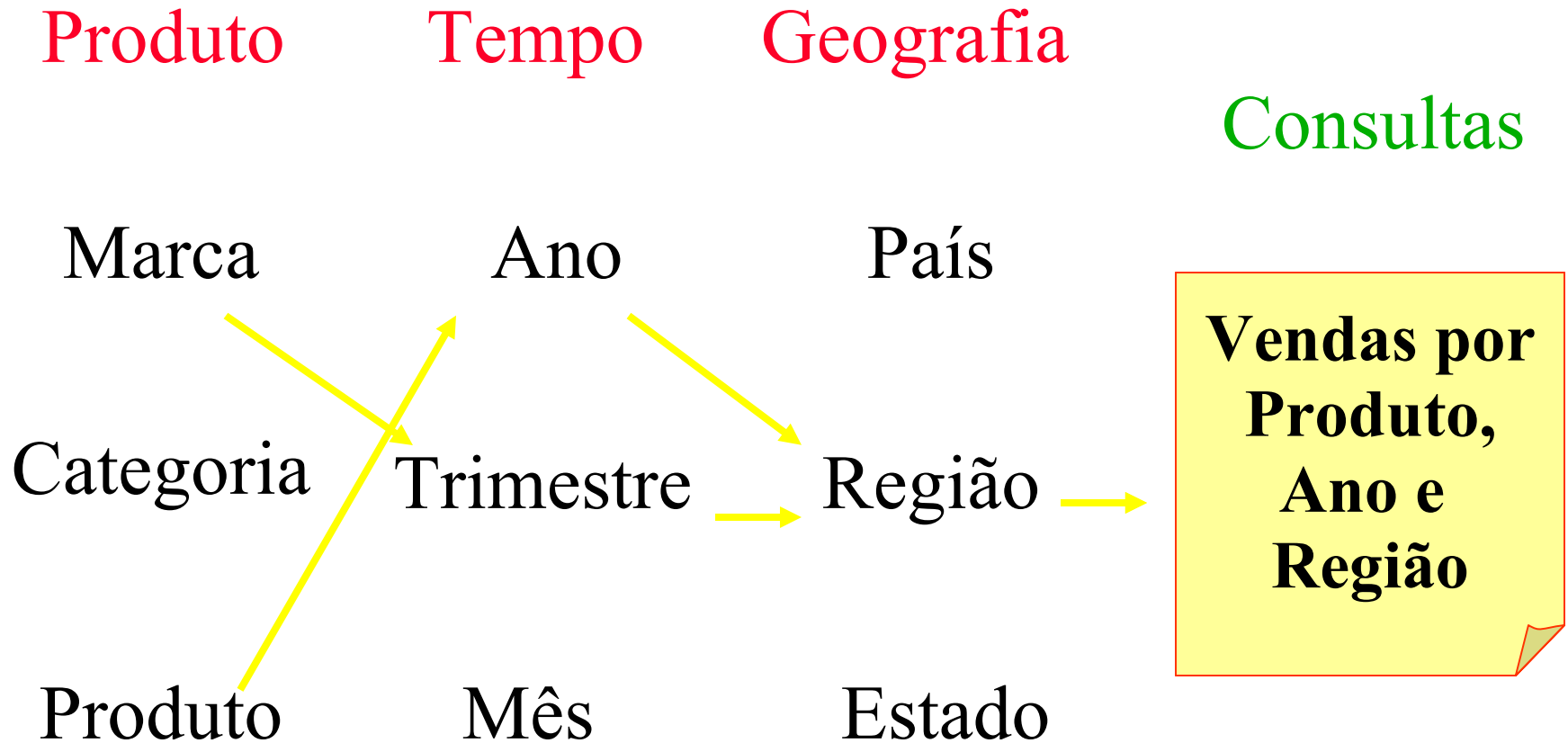


(4 níveis em cada dimensão)

Agregados

- **As hierarquias permitem que o usuário possa ter acesso a dados com maior ou menor detalhe**
- **Os valores apresentados quando o analista consulta dados em níveis hierárquicos mais altos são valores agregados**

Hierarquias e Agregados



Ferramentas OLAP

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

“Doze Regras de Codd” para ferramentas OLAP

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

“Doze Regras de Codd” para ferramentas OLAP

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

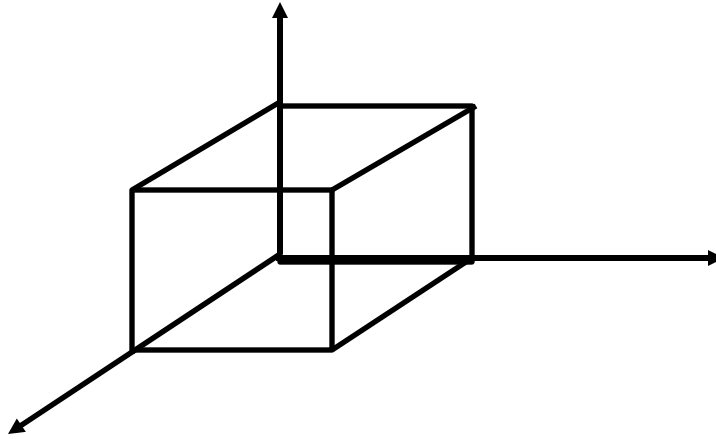
“Doze Regras de Codd” para ferramentas OLAP

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

Operações OLAP

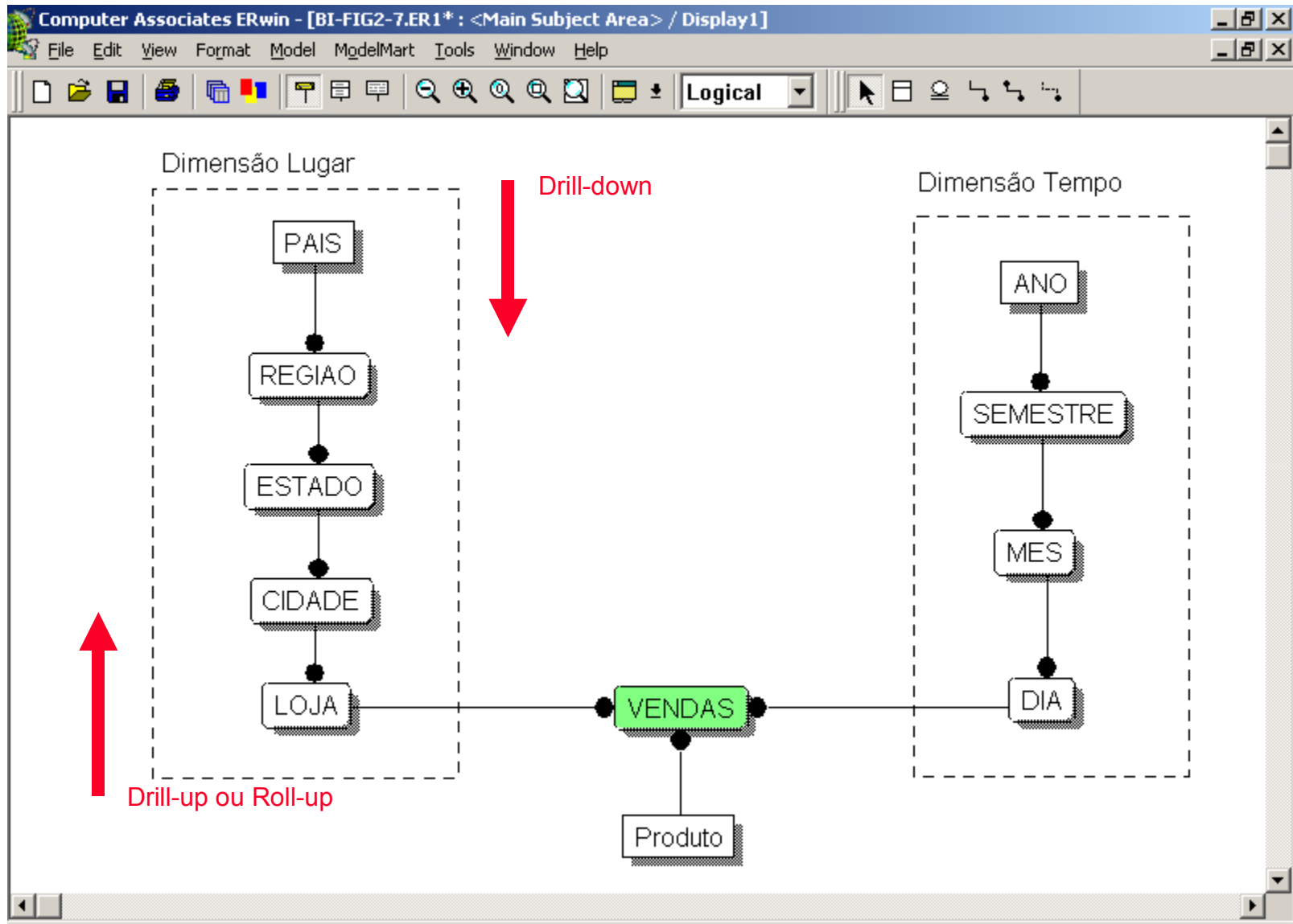
- Ferramentas OLAP fornecem suporte para funções de análise de dados, típicas de aplicações avançadas de planilhas eletrônicas.
- Operações dimensionais de ferramentas OLAP:
 - *Slice and Dice (Ponto, Plano, Cubo)*
 - *Rotation (Rotação ou Pivotamento)*
 - *Drilling*
 - » *Drill Down*
 - » *Drill Across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

Operadores Dimensionais

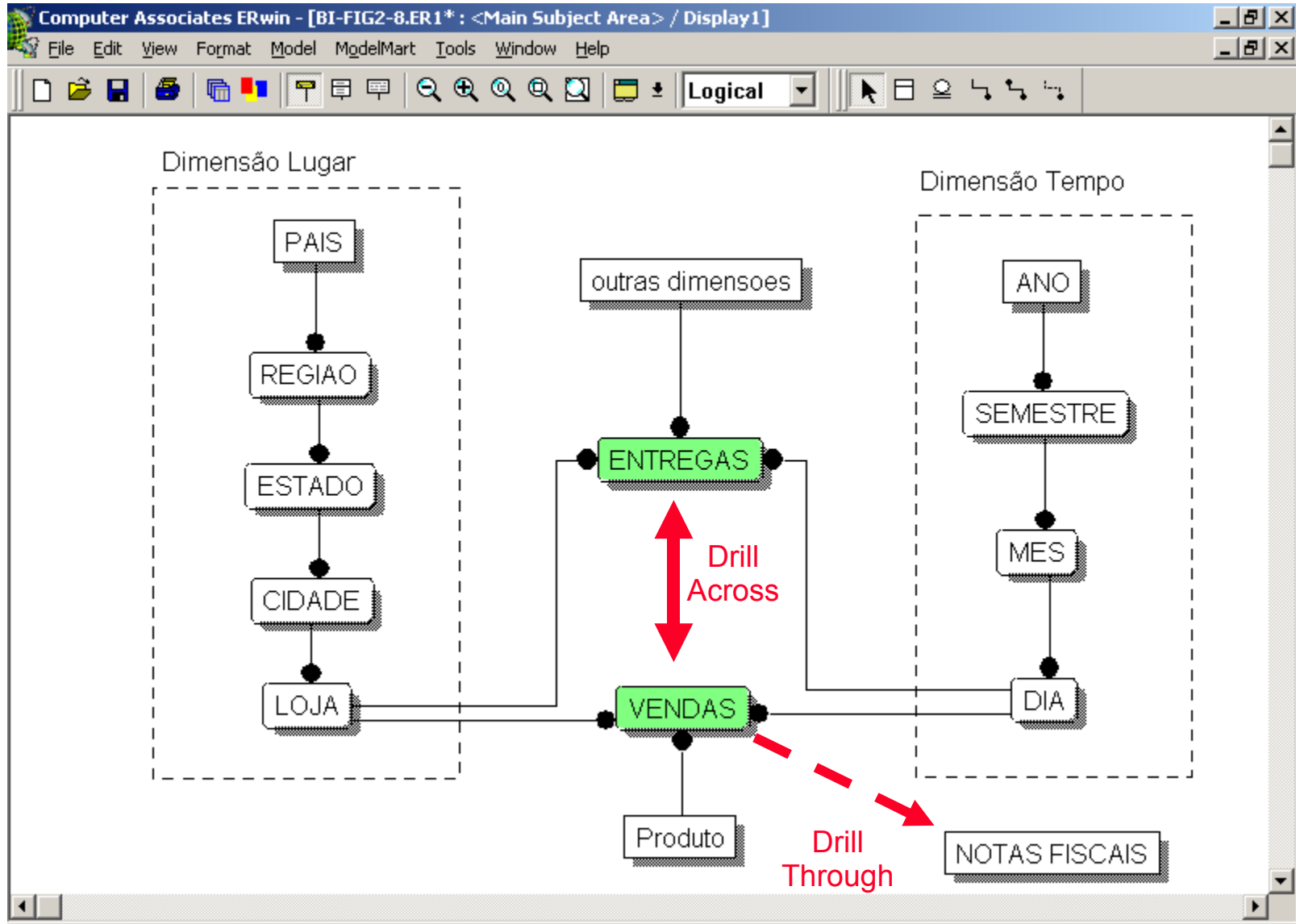


- Ponto - Valor pontual
 - Interseção de valores (Fato) com relação aos eixos (Dimensões)
- Plano – Slicing
 - Duas dimensões variando com outras fixas.
- Cubo – Dicing
 - Todas as dimensões variando
- Rotação – Pivotamento
 - Mudança dos eixos das dimensões, para fins de visualização
 - Vide tabelas dinâmicas no MS Excel

Operadores Drilling



Operadores Drilling



Tipos de ferramentas OLAP

- **OLAP Multidimensional (MOLAP)**

- Utilizam estrutura de dados multidimensional e permitem a navegação pelos níveis de detalhamento em tempo real.
- O BD e o SGBD são multidimensionais
- Estrutura de dados é um array com um número fixo de dimensões. O (hiper)cubo é uma metáfora visual, onde as dimensões coexistem para todo ponto e são independentes entre si.

- **OLAP RELACIONAL (ROLAP)**

- Decorrência do uso consagrado de SGBDs relacionais nos BDs operacionais (transacionais), com as vantagens da tecnologia aberta e padronizada (SQL).
- Utiliza os metadados no apoio à descrição do modelo de dados e na construção de consultas. Através de uma camada semântica acima do esquema relacional, os dados são apresentados ao usuário com visão multidimensional.

- **OLAP HÍBRIDO (HOLAP)**

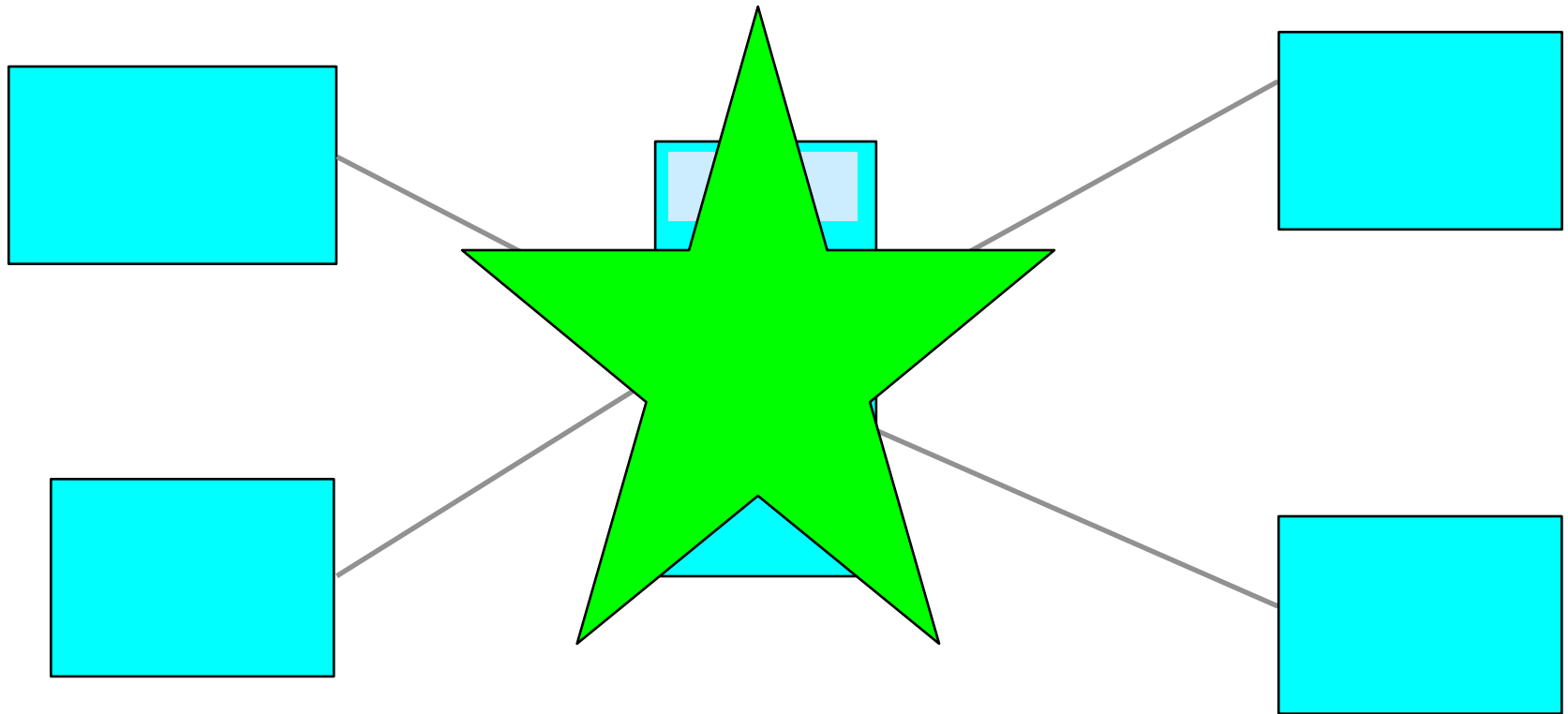
- Tendência dos modernos SGBDs relacionais de adicionar uma arquitetura multidimensional para prover facilidades a ambientes de suporte a decisão.
- Proporciona o desempenho e flexibilidade de um BD multidimensional e mantém a gerenciabilidade, escalabilidade, confiabilidade e acessibilidade conquistadas pelos BDs relacionais.

Modelagem Multidimensional

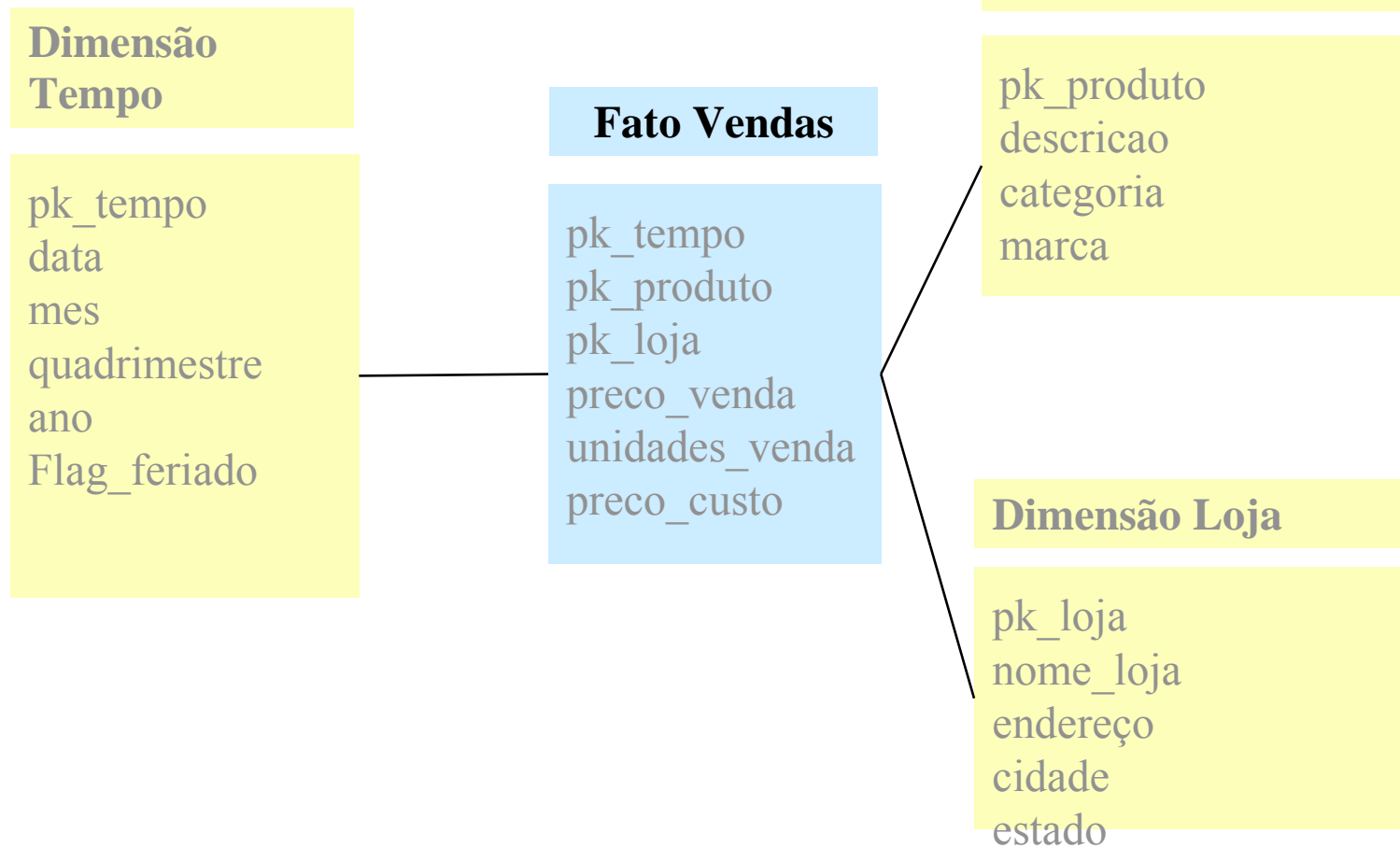
- Proposto por Ralph Kimball para projeto de DW
 - Dimensional Modeling Manifesto, 1997
 - O próprio Kimball atribui a origem a um projeto conduzido por uma empresa (General Mills) e uma universidade (Dartmouth) nos anos 1960s.
- Dominante no projeto de DW
 - Para Kimball, em todo o DW
 - Para Inmon, nos data marts
- Características:
 - Distingue melhor as dimensões dos fatos medidos
 - Simplifica a visualização dimensional (essencial em consultas OLAP)
 - Na verdade é uma mistura de modelagem conceitual com modelagem lógica, pois já é bastante voltada para a abordagem relacional (a literatura fala sempre em *tabelas*)

Esquema Estrela

Uma tabela de fatos cercada de tabelas de dimensões



Esquema Estrela - Exemplo



Exemplo Consultas

“Vendas por categoria de produto sobre os últimos seis meses”

“Vendas por marca entre 1990 e 1995”

Dimensão Loja

Dimensão Produto

Dimensão Tempo

Colunas da chave composta ligando a tabela de fatos às tabelas de dimensão

Medidas Numéricas

pk_tempo	pk_produto	pk_loja	preco_venda	unidades_venda	preco_custo

Tabelas de Dimensão

Tabela de Fatos

...

Consulta SQL sobre um esquema estrela

```
select
  [Loja].[NomeLoja], [Tempo].[DataCompleta],
  [Produto].[Descricao],
  Sum( [Vendas].[Unidades_Venda]) as Total
from
  [Vendas], [Tempo], [Produto], [Loja]
where
  [Vendas].[CodTempo] = [Tempo].[CodTempo] and
  [Vendas].[CodProduto] = [Produto].[CodProduto] and
  [Vendas].[CodLoja] = [Loja].[CodLoja]
group by
  [Loja].[NomeLoja], [Tempo].[DataCompleta], [Produto].[Descricao]
order by
  [Tempo].[DataCompleta], [Loja].[NomeLoja],
  [Produto].[Descricao]
```

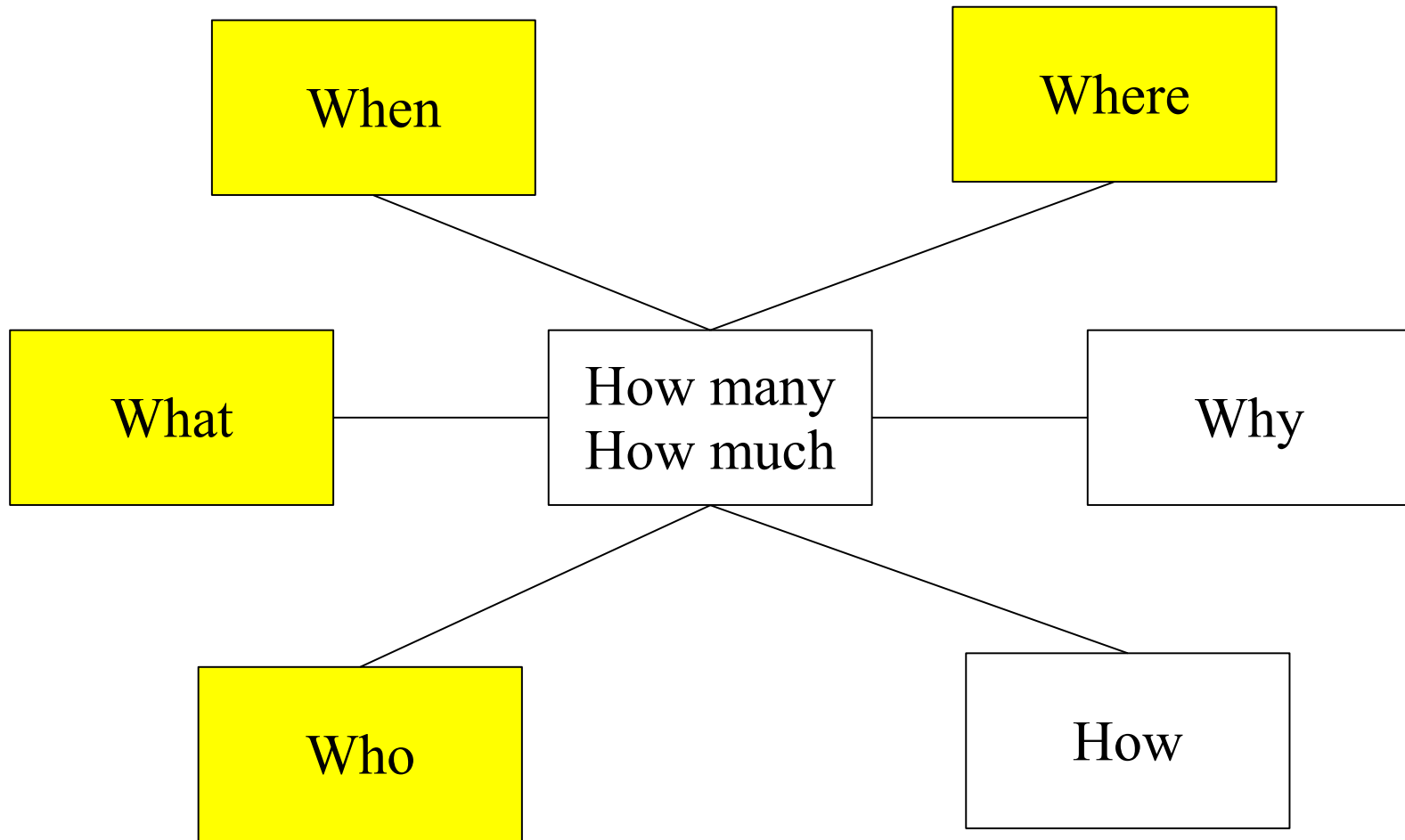
**Qtd Vendida
de cada Produto
por Loja e
por Data**

Resultados

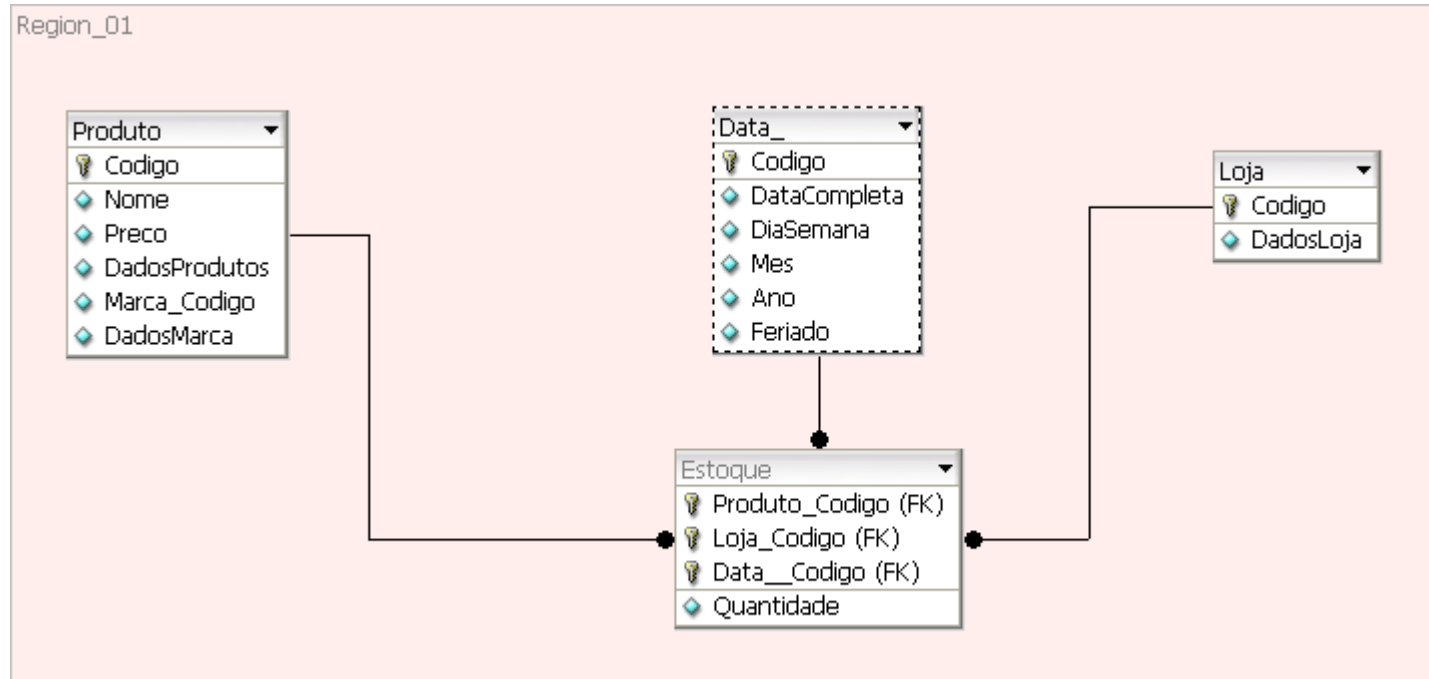
NomeLoja	DataCompleta	Descricao	Total
East Loja	Oct 1, 1994	Athletic Drink	57
East Loja	Oct 1, 1994	Beef Stew	128
East Loja	Oct 1, 1994	Buffalo Jerky	202
East Loja	Oct 1, 1994	Chicken Dinner	161
East Loja	Oct 1, 1994	Clear Refresher	73
East Loja	Oct 1, 1994	Dried Grits	102
East Loja	Oct 1, 1994	Dry Tissues	16
East Loja	Oct 1, 1994	Extra Nougat	442
East Loja	Oct 1, 1994	Fizzy Classic	46
East Loja	Oct 1, 1994	Fizzy Light	65
East Loja	Oct 1, 1994	Lasagna	162
East Loja	Oct 1, 1994	Lots of Nuts	248
East Loja	Oct 1, 1994	Onion Slices	120

Esquema Estrela de DW

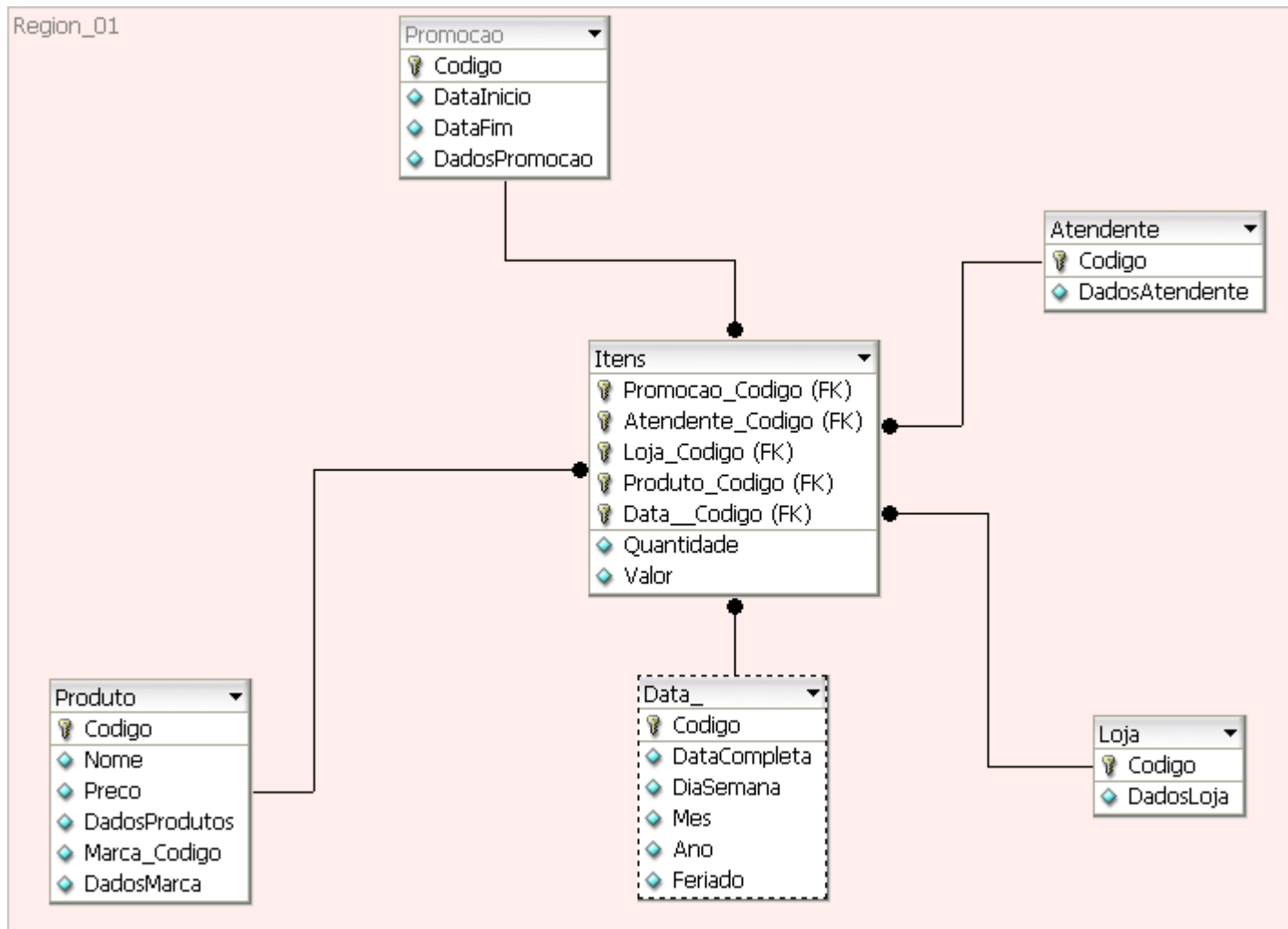
5 W e 3 H



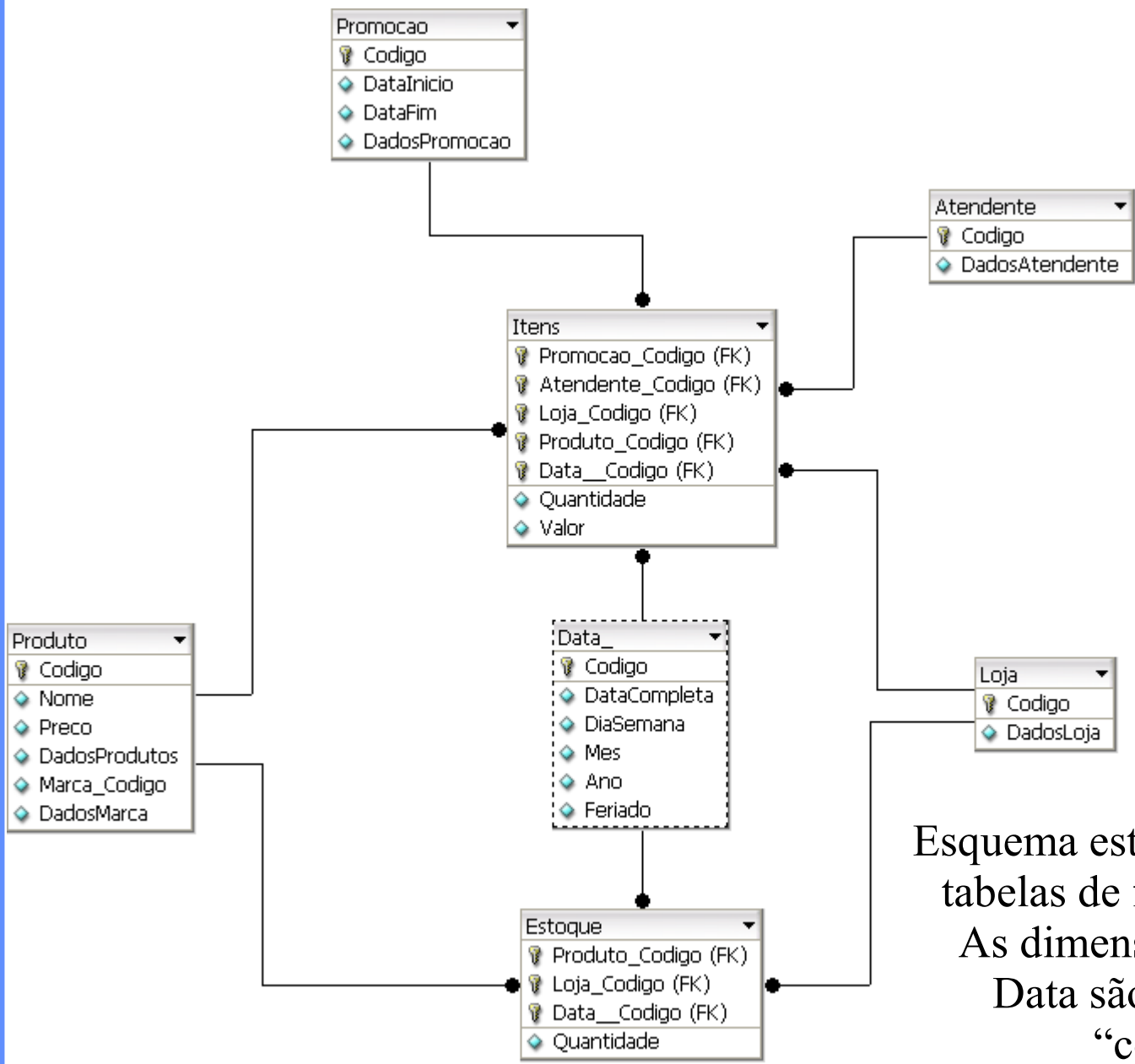
 Tipos de dimensão mais comuns



Esquema estrela com tabela de fatos Estoque e dimensões Produto, Loja e Data



Esquema estrela com tabela de fatos Itens e dimensões Produto, Promoção, Atendente, Loja e Data



Esquema estrela (constelação) com tabelas de fatos Itens e Estoque. As dimensões Produto, Loja e Data são compartilhadas e “conformadas”

Modelagem Dimensional

- **Esquema Estrela é simétrico**
 - Comparado a esquemas ER típicos
- **Tabela de Fatos**
 - Expressa relacionamento M:N entre dimensões
 - Tabela dominante
 - » Usualmente com grande volume de dados; ocupam 90% do espaço em um DW típico
 - » Tendem a ter muitas linhas e poucas colunas
- **Tabelas de Dimensões**
 - Tabelas que “qualificam” os fatos, com muitos campos descritivos (é comum ter dimensões com dezenas de colunas)
 - Dimensões apresentam-se em consultas qualificadas como “por dimensão” (vendas “por semana” “por marca” “por loja”) e são as bases para agregações e agrupamentos.
 - Uma junção liga cada tabela de dimensão à tabela de fatos
 - Volume bem menor que as tabelas de fatos
 - O poder de um DW é diretamente proporcional à qualidade e profundidade dos atributos das dimensões.

(Mais uma) Comparação entre Modelagem ER e Multidimensional

ER	Multidimensional
1 diagrama (vários processos de negócio)	Vários diagramas dimensionais (1 para cada processo de negócio)
Usuários acham difícil entender e navegar pelo modelo	Usuários reconhecem “o seu negócio”
Muitas junções para responder a consultas	Poucas junções
Dados atômicos	Dados atômicos e agregados
Planos de consultas extremamente distintos e específicos para as consultas previstas	Planos de consultas “genéricos” (simetria do modelo)

Tabela de Fatos

Ex: Tabelas Itens e Estoque

- Tabela de fatos normalizada em 3ª forma normal
- Chave primária composta por um subconjunto das chaves das dimensões (subconjunto que garanta unicidade – às vezes todas as chaves)
 - Vide por exemplo a tabela Itens, se houvesse uma dimensão CupomFiscal (bastariam as chaves de CupomFiscal e de Produto como chave primária)
- Por ser o DW histórico, a tabela de fatos tem muitas linhas (milhões, bilhões) e poucas colunas (chaves das dimensões e medidas dos fatos).
- Medidas do fatos são usualmente numéricas, mas podem ser não numéricas ou sem medida (tabelas sem fato)
- Fatos são tipicamente aditivos, mas podem ser
 - Semi-aditivos ou mesmo Não aditivos

Fatos Aditivos

Ex: Quantidade, Valor na Tabela Itens

- **São numéricos e podem ser somados em relação às dimensões existentes**
 - Ex: quantidade e valor podem ser somados ao longo de qualquer dimensão (Produto, Promoção, Atendente, Loja e Data)
- **Sempre que, em uma modelagem, um dado numérico for apresentado, então este será um bom indício de um atributo em fatos.**
- **Em geral, fatos aditivos representam medidas de atividade do negócio, ligadas aos seus indicadores de desempenho (KPI – Key performance indicators).**

Fatos Semi-Aditivos

Ex: Quantidade na Tabela Estoque

- **Também são numéricos, mas não podem ser somados em relação a todas as dimensões existentes (a semântica não permite)**
 - Ex: quantidade em estoque só pode ser somada ao longo da dimensão Produto. Nas dimensões Loja e Data, a soma não faria muito sentido (especialmente nesta última, nenhum sentido)
- **Em geral, fatos semi-aditivos representam leituras medidas de intensidade do negócio.**
 - São snapshots destas leituras que entram no DW.
 - O valor atual já leva em consideração valores passados.
- **Fatos semi-aditivos típicos: Níveis de Estoque, Saldos, Fechamento diário/mensal de conta, etc...**

Fatos Não-Aditivos

- **Algumas observações não numéricas podem eventualmente ser fatos.**
 - Ex: DW de registro de acidentes de trânsito
 - » Atributos: carro1, carro2, motorista1, motorista2, descrição do acidente, descrição do tempo e descrição da pista.
- **Informações textuais são fatos que só permitem contagem e estatísticas associadas a contagens. Alternativamente, poderiam ser modeladas como dimensões ligadas a uma tabela de fatos “sem fatos”, isto é, só para contagem.**
 - Ex: DW de registro de inscrições em turmas por disciplina, por semestre, por curso, por aluno, por professor.

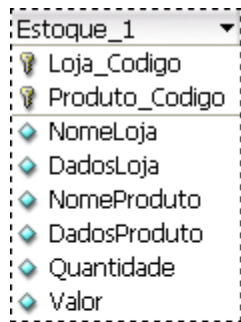
Tabelas de Dimensões

- **Objetivo:**
 - Contém descrições textuais do negócio (fato)
 - Atributos de dimensões servem como cabeçalho das linhas e colunas das análises e filtro nas consultas e relatórios
- **Características:**
 - Chaves simples (em geral, artificiais: “surrogate keys”)
 - » Números inteiros de 4 bytes: $2^{32} > + 2$ bilhões
 - Muitas colunas (dezenas); poucas linhas (centenas ou milhares) se comparadas com tabelas de fatos
 - Usualmente não dependente do tempo
 - » Tempo é outra dimensão (quase sempre presente)
 - Desnormalizada (em geral, na 2ª forma normal)
 - Hierarquias implícitas (à custa da 3ª forma normal)

Dívida: Segunda forma normal

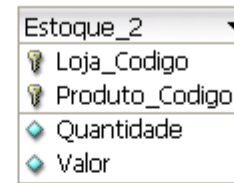
- **Informalmente:**

- *Uma relação está em 2FN se todo atributo não-primário (isto é, que não seja membro de chave) for totalmente dependente de qualquer chave.*



Estoque_1	
Loja_Codigo	Primary Key
Produto_Codigo	Primary Key
NomeLoja	Non-Primary
DadosLoja	Non-Primary
NomeProduto	Non-Primary
DadosProduto	Non-Primary
Quantidade	Non-Primary
Valor	Non-Primary

Tabela Estoque_1
não está em 2FN



Estoque_2	
Loja_Codigo	Primary Key
Produto_Codigo	Primary Key
Quantidade	Non-Primary
Valor	Non-Primary

Tabela Estoque_2
está em 2FN. Na
verdade, está
também em 3FN

Em geral, as tabelas de fatos são normalizadas em 3FN

Dívida: Terceira forma normal

- **Informalmente:**

- *Uma relação está em 3FN se estiver em 2FN e nenhum atributo não-primo (isto é, que não seja membro de uma chave) for transitivamente dependente da chave.*

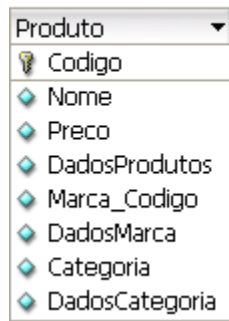


Tabela Produto_1
não está em 3FN

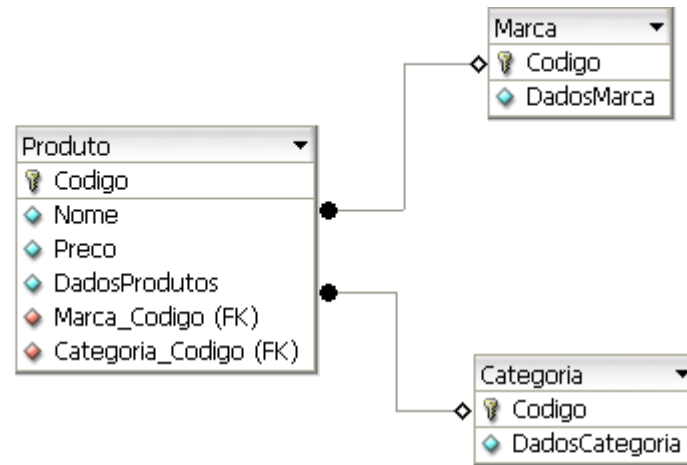


Tabela Produto_2
está em 3FN.

Num esquema estrela, as tabelas de dimensões não são normalizadas em 3FN; estão apenas em 2FN.

Hierarquias de Dimensões

- **Uma dimensão pode ter múltiplas hierarquias além de outros atributos descritivos**
- **Exemplos:**
 - Para a dimensão Loja
 - » Geografia física: CEP, cidade, estado, região, país
 - » Geografia de vendas: território, região, zona
 - » Geografia de distribuição: área primária, região
 - Para a dimensão Produto
 - » Hierarquia de Marcas
 - » Hierarquia de Categorias
 - » Hierarquia de Tipo de Armazenamento

Tabelas de Dimensão

Segundo KIMBALL, as tabelas de dimensão não devem ser normalizadas pois:

- 1) não há atualização freqüente nas bases;**
- 2) o espaço em disco economizado é relativamente pequeno;**
- 3) esse ganho de espaço não justifica a perda de performance na realização de consultas por conta das junções necessárias em caso de normalização.**

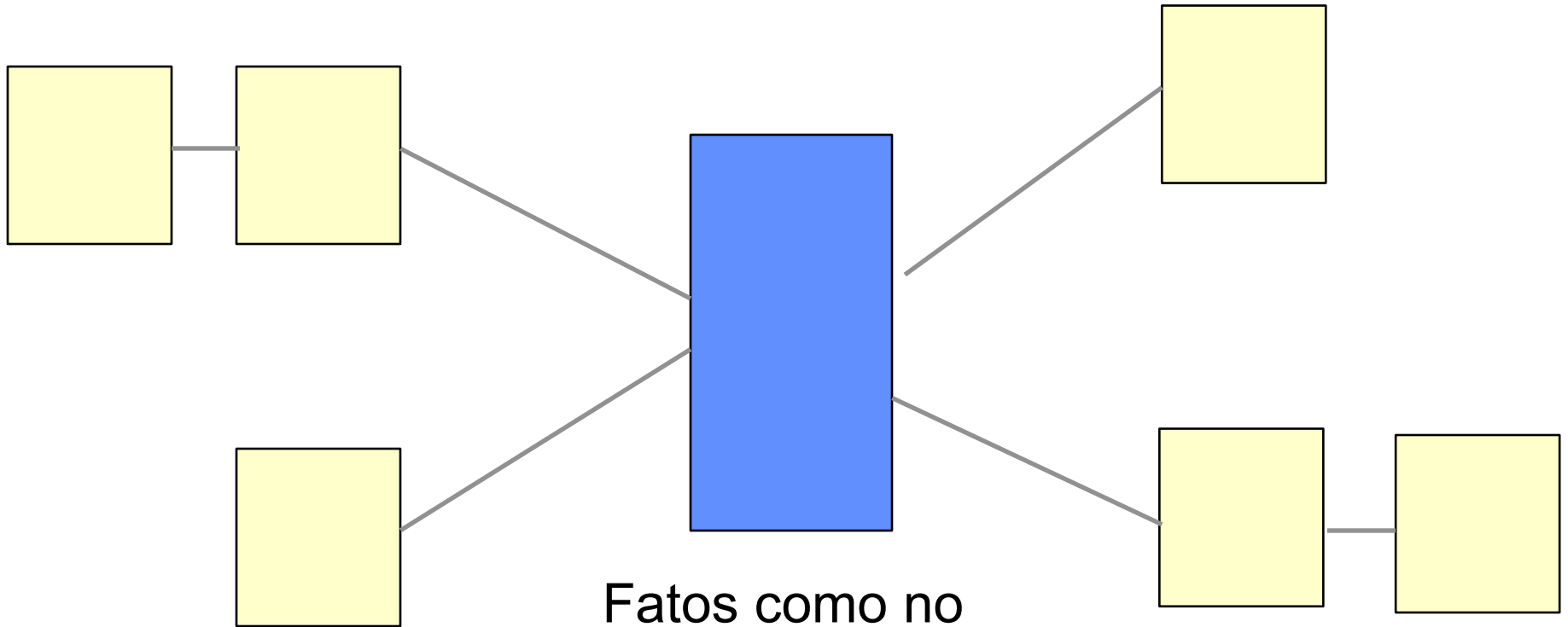
Variações do Esquema Estrela

Esquema floco de neve

- O esquema floco de neve é uma variação do esquema estrela no qual todas as tabelas dimensão são normalizadas na terceira forma normal (3FN)**
- Reduzem a redundância mas aumentam a complexidade do esquema e conseqüentemente a compreensão por parte dos usuários**
- Dificultam as implementações de ferramentas de visualização dos dados**

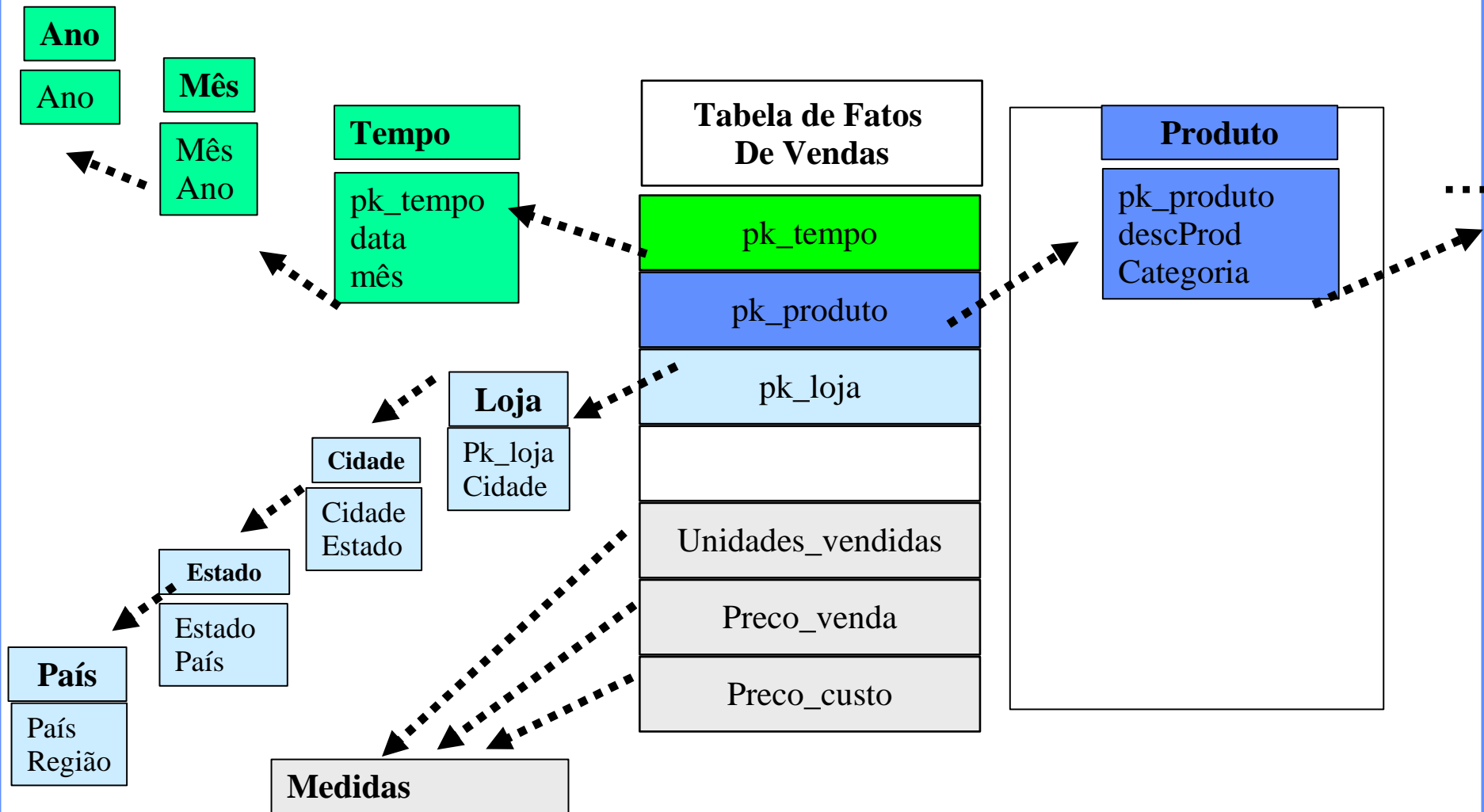
Esquema Floco de Neve

Dimensões normalizadas



Fatos como no
Esquema estrela

Esquema Flocos de Neve - Exemplo



Mitos sobre Modelagem Dimensional

- 1. Modelos dimensionais e Data Marts são para dados sumarizados somente**
 - 2. Modelos dimensionais e Data Marts são soluções departamentais, não empresariais**
 - 3. Modelos dimensionais e Data Marts não são escaláveis**
 - 4. Modelos dimensionais e Data Marts são apropriados somente quando há um padrão de uso previsível**
 - 5. Modelos dimensionais e Data Marts não podem ser integrados e levam a soluções isoladas**
- Ralph Kimball; Margy Ross. The Data Warehouse Toolkit. John Wiley, 2002 – Cap. 1
 - Margy Ross & Ralph Kimball Fables and Facts: Do you know the difference between dimensional modeling truth and fiction? Oct 2004
<http://www.intelligententerprise.com/showArticle.jhtml?articleID=49400912>

Data Warehouse Bus Architecture

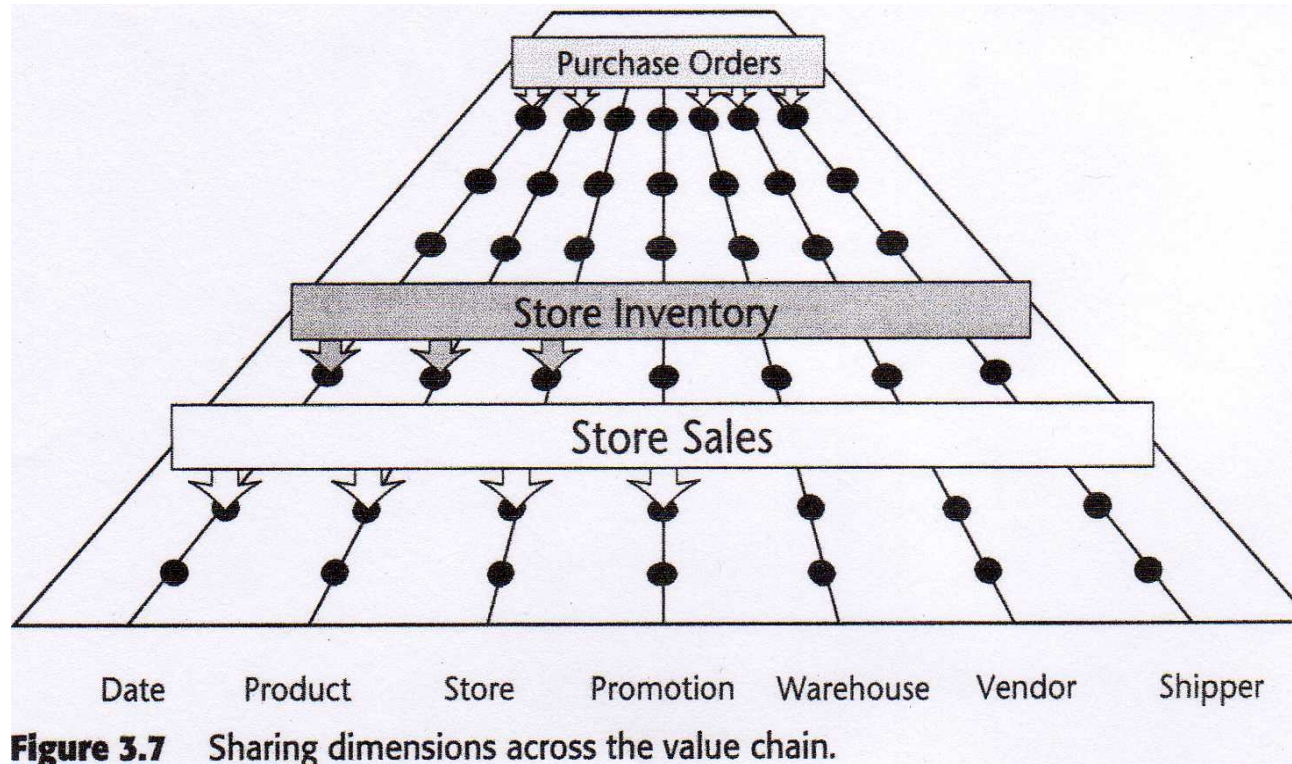


Figure 3.7 Sharing dimensions across the value chain.

Definindo um barramento padrão para o ambiente de DW, data marts separados podem ser implementados por grupos diferentes em tempos diferentes. Todos os processos da cadeia de valores da organização criarão uma família de modelos dimensionais que compartilham um conjunto completo de dimensões comuns e conformadas.

Data Warehouse Bus Matrix

BUSINESS PROCESSES	COMMON DIMENSIONS							
	Date	Product	Store	Promotion	Warehouse	Vendor	Contract	Shipper
Store Sales	X	X	X	X				
Store Inventory	X	X	X					
Store Deliveries	X	X	X					
Warehouse Inventory	X	X			X	X		
Warehouse Deliveries	X	X			X	X		
Purchase Orders	X	X			X	X	X	X

As linhas da matriz correspondem a data marts e as colunas a dimensões conformadas. A matriz é a ferramenta usada para criar, documentar, gerenciar e comunicar a arquitetura de barramento. Segundo Kimball, é o artefato de análise mais importante do desenvolvimento de um DW. É uma ferramenta híbrida, que serve para design técnico, para gerência de projeto e como forma de comunicação organizacional.

Quatro Passos da Modelagem Dimensional

1. Selecionar o processo de negócio a modelar

- Um processo é uma atividade de negócio natural da organização que tipicamente é suportada por um sistema fonte de coleções de dados.
- Exemplos: vendas, compras de matéria prima, pedidos, expedições, faturamento, inventário, contas a pagar/receber.

2. Declarar o grão do processo de negócio

- Significa especificar *exatamente* o que uma linha da tabela fato representa.
- Exemplos: uma linha de um cupom fiscal, um cartão de embarque individual, um nível diário de estoque de cada produto, um saldo mensal de cada conta bancária.

3. Escolher as dimensões que se aplicam a cada linha da tabela de fatos

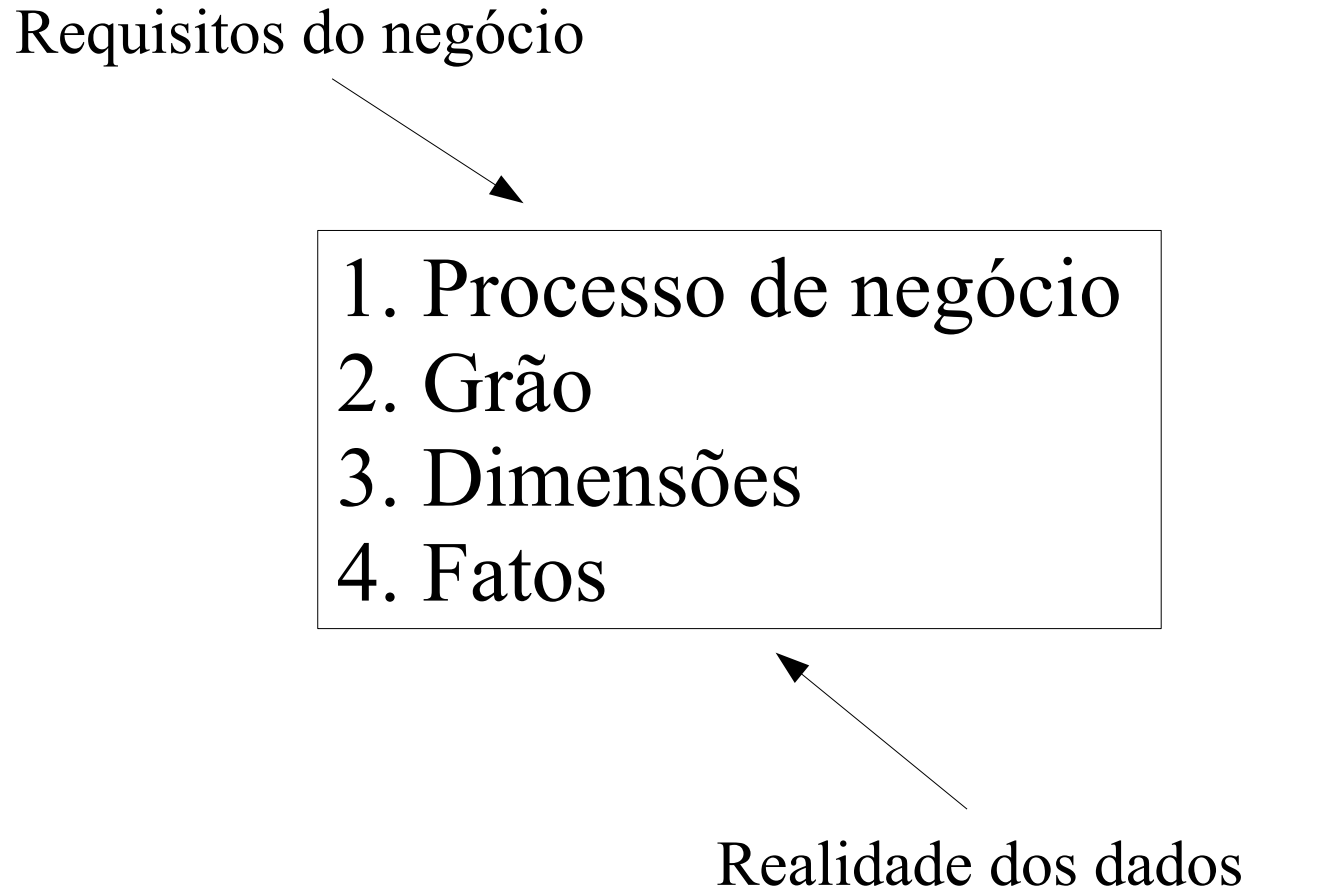
- Implica em responder à pergunta: “Como o pessoal do negócio descreve os dados que resultam do processo de negócio?”
- Exemplos: data, produto, cliente, tipo de transação, status de pedido.

4. Identificar os fatos que irão popular cada linha da tabela de fatos

- Implica em responder à pergunta: “O que nós estamos medindo?” Os fatos candidatos devem ser coerentes com o grão declarado no passo 2.
- Exemplos: quantidade, valor.

Quatro Passos da Modelagem Dimensional

Requisitos do negócio

- 
1. Processo de negócio
 2. Grão
 3. Dimensões
 4. Fatos

Realidade dos dados

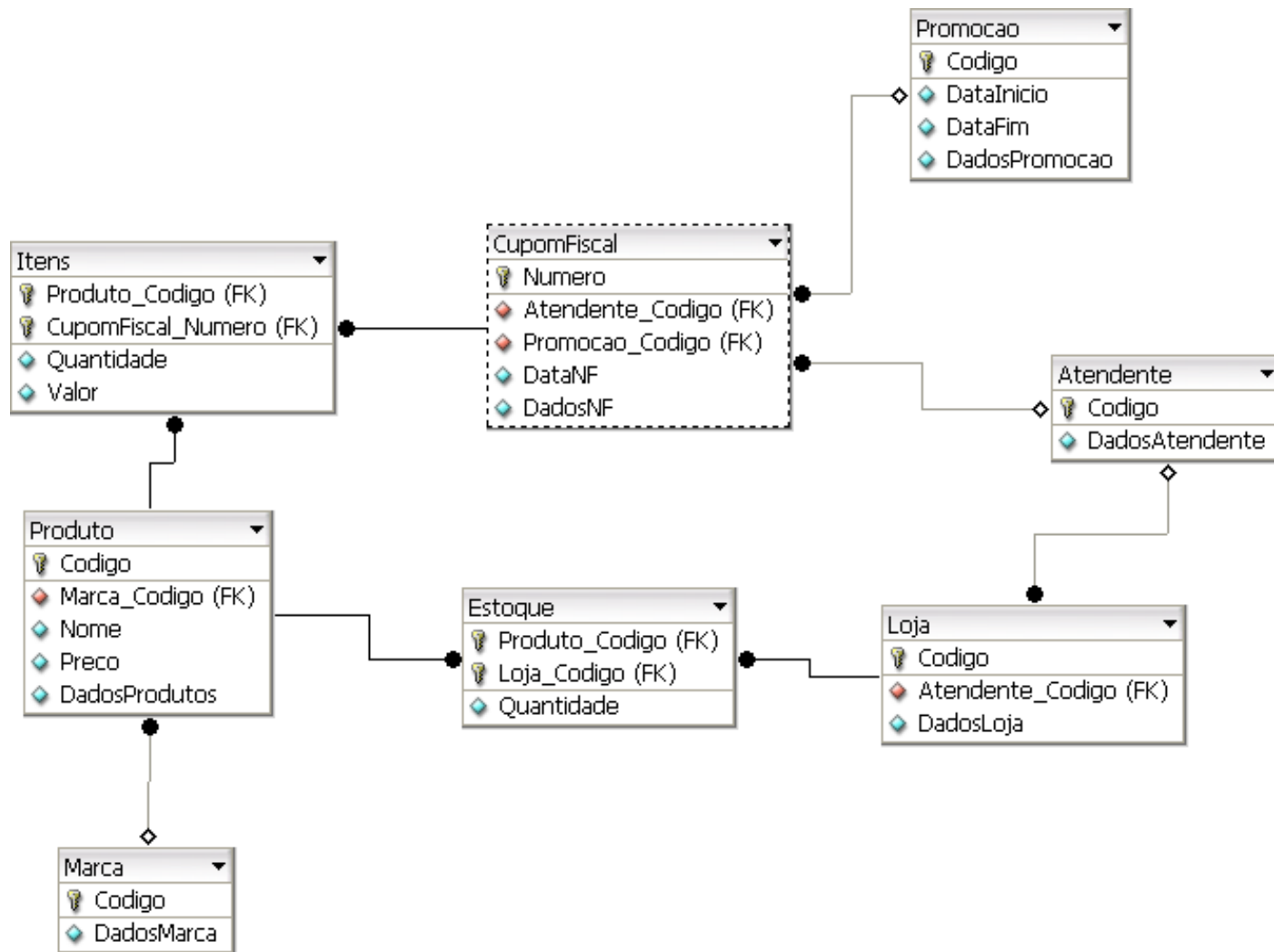
Dicas importantes na Modelagem Dimensional

- ***Resista à tentação de simplesmente examinar as fontes de dados somente: não há substituto para o input dos usuários do negócio.***
- Caso exista, use um modelo de dados convencional E-R como ponto de partida para o trabalho de modelagem dimensional.
 - Observe os relacionamentos 1:N existentes. Eles podem sugerir dimensões.
 - Observe as entidades fortes. Elas também podem sugerir dimensões.
 - Observe as entidades que expressam documentos como Nota Fiscal, Pedido, Ordem de Compra, etc. Elas podem sugerir fatos.
 - Observe os relacionamentos M:N. Na sua interseção, pode haver valores numéricos. Isto sugere fatos.
 - Observe os atributos que estarão nas tabelas de dimensões. Analise a relação de hierarquias entre esses atributos de dimensão. Atente para os relacionamentos M:N entre eles. Isto pode definir granularidade.

Dicas importantes na Modelagem Dimensional

- As tabelas FATOS, tipicamente, armazenam dados, valores atômicos ou agregados obtidos a partir destes.
- As medidas das tabelas FATOS são normalmente aditivas em certas dimensões (ou em todas).
- As tabelas FATOS possuem chaves que as conectam às diferentes DIMENSÕES que as circundam. Essa conexão se dá num nível de granularidade compatível entre elas (FATO e DIMENSÃO).
- As tabelas DIMENSÃO armazenam os valores de filtro, acesso e textos que caracterizam os dados trabalhados.
- As tabelas FATOS são normalmente normalizadas (3ª forma normal).
- As tabelas DIMENSÕES são normalmente desnormalizadas (2ª forma normal - Esquema Estrela).
- A granularidade combinada da tabela FATO com a de suas tabelas DIMENSÕES determina o número de linhas das tabelas do projeto.

Modelo Entidades Relacionamentos



Exemplo de Modelagem

1. Selecionar o processo de negócio

- *Vendas no caixa da loja*

2. Declarar o grão

- *Venda individual de cada produto por loja (isto é, uma linha de cada cupom fiscal de venda)*

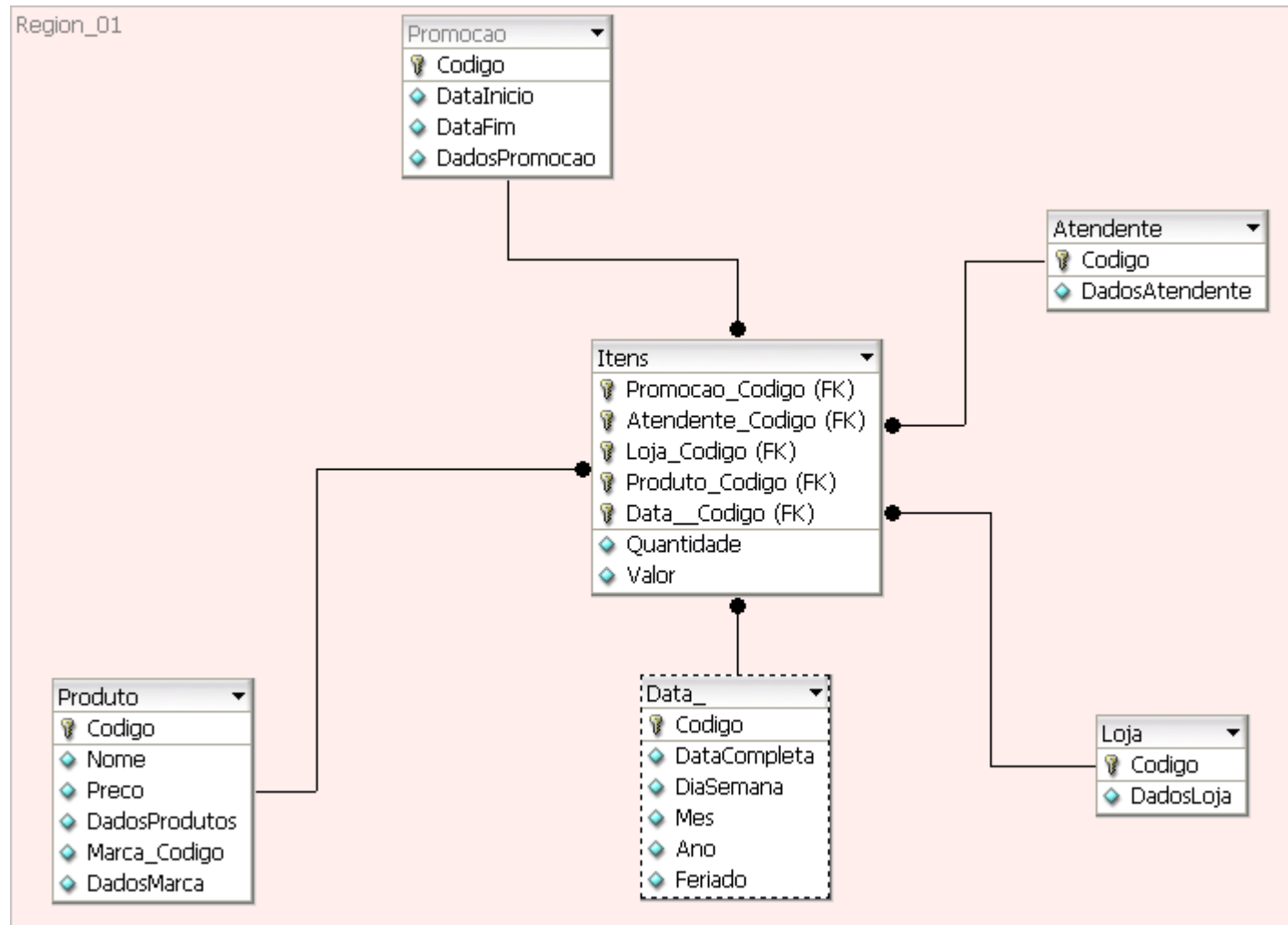
3. Escolher as dimensões

- *Dimensões principais*
 - *Data, Produto, Loja*
- *Outras dimensões descritivas relevantes possíveis (compatíveis com o grão escolhido)*
 - *Promoção, Atendente*

4. Identificar os fatos

- *Quantidade e Valor de cada venda*

Exemplo: esquema resultante



Dez Armadilhas a evitar no projeto de DW

(a maioria válida para qualquer projeto de sistema)

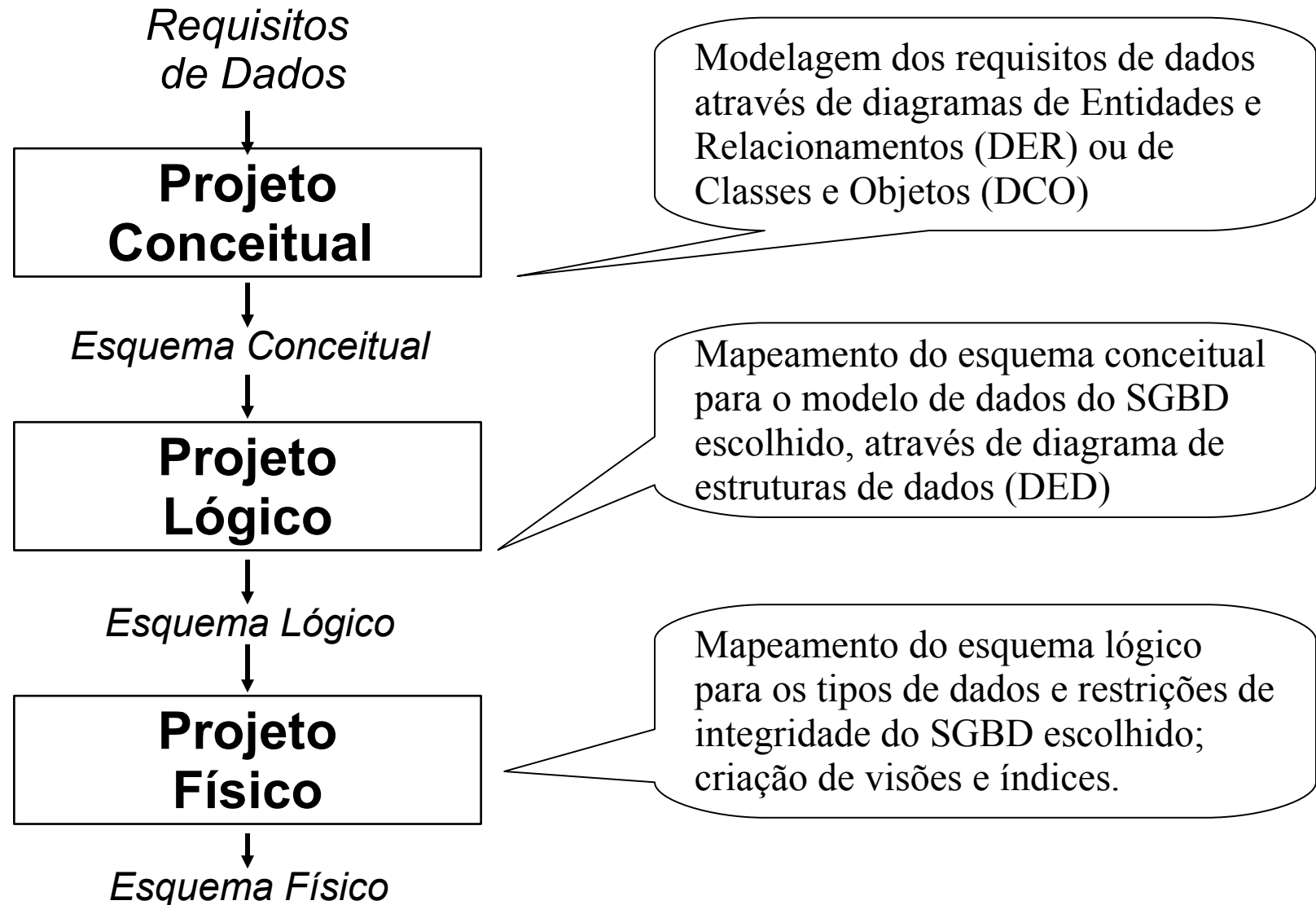
1. Negligenciar o reconhecimento de que o sucesso do DW está amarrado à aceitação do usuário.
2. Presumir que o negócio, seus requisitos e análises, assim como os dados subjacentes e a tecnologia, são estáticos.
3. Carregar somente dados sumarizados nas estruturas dimensionais da área de apresentação.
4. Popular modelos dimensionais de forma isolada, sem levar em conta a arquitetura que os amarra juntos usando dimensões compartilhadas e conformadas
5. Tornar os dados supostamente consultáveis na área de apresentação desnecessariamente complexos.

Dez Armadilhas a evitar no projeto de DW

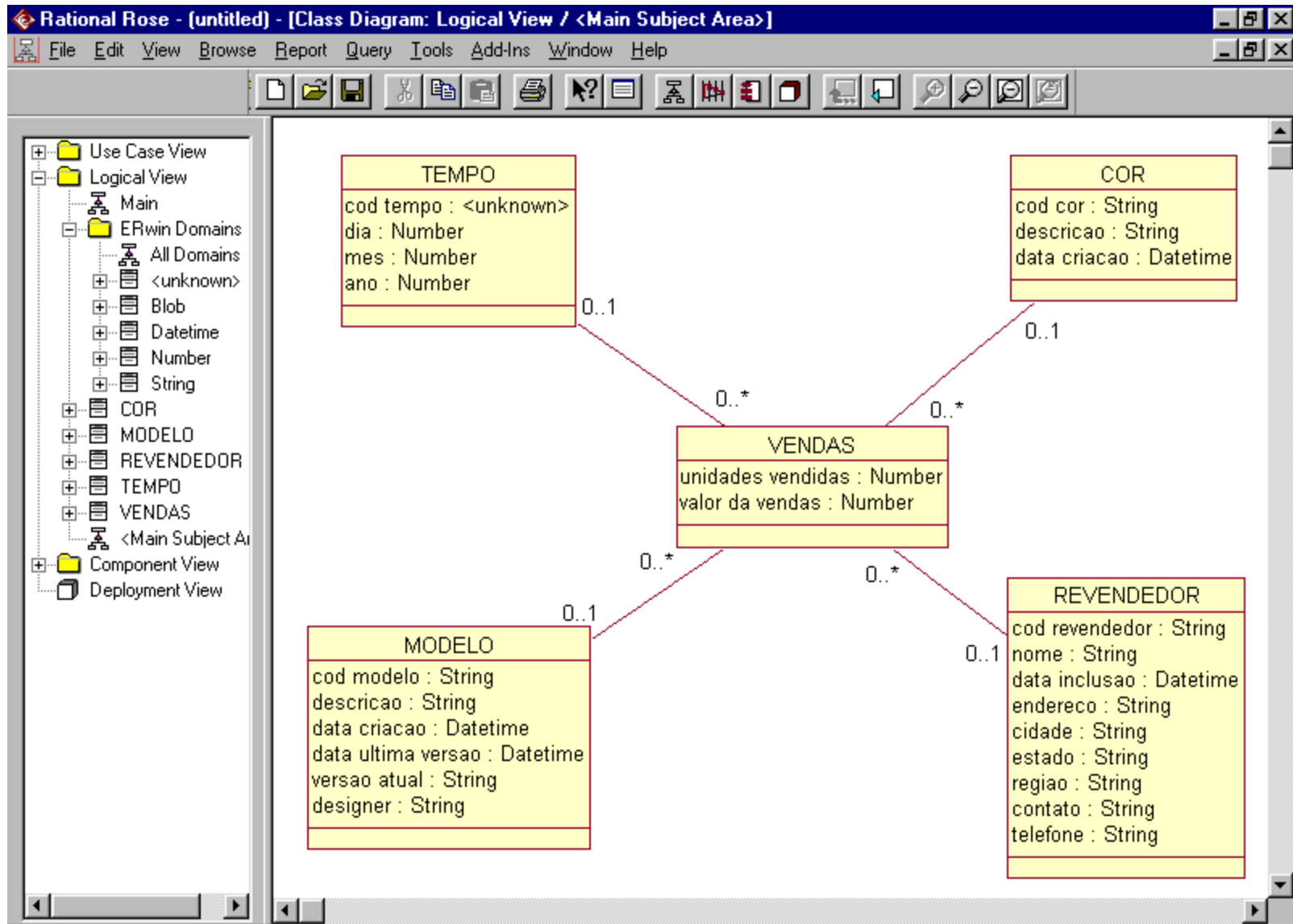
(a maioria válida para qualquer projeto de sistema)

6. Prestar mais atenção no desempenho operacional e na facilidade de desenvolvimento do “back-room” do que no desempenho de consultas e facilidade de uso do “front-room”.
7. Alocar energia para construir uma estrutura de dados normalizada, mesmo estourando o orçamento, do que para construir um área de apresentação viável baseada no modelo dimensional.
8. Atacar um projeto galático plurianual ao invés de perseguir esforços de desenvolvimento mais gerenciáveis, porém ainda desafiadores e iterativos.
9. Falhar em identificar e adotar uma gerência influente, acessível e razoavelmente visionária como patrocinador do negócio.
10. Tornar-se enamorado da tecnologia e dos dados ao invés de focar nos requisitos e objetivos do negócio.

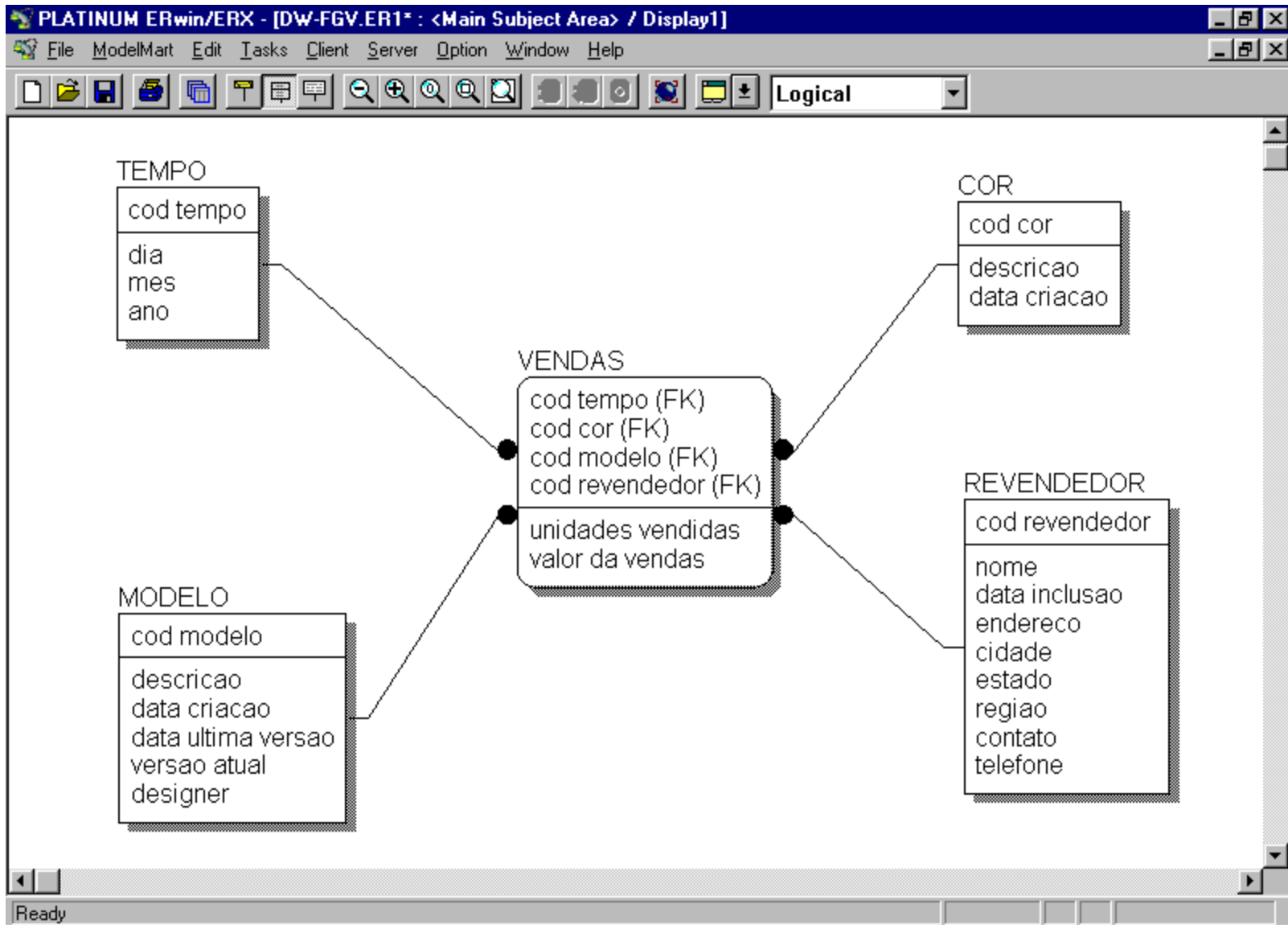
Projeto de Data Warehouse = Projeto de Bancos de Dados



Esquema Estrela - Conceitual



Esquema Estrela - Lógico



Implementação do Modelo Dimensional

- **SGBDs multidimensionais**

- Implementam fisicamente o modelo dimensional
- Problemas de desempenho, segurança e confiabilidade
- Problema de esparsidade: células onde não há dados (nulos)

- **SGBDs relacionais**

- Maior aceitação (força do mercado de BD relacional)
- Exige mapeamento (como qualquer projeto de BD relacional)

Escolha do SGBD

Computer Associates ERwin -- Target Server

Target SQL DBMS

AS/400 Ingres ORACLE SAS Ieradata

DB2 InterBase PROGRESS SQL Server WATCOM/SQL Anywhere

HiRDB ODBC/Generic Rdb SQLBase SYBASE

Red Brick

Target Desktop DBMS

Access FoxPro

Clipper dBASE IV Paradox

Red Brick Version

5.x

Default Red Brick Datatype

char(18)

Default Non-Key Null Option

NOT NULL

OK

Cancel

Esquema Estrela - Físico (Dimensional)

