



SISTEMAS DE INFORMAÇÃO

ENTERPRISE ANALYTICS AND DATA WAREHOUSING

PROF^º FABIANO J. CURY MARQUES

ENTERPRISE ANALYTICS AND DATA WAREHOUSING

MODELAGEM DIMENSIONAL PARTE II



AGENDA

X Modelagem Dimensional

X Conceitos Avançados

X Exercício Prático

X Referências



MODELAGEM DIMENSIONAL

Passo a passo da modelagem dimensional

PROCESSO DO PROJETO DIMENSIONAL



*Business
Requirements*



Dimensional Model

- 1. Business Process**
- 2. Grain**
- 3. Dimensions**
- 4. Facts**



*Data
Realities*

PROCESSO DO PROJETO DIMENSIONAL



1. Selecione o **processo de negócio** que será modelado:
 - X Um processo é uma **atividade de negócio** da organização que tipicamente é suportado por sistemas transacionais que coletam os dados;
 - X Ouvir os usuários é a maneira mais eficiente de selecionar os processos de negócio;
 - X Exemplos em uma empresa de Vendas no Varejo (compras, pedidos, faturamento, entregas, inventário etc.);
 - X Quando se fala em processo de negócio, não devemos nos confundir com um departamento ou função da empresa;
 - X Isto **evita a duplicação de dados** com diferentes terminologias para dois departamentos diferentes que desejam ver os mesmos dados;
 - X Outro fator importante é o econômico, **menos ETL, gestão, manutenção e espaço em disco**;

PROCESSO DO PROJETO DIMENSIONAL



2. Declare o **grão** do processo de negócio:

X Definir exatamente o que uma única linha da tabela fato representa

X Exemplos:

- Um **item** da nota fiscal do **cliente** de um **varejista**;
- Um **item** de uma **conta** em um **restaurante**;
- Uma **passagem** para um **vôo**;
- Uma **foto diária** dos **níveis de estoque** para **cada produto** em um **armazém**;
- Uma **foto mensal** para **cada conta** bancária;

PROCESSO DO PROJETO DIMENSIONAL



3. Escolha as **dimensões** que se aplicam a cada linha da tabela fato:
- X A ideia é responder: Como os usuários descrevem os dados que resultam do processo de negócio?
 - X Deseja-se relacionar as tabelas fatos com um conjunto robusto de dimensões que representam todas as possíveis descrições no contexto de cada medição
 - X Se o grão estiver claro, as dimensões podem ser identificadas facilmente
 - X Com a escolha de cada dimensão, deve-se definir todos seus atributos
 - X Exemplos de dimensões comuns: **data, produto, cliente, tipo de transação, status**

PROCESSO DO PROJETO DIMENSIONAL



4. Identifique os **fatos** numéricos que irão popular cada linha da tabela fato:

- X A ideia é responder: O que estamos medindo?
- X Todos candidatos a fato devem ser verdadeiros para o grão definido no passo 2;
- X Fatos que claramente pertencem a diferentes grãos devem estar em tabelas fatos diferentes;
- X Fatos típicos são numéricos e aditivos como quantidade pedida ou total de custo em reais;

ESTUDO DE CASO – VAREJISTA



- X Imagine-se trabalhando em uma grande cadeia varejista (supermercados). Nosso negócio tem **100 lojas** espalhadas em cinco estados.
- X Cada loja tem diversos departamentos como: **doces, comidas congeladas, açougue, flores, limpeza doméstica** etc.
- X Cada loja tem aproximadamente **60.000 produtos específicos (SKU)** para venda em suas prateleiras. Destes, 55.000 são de manufaturas externas e 5.000 são de departamentos como açougue, panificadora etc.
- X Os dados são coletados em diversos lugares, mas o mais importante para o negócio são os dados de produtos comprados pelo cliente **coletado pelo scanner dos caixas (POS)**
- X Para o lucro da empresa trabalha-se **baixando custos de aquisição e processamento** e ao mesmo tempo atraindo tantos clientes quanto possível em um ambiente de preços altamente competitivos
- X Algumas das decisões gerenciais mais importantes levam em conta **preços e promoções**

SEGUINDO O
PASSO A
PASSO
APRESENTADO

...

ESTUDO DE CASO – VAREJISTA



1. Selecionando o **processo de negócio**:

- X Nesse estudo de caso a gestão quer entender melhor as compras dos clientes de acordo com o capturado nos POS (caixas do supermercados)
- X Assim, o processo de negócio que será modelado é **Vendas a Varejo no POS**
- X Estes dados permitirão analisar **quais produtos** são vendidos em **quais lojas** em **que dias** e atrelados à **quais promoções**



2. Declarando o Grão:

- X Preferencialmente, deve-se desenvolver modelos dimensionais para as informações mais atômicas capturadas por um processo de negócio;
- X Neste estudo de caso, o dado mais granular é um **item específico em uma transação do POS (uma compra no caixa)**;
- X Para garantir maior dimensionalidade e flexibilidade, vamos utilizar este grão;



3. Escolhendo as Dimensões:

- X Com o grão da tabela fato definido, algumas dimensões já aparecem imediatamente: **data**, **produto** e **loja**
- X Além das dimensões que são claras de acordo com o grão (dimensionalidade primária), pode-se adicionar outras dimensões
- X Neste estudo de caso, vamos adicionar também a dimensão **promoção**
- X Além disso, uma dimensão especial **Número de Transação POS** será adicionada e falaremos a seguir sobre ela;



ESTUDO DE CASO – VAREJISTA

3. Escolhendo as Dimensões:



TBD = To Be Defined;



DIMENSÃO GENERALIZADA

A tabela fato de vendas do estudo de caso contém o número da transação POS para cada item de uma compra (cada linha da tabela)

O número da transação POS ainda é útil porque funciona como uma chave de agrupamento para obtermos todos os produtos comprados em uma única transação

Em uma base normalizada, esse número seria a chave para o registro do cabeçalho da venda contendo toda a informação válida para a transação como um todo, com a data da compra, a loja etc.

Entretanto, no modelo dimensional, todas estas informações já foram extraídas para outras dimensões



DIMENSÃO GENERALIZADA

Embora o número da transação POS pareça uma chave de dimensão na tabela fato, todos seus atributos descritivos foram retirados para outras dimensões

Estas dimensões são muito comuns nos casos em que o grão da tabela representa uma única transação ou um item da transação

Assim, como a dimensão resultante está vazia (sem atributos), refere-se a esta dimensão como uma **dimensão degenerada**

O dimensão degenerada existe apenas na própria tabela fato, sem uma tabela de dimensão para relacioná-la



4. identificando os **Fatos**:

- X Os fatos devem ser verdadeiros para o grão, no caso, um item em uma transação do POS
- X Os fatos coletados no POS incluem:
 - **quantidade vendida** (ex: número de latas de cerveja)
 - **preço unitário de venda** ????
 - **total de venda** (quantidade x preço unitário)
 - **total de custo**
- X Pode-se calcular então:
- X **lucro bruto** (total de venda – total de custo)
- X Além disso, uma dimensão especial **Número de Transação POS** será adicionada e falaremos a seguir sobre ela;

ESTUDO DE CASO – VAREJISTA

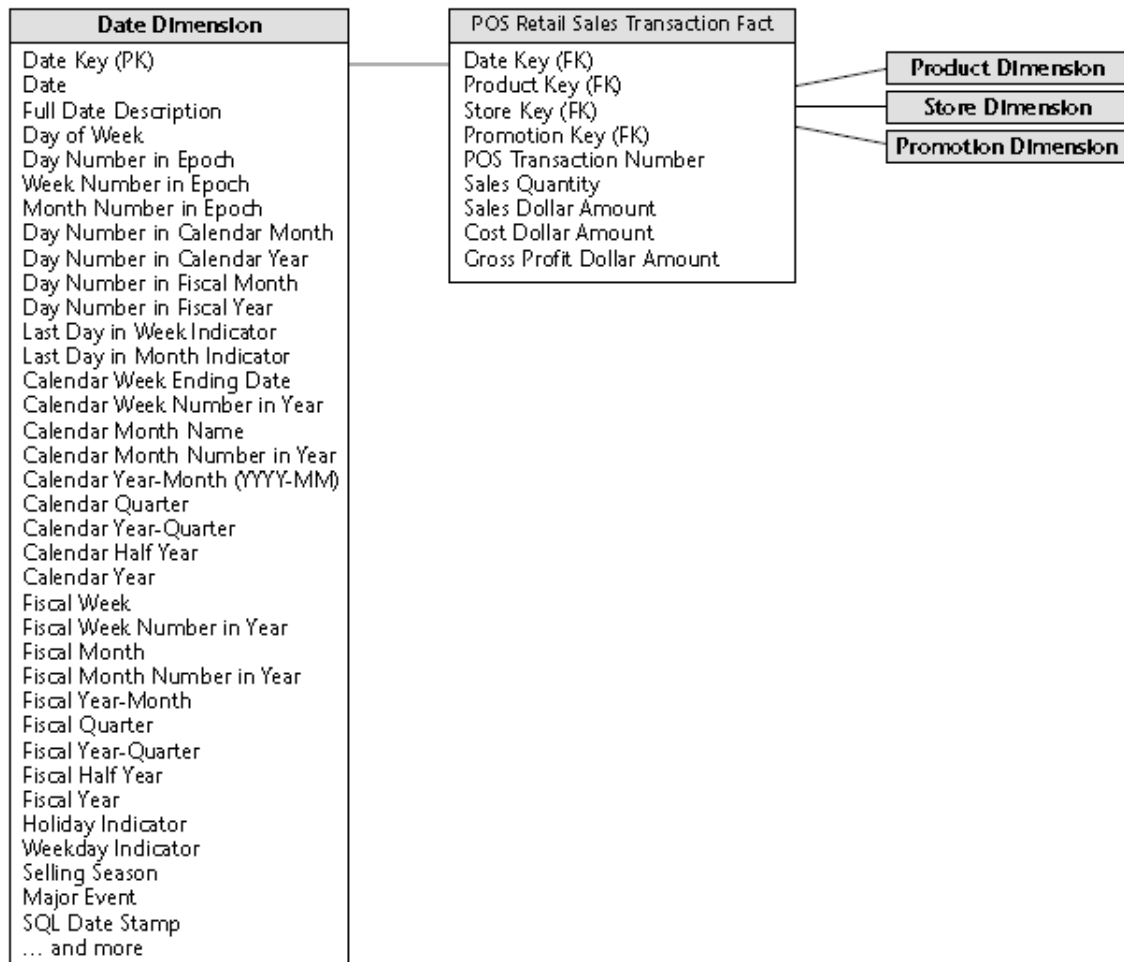


4. identificando os **Fatos**:

- X Quatro destes fatos são totalmente **aditivos** (quantidade, total de venda, total de custo e até lucro bruto)
- X Isso significa que qualquer soma destes fatos, não importa em qual combinação de dimensões, é válida e correta
- X Preço unitário é um fato **não-aditivo** neste caso (ou seria parte de uma dimensão Produto?)
- X Somar os preços unitários em qualquer dimensão resulta em um número sem significado
- X **Fatos não-aditivos** são aqueles que, se somados, não significam nada, não importa qual combinação de dimensões seja utilizada
- X **Fatos semi-aditivos** são aqueles que dependendo da combinação de dimensões pode ou não resultar em um número válido

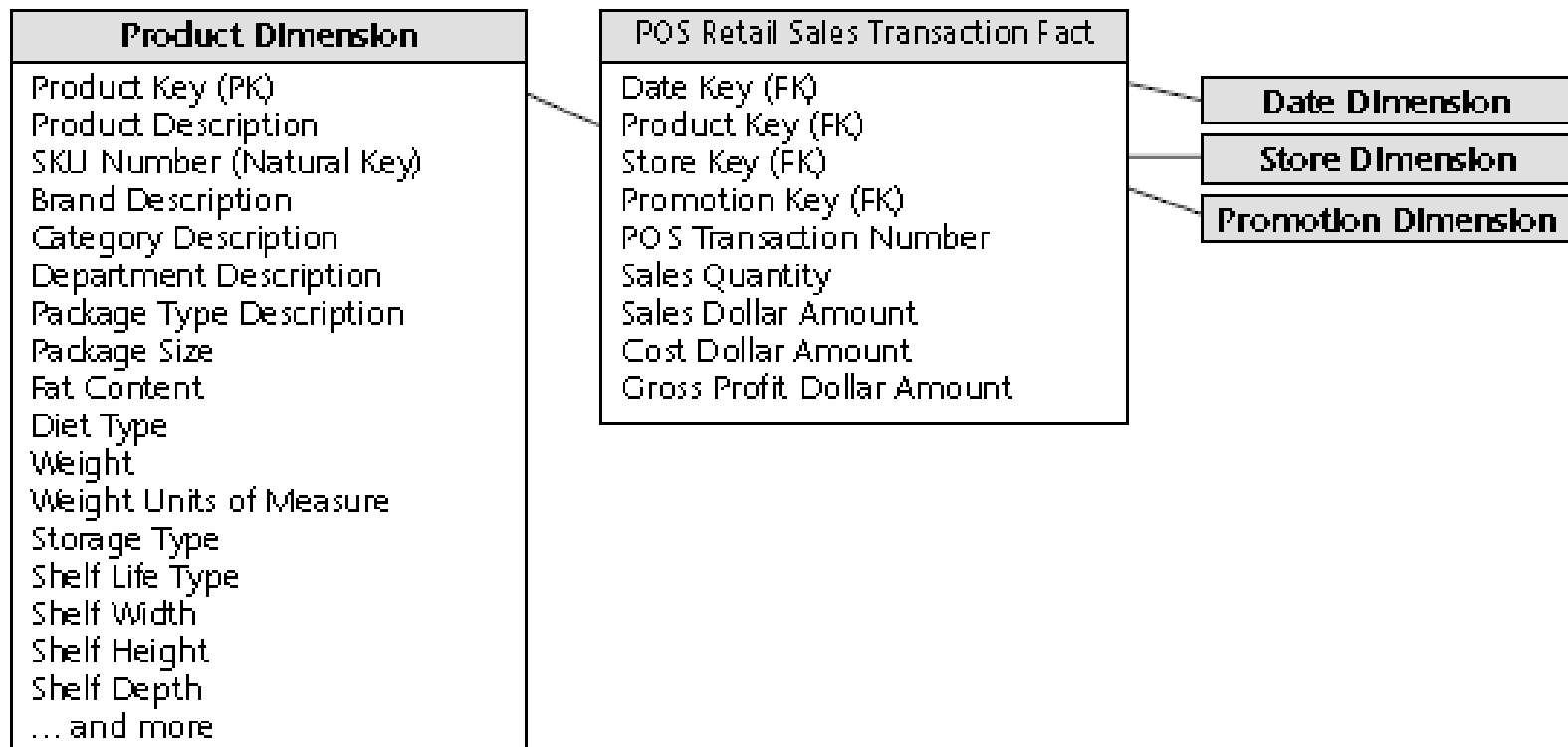
COMO FICAM
AS DIMENSÕES
MODELO/DADOS
?

DIMENSÃO DATA – MODELO





| Date Key | Date | Full Date Description | Day of Week | Calendar Month | Calendar Year | Fiscal Year-Month | Holiday Indicator | Weekday Indicator |
|----------|------------|-----------------------|-------------|----------------|---------------|-------------------|-------------------|-------------------|
| 1 | 01/01/2002 | January 1, 2002 | Tuesday | January | 2002 | F2002-01 | Holiday | Weekday |
| 2 | 01/02/2002 | January 2, 2002 | Wednesday | January | 2002 | F2002-01 | Non-Holiday | Weekday |
| 3 | 01/03/2002 | January 3, 2002 | Thursday | January | 2002 | F2002-01 | Non-Holiday | Weekday |
| 4 | 01/04/2002 | January 4, 2002 | Friday | January | 2002 | F2002-01 | Non-Holiday | Weekday |
| 5 | 01/05/2002 | January 5, 2002 | Saturday | January | 2002 | F2002-01 | Non-Holiday | Weekend |
| 6 | 01/06/2002 | January 6, 2002 | Sunday | January | 2002 | F2002-01 | Non-Holiday | Weekend |
| 7 | 01/07/2002 | January 7, 2002 | Monday | January | 2002 | F2002-01 | Non-Holiday | Weekday |
| 8 | 01/08/2002 | January 8, 2002 | Tuesday | January | 2002 | F2002-01 | Non-Holiday | Weekday |



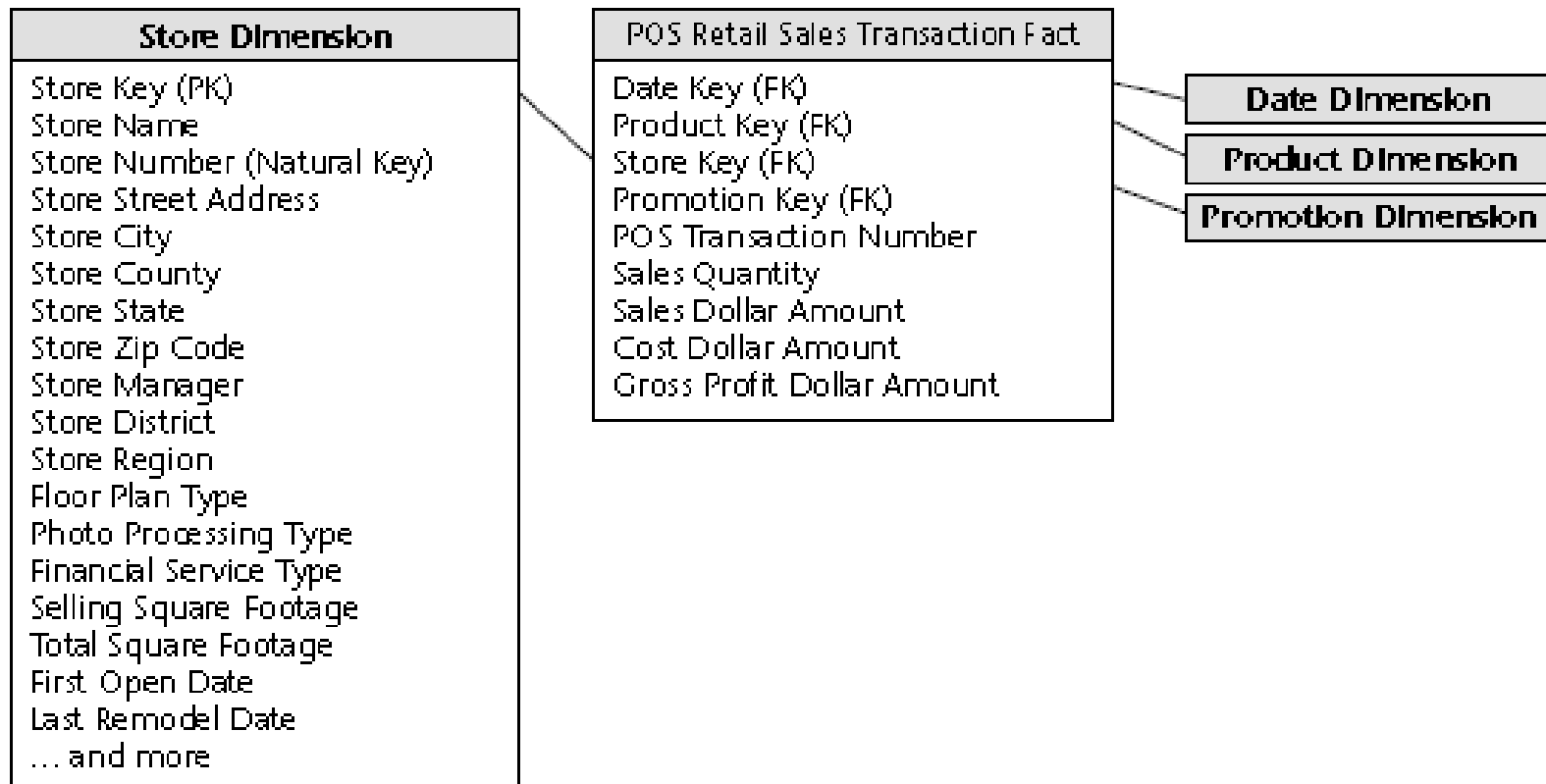
RELATÓRIO BASEADO NA DIMENSÃO PRODUTO



| Department Description | Sales Dollar Amount | Sales Quantity |
|---------------------------|------------------------|----------------|
| Bakery | \$12,331 | 5,088 |
| Frozen Foods | \$31,776 | 15,565 |

drill down (hierárquico) ↓

| Department Description | Brand Description | Sales Dollar Amount | Sales Quantity |
|---------------------------|----------------------|------------------------|-------------------|
| Bakery | Baked Well | \$3,009 | 1,138 |
| Bakery | Fluffy | \$3,024 | 1,476 |
| Bakery | Light | \$6,298 | 2,474 |
| Frozen Foods | Coldpack | \$5,321 | 2,640 |
| Frozen Foods | Freshlike | \$10,476 | 5,234 |
| Frozen Foods | Frigid | \$7,328 | 3,092 |
| Frozen Foods | Icy | \$2,184 | 1,437 |
| Frozen Foods | QuickFreeze | \$6,467 | 3,162 |



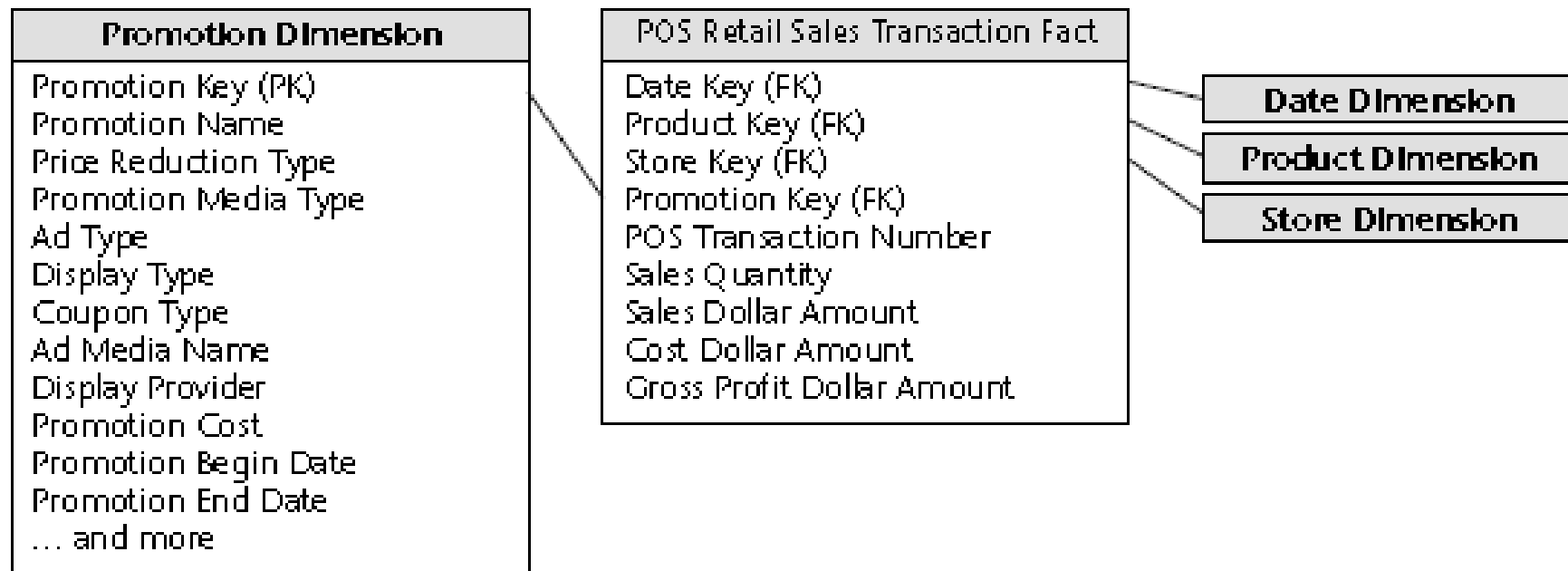


DIMENSÃO PROMOÇÃO

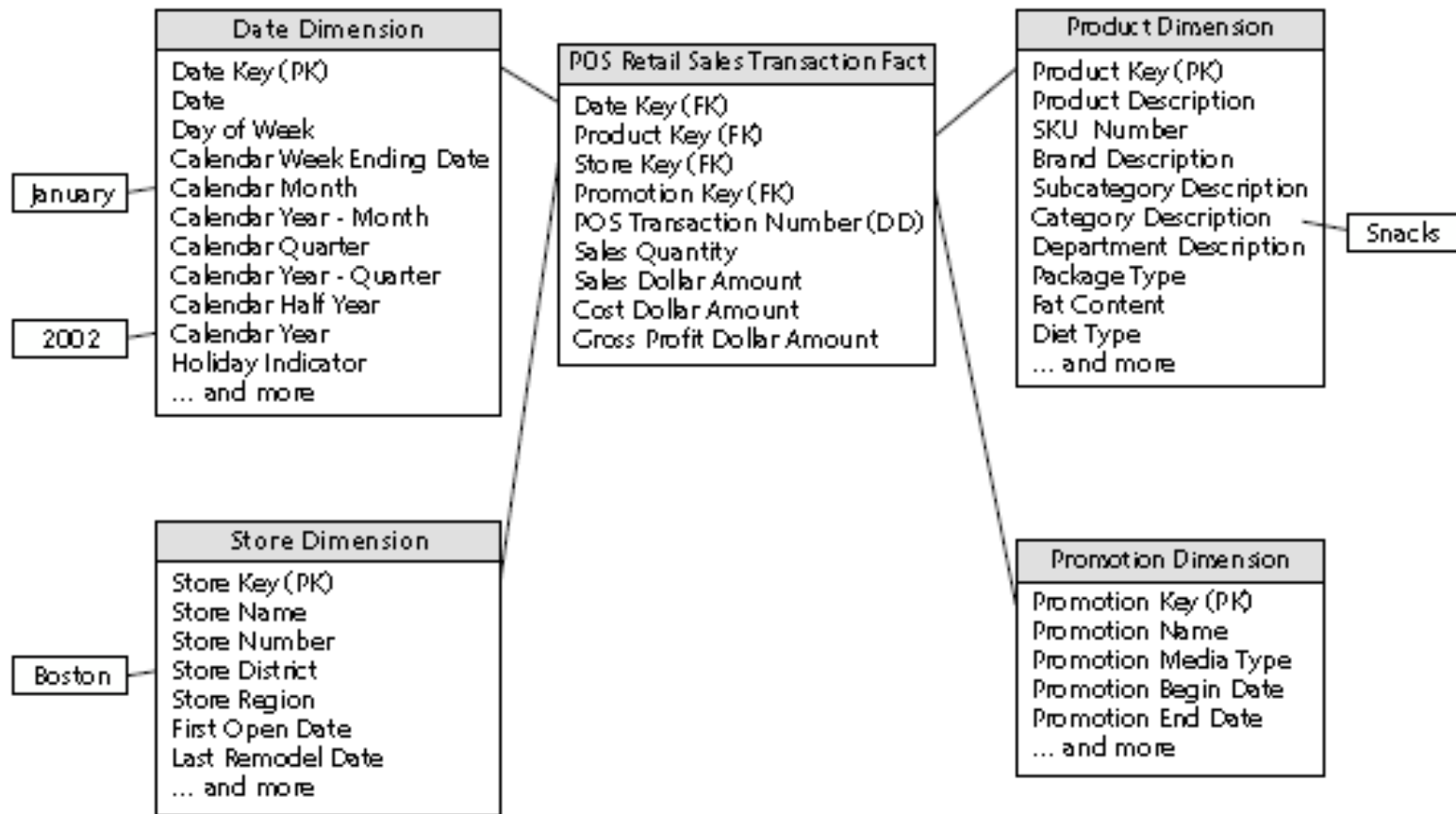
Esta dimensão descreve as condições de promoção sob a qual um produto foi vendido.

Dimensões como essa geralmente são **chamadas de dimensões causais** pois elas descrevem fatores que causam mudanças na venda de um produto

Tipicamente, muitos itens da transação envolvem produtos que não estão em promoção. É necessário então incluir uma linha na dimensão promoção com sua própria chave para identificar **“Nenhuma Promoção”** e evitar colocar valores **NULL** na tabela fato



CONSULTANDO O DATA MART CRIADO



CONSULTANDO O DATA MART CRIADO



| Calendar Week Ending Date | Promotion Name | Sales Dollar Amount |
|------------------------------|----------------------|------------------------|
| January 6, 2002 | No Promotion | 22,647 |
| January 13, 2002 | No Promotion | 4,851 |
| January 20, 2002 | Super Bowl Promotion | 7,248 |
| January 27, 2002 | Super Bowl Promotion | 13,798 |

| Calendar Week Ending Date | Super Bowl Promotion Sales Dollar Amount | No Promotion Sales Dollar Amount |
|------------------------------|--|--|
| January 6, 2002 | 0 | 22,647 |
| January 13, 2002 | 0 | 4,851 |
| January 20, 2002 | 7,248 | 0 |
| January 27, 2002 | 13,798 | 0 |



ESTUDO DE CASO CONCLUÍDO!





REFERÊNCIAS

- X KIMBALL, R., ROSS, M. The Data Warehouse Toolkit. 2^a ed., John Wiley Professional, 2002.
- X MACHADO, F. N. R. Tecnologia e Projeto de Data Warehouse. 1^a ed., São Paulo: Ed. Érica, 2004.



OBRIGADO!

Copyright © 2019 Prof. MSc. Eng. Wakim B. Saba

<https://br.linkedin.com/in/wakimsaba>

Todos direitos reservados. Reprodução ou divulgação total ou parcial deste documento é expressamente proibido sem o consentimento formal, por escrito, do Professor (autor).