

Conceitos Básicos

Processamento Analítico de Dados
Profa. Dra. Cristina Dutra de Aguiar Ciferri
Prof. Dr. Ricardo Rodrigues Ciferri

Data Warehousing

Engloba **arquiteturas**, **algoritmos** e **ferramentas** que possibilitam que dados selecionados de **provedores de informação** autônomos, heterogêneos e distribuídos sejam **integrados** em uma única base de dados, conhecida como **data warehouse (DW)**

Acesso às Informações

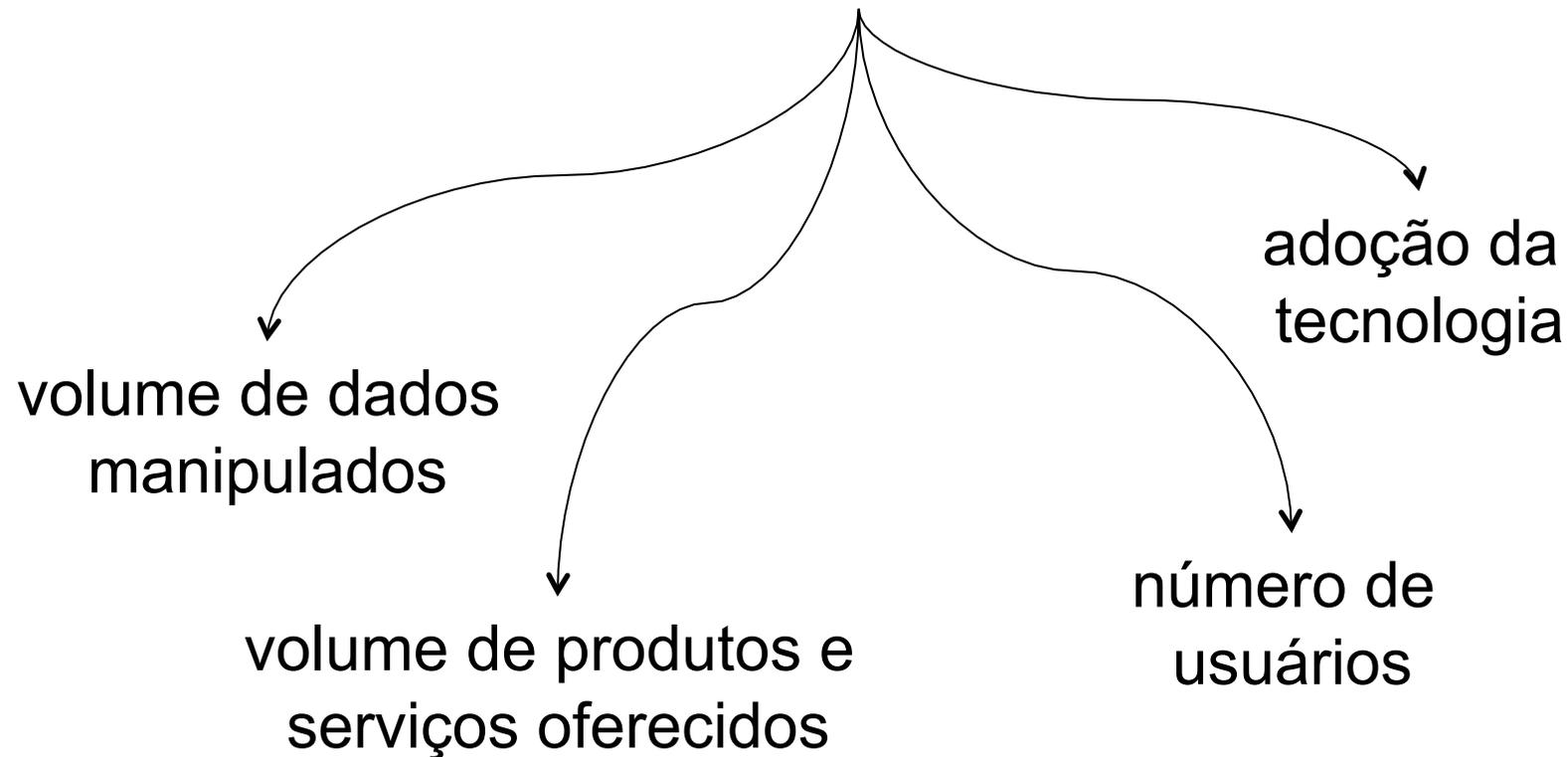
- Duas etapas
 - a informação de cada provedor é extraída previamente, devendo ser traduzida, filtrada, integrada à informação relevante de outros provedores e finalmente armazenada no **DW**
 - as consultas, quando realizadas, são executadas diretamente no **DW**, sem acessar os provedores de informação originais

Exemplos de Análises

- Análises de tendências simples
 - *Quais as vendas mensais de um certo produto no ano de 1998?*
- Análises comparativas
 - *Quais as vendas mensais dos produtos de uma dada marca nos últimos 3 anos?*
- Análises de tendência múltiplas
 - *Quais as vendas mensais dos produtos de uma dada marca nos últimos 3 anos, de acordo com as promoções de Natal?*

Visão do Mercado

Crescimento explosivo do uso da tecnologia de data warehousing



Ambiente Operacional *versus* Ambiente Informacional

	Ambiente Operacional	Ambiente Informacional
Principal Característica	voltado ao processamento de transações OLTP	voltado ao processamento de consultas OLAP
Tipos de Operação mais Frequentes	atualização remoção inserção	leitura (consulta)

o termo OLAP (on-line analytical processing) foi introduzido em 1993 por Codd *et al.* para definir a categoria de processamento analítico sobre um banco de dados histórico voltado para os processos de gerência e tomada de decisão

Ambiente Operacional *versus* Ambiente Informacional

	Ambiente Operacional	Ambiente Informacional
Volume de Transações	relativamente alto	relativamente baixo
Características das Transações	pequenas e simples, acessam poucos registros por vez	longas e complexas, acessam muitos registros por vez e realizam várias varreduras e junções de tabelas

Ambiente Operacional *versus* Ambiente Informacional

	Ambiente Operacional	Ambiente Informacional
Tipos de Usuários	administradores do sistema, projetistas, usuários de entrada de dados	usuários de SSD por exemplo: executivos, analistas, gerentes, administradores
Número de Usuários Concorrentes	grande (geralmente milhares)	relativamente pequeno (geralmente centenas)
Interações com os Usuários	pré-determinadas estáticas	<i>ad-hoc</i> dinâmicas

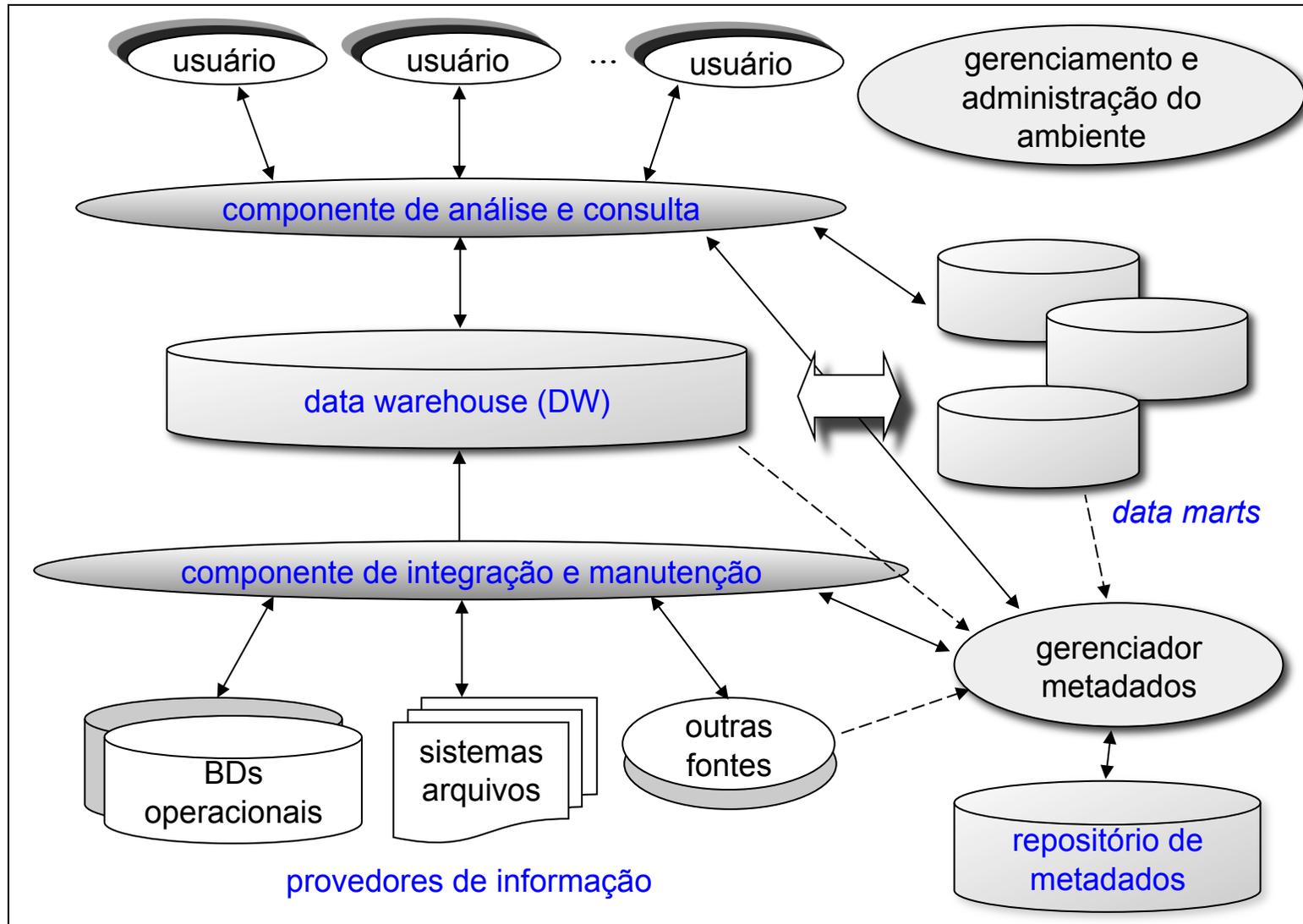
Ambiente Operacional *versus* Ambiente Informacional

	Ambiente Operacional	Ambiente Informacional
Volume de Dados	<i>megabytes a gigabytes</i>	<i>gigabytes a terabytes</i>
Projeto do Banco de Dados	normalizado para suporte às propriedades ACID	multidimensional, refletindo as necessidades de análise dos usuários de SSD
Granularidade dos Dados	detalhado	detalhado e agregado

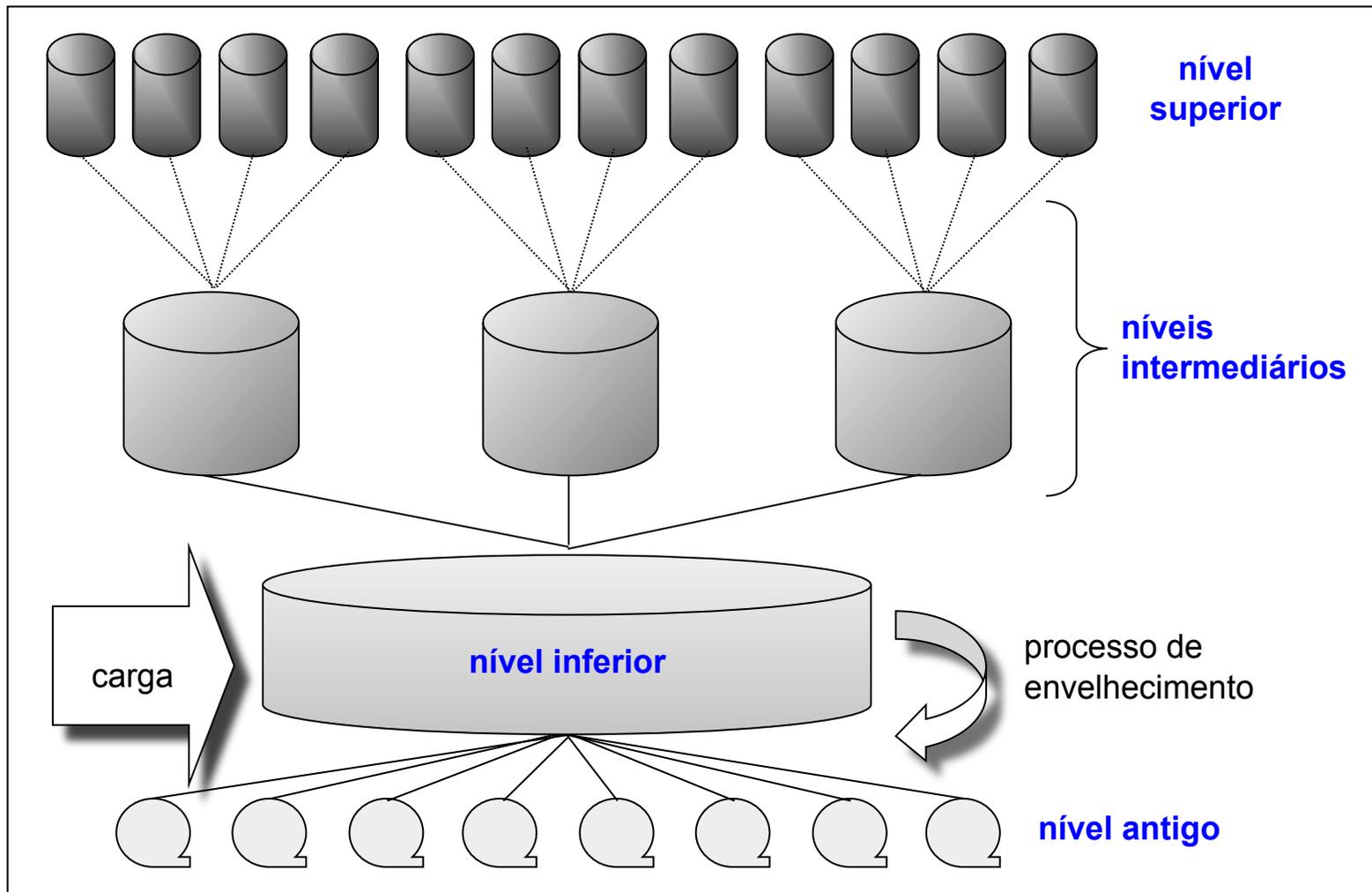
Ambiente Operacional *versus* Ambiente Informacional

	Ambiente Operacional	Ambiente Informacional
Principal Questão de Desempenho	produtividade da transação	produtividade da consulta
Tempo de Resposta	geralmente poucos segundos	de minutos a horas
Exemplos de aplicações	transações bancárias, empréstimos de livros, contas a pagar	planejamento de <i>marketing</i> , análise financeira

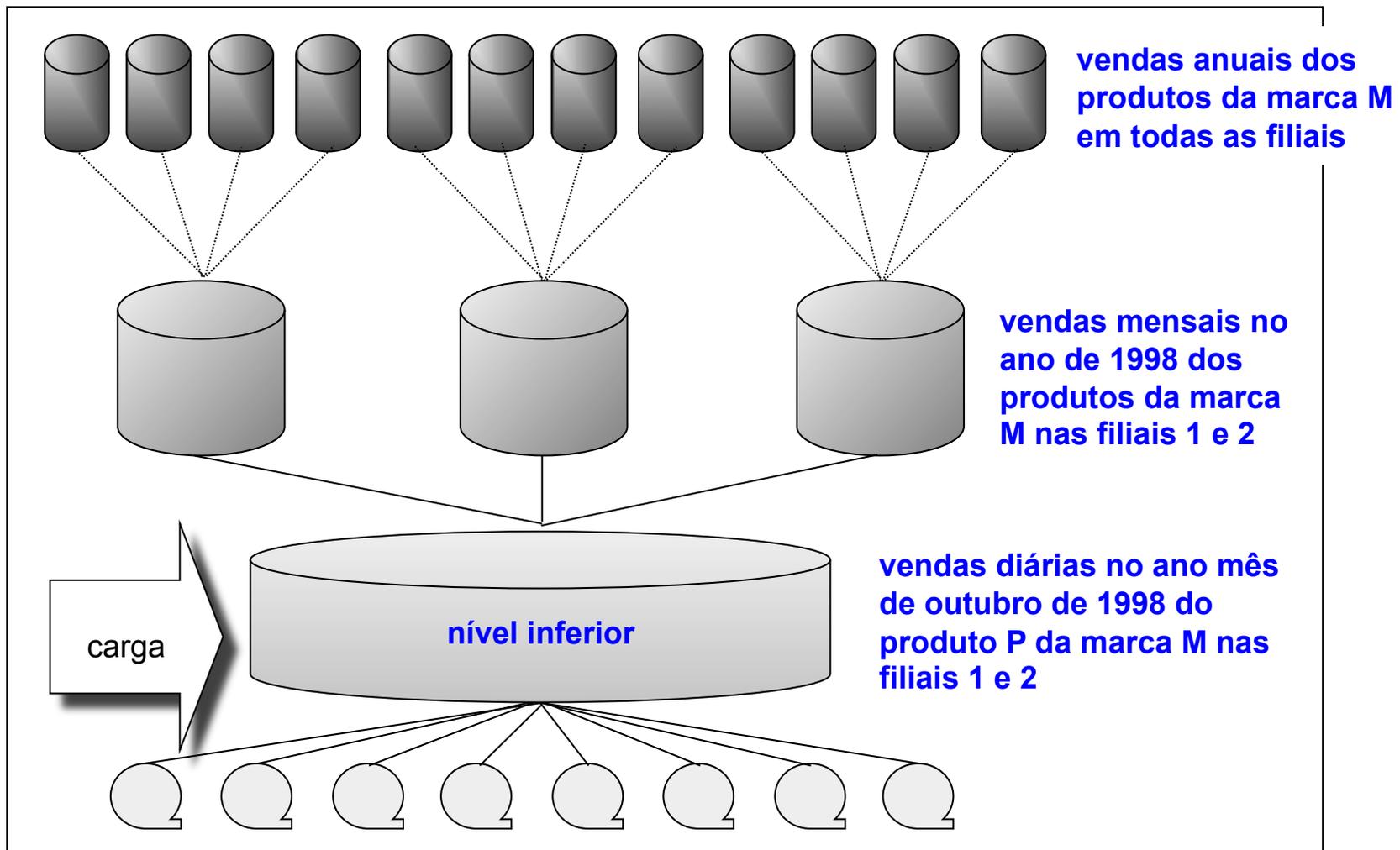
Arquitetura Típica



Níveis de Agregação



Níveis de Agregação

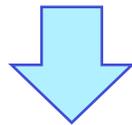


Modelagem Multidimensional

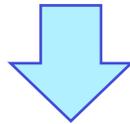
- Análises dos usuários de SSD
 - representam requisições multidimensionais aos dados do DW
 - permitem a identificação de problemas e de tendências
- Principais enfoques
 - modelo de dados multidimensional
 - abordagens para a representação lógica

Modelo de Dados Multidimensional

aspectos **estáticos**

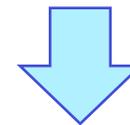


modelagem dos dados

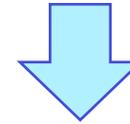


dimensões (atributos)
medidas numéricas

aspectos **dinâmicos**



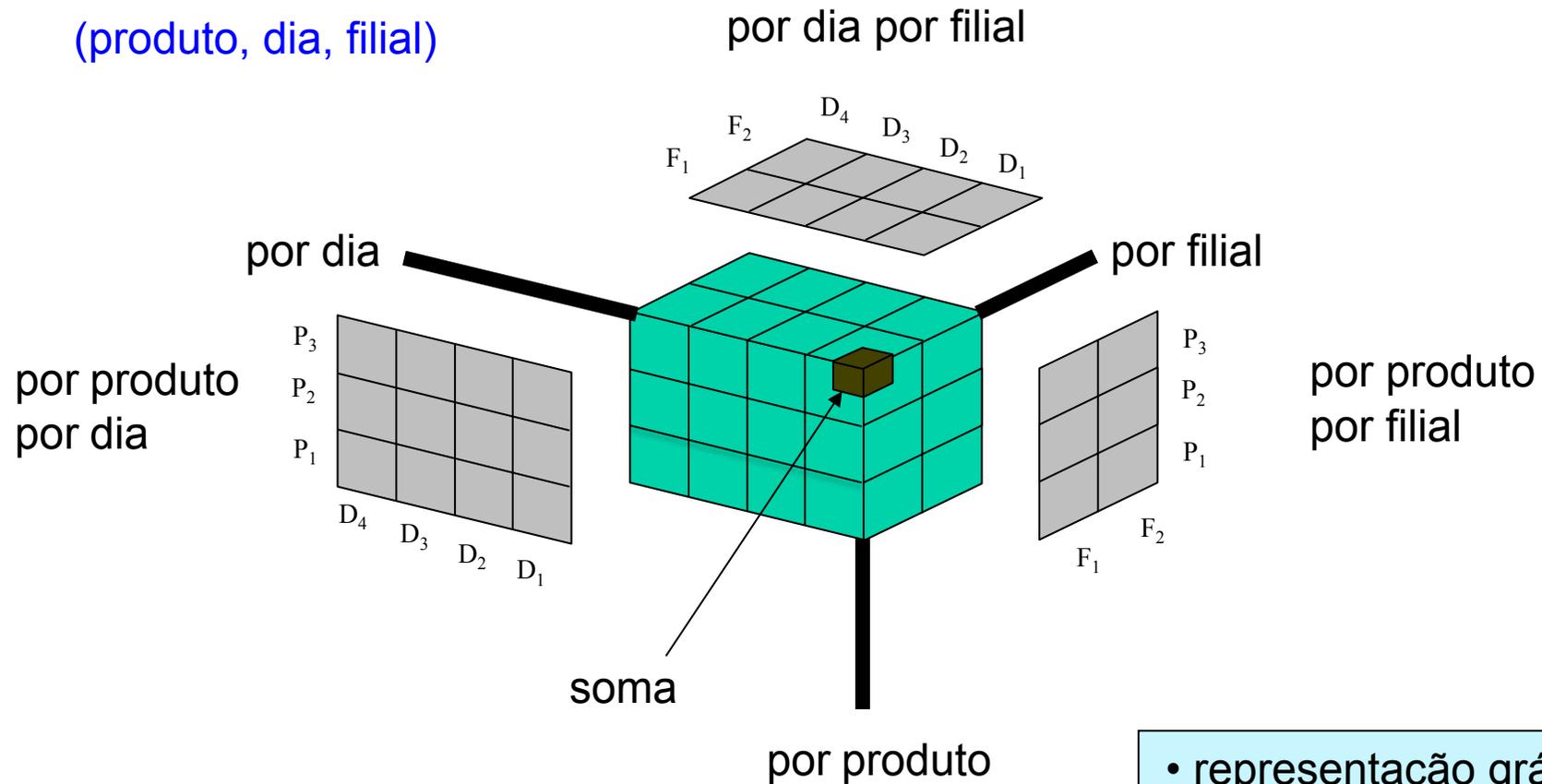
operações analíticas



drill-down/roll-up
slice and dice
pivot
drill-across ...

Cubo de Dados Multidimensional

(produto, dia, filial)



- representação gráfica
- semântica subjacente

Dimensão

- Representa uma perspectiva de análise dos usuários de SSD
- Composta por atributos
- *Exemplo*: dimensão **filial**
 - *atributos*: **cidade**, **estado**, **região**, **país**
 - *semântica*: a filial “Primeira Filial” está localizada na cidade de “São Carlos”, estado de “São Paulo”, região “Sudeste” do país “Brasil”

Hierarquia de Atributos

- Definição
 - permite que atributos de uma dimensão relacionem-se com outros atributos da mesma dimensão
 - especifica níveis de agregação e, portanto, granularidade dos itens de dados
- *Exemplo*: dimensão **filial**
 - cidade → estado → região → país
 - *hierarquia de nível quatro na dimensão filial*

Definição formal: grafo de derivação

Medida Numérica

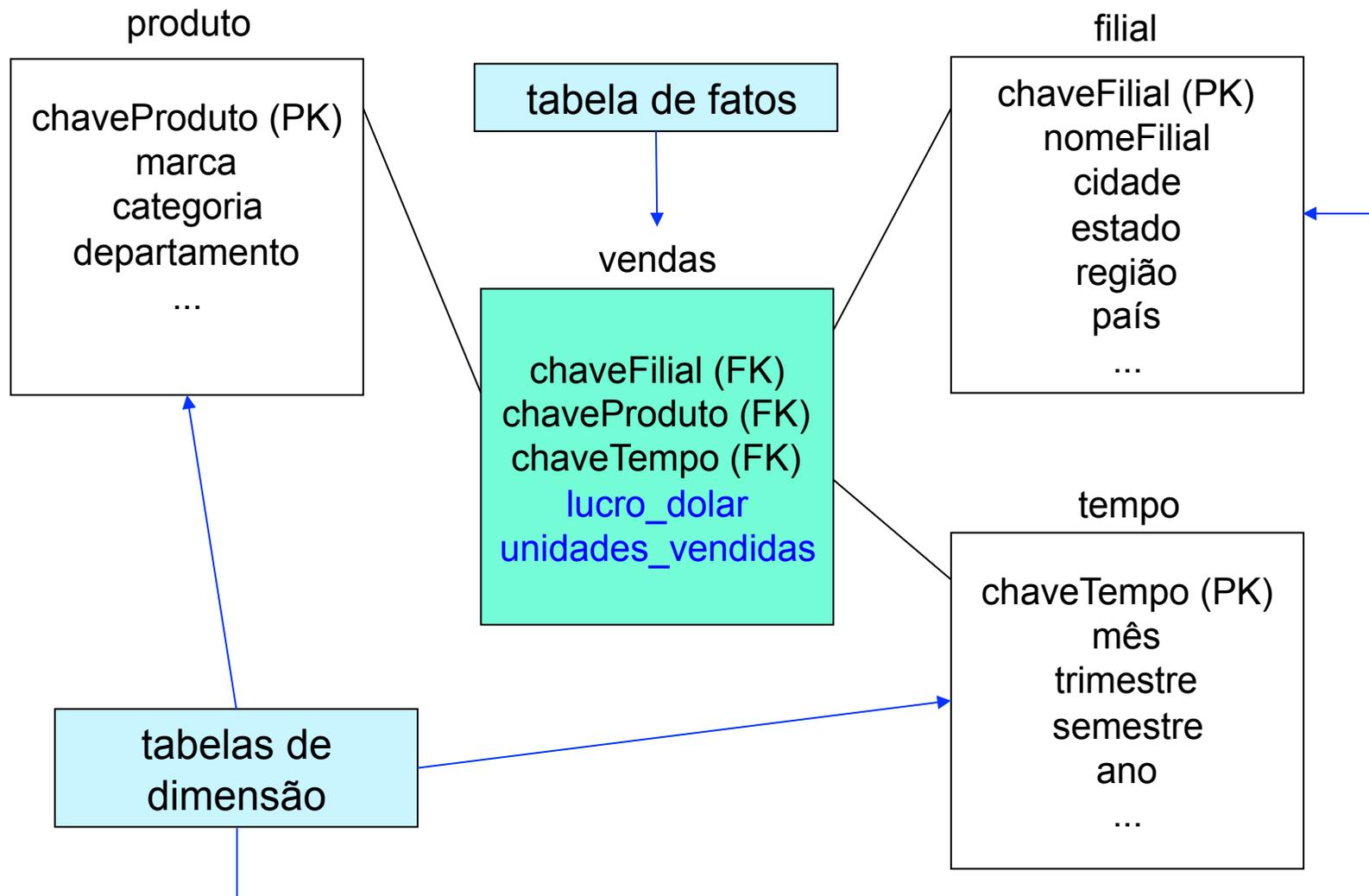
- Objeto de análise relevante ao negócio
- Definida como uma função de suas dimensões correspondentes

Classificação	Definição	Exemplo
aditiva	<i>somada</i> através de todas as suas dimensões	unidades-vendidas
semi-aditivas	<i>somadas</i> somente através de algumas de suas dimensões	número-clientes
não aditivas	não podem ser somadas	preço

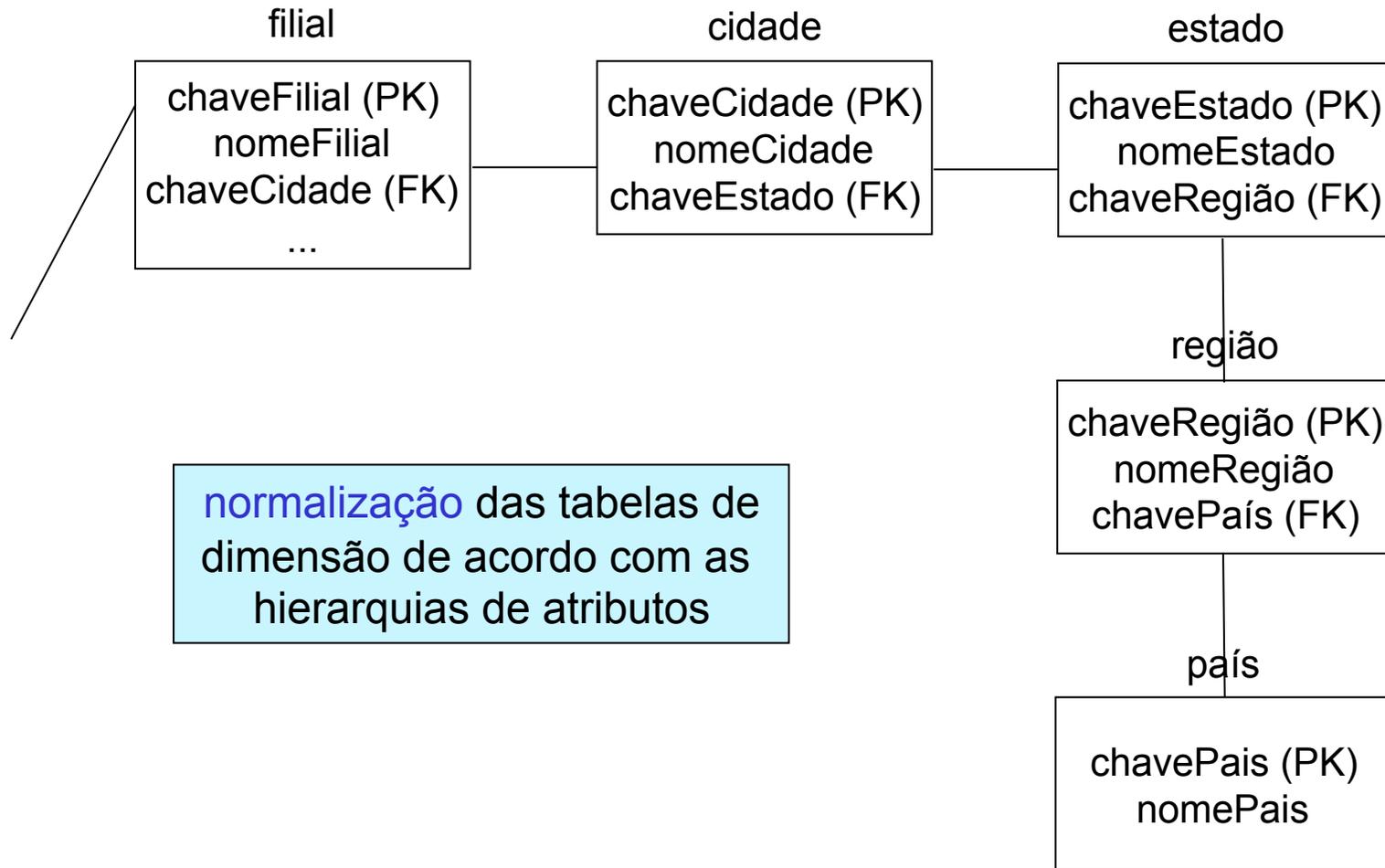
Operações Analíticas

Operação	Definição
<i>drill-down</i>	analisa os dados em níveis de agregação progressivamente mais detalhados, ou de menor granularidade
<i>roll-up</i>	analisa os dados em níveis de agregação progressivamente menos detalhados, ou de maior granularidade
<i>slice and dice</i>	restringe os dados sendo analisados a um subconjunto destes dados <i>slice</i> : corte para um valor fixo <i>dice</i> : seleção de faixas de valores
<i>pivot</i>	reorienta a visão multidimensional dos dados, oferecendo diferentes perspectivas dos mesmos dados
<i>drill-across</i>	compara medidas numéricas distintas que são relacionadas entre si através de pelo menos uma dimensão em comum

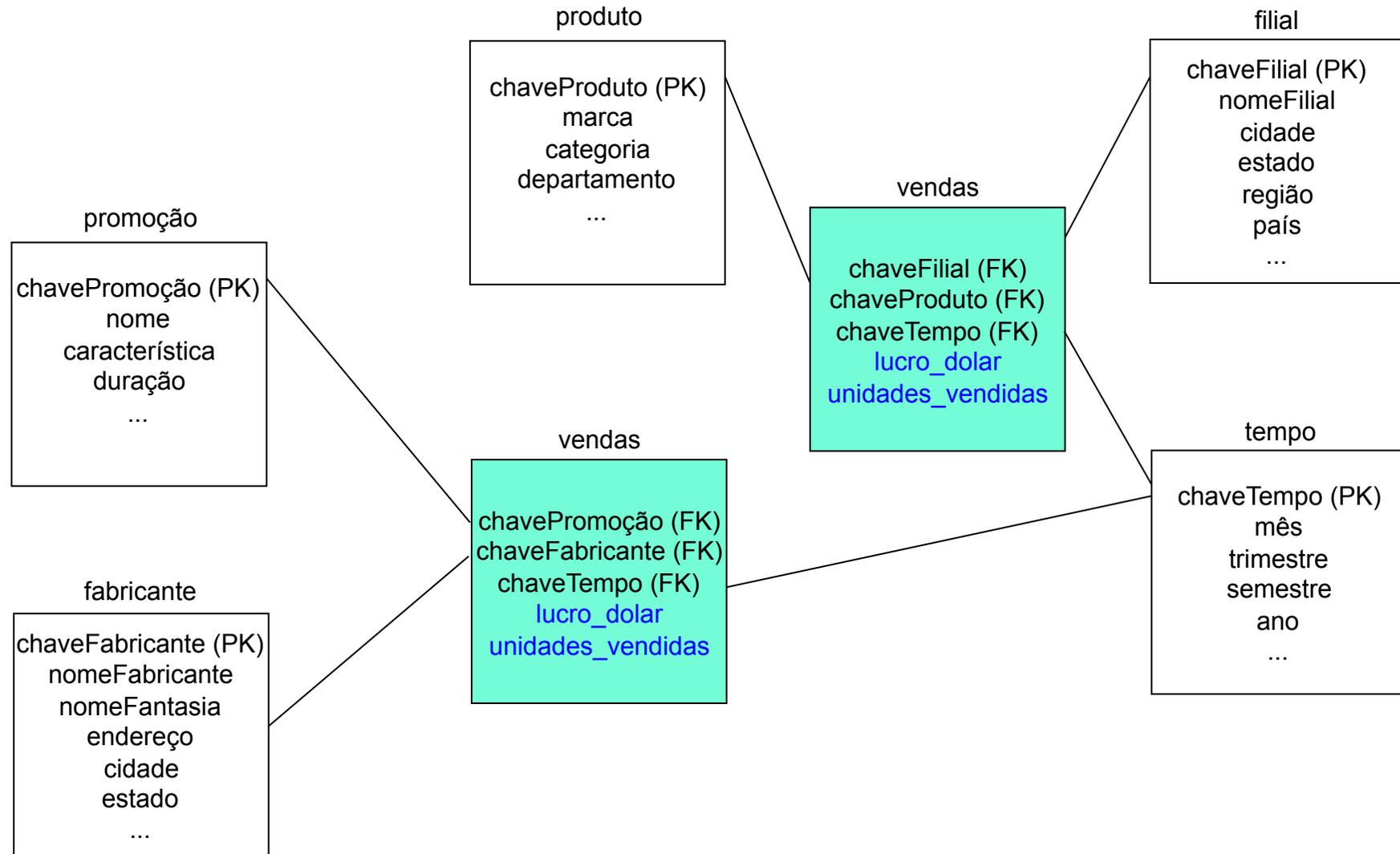
Esquema Estrela



Esquema Floco de Neve



Constelação de Fatos



Visão Materializada

- Definição
 - especificação: intenção
 - dados: extensão
- Utilidade em ambientes de DWing
 - aumento no desempenho de processamento de consultas
 - diminuição dos custos relacionados à atualização de outras visões materializadas

Visões Materializadas & Níveis de Agregação

- Nível inferior
 - conjunto de visões materializadas no qual as **relações base** residem nos **provedores** de informação
- Demais níveis
 - conjunto de visões materializadas no qual as **relações base** são as do **nível** imediatamente **subjacente**