

Arquitetura de *Data Warehouse* Aula 04 – Modelagem de DW

Conteúdo Programático desta aula

- Modelagem Multidimensional
- Modelo Esquema Estrela e Modelo Floco de Neve

Modelagem de Dados para o DW

A modelagem de dados para DW é diferente da utilizada para sistemas OLTP.

Se simplesmente movermos o modelo de dados transacional para um banco de dados separado, inserirmos dados históricos e chamarmos de DW não conseguiremos trabalhar com estes dados por duas razões principais:

Modelagem de Dados para o DW

- 1) Os usuários não ficarão felizes porque a complexidade será muito alta, para que eles próprios possam realizar consultas *ad hoc*
- 2) Modelos de dados transacionais são construídos respeitando a Terceira Forma Normal e não respondem com rapidez a questões típicas de *queries* de apoio á decisão

Modelagem de Dados para o DW

A Modelagem Multidimensional é uma técnica de concepção e visualização de um Modelo de Dados de um conjunto de medidas que descrevem aspectos comuns de negócios.

É utilizada especialmente para sumarizar e reestruturar dados e apresentá-los em visões que suportem a análise dos valores desses dados.

Modelagem de Dados para o DW

Um modelo multidimensional é formado por três elementos básicos:

- Fatos;
- Dimensões;
- Medidas (Variáveis)

Modelagem de Dados para o DW

FATO

- Representação de um assunto
- Um assunto pode ser um dado operacional, uma transação do negócio ou um evento
- Um fato é composto por dimensões e medidas
- Exemplos (domínio Loja):
 - vendas (transação do negócio)
 - promoções (evento)
 - produtos e estoque (dado operacional)

Modelagem de Dados para o DW

DIMENSÃO

Representação dos contextos relevantes para a análise de um fato.

- Exemplo:
 - Fato: Vendas
 - Dimensões: Clientes, Produtos, Tempo, Locais, ...
- Uma dimensão pode:
 - conter membros
 - ser organizada em hierarquias

Modelagem de Dados para o DW

DIMENSÃO

São as possíveis formas de visualizar os dados, ou seja, são os “por” dois dados: “por mês”, “por produto”, “por país”, “por região”, etc.

Normalmente não possuem atributos numéricos, são somente descritivas e classificatórias dos elementos que participam de um fato.

Modelagem de Dados para o DW

MEDIDA

- Representação de atributos (variáveis) relevantes para a análise de um fato
 - valores numéricos em geral
 - indicadores de desempenho para a análise
- Uma medida é determinada pela associação de dimensões
 - produto X tempo.ano;
 - local.estado X cliente.classe X tempo.semana.dia
- Exemplo:
 - Fato: Vendas
 - Medidas: quantidade vendida, valor da venda

Modelagem de Dados para o DW

MEDIDA

São os atributos numéricos que representam um fato, a performance de um indicador de negócios relativo às dimensões que participam desse fato.

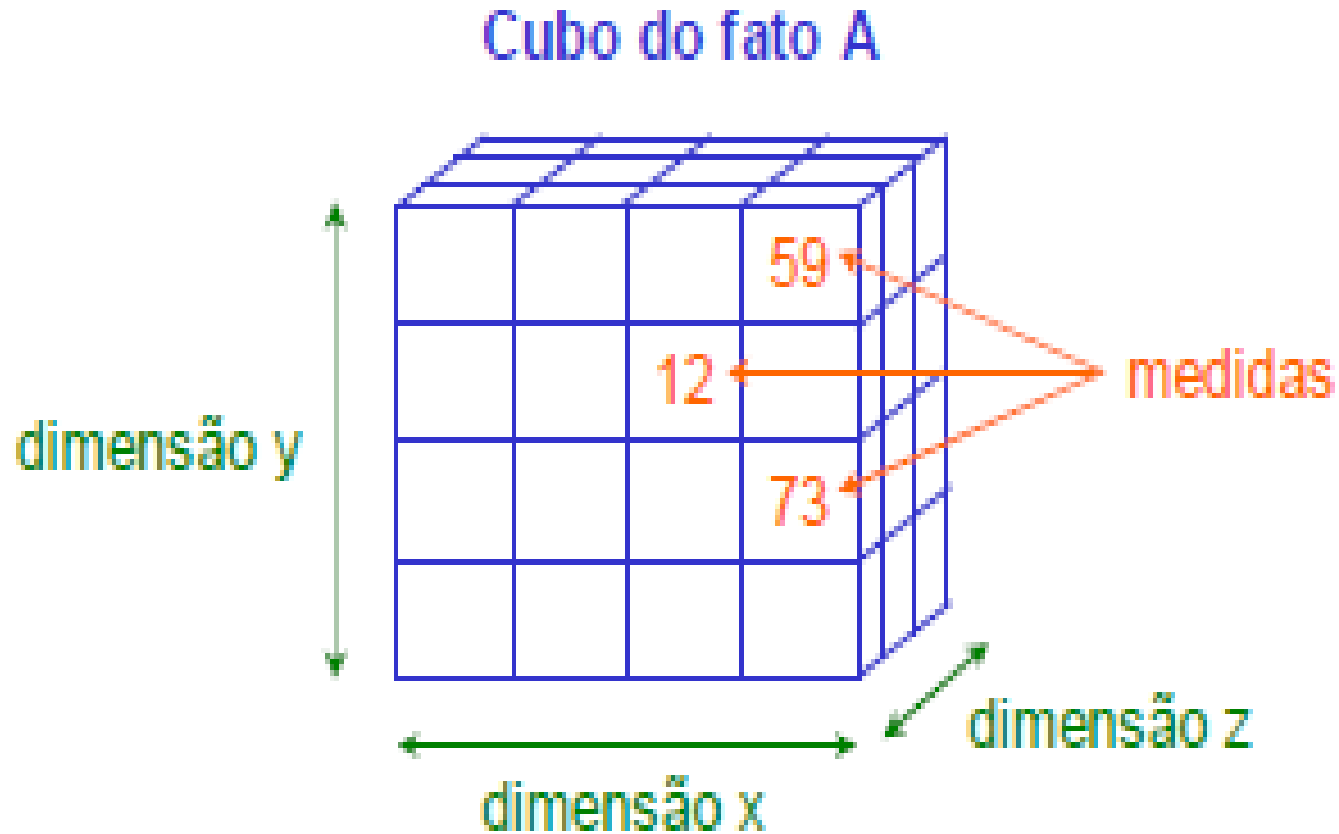
Uma medida é determinada pela combinação das dimensões que participam de um fato, e estão localizadas como atributos de um fato.

Uma Nova Forma de Ver os Dados

Uma visão multidimensional é usualmente representada por um cubo

- A metáfora do cubo dá a impressão de múltiplas dimensões
- Os “cubos” podem ter 2, 3, 4 ou mais dimensões
- O usuário pode rolar e cortar (“*slice and dice*”) o cubo, escolhendo qual a dimensão(ões) que será usada em cada consulta (*query*)

Uma Nova Forma de Ver os Dados



Uma Nova Forma de Ver os Dados

Visualização - Exemplo

Cubo do Fato Vendas

Volume de vendas (milhares de R\$)

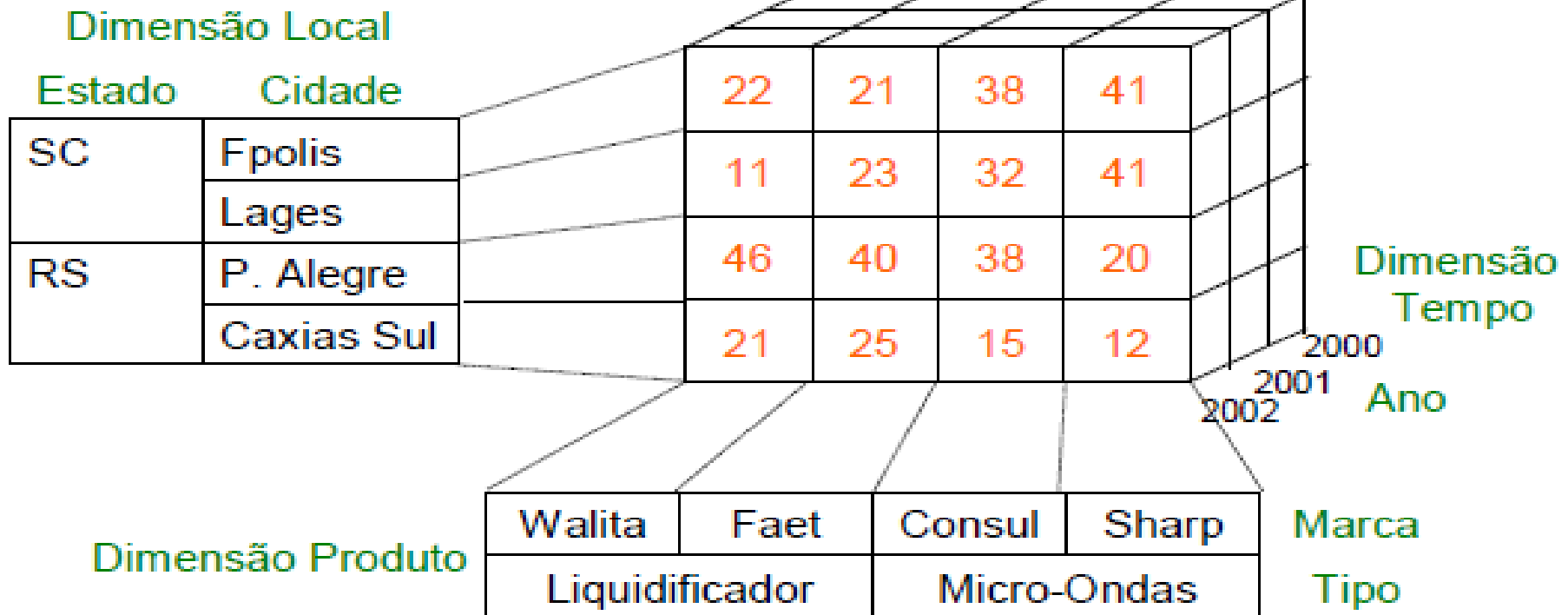


Diagrama Estrela (*Star Schema*)

Star Schema é o termo comum para designação de modelos de dados multidimensionais;

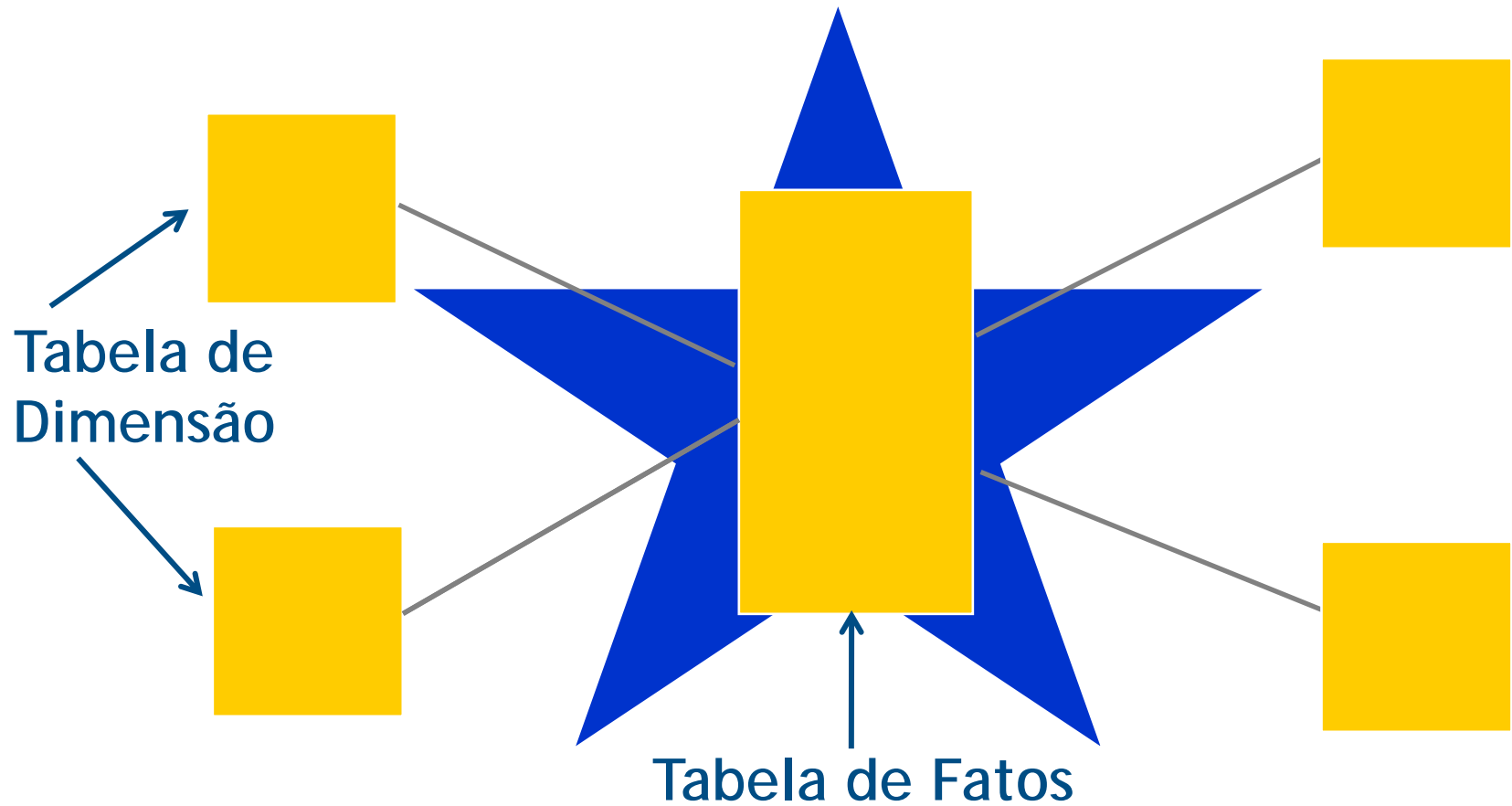
Como o termo esquema nos leva diretamente a uma ideia de modelo físico, o mais usual é denominarmos de modelo estrela;

Diagrama Estrela (*Star Schema*)

O modelo estrela é a estrutura básica de um modelo de dados multidimensional

Sua composição típica possui uma grande entidade central denominada fato e um conjunto de entidades menores denominadas dimensões, arranjadas ao redor dessa entidade central, formando uma estrela

Diagrama Estrela (*Star Schema*)



(Fonte: Campos. M.L.C., material de aula - UFRJ)

Diagrama Estrela (*Star Schema*)

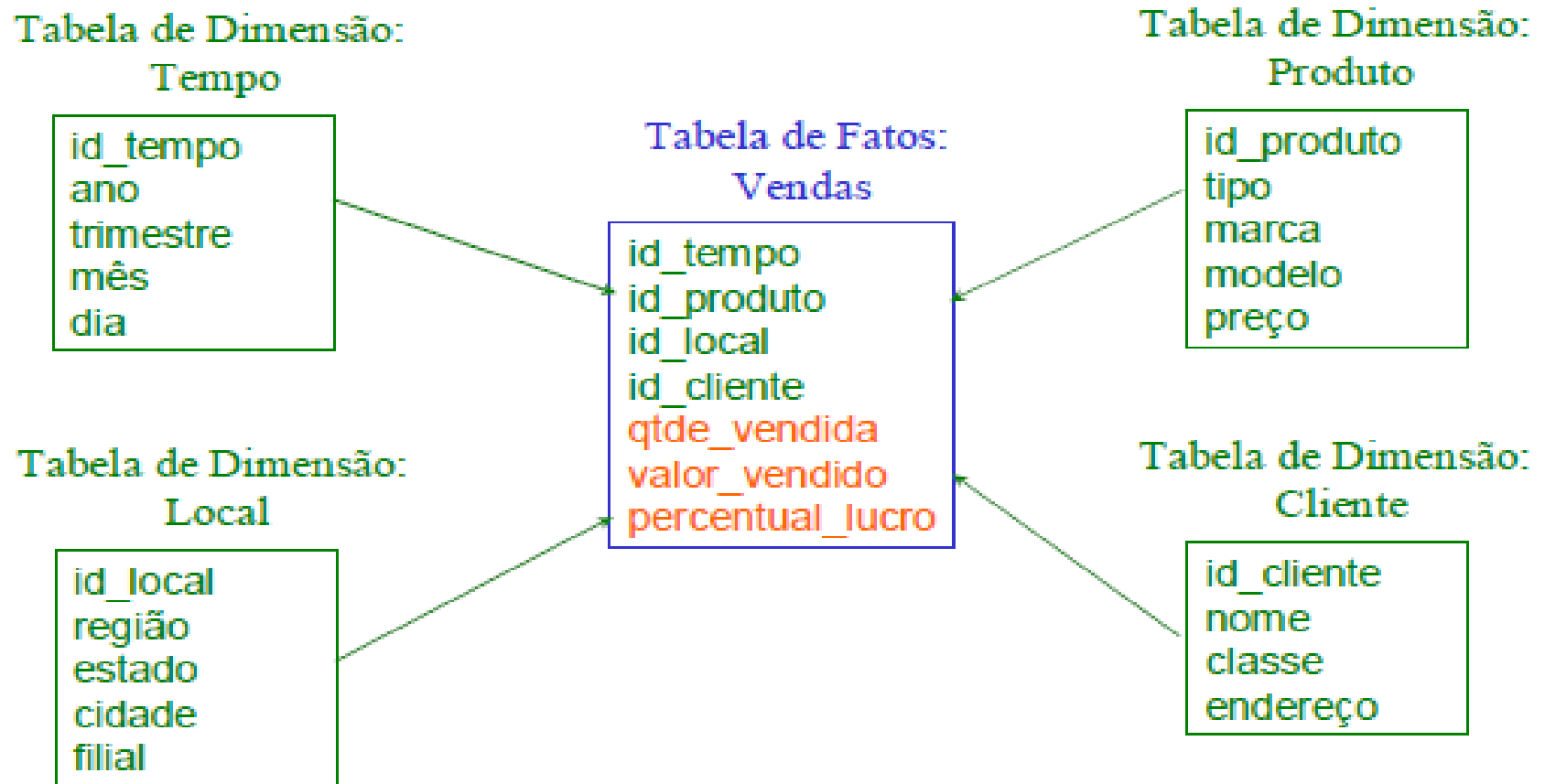


Diagrama Floco de Neve (*Snowflake*)

O modelo *Snowflake* é o resultado da decomposição de uma ou mais dimensões que possuam hierarquias entre seus membros;

Podemos definir relacionamento muitos para um entre os membros em uma dimensão, formando uma hierarquia por meio desses relacionamentos entre entidades dimensões;

➤ O modelo floco de neve é resultado da aplicação da terceira forma normal sobre as entidades dimensão

Diagrama Floco de Neve (Snowflake)

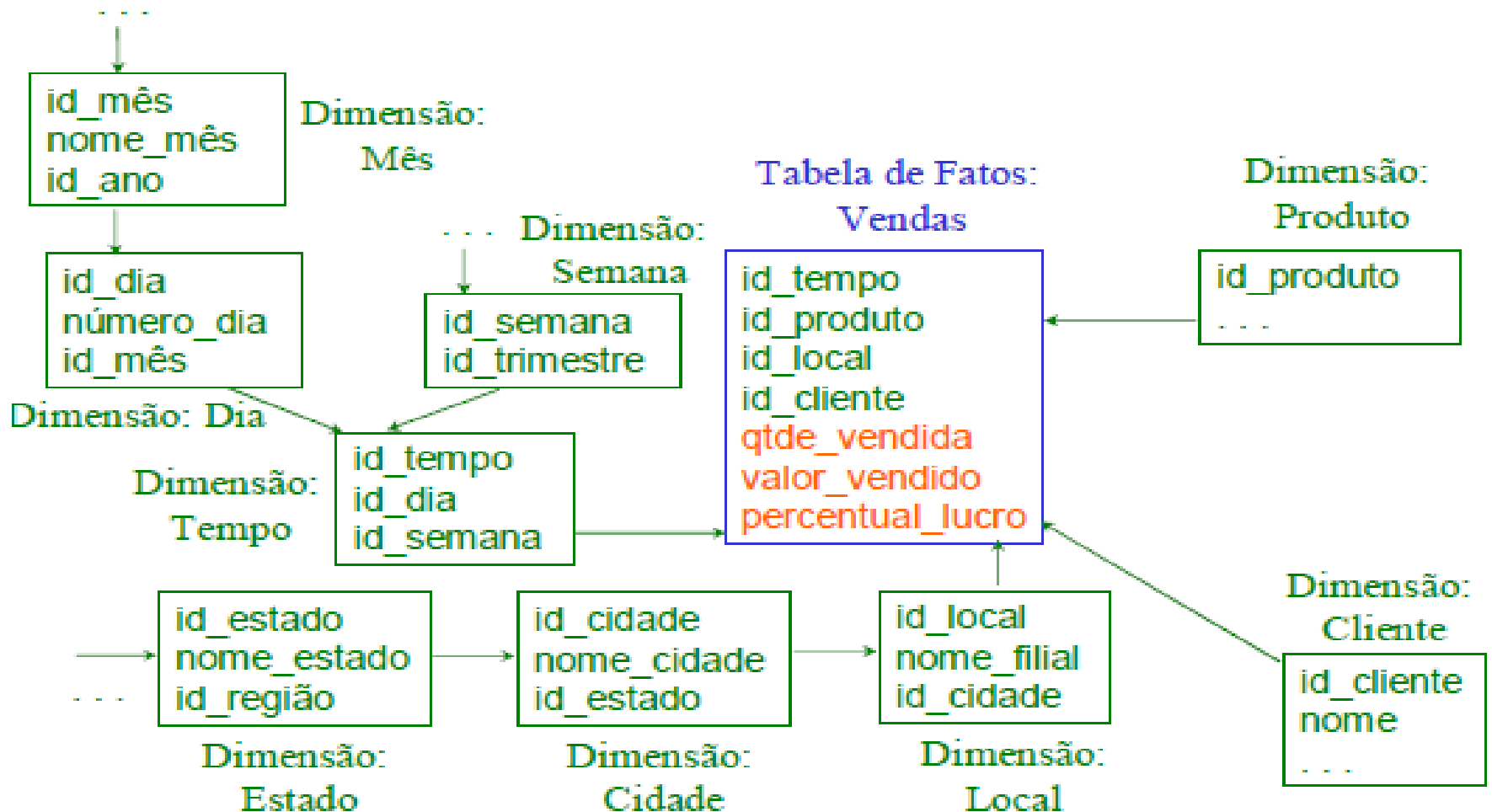


Diagrama Floco de Neve (*Snowflake*)

- Variante do modelo estrela

 - Tabelas de dimensões estão normalizadas (3FN)

 - Representação explícita das hierarquias

- Vantagem

 - Evita redundância (mais economia de espaço)

Diagrama Floco de Neve (*Snowflake*)

- Desvantagem
 - maior número de tabelas (maior número de junções)
- Modelo estrela é considerado mais apropriado
 - desperdício de espaço é menos crítico que tempo para recuperação de dados em um DW

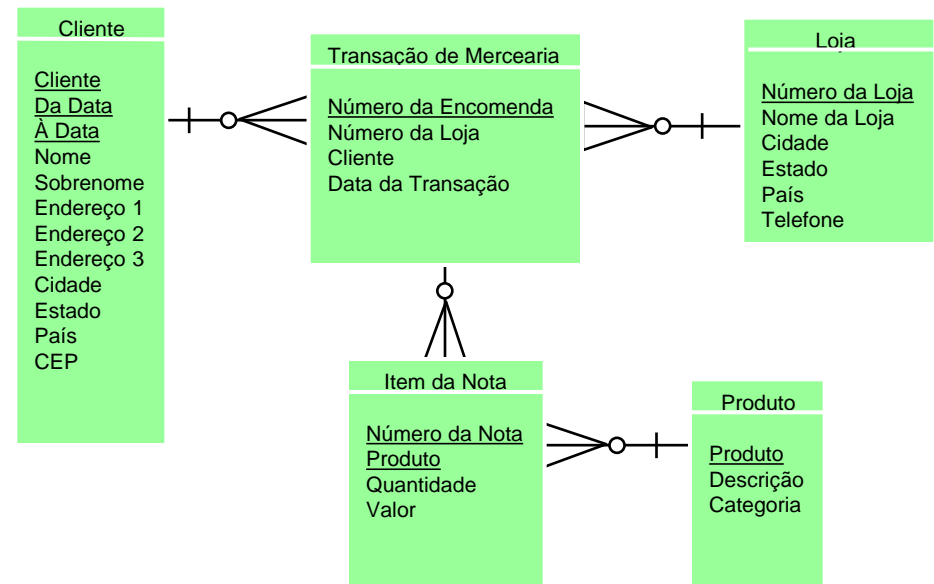
Diagrama Floco de Neve (Snowflake)

Estrela



Altamente desnormalizado para melhor performance
Os dados transacionais são armazenados em Tabelas Fato
e os dados de referência são armazenados em Tabelas Dimensão

Relacional Normalizado

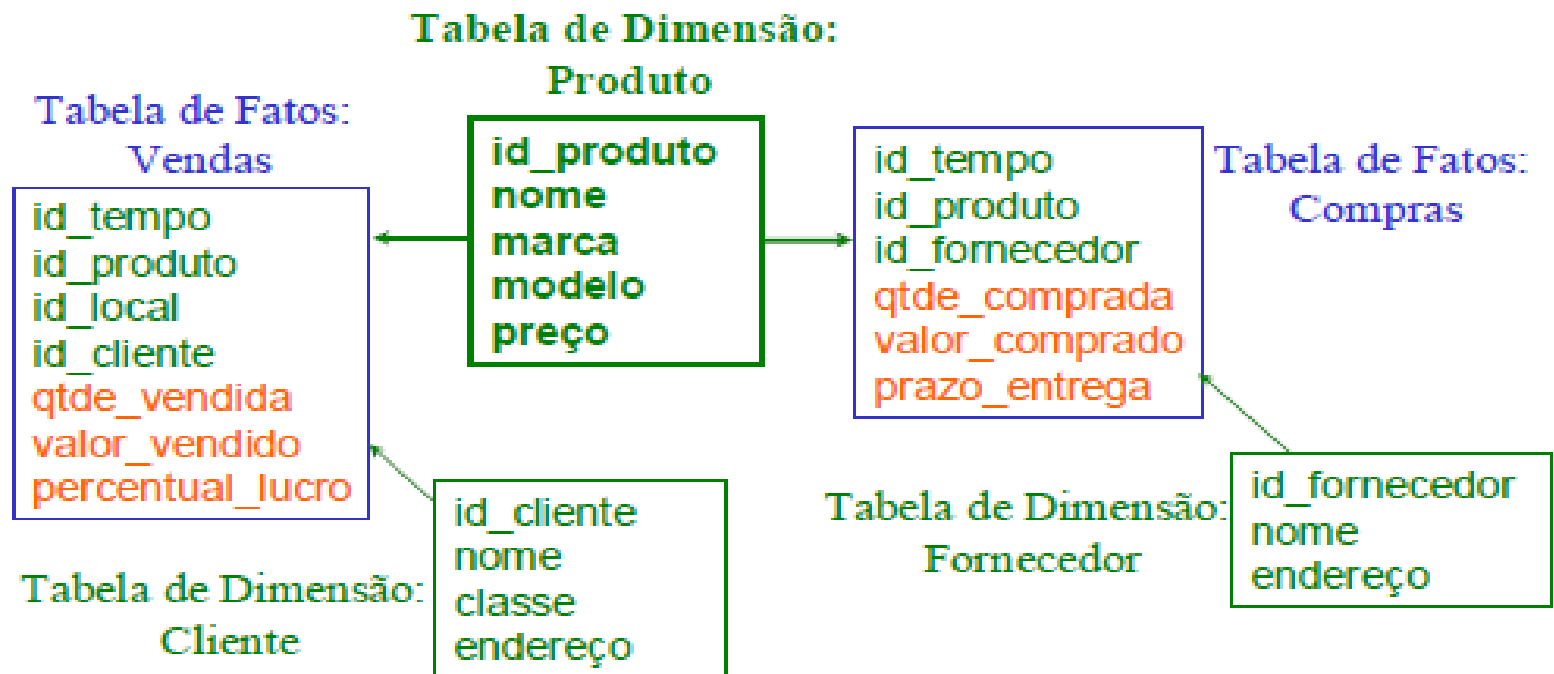


Minimiza a redundância de dados
Os dados transacionais podem ser armazenados em qualquer tabela

Diagrama Constelação

União de esquemas estrela e/ou floco de neve

- esquema estrela: modela em geral um Data Mart
- esquema constelação: modela o DW



Exemplo Modelagem

Uma rede de lojas de departamento possui diversas lojas distribuídas em diferentes estados do Brasil. O sistema transacional gerencia todas as vendas efetuadas em cada uma das lojas.

O modelo relacional utilizado nesse sistema possui:

- a) uma tabela Fatura que contém todos os dados relacionados a uma fatura, com seus itens vendidos sendo armazenados na tabela Item_Fatura.
- b) cadastros de Lojas, de Clientes, de Funcionários, de Produtos, de Categorias e de Fornecedores.

Exemplo Modelagem

A área de negócios solicitou a construção de um Data Mart que permitisse consultas analíticas, como por exemplo:

- ✓ qual o faturamento com a venda de cosméticos da "Johnson & Johnson", em lojas da zona sul da cidade de São Paulo no período do dia das mães ?
- ✓ quais foram as vendas no período de janeiro a março deste ano na região sudeste e nordeste ?
- ✓ quais os produtos mais rentáveis ?
- ✓ quem foram os maiores clientes ?

Exemplo Modelagem

Quem é o fato?

Neste exemplo o fato é VENDAS

Quem são as dimensões?

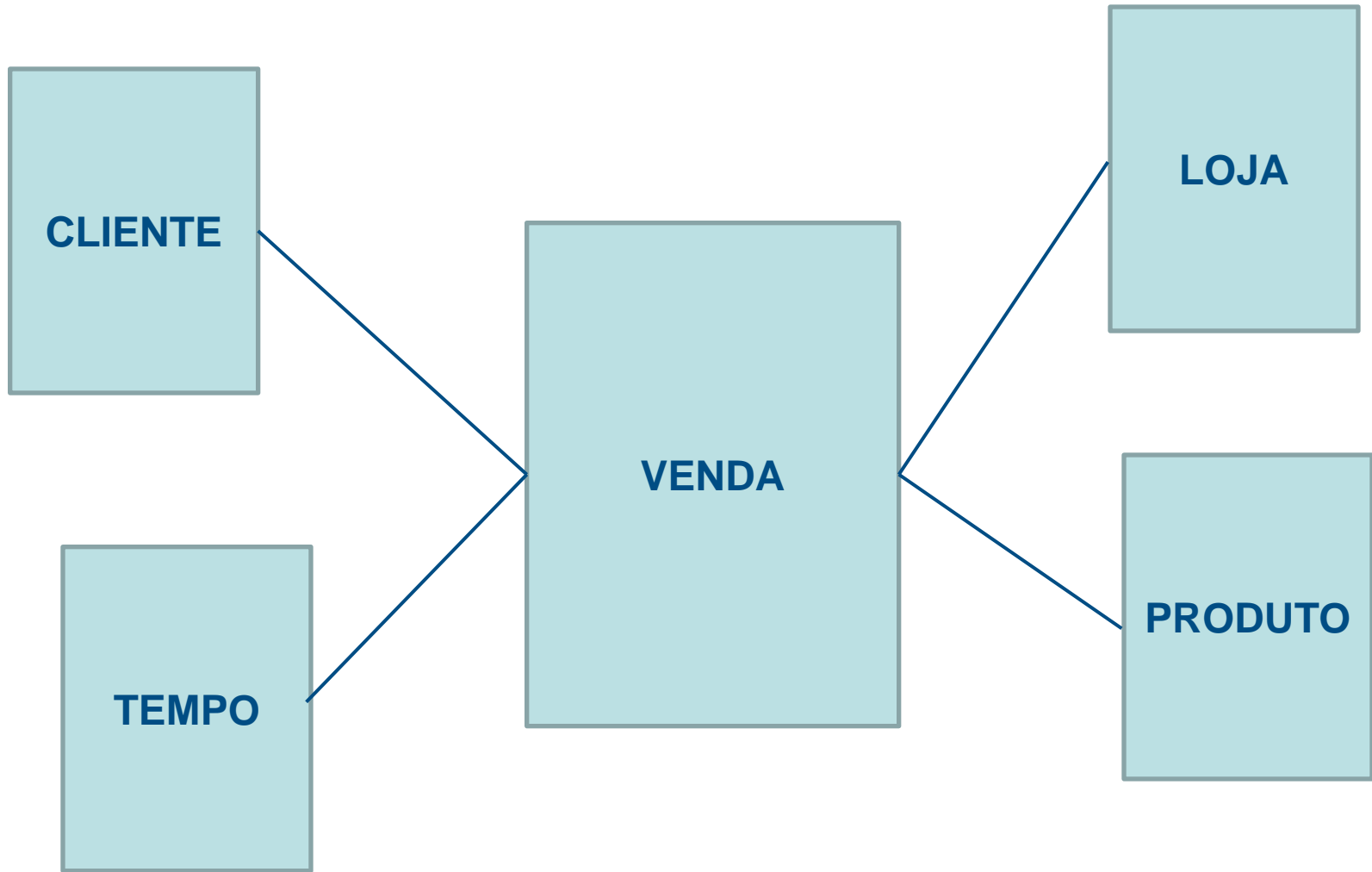
PRODUTO

LOJA

CLIENTE

TEMPO

Exemplo Modelagem



Obrigado!

...e agora suas perguntas?



ricardo.avila@outlook.com.br



@theavila