

Business Intelligence

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



Modelagem Dimensional – Conceitos Básicos

Material baseado em originais de Maria Luiza Campos – NCE/UFRJ

Livro-Texto: [The Data Warehouse Toolkit, Third Edition: The Definitive Guide to Dimensional Modeling](#)

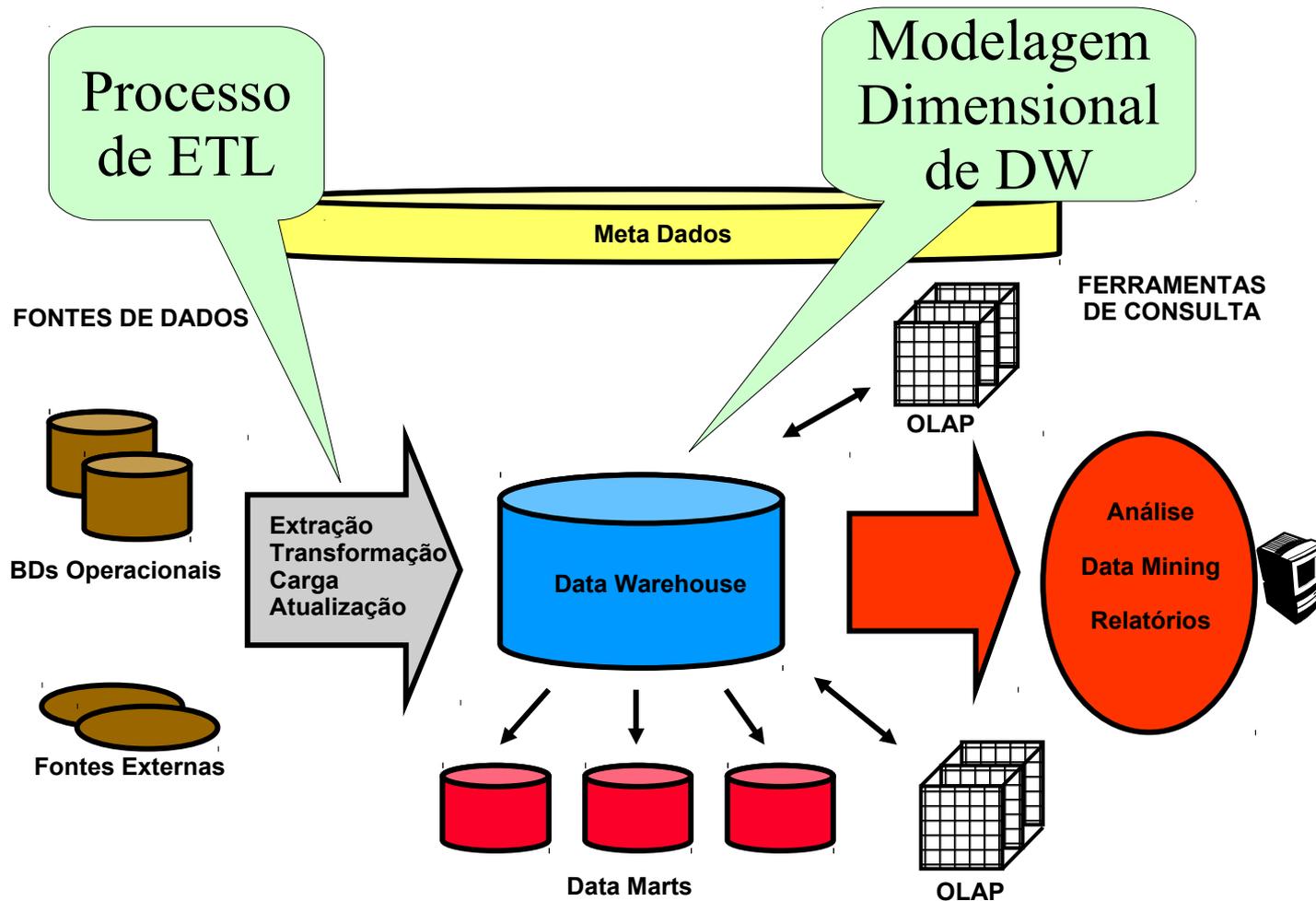
Ralph Kimball and Margy Ross, Wiley, 2013

Complementado com referências atuais de Ralph Kimball (<http://www.kimballgroup.com/>)

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



A seguir ???



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

Introdução à Modelagem Dimensional de DW

- Conceitos de modelagem dimensional
 - Data Warehouse(ing) segundo Kimball
 - 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 de DW como projeto de BD

Modelagem de DW para OLAP

Data Warehouses

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

▶ **MODELAGEM DIMENSIONAL**

- facilita o desenvolvimento iterativo e incremental

“pensar grande, começar pequeno e evoluir rápido”

Modelagem Multidimensional

- Proposta por Ralph Kimball para projeto de DW
 - Dimensional Modeling Manifesto, 1997
 - <http://www.kimballgroup.com/1997/08/02/a-dimensional-modeling-manifesto/>
 - O próprio Kimball credita origem a projeto de uma empresa (General Mills) com uma universidade (Dartmouth) nos anos 1960s.
- Dominante no mercado, 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*)

Data Warehouse segundo Kimball

Modelagem
Dimensional
de DW

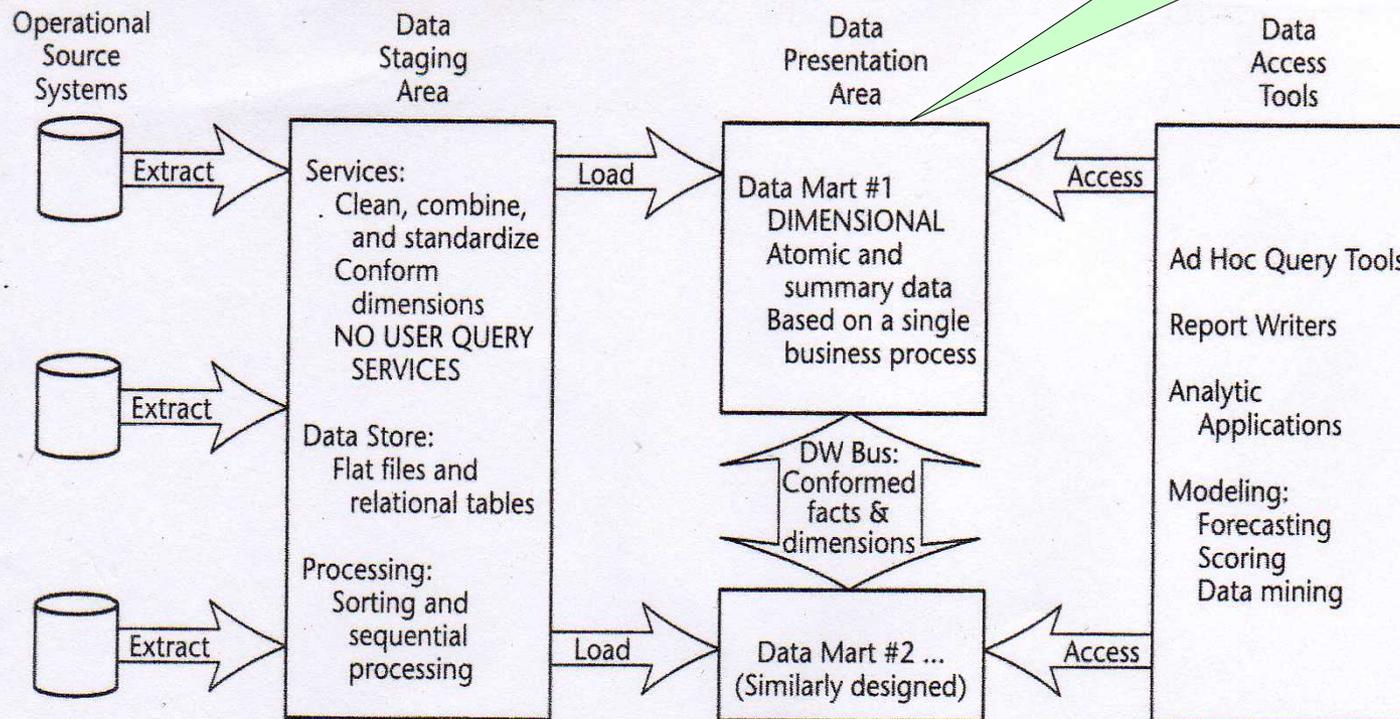
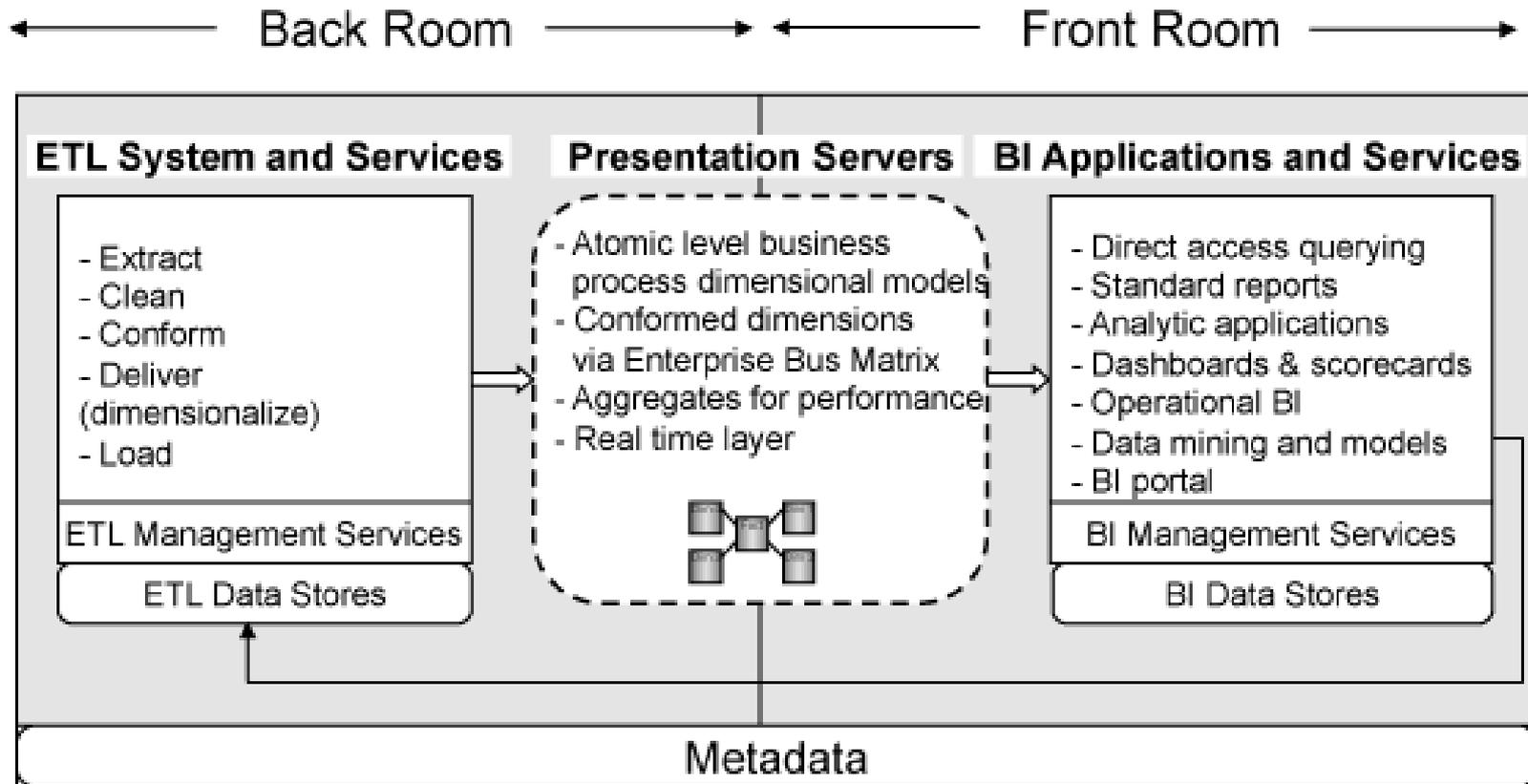


Figure 1.1 Basic elements of the data warehouse.

Segundo Kimball, DW = Conjunto de Data Marts Integrados, com Fatos e Dimensões Conformados
Ralph Kimball, Margy Ross: The Data Warehouse Toolkit, 2ª Edição, Wiley, 2002

Data Warehouse (Kimball 2013)

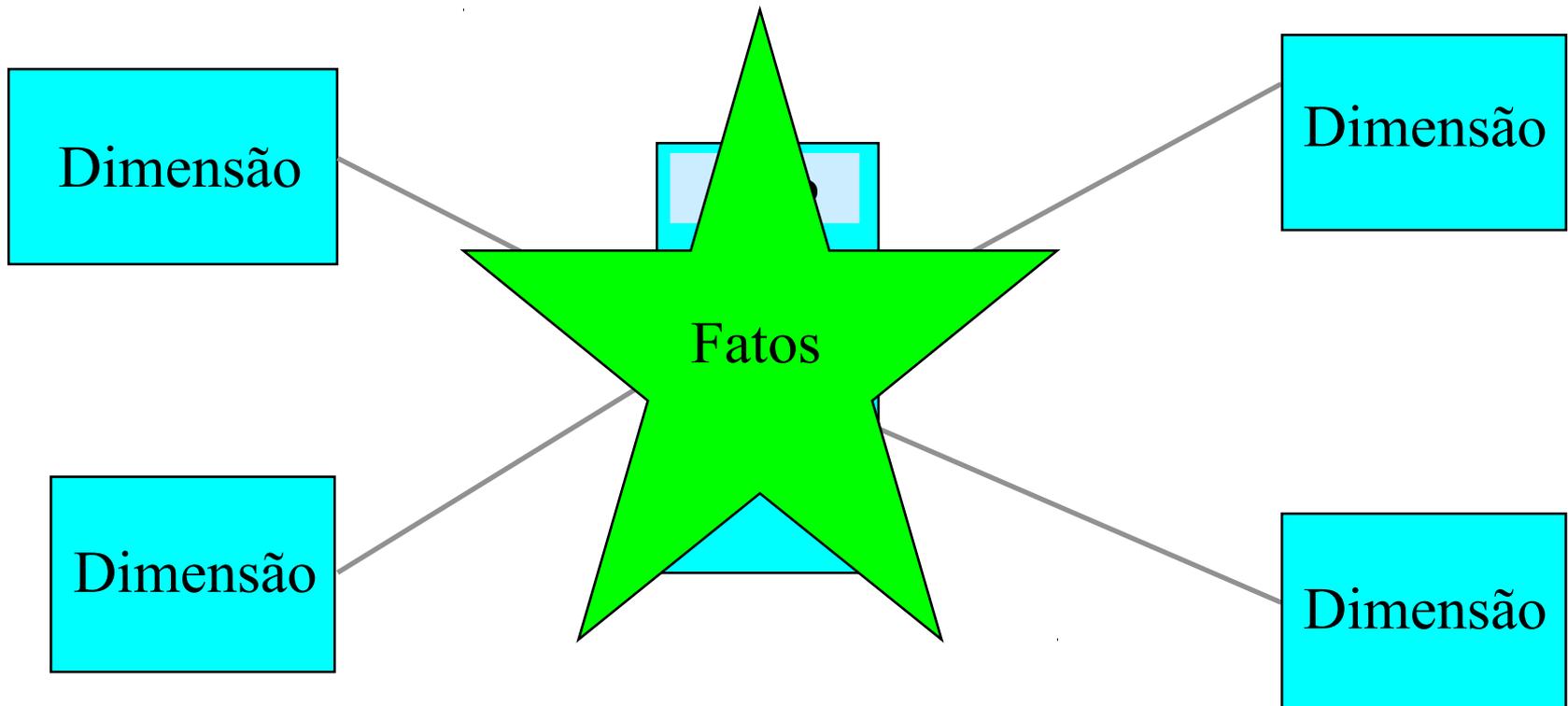


<http://www.kimballgroup.com/data-warehouse-and-business-intelligence-resources/kimball-core-concepts/>

<http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/technical-dw-bi-system-architecture/>

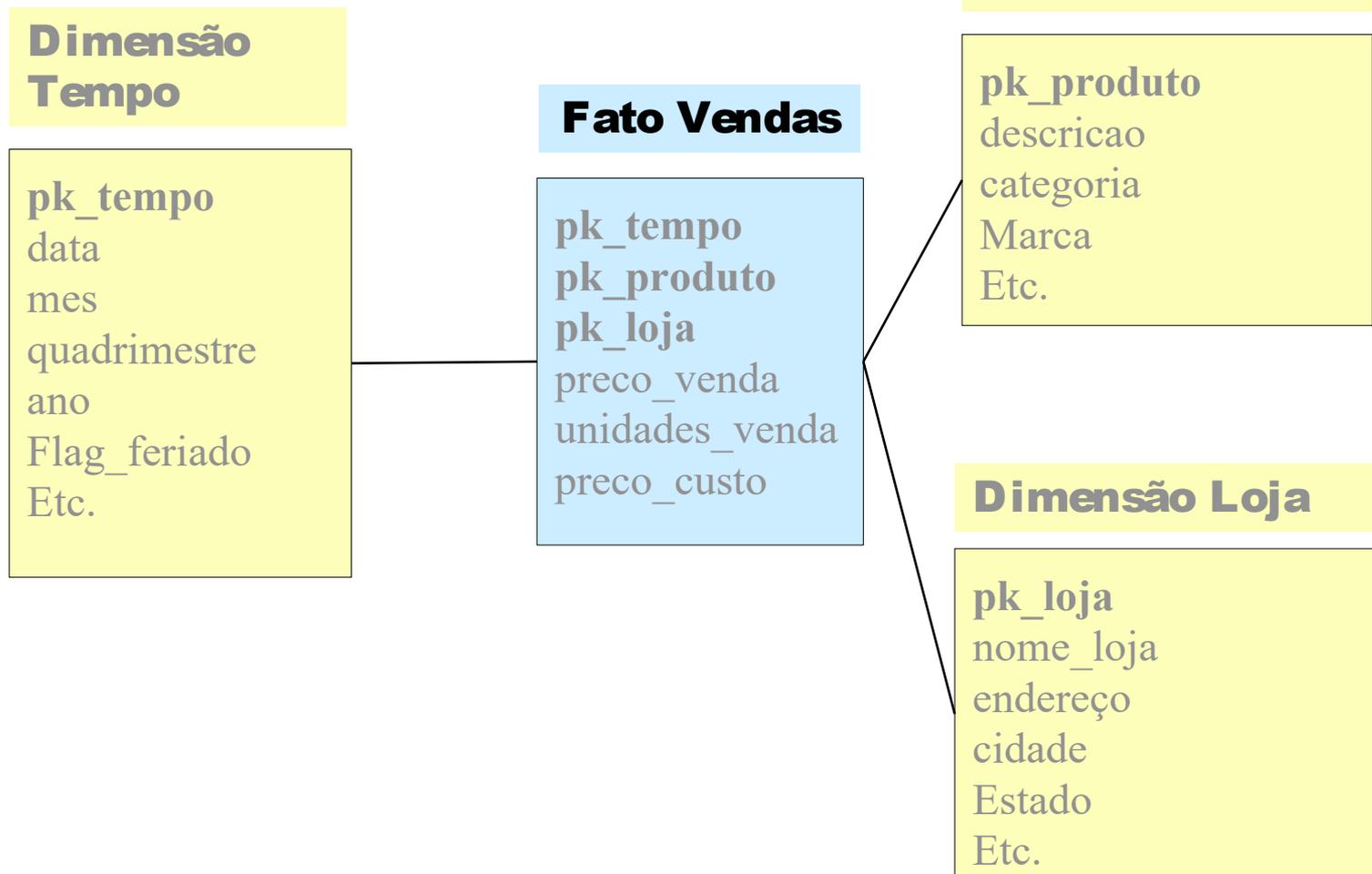
Esquema Estrela (Star Schema)

Uma tabela de fatos cercada de tabelas de dimensões



Também conhecido como “Star Join”, no modelo relacional

Esquema Estrela - Exemplo



Exemplo Consultas

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

“Vendas por marca entre 2010 e 2013”

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

Modelo ER x Dimensional

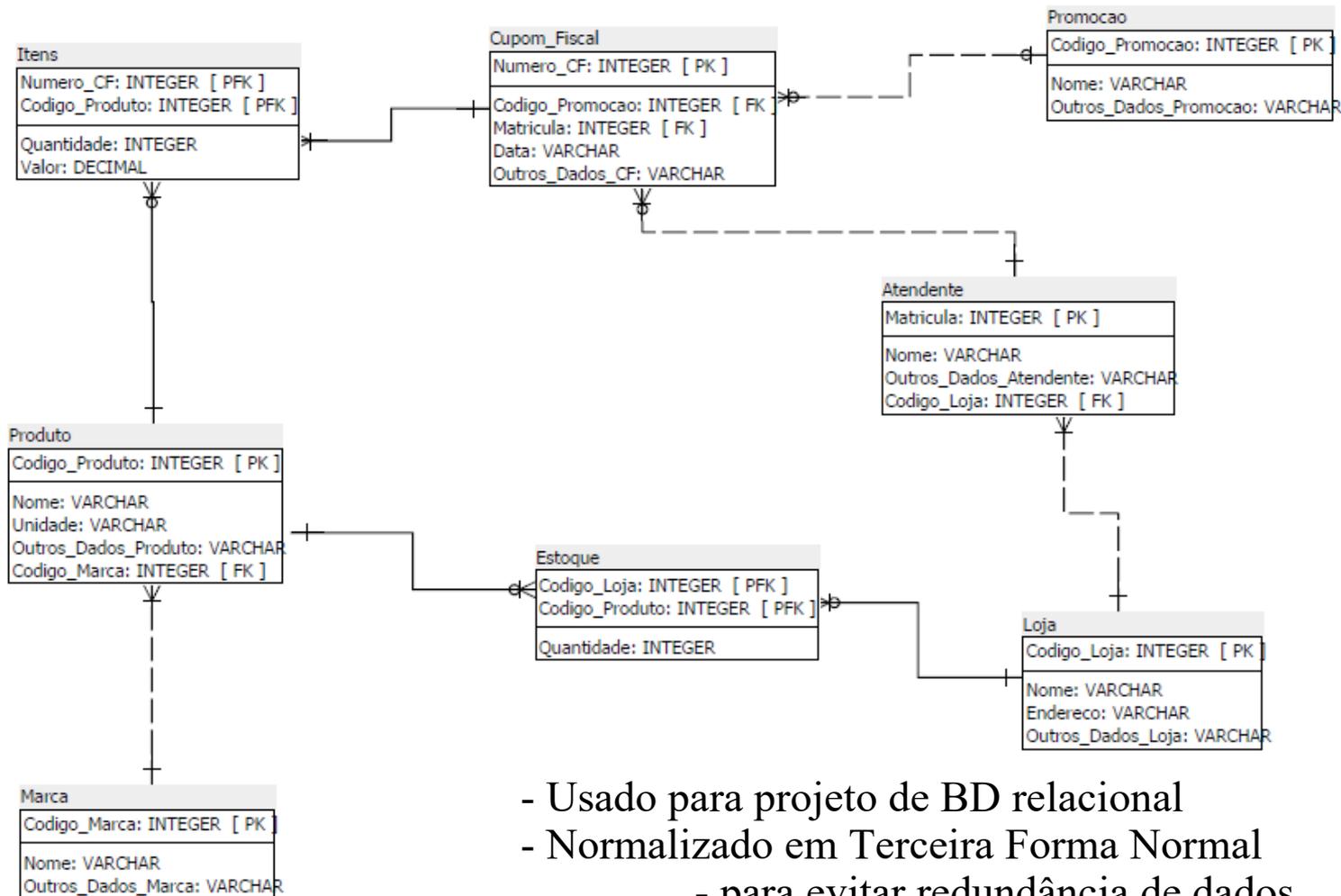
• Modelo ER/Relacional

- Mais complexo
- Anos 70 – BD relacional
- Tabelas representam conjuntos de entidades e relacionamentos
- Tabelas resultantes naturalmente normalizadas (até 3ª Forma Normal)
- Tabelas acessadas indistintamente de filtro inicial
- Maior necessidade e dificuldade de junção
- Maior dificuldade de leitura e de consulta por usuário não especializado

• Modelo Dimensional

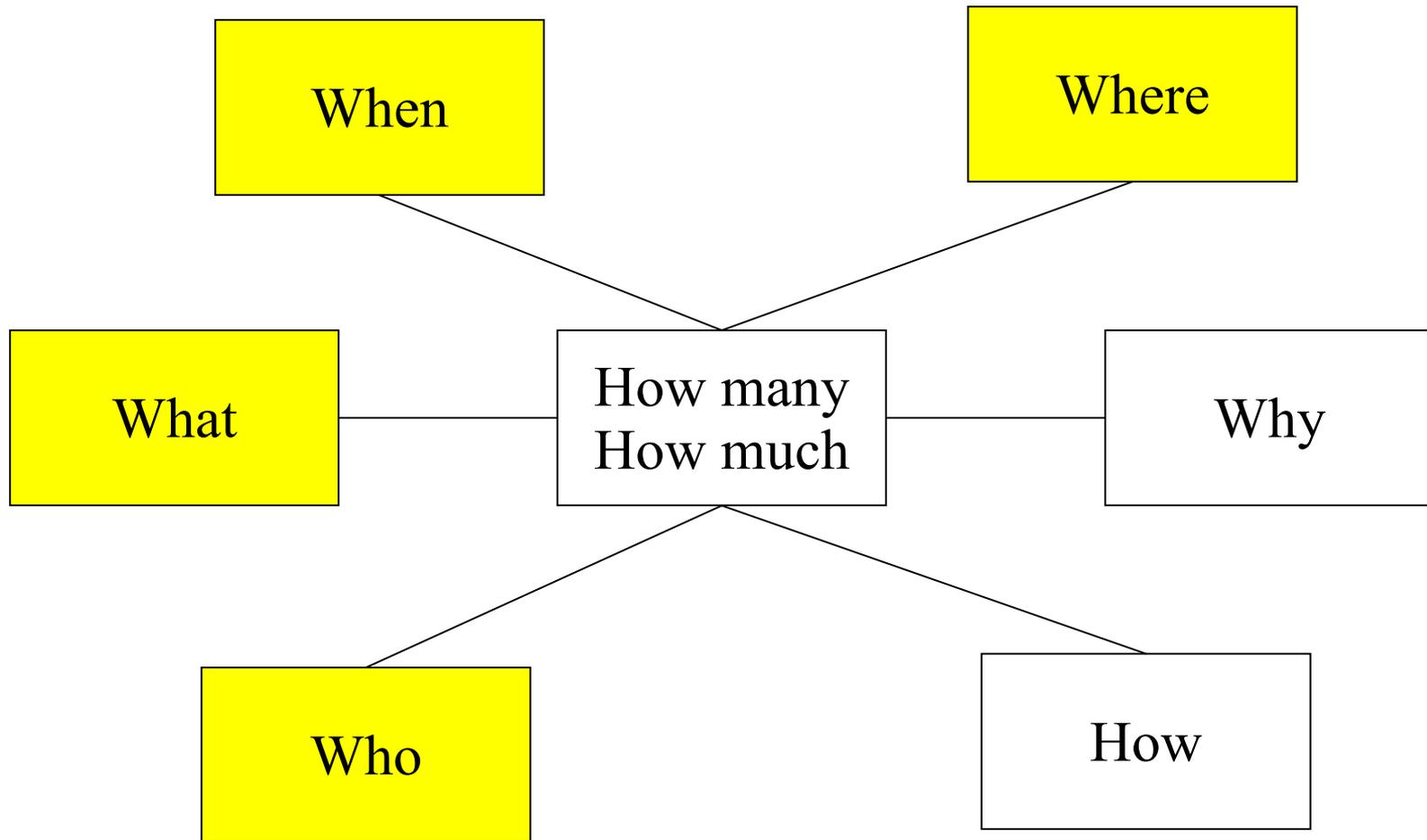
- Estrutura mais fácil e intuitiva
- Anterior ao ER, recriada por Kimball
- Tabelas representam Fatos e Dimensões
- Tabelas Fato normalizadas, Tabelas Dimensão podem não ser normalizadas
- Tabelas Dimensão são pontos de entrada para acesso
- Junções só ocorrem entre Tabelas Fato e Dimensões
- Leitura e consulta mais fáceis para usuários não especializados

Modelo Entidades Relacionamentos



- Usado para projeto de BD relacional
- Normalizado em Terceira Forma Normal
 - para evitar redundância de dados
- Adequado para aplicações operacionais (OLTP)
 - otimizado para transações
 - armazena valores correntes

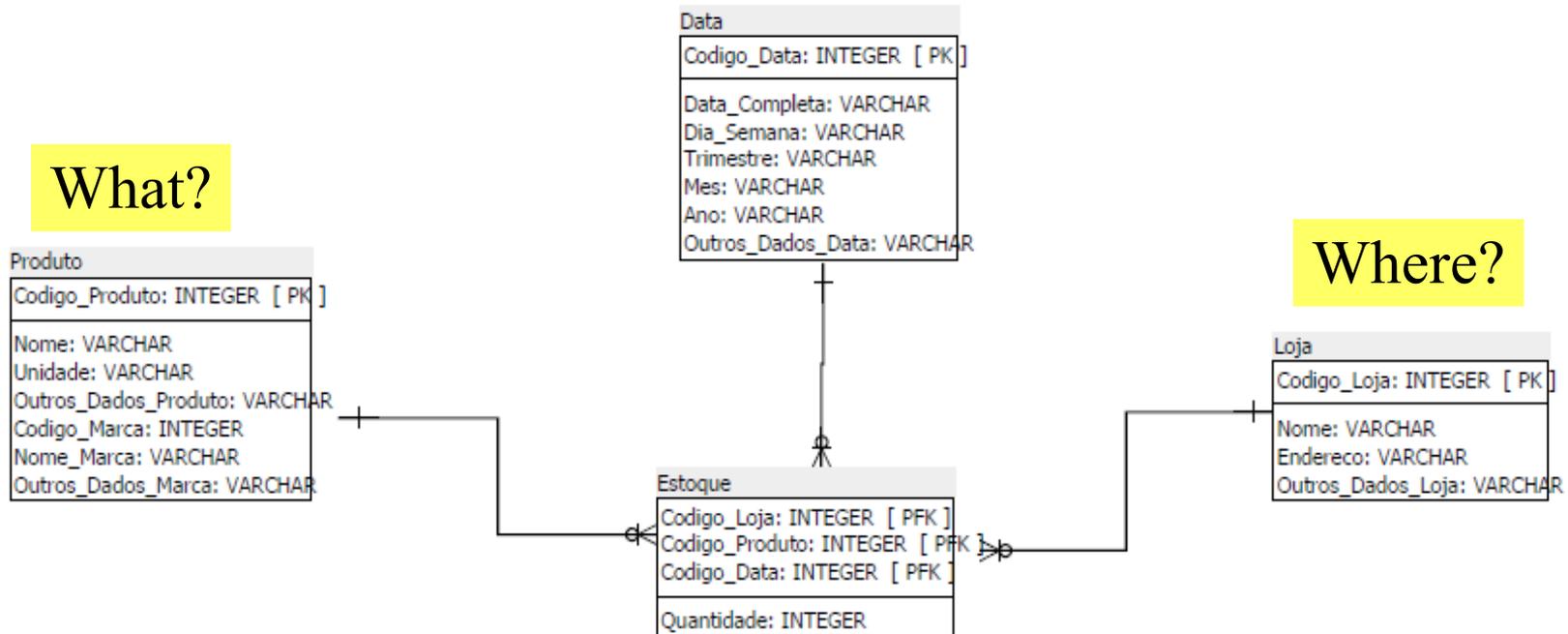
Esquema Estrela de DW 5 W e 3 H (vide BPM)



 Tipos de dimensões mais comuns

When?

What?

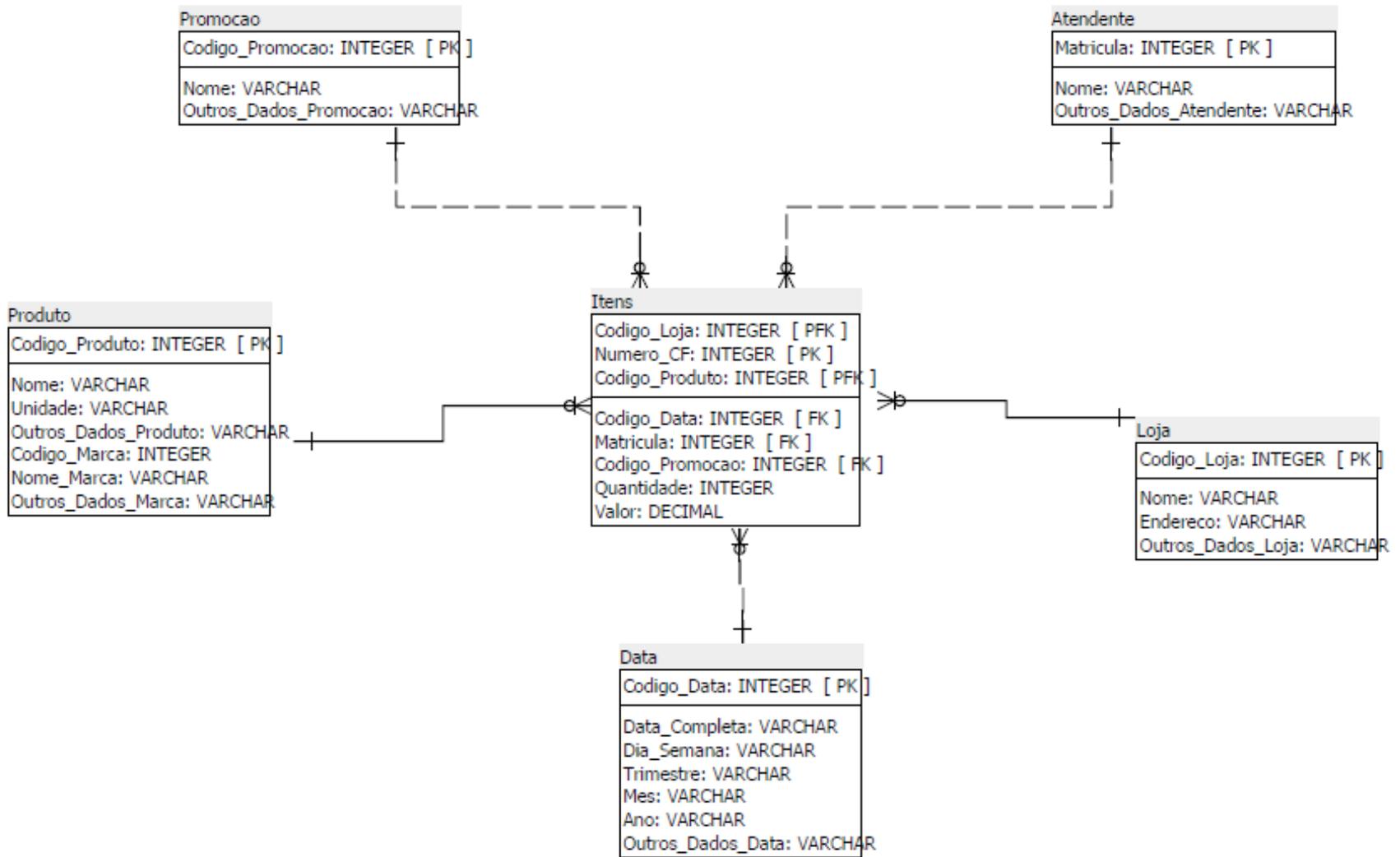


Where?

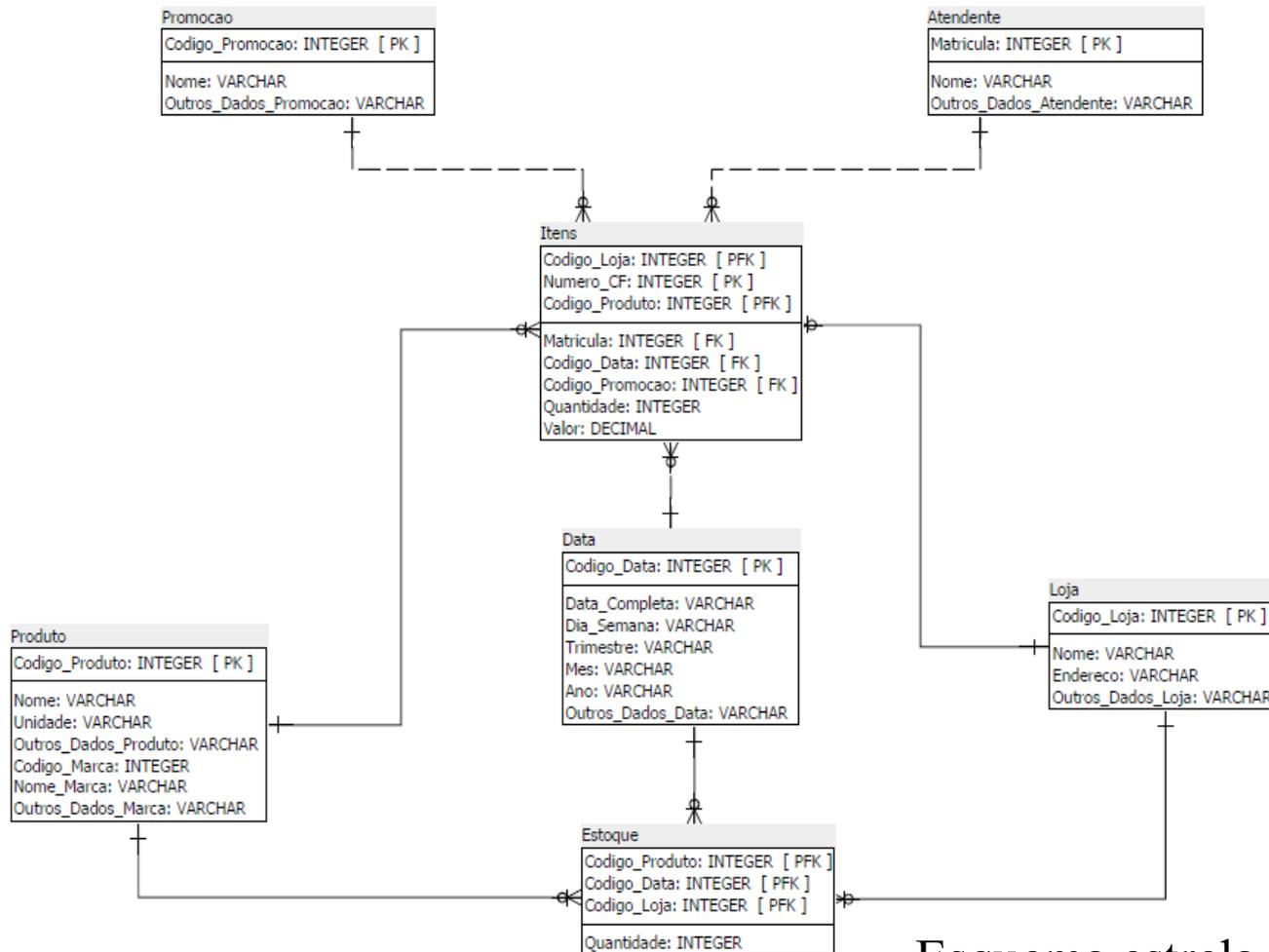
Esquema estrela com tabela de fatos Estoque e dimensões Produto, Loja e Data

- Contém dados redundantes nas tabelas de dimensões (desnormalizadas)
- Otimizado para consultas
- Armazena dados históricos, não somente os valores correntes

Note que a tabela de fatos Estoque é uma tabela associativa (relacionamento M:N) no modelo operacional



Esquema estrela com tabela de fatos Itens e dimensões Produto, Promoção, Atendente, Loja e Data
Note que a tabela de fatos Itens é uma tabela associativa (relacionamento M:N) no modelo operacional



Esquema estrela (constelação) com tabelas de fatos Itens e Estoque. As dimensões Produto, Loja e Data são compartilhadas e “conformadas” no mesmo nível de detalhamento

Mais comparações entre Modelagem ER e Multidimensional

Entidades e Relacionamentos	Multidimensional
1 diagrama representando vários processos de negócio que, em geral, só podem ser reconhecidos por navegação.	Vários diagramas dimensionais, um para cada processo de negócio, podendo estar conectados através de dimensões comuns.
Usuários acham difícil entender e navegar pelo diagrama.	Usuários reconhecem “o seu negócio” no diagrama.
Requer muitas junções para responder a consultas complexas.	Requer poucas junções, em geral, entre tabelas de fatos e tabelas de dimensões relevantes à consulta.
Dados atômicos e atuais (transações, em geral, sobrescrevem o dado anterior).	Dados atômicos históricos e alguns sumarizados (agregados) para consulta direta.
Planos de consultas extremamente distintos e específicos para as consultas e transações previstas. Otimizado para transações.	Planos de consultas “genéricos” possibilitados pela simetria do modelo. Otimizado para consultas.

Simetria do Esquema Estrela

- **Esquema Estrela é simétrico**
 - Diferentemente de esquemas ER operacionais, que são assimétricos.
 - Todas as dimensões servem como pontos de entrada simetricamente iguais para a tabela de fatos.
 - Esta simetria permite um padrão previsível do esquema “Star Join” que resiste a mudanças não previstas do comportamento do usuário ao consultar.
 - *“The user interfaces are symmetrical, the query strategies are symmetrical, and the SQL generated against the dimensional model is symmetrical.”*

Esquema Estrela - Fatos

- **Tabela de Fatos**

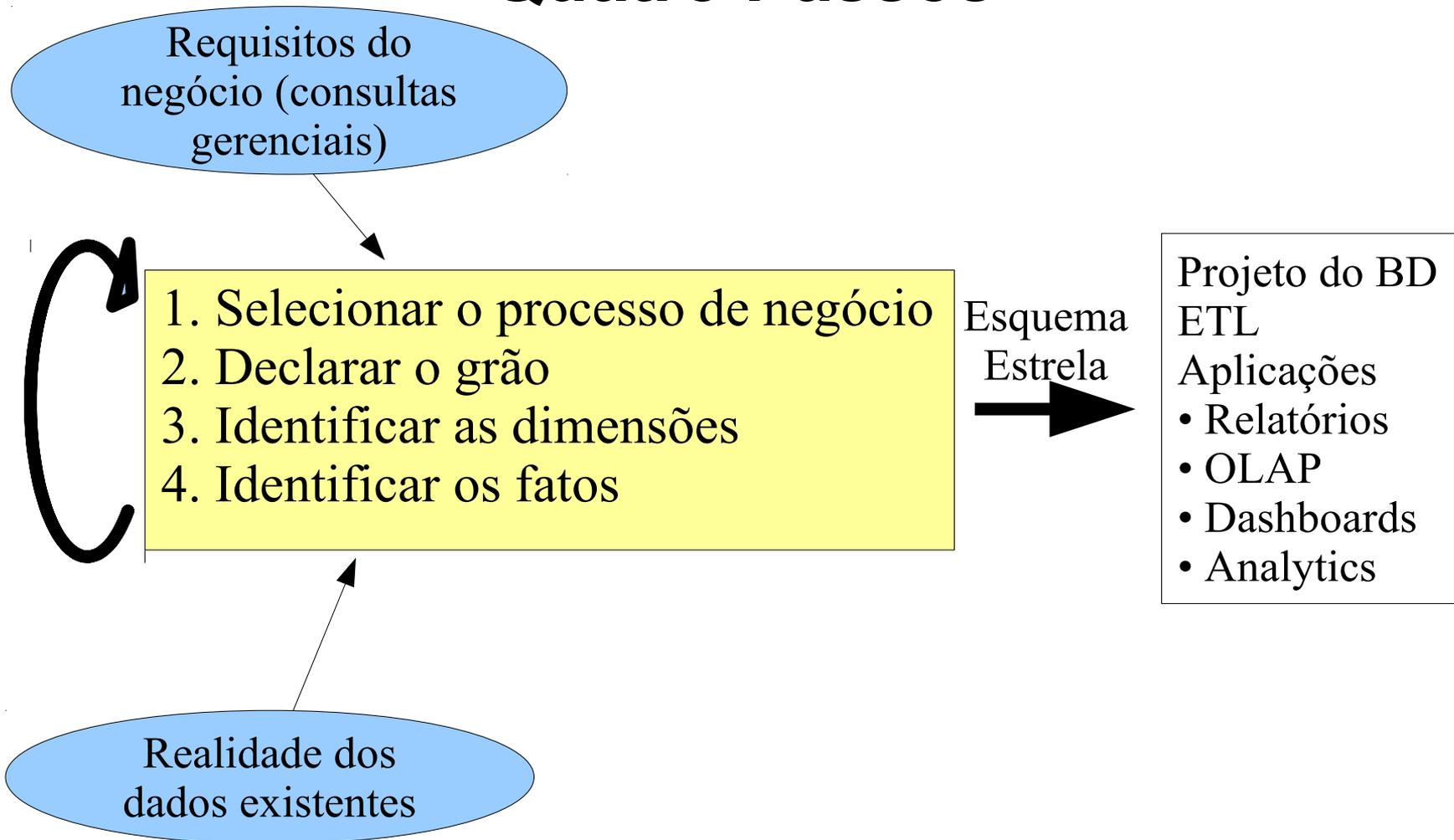
- Expressa relacionamento M:N do modelo ER operacional
 - » Uma espécie de tabela associativa entre dimensões
 - » Em geral, mas não necessariamente, os fatos possuem medidas armazenadas nesta tabela
 - » Guardam o histórico dos fatos ocorridos no ambiente operacional
- Tabela de fatos é dominante no DW
 - » Usualmente possuem grande volume de dados; ocupam 90% do espaço em um DW típico
 - » Tendem a ter muitas linhas e poucas colunas

Esquema Estrela – Dimensões

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 da forma “por dimensão” (vendas “por semana”, “por marca”, “por loja”) e são as bases para agregações e agrupamentos.
- Grau de detalhamento da dimensão depende da granularidade (grão) do fato.
- Volume de dados (linhas) bem menor que nas tabelas de fatos (dimensões não guardam histórico)
- O poder de análise de um DW é diretamente proporcional à qualidade e profundidade dos atributos das dimensões.

Processo da Modelagem Dimensional Quatro Passos



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 (ou módulo de um ERP).
- Exemplos: vendas, compras, pedidos, controle de estoque, 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 e em que nível de detalhe o pessoal do negócio descreve os dados que resultam do processo de negócio?”
- Exemplos de dimensões: data, produto, loja, 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.

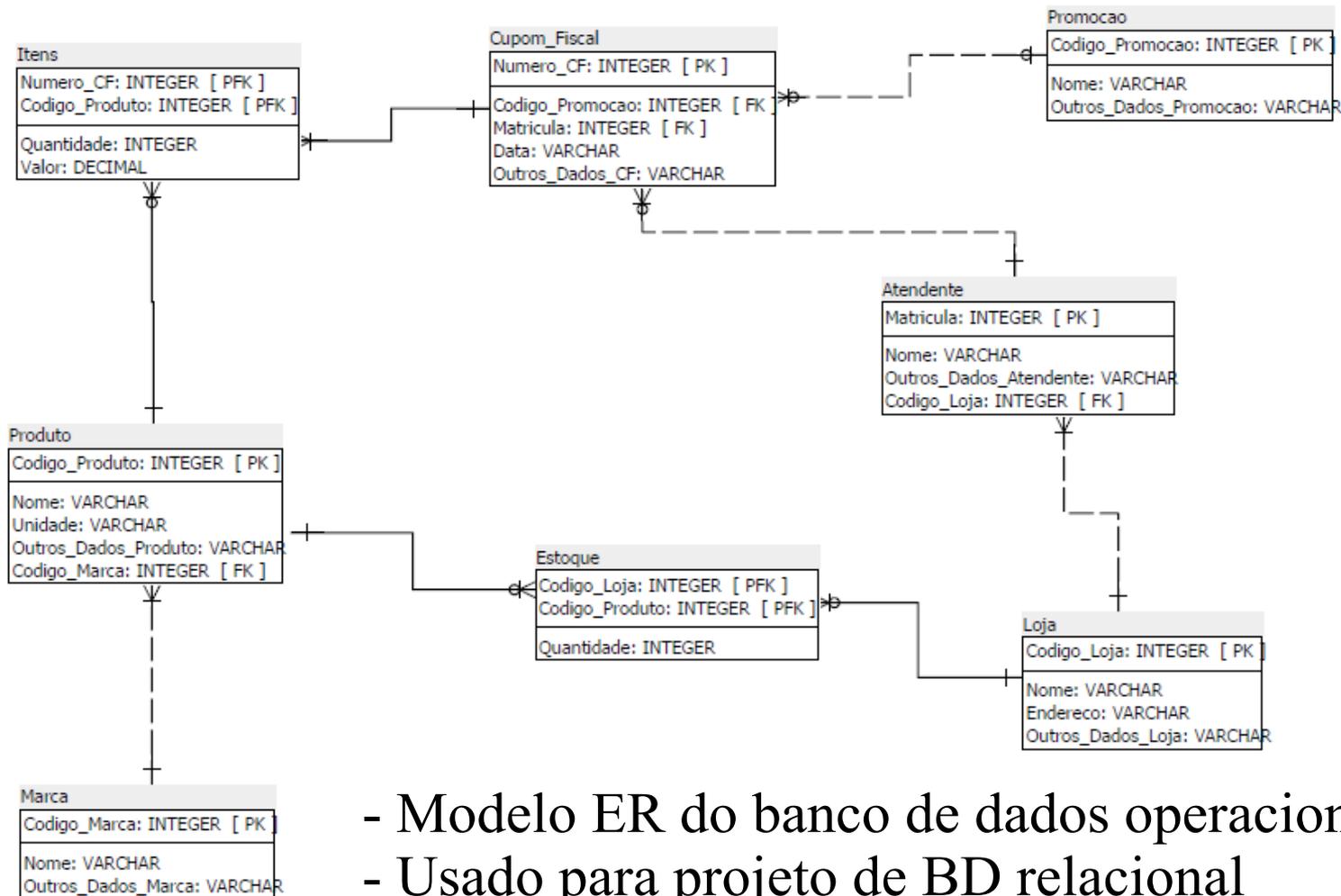
Orientações (Guidelines) para Modelagem Dimensional

- ***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. Se não existir, faça uma engenharia reversa do banco de dados operacional.
 - 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 (tabelas associativas), 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.

Orientações (Guidelines) para 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.

Exemplo de processo de modelagem dimensional



- Modelo ER do banco de dados operacional
- Usado para projeto de BD relacional
- Normalizado em Terceira Forma Normal
- Adequado para aplicações operacionais (OLTP)

Exemplo de Modelagem Dimensional – Quatro passos

1. Selecionar o processo de negócio

- *Vendas no caixa de loja*

2. Declarar o grão

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

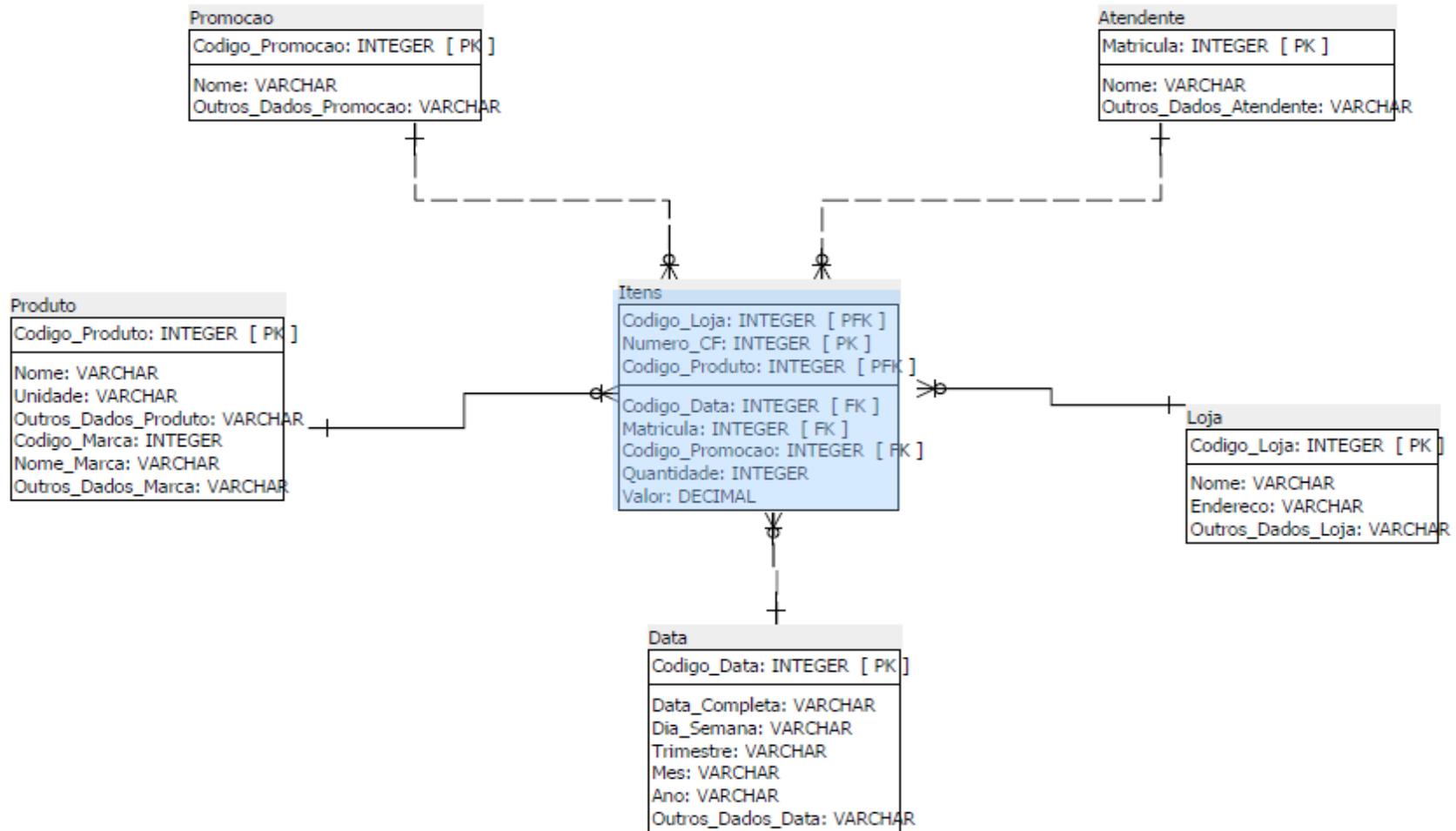
3. Escolher as dimensões

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

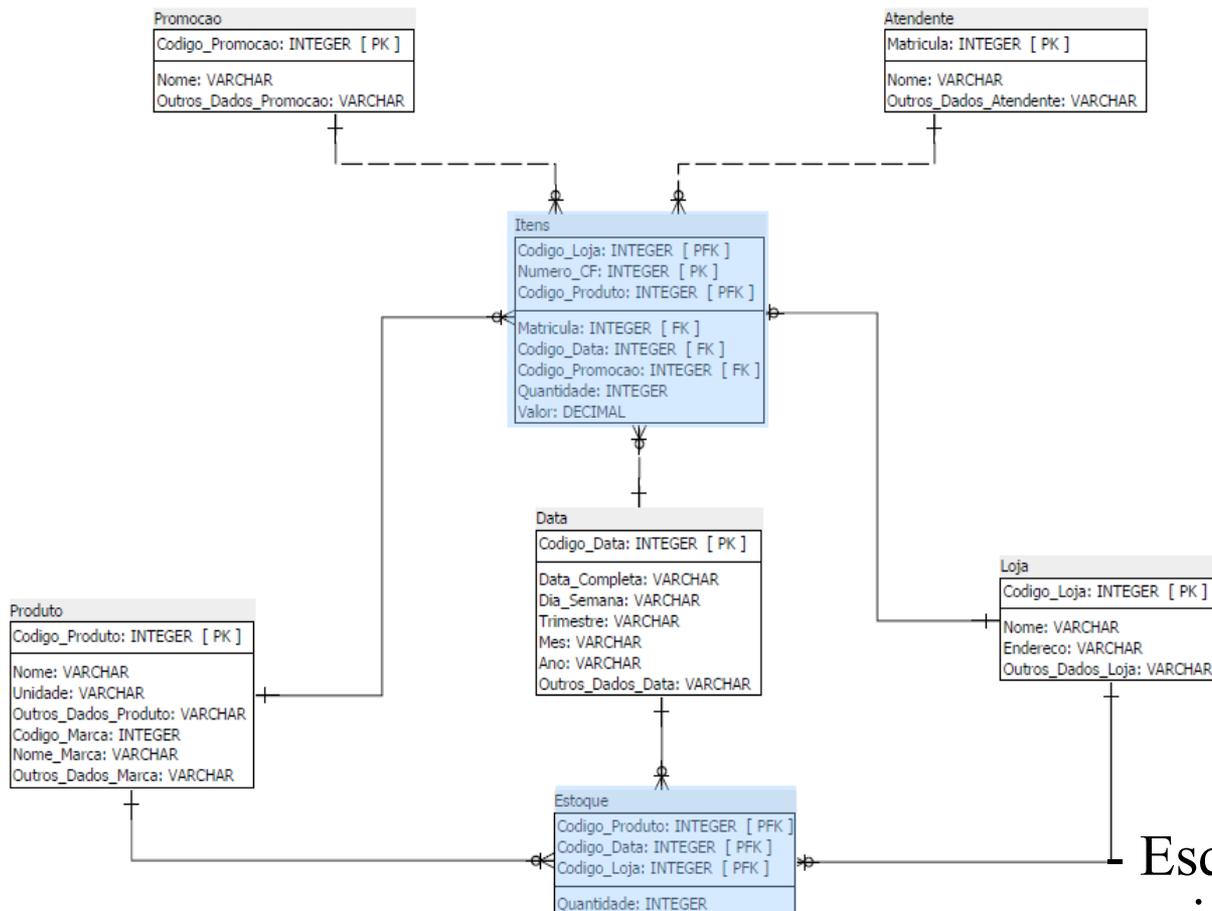
4. Identificar os fatos

- *Quantidade e Valor de cada venda*

Esquema resultante para o Processo de Negócio escolhido



Modelo Dimensional (Esquema Estrela)



Esquema dimensional
equivalente ao ER anterior
- Em geral desnormalizado
- Adequado para aplicações
analíticas (OLAP)

Data Warehouse Bus Architecture

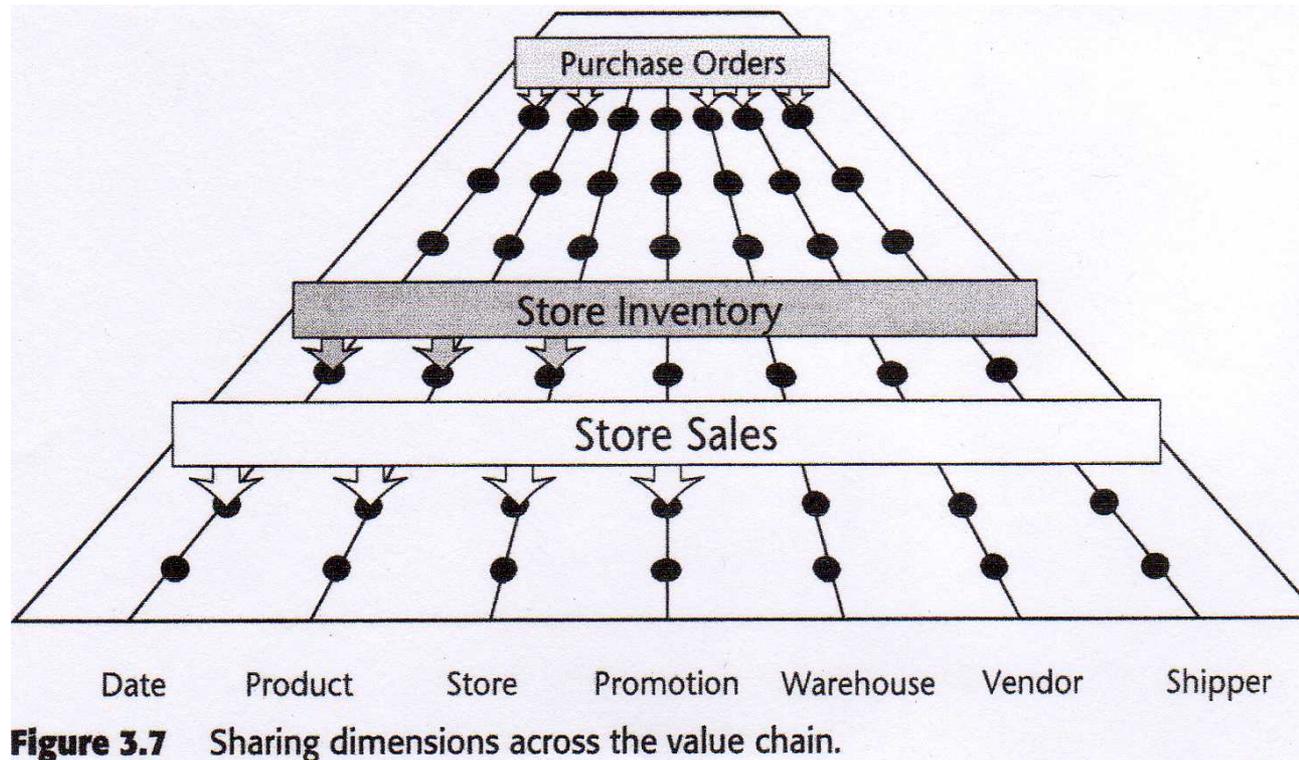


Figure 3.7 Sharing dimensions across the value chain.

Como compartilhar dimensões por diversos processos de negócio

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.

Tabela de Fatos – Conceitos Básicos

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)
 - A tabela fato Estoque requer que todas as chaves estrangeiras de dimensões façam parte da chave primária. O mesmo não acontece com a tabela fato Itens, devido à presença do atributo Numero_CF.
- 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 – há controvérsia sobre *quantidade* (depende do negócio)
- **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 pode ser somada ao longo da dimensão Produto e Loja (dependendo do negócio). Ao longo de Data, a soma não faria 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

- **Alguns atributos numéricos podem não ser somados.**
 - Ex: preço unitário, taxas e percentagens.
- **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

Conceitos Básicos

- **Objetivo das dimensões**

- 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 (> 4 bilhões se for UNSIGNED INTEGER)
- 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	Chave Primária
Produto_Codigo	Chave Primária
NomeLoja	Dependente de Loja_Codigo
DadosLoja	Dependente de Loja_Codigo
NomeProduto	Dependente de Produto_Codigo
DadosProduto	Dependente de Produto_Codigo
Quantidade	Dependente de Loja_Codigo e Produto_Codigo
Valor	Dependente de Loja_Codigo e Produto_Codigo

Tabela Estoque_1 não está em 2FN

Loja_Codigo → {Nome_Loja, Dados_Loja}

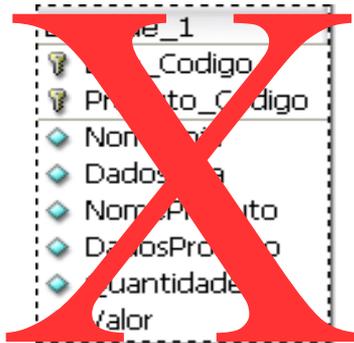
Produto_Codigo → {Nome_Produto, Dados_Produto}

Estoque_2	
Loja_Codigo	Chave Primária
Produto_Codigo	Chave Primária
Quantidade	Dependente de Loja_Codigo e Produto_Codigo
Valor	Dependente de Loja_Codigo e Produto_Codigo

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

Formas normais & Esquema Estrela

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

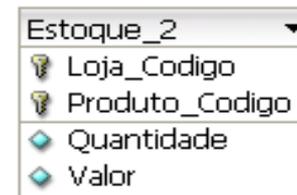


Loja_Codigo	Produto_Codigo	Nome_Produto	Dados_Produto	Quantidade	Valor
-------------	----------------	--------------	---------------	------------	-------

Tabela Estoque_1 não está em 2FN

Loja_Codigo → {Nome_Loja, Dados_Loja}

Produto_Codigo → {Nome_Produto, Dados_Produto}



Loja_Codigo	Produto_Codigo	Quantidade	Valor
-------------	----------------	------------	-------

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

PRODUTO
Cod_Produto: INTEGER [PK]
Descricao: VARCHAR
Unidade: VARCHAR
Peso: DECIMAL
Outros_Dados_Produto: VARCHAR
Cod_Marca: INTEGER
Nome_Marca: VARCHAR
Outros_Dados_Marca: VARCHAR
Cod_Categoria: INTEGER
Nome_Categoria: VARCHAR
Outros_Dados_Categoria: VARCHAR

Tabela Produto está em 2FN
mas não em 3FN

Cod_Marca → {Nome_Marca, Outros_Dados_Marca}

Cod_Categoria → {Nome_Categoria, Outros_Dados_Categoria}

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.*

PRODUTO
Cod_Produto: INTEGER [PK]
Descricao: VARCHAR
Unidade: VARCHAR
Peso: DECIMAL
Outros_Dados_Produto: VARCHAR
Cod_Marca: INTEGER
Nome_Marca: VARCHAR
Outros_Dados_Marca: VARCHAR
Cod_Categoria: INTEGER
Nome_Categoria: VARCHAR
Outros_Dados_Categoria: VARCHAR

PRODUTO
Cod_Produto: INTEGER [PK]
Descricao: VARCHAR
Unidade: VARCHAR
Peso: DECIMAL
Outros_Dados_Produto: VARCHAR
Cod_Marca: INTEGER [FK]
Cod_Categoria: INTEGER [FK]

Marca
Cod_Marca: INTEGER [PK]
Nome_Marca: VARCHAR
Outros_Dados_Marca: VARCHAR

Categoria
Cod_Categoria: INTEGER [PK]
Nome_Categoria: VARCHAR
Outros_Dados_Categoria: VARCHAR

Tabela Produto agora está em 3FN.

Tabela Produto está em 2FN
mas não em 3FN

Cod_Marca → {Nome_Marca, Outros_Dados_Marca}

Cod_Categoria → {Nome_Categoria, Outros_Dados_Categoria}

Dívida: Terceira forma normal

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

PRODUTO
Cod_Produto: INTEGER [PK]
Descricao: VARCHAR
Unidade: VARCHAR
Peso: DECIMAL
Outros_Dados_Produto: VARCHAR
Cod_Marca: INTEGER
Nome_Marca: VARCHAR
Outros_Dados_Marca: VARCHAR
Cod_Categoria: INTEGER
Nome_Categoria: VARCHAR
Outros_Dados_Categoria: VARCHAR

PRODUTO
Cod_Produto: INTEGER [PK]
Descricao: VARCHAR
Unidade: VARCHAR
Peso: DECIMAL
Outros_Dados_Produto: VARCHAR
Cod_Marca: INTEGER [FK]
Cod_Categoria: INTEGER [FK]

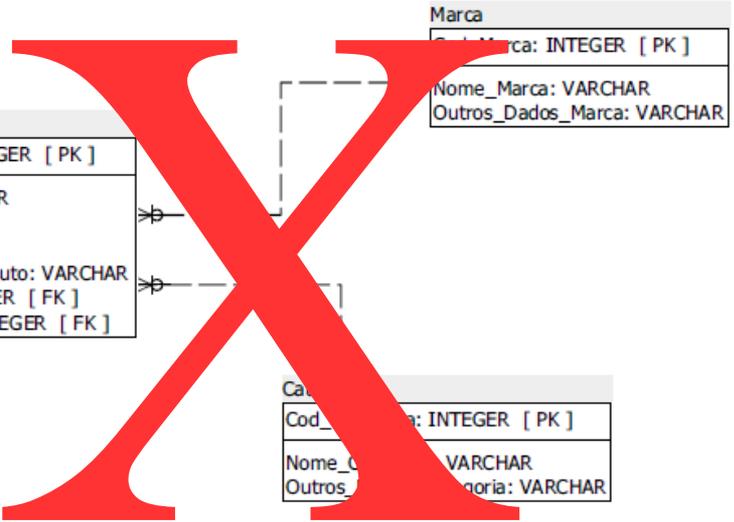


Tabela Produto está em 2FN
mas não em 3FN

- Cod_Marca → {Nome_Marca, Outros_Dados_Marca}
- Cod_Categoria → {Nome_Categoria, Outros_Dados_Categoria}

Tabela Produto agora
está em 3FN.

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
- No esquema estrela, as hierarquias estão implícitas na tabela de dimensão.

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.**

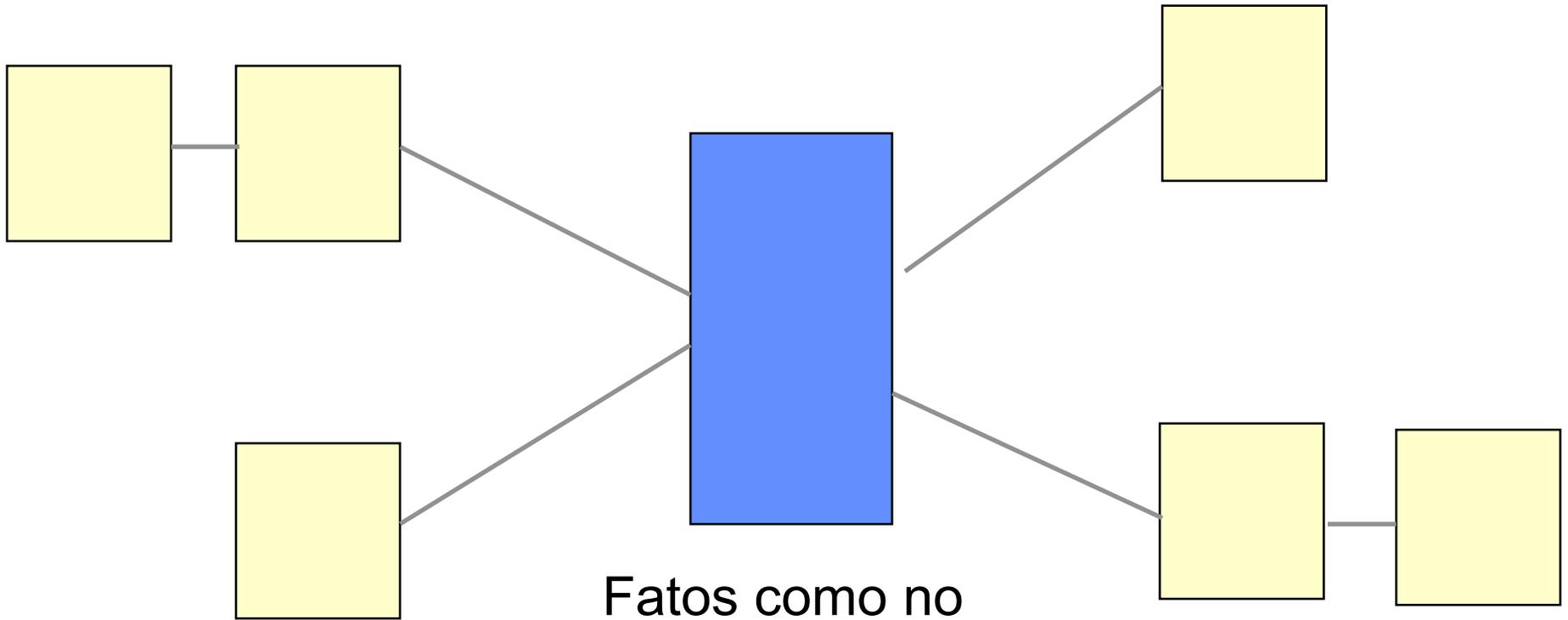
Variações do Esquema Estrela

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**

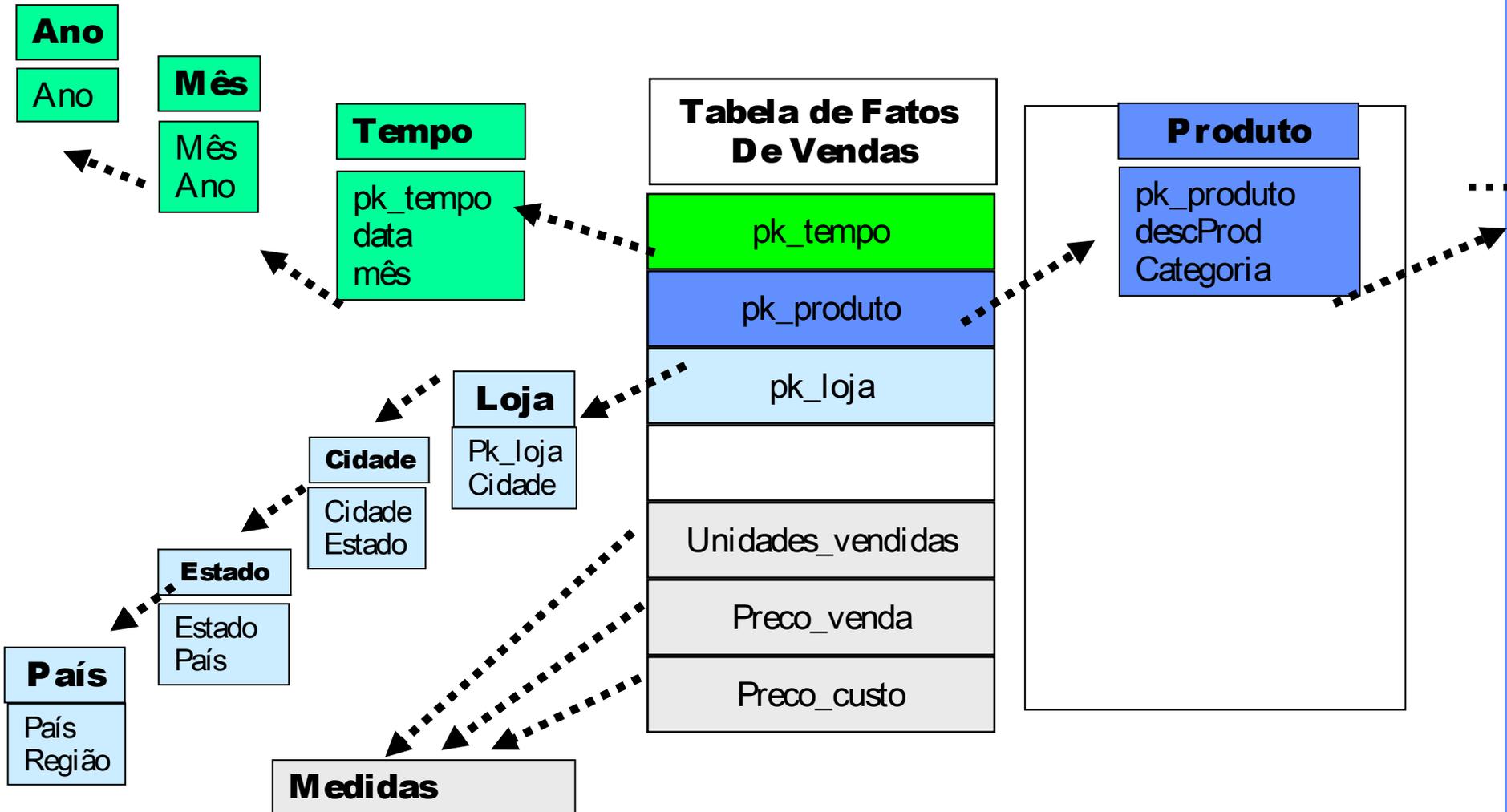
Esquema Floco de Neve

Dimensões normalizadas



Fatos como no
Esquema estrela

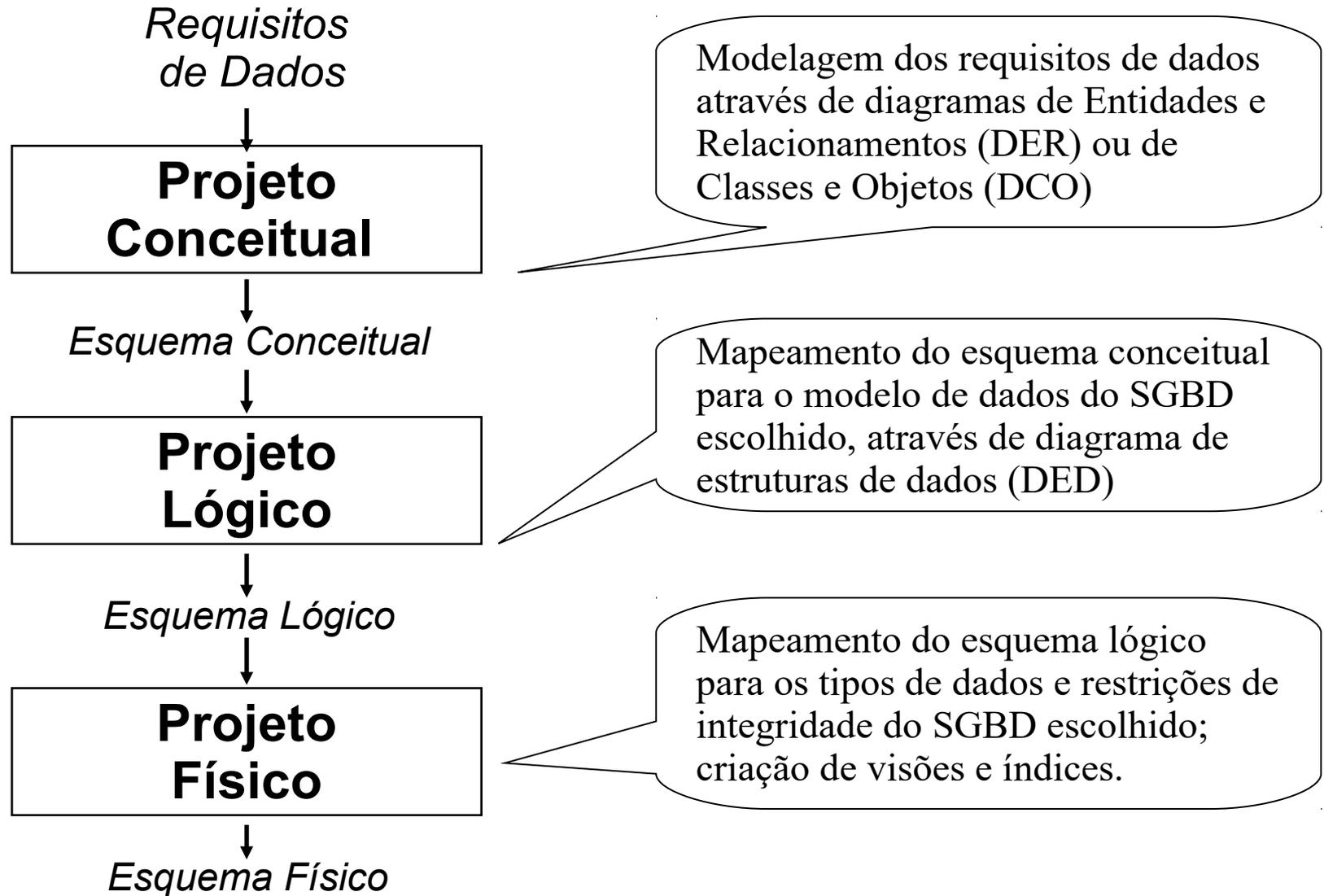
Esquema Flocos de Neve - Exemplo



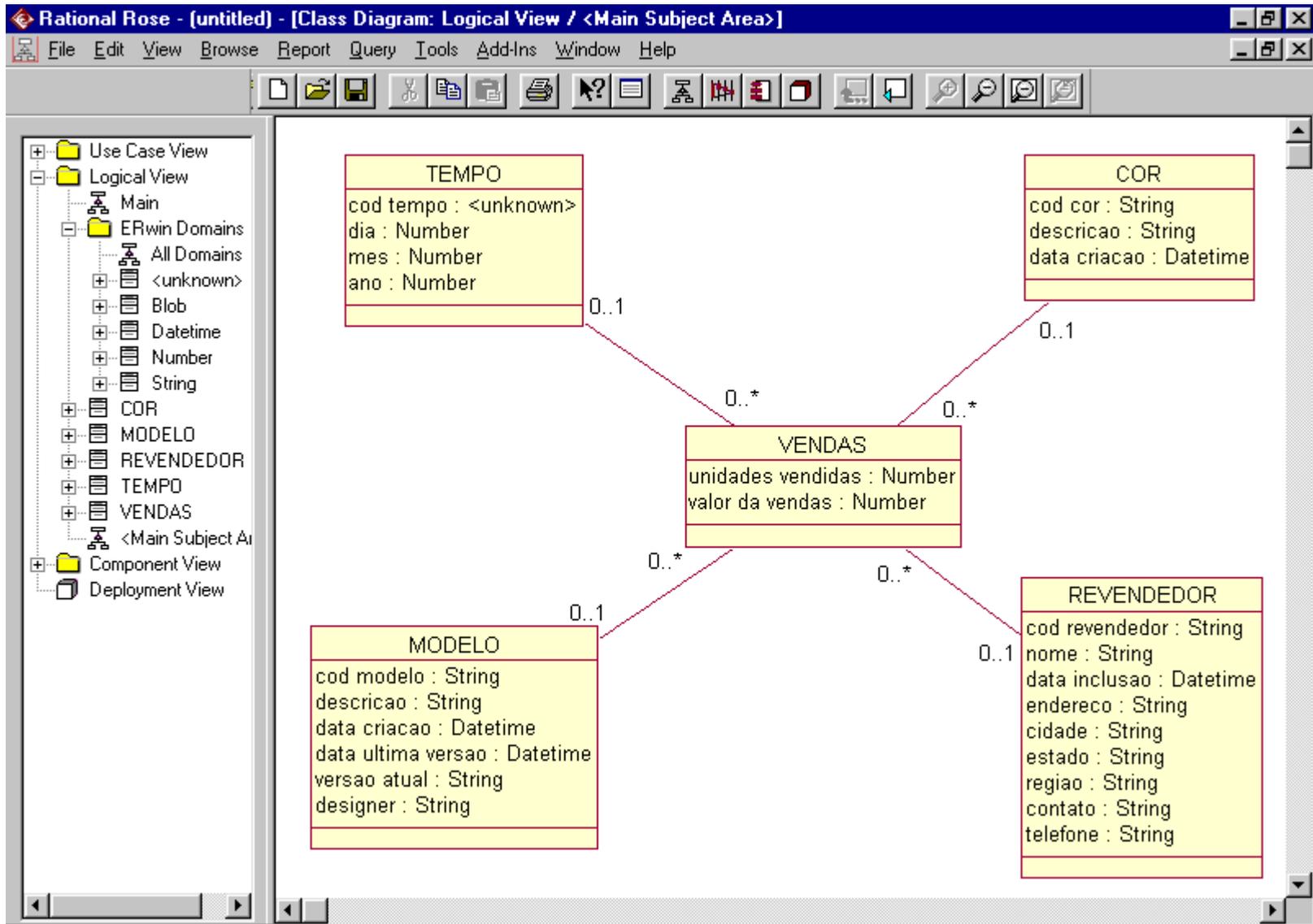
Mitos sobre Modelagem Dimensional

- 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, 2013 – Cap. 1
 - [Margy Ross & Ralph Kimball Fables and Facts](#): Do you know the difference between dimensional modeling truth and fiction? Oct 2004

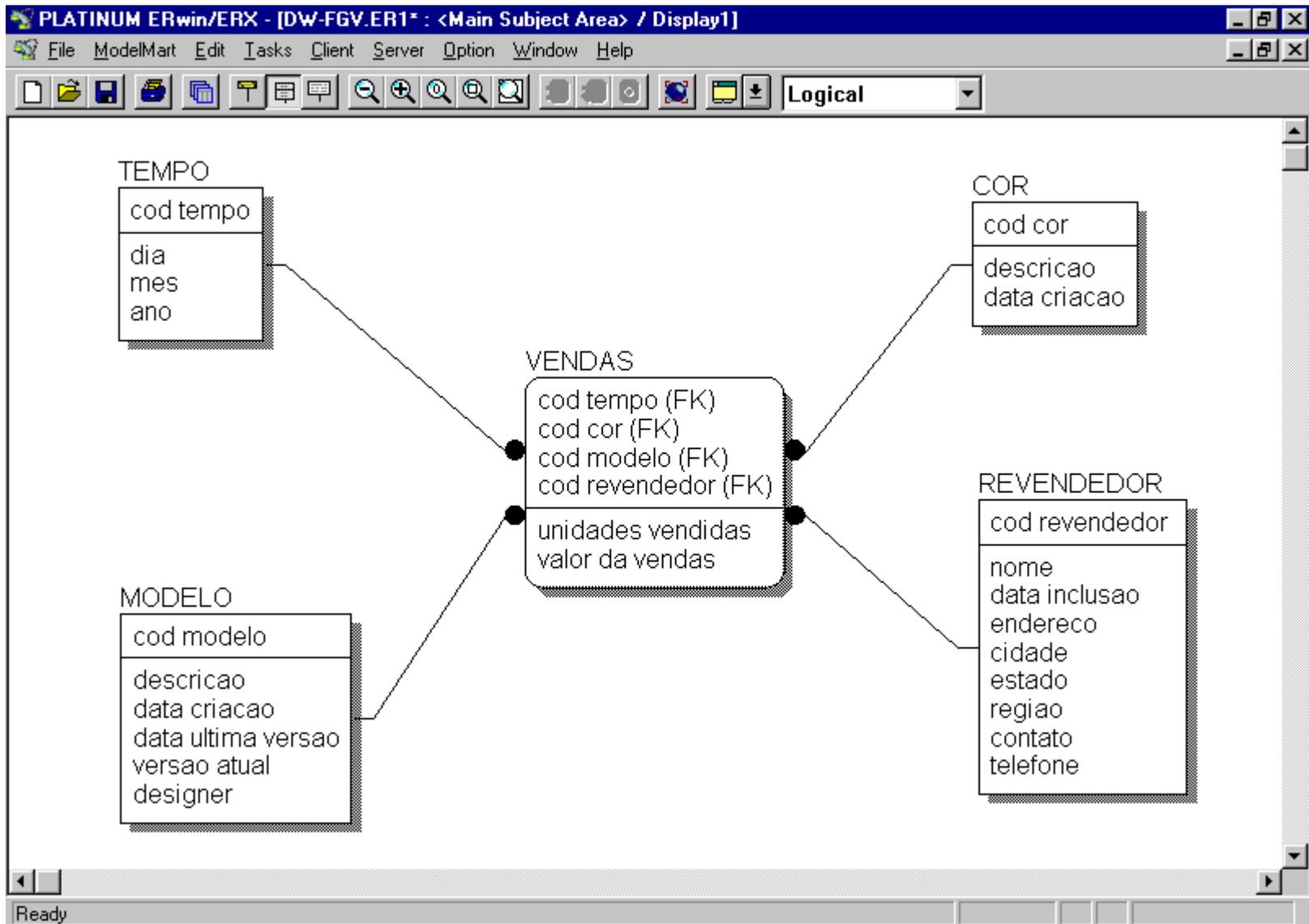
Projeto de Data Warehouse = Projeto de Bancos de Dados



Esquema Estrela – Conceitual (pouco usado, pois o esquema estrela é lógico)



Esquema Estrela - Lógico



Implementação do Modelo Dimensional

- **SGBDs relacionais**
 - Maior aceitação (força do mercado de BD relacional)
 - Exige mapeamento (como qualquer projeto de BD relacional)
- **SGBDs analíticos (NoSQL)**
 - Multidimensionais (OLAP)
 - » Implementam fisicamente o modelo dimensional
 - » Problemas de desempenho, segurança e confiabilidade
 - » Problema de esparsidade: células onde não há dados (nulos)
 - Outros tipos:
 - » Columnar; DW Appliances; In Memory; MPP (Massively Parallel Processing)

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 SQL Anywhere

INFORMIX Red Brick SYBASE

Target Desktop DBMS

Access FoxPro

Clipper dBASE IV Paradox

Red Brick Version

5.x

Default Red Brick Datatype

char(18)

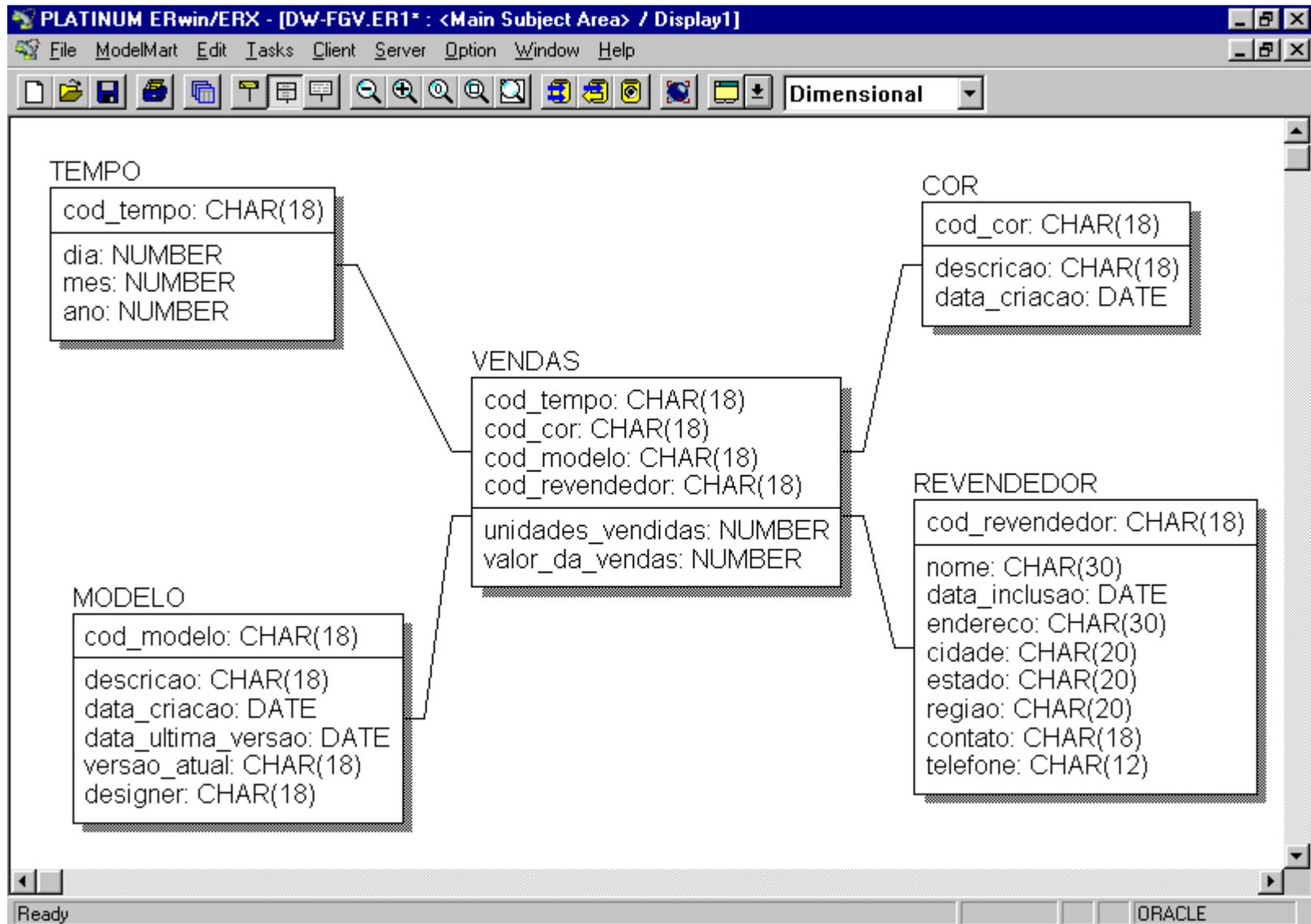
Default Non-Key Null Option

NOT NULL

OK

Cancel

Esquema Estrela - Físico (Dimensional)



Exercício de modelagem dimensional

Vide Tarefa no Moodle

- Dado um banco de dados operacional, obter o seu modelo relacional por engenharia reversa (SQL Power Architect)
- A partir do modelo do banco de dados operacional e uma descrição do negócio, identificar os processos de negócio envolvidos.
- Para cada processo de negócio, projetar um Data Mart aplicando os quatro passos da modelagem dimensional.
- Com cada Data Mart e suas respectivas dimensões identificadas, propor a matriz de barramento do Data Warehouse correspondente.