



Modelagem Multidimensional

GRIMALDO OLIVEIRA

Agenda

➤ Objetivo

- Apresentar os principais assuntos na construção de um projeto de Data Warehouse
 - ▶ Dimensões e Fatos
 - ▶ Modelagem Multidimensional
 - ▶ Fundamentos da Modelagem Multidimensional
 - ▶ Desnormalização: anti-forma normal.
 - ▶ Tipos de Dimensões
 - ▶ Dimensões: Chaves Artificiais e Histórico
 - ▶ Dimensões Hierárquicas
 - ▶ Tipos de Métricas
 - ▶ Tempo
 - ▶ Granularidade

Dimensão e Fato

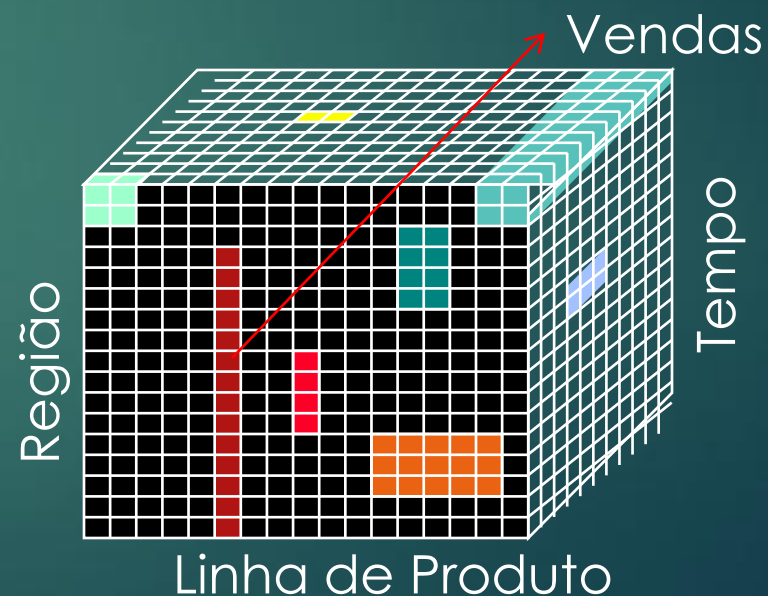
➤ DIMENSÃO

- Contém os descritores textuais do negócio.
- Exemplo : Tempo, Cliente, Produto, Tipo de Embalagem, Situação, etc.

➤ FATO

- Termo utilizado para a medição do negócio.
- Exemplo: quantidade de produtos vendidos, preço de compra, preço de venda, lucro, etc.

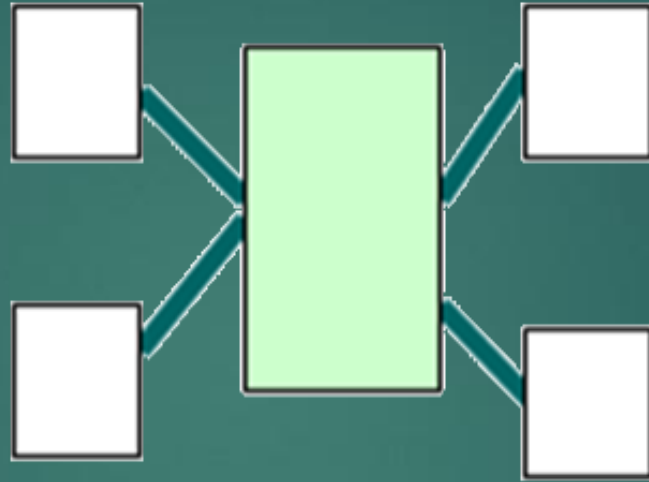
- Dimensões
 - Região
 - Linhas de Produto
 - Tempo (obrigatória)
- Fato
 - Vendas



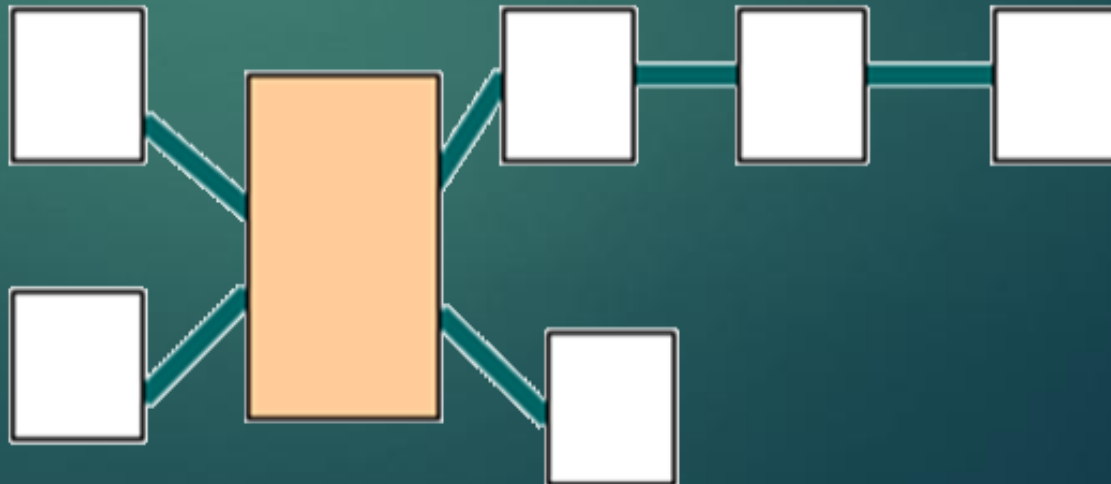
Modelagem Multidimensional

➤ MODELAGEM MULTIDIMENSIONAL

STAR SCHEMA

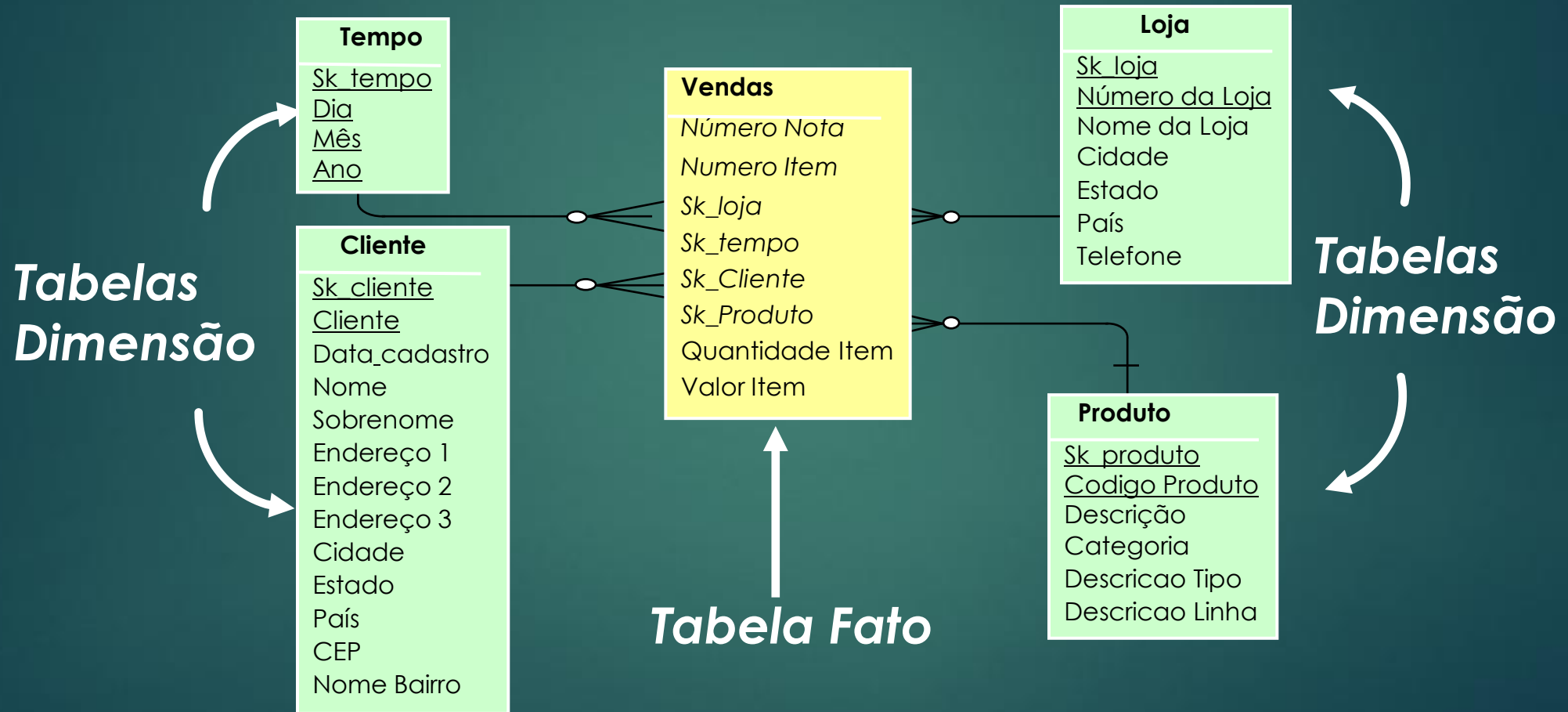


SNOW FLAKE



Modelo Multidimensional – Star Schema

➤ MODELO MULTIDIMENSIONAL – STAR SCHEMA



Desnormalização

- ▶ Só a 1ºFN deve ser respeitada. As demais FN obrigatoriamente devem ser ignoradas.
- ▶ Vantagens: Excelente tempo de 'query response'.
- ▶ Exemplo:

CLIENTE

Cod_cliente

@Cod_cliente

Nom_cliente

Nom_Cliente

Cod_tipo_cliente

Cod_tipo_cliente

TIPO CLIENTE

Des_tipo_cliente

Cod_tipo_cliente

Cod_condicao_tipo_cliente

Des_tipo_cliente

Des_condicao_tipo_cliente

Cod_condicao_tipo_cliente

Considerações sobre espaço

A desnormalização simplifica o modelo, otimizando as consultas geradas pela ferramenta OLAP. Porém, causa excesso de uso em disco.

Exemplo: Clientes do Bank Money

	<u>Normalizado</u>	<u>Desnormalizado</u>
Tamanho	400Mb	2Gb
Qtd. Campos	~70	~180
Consulta Estatística	> 10hs	<15 min

Chaves Artificiais

- ▶ Permite que o controle do histórico dos dados seja facilmente implementado.
- ▶ Gera independência de relacionamento com outras tabelas.
- ▶ Relacionam as Dimensões e Fatos.
- ▶ Devem ser apenas números e não carregar em si nenhum significado.
 - ▶ Exemplo: 132453
 - ▶ Contra Exemplo: Jan/99 , IFZ01

Tipos de Dimensões

- ▶ Slow Changing Dimensions
- ▶ Fast Changing Monster Dimensions
- ▶ Dimensões Degeneradas
- ▶ Dimensão Ponte (Bridge Table)

Slow Changing Dimensions

▶ **Tipo 1: Sobrescrever os Dados**

- ▶ O novo registro substitui o registro original. Só existe um registro no banco de dados - os dados atuais.
- ▶ Não mantém histórico.

Tipo1 : Sobrescrever os Dados

- Um exemplo seria de uma tabela de banco de dados que mantém as informações do fornecedor.

Supplier_key	Supplier_Name	Supplier_State
001	Phlogistical Sociedade de Abastecimento	CA

Supplier_key	Supplier_Name	Supplier_State
001	Phlogistical Sociedade de Abastecimento	IL

Slow Changing Dimensions

▶ **Tipo 2: Controle de Versão**

- ▶ Mantém histórico
- ▶ Um novo registro é adicionado na tabela de Dimensão. Dois registros existentes no banco de dados - os dados atuais e dados da história anterior.
- ▶ É recomendável para a maioria dos casos.

Tipo 2 : Controle de Versão

- ▶ Exemplo seria o de uma tabela de banco de dados que mantém as informações do fornecedor.

Supplier_key	Supplier_Code	Supplier_Name	Supplier_State	Data_inicial	Data_final
001	ABC	Phlogistical Sociedade de Abastecimento	CA	01 de janeiro-2000	21-Dez-2004
002	ABC	Phlogistical Sociedade de Abastecimento	IL	22-Dez-2004	

Exemplo : DIM_HOSPEDE

Seq_cliente

Cod_hospede

Nom_hospede

Des_endereco

Dtc_inicio

Dtc_fim

Sts_corrente

***Em branco: campos de controle
utilizados pela modelagem
multidimensional***

***Em verde: campos do sistema
fonte.***

Exemplo : DIM_HOSPEDE

Seq_cliente : 1
Cod_cliente : X1
Nom_cliente : Antônio dos Santos
Des_Endereco : Rua das Margaridas
Dtc_inicio : 01/03/2010
Dtc_fim : <null>
Sts_corrente : Ativo

E se o cliente
vier a trocar de
endereço?

Exemplo : DIM_HOSPEDE

Seq_cliente : 3435

Cod_cliente : X1

Nom_cliente : Antônio dos Santos

Des_Endereco : Rua das Rosas

Dtc_inicio : 01/03/2012

Dtc_fim : <null>

Sts_corrente : Ativo

Um **novo** registro é
inserido na base de
dados.

E o registro anterior?

Exemplo : DIM_HOSPEDE

Seq_cliente : 1

Cod_hospede : X1

Nom_hospede : Antonio dos Santos

Des_Endereco : Rua das Margaridas

Dtc_inicio : 01/03/2010

Dtc_fim : 01/03/2012

Sts_corrente : Inativo

Ambos os registros
estão no banco de
dados.

Considerações

- ▶ O uso de chaves artificiais permite o controle do histórico e facilita o relacionamento com as tabelas Fato.
- ▶ Os campos de controle de data (Dtc_inicio e Dtc_fim) permitem saber quando a informação refletia o operacional.
- ▶ O campo “*Sts_corrente*” permite a fácil identificação do registro que contém os dados mais recentes.

Fast Changing Dimensions

- ▶ Algumas dimensões que possuem grande volume de registros e muitos campos, crescem rapidamente, explodindo o espaço físico de armazenamento.
- ▶ Solução: Colocar campos que trocam de valores mais rapidamente em outra tabela, sem alterar a versão do registro.

Dimensões Degeneradas

- ▶ É caracterizada por não ter a sua própria Dimensão e sim por estar presente na tabela Fato.
- ▶ Exemplo: Considere que uma tabela Fato com os itens das notas fiscais de uma empresa varejista. O que fazer com o número da nota fiscal? Ele em si, não representa nada, apenas serve para agrupar os itens de uma mesma compra. Não existe fisicamente uma Dimensão nota fiscal, embora exista uma coluna na tabela Fato com o número da nota propriamente dito.

Dimensão Ponte (Bridge Table)

- Uma tabela com chave composta capturando um relacionamento muitos-para-muitos que não pode ser acomodado pela granularidade natural de uma tabela Fato ou de Dimensão. Serve como uma ponte entre a tabela Fato e de Dimensão.



- Outros exemplos de dimensões multivaloradas: titulares de conta bancária.

Dimensões Hierárquicas

- ▶ Exemplo : País → Região → UF → Cidade
- ▶ As tabelas de País, Região, UF e Cidade são armazenadas separadamente no sistema fonte (normalizado). No Data Warehouse, elas compõem uma única tabela, a Dimensão Geografia. Cada nível da hierarquia deve ser representado individualmente.
- Usada para Permitir Drill:
 - **Drill Down:** Detalha a informação
 - **Drill Up:** Sumariza a informação
 - **Drill Across:** Muda de Dimensão, mantém Fato.
 - **Drill Through:** Vê registros do ambiente transacional que originaram aquela ocorrência

Tipos de Métricas

Aditivas: Podem ser utilizadas para sumarizações.

Ex.: Valor de Venda

Semi-Aditivas: Podem ser utilizadas para sumarizações, com ressalvas.

Ex.: Saldo Bancário – Faz sentido somar o seu saldo caso ele tenha conta em mais de um banco, mas não faz sentido somar seu saldo todos os dias de uma mesma conta.

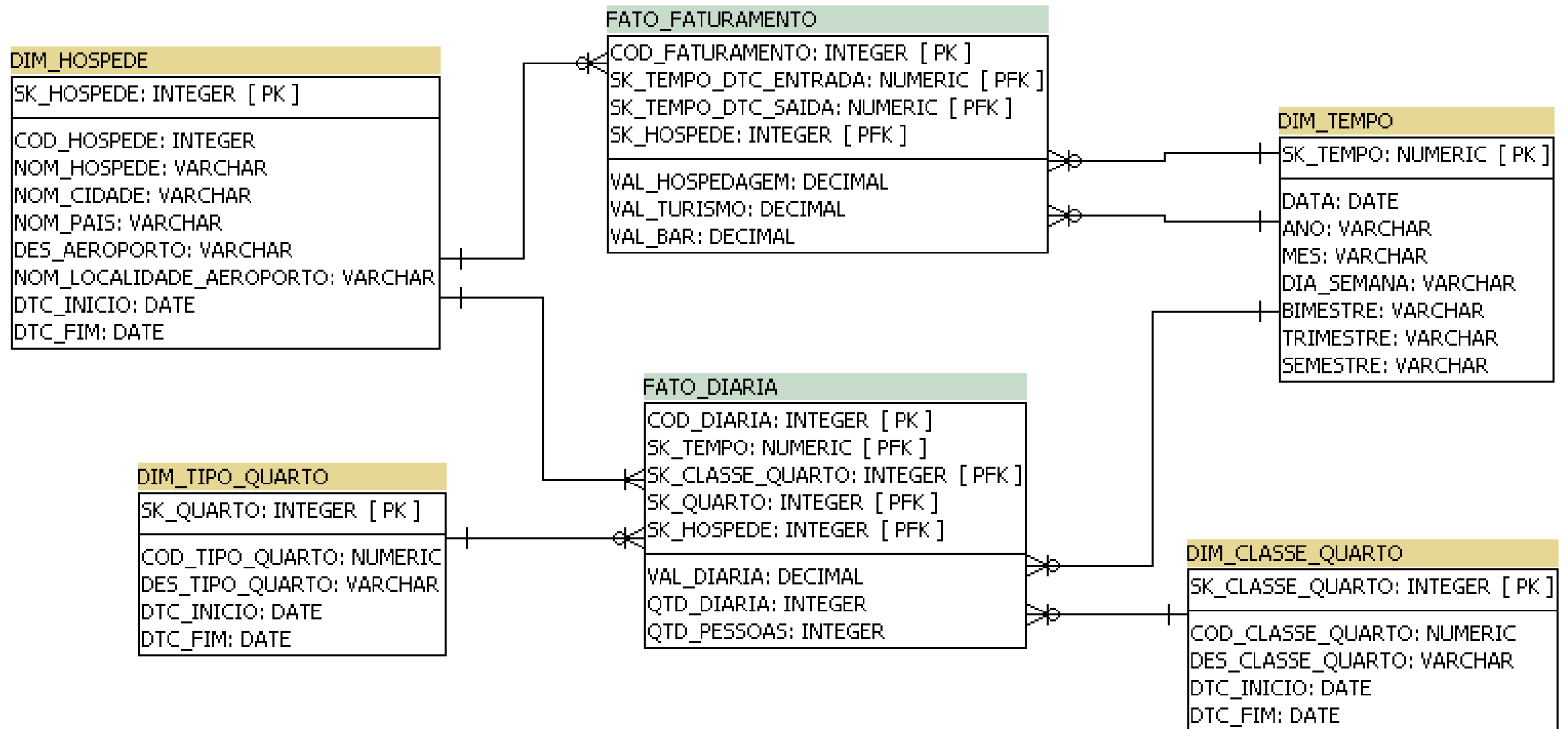
Não-Aditivas: Não podem ser utilizadas para sumarização.

Ex.: % margem de lucro

Grão

- ▶ Conceito que identifica a unidade de medida das métricas.
- ▶ Nível de detalhe dos dados.
- ▶ Menor Grão: Mais detalhe -> Mais dados -> Análise mais longa -> Informação mais detalhada.
- ▶ Maior Grão: Menos detalhe -> Menos dados -> Análise mais rápida -> Informação menos detalhada.

Modelo Multidimensional: Hotel Dallas



Projeto completo BI

Finalidade: Coleta de dados com os gestores para a construção do BI.

Fatos		Diária
Dimensões		
Hóspede		✓
Tipo Quarto		✓
Código Tipo Quarto		
Tipo Quarto	HISTÓRICO	
Classe Quarto		✓
Tempo (Data Registro Primeira Diária)		✓

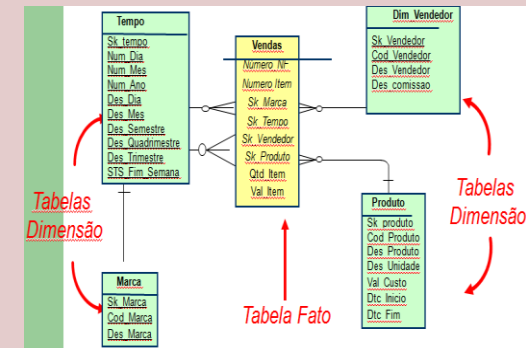
Matriz de
Necessidades

Finalidade: Levantamento dos relacionamentos e objetos que armazenam os dados da empresa.

DIMENSÕES	ORIGEM	
	TABELA/VISÃO	CAMPO
Hóspede		
Nome Hóspede	HOSPEDE	NOM_HOSPEDE
Cidade Hóspede	CIDADE_ORIGEM	NOM_CIDADE
País Hóspede	PAIS_ORIGEM	NOM_PAIS
Aeroporto Hóspede	AEROPORTO_SAIDA	DES_AEROPORTO
Local Aeroporto Saída	AEROPORTO_SAIDA	NOM_LOCALIDADE
Código Hóspede	HOSPEDE	COD_HOSPEDE

Fonte de Dados

Finalidade: Modelo adequado para realizar as consultas nas bases que servirão ao BI



Modelagem
Multidimensional