

Modelagem Multidimensional & Visões Materializadas

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

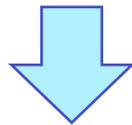
Modelagem Multidimensional

Modelagem Multidimensional

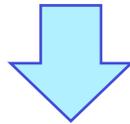
- Análises dos usuários de SSD
 - representam requisições multidimensionais aos dados do DW
 - permitem a identificação de problemas e de tendências
- Principais enfoques
 - modelo de dados multidimensional
 - abordagens para a representação lógica

Modelo de Dados Multidimensional

aspectos **estáticos**

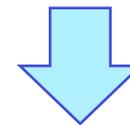


modelagem dos dados

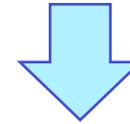


dimensões (atributos)
medidas numéricas

aspectos **dinâmicos**

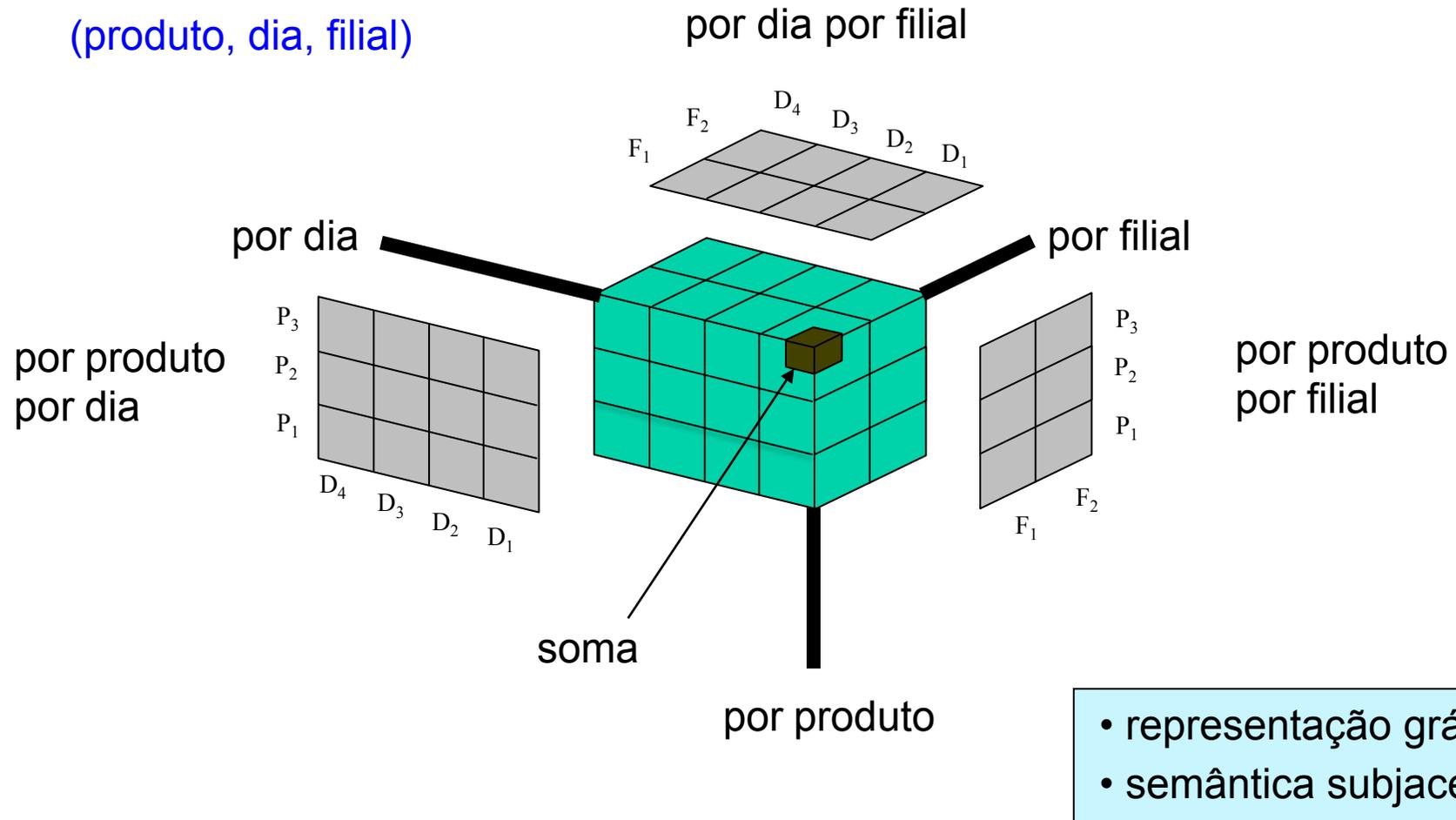


operações analíticas



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

Cubo de Dados Multidimensional



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

Representação Lógica

sistemas ROLAP

(OLAP relacional)



estruturas relacionais



dados armazenados em
tabelas e SGBD
relacional estendido para
oferecer suporte às
operações OLAP

sistemas MOLAP

(OLAP multidimensional)



estruturas de dados
especializadas



dados armazenados em
matrizes e operações
OLAP diretamente
implementadas sobre
essas matrizes

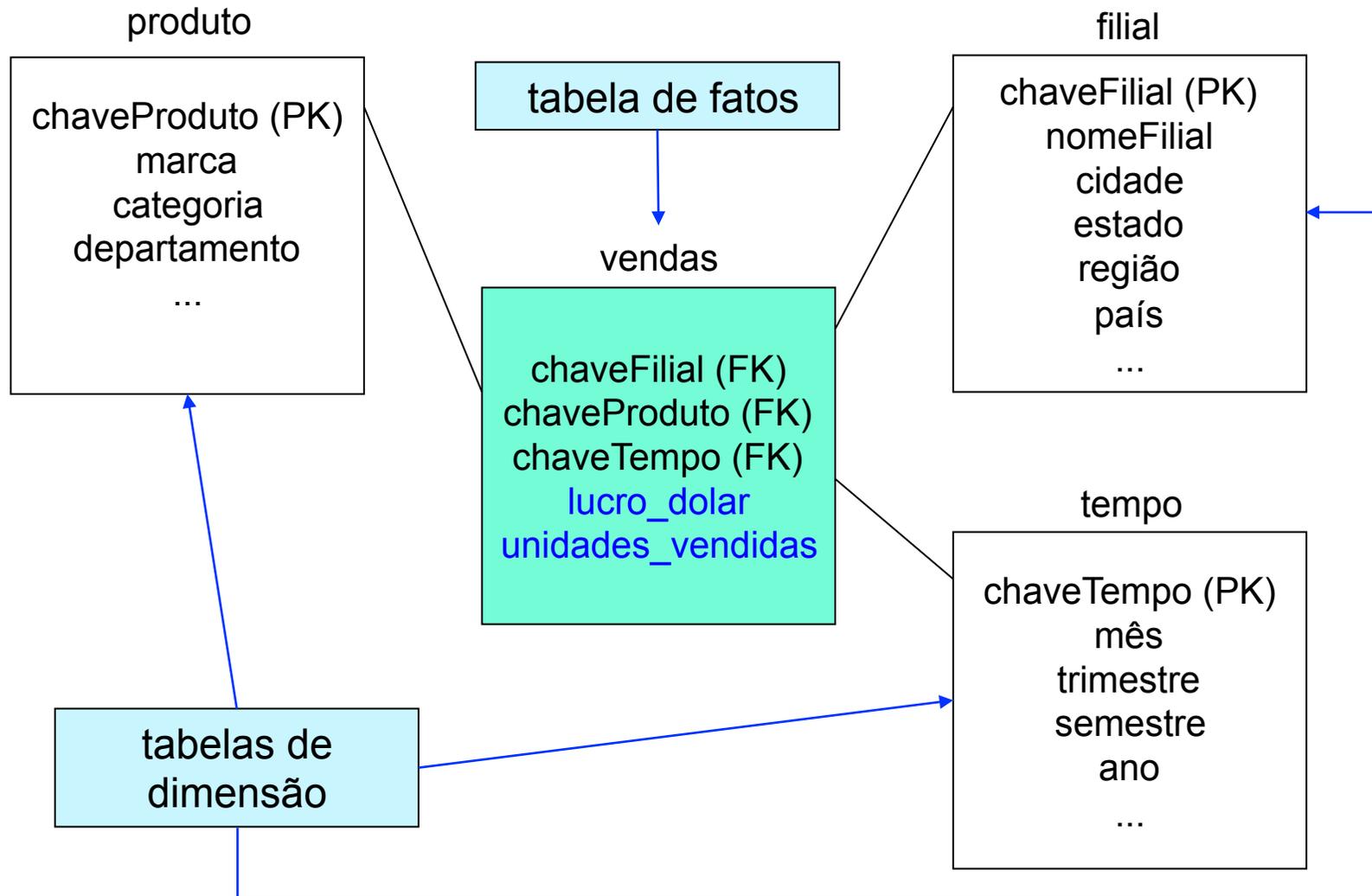
Sistemas ROLAP: Componentes

- Máquina relacional
 - SGBD adaptado para processamento analítico
 - armazenamento dos dados
 - esquema estrela
 - esquema floco de neve
 - funcionalidades adicionais
 - novos operadores SQL
 - estruturas de indexação especializadas
 - otimização de consultas complexas
 - execução paralela de consultas complexas

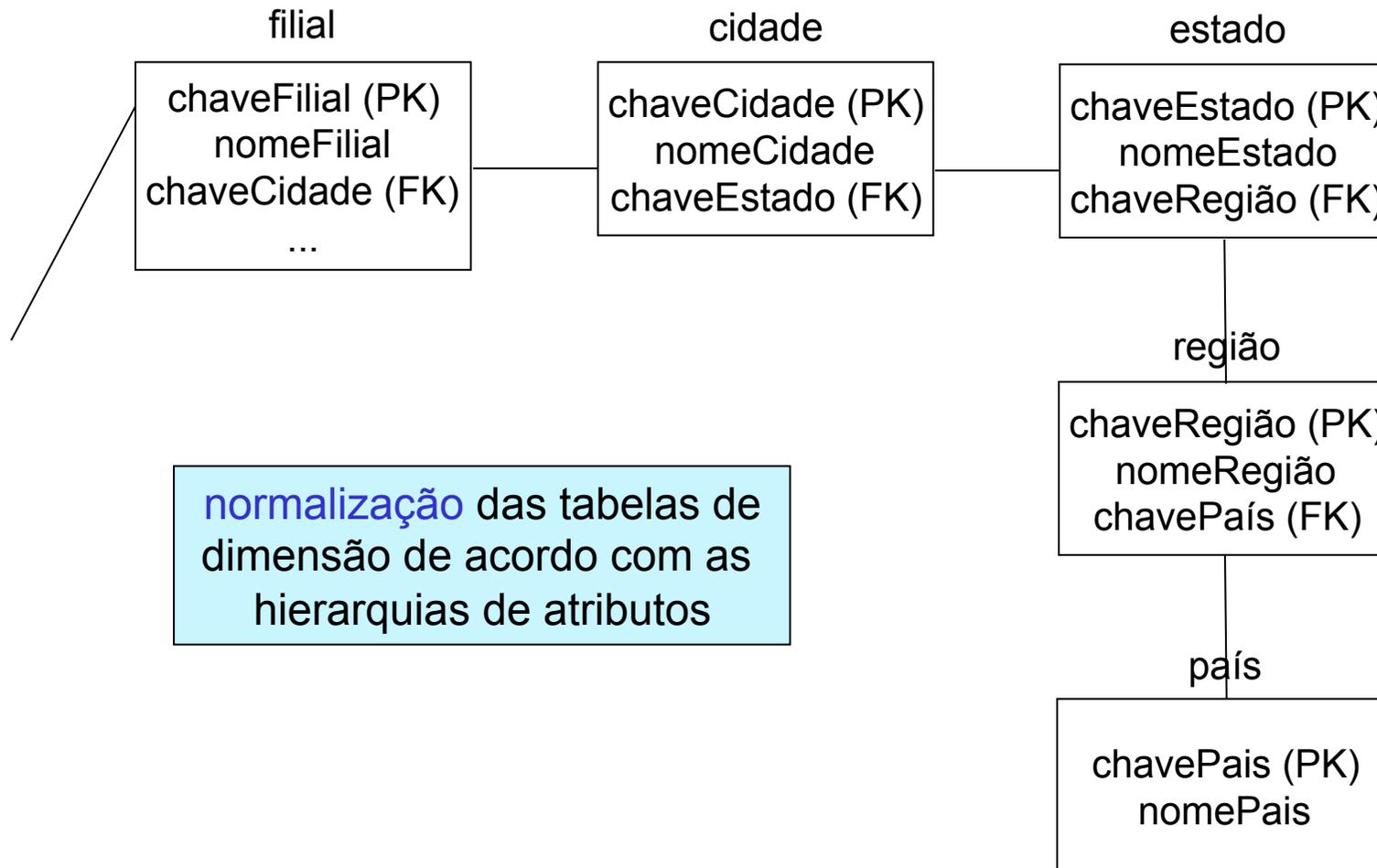
Sistemas ROLAP: Componentes

- Máquina ROLAP
 - camada de *software* que complementa a máquina relacional
 - objetivos
 - suprir as limitações da máquina relacional
 - apresentar os dados multidimensionalmente
 - funcionalidade adicional
 - materialização de visões

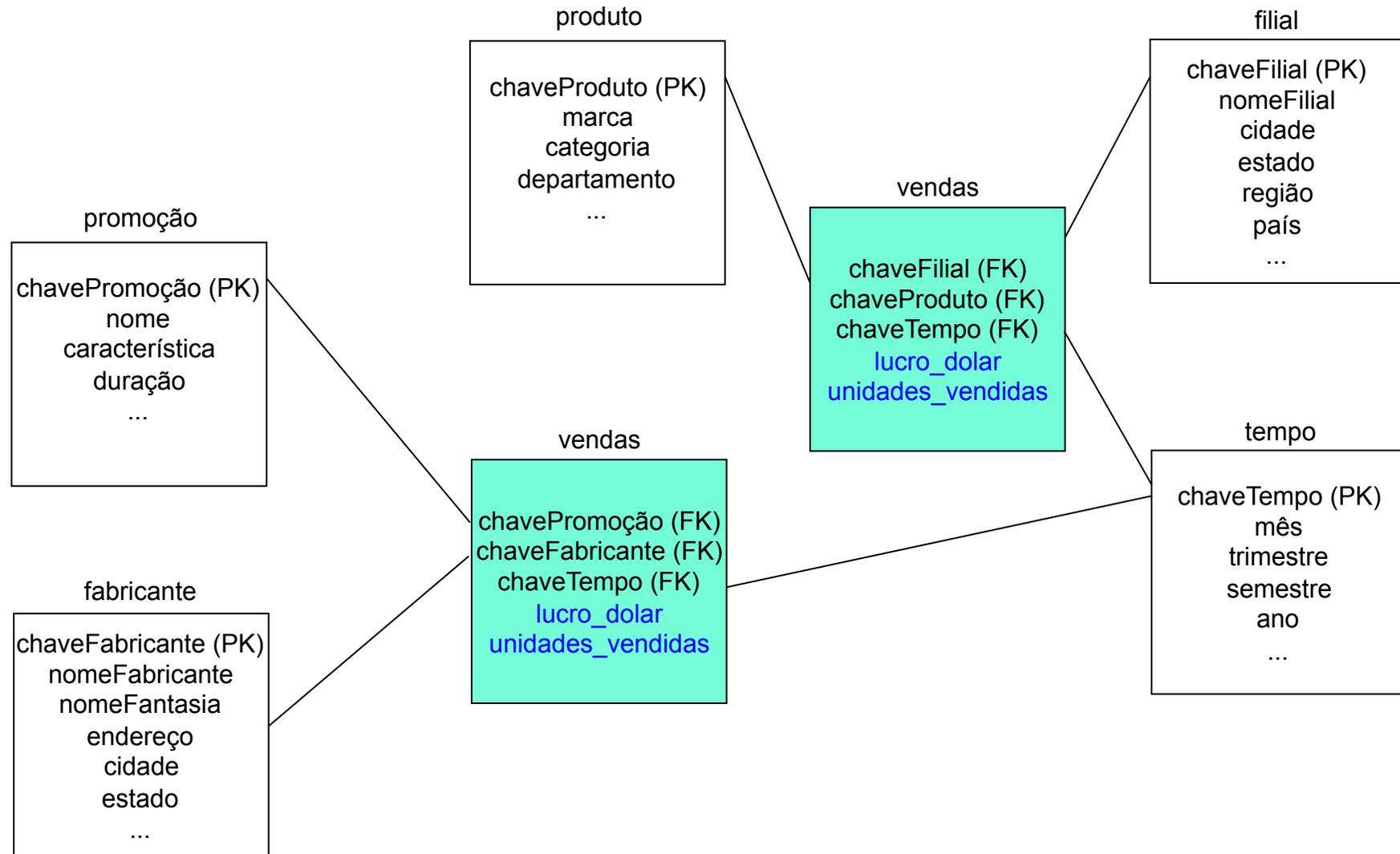
Esquema Estrela



Esquema Floco de Neve



Constelação de Fatos



Visões

Materializadas

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**

Grafo Orientado

- Par (V, E) de conjuntos disjuntos de vértices V e arestas E
- Mapeamentos
 - $\text{inic}: E \rightarrow V$ e $\text{term}: E \rightarrow V$
 - cada aresta e sai de um vértice inicial $\text{inic}(e)$ e chega a um vértice terminal $\text{term}(e)$
 - e é direcionada de $\text{inic}(e)$ para $\text{term}(e)$.
- Característica
 - não possui ciclos nem arestas múltiplas

Grafo de Derivação

Lattice (de *visões* ou *agregações*)

- Grafo orientado, com as propriedades:
 - existe uma ordenação parcial \preceq entre as visões
 - $v \preceq u$ se e somente se v pode ser determinada usando somente os resultados de u
 - existe uma visão **topo**, da qual cada outra visão agregada é dependente
 - existe uma visão completamente agregada **vazio** que pode ser calculada a partir de qualquer outra visão

Grafo de Derivação

- Para três visões v , w , u , podem ser definidas as seguintes funções:

$$\text{ancestrais}(v) = \{ w \mid v \preceq w \} .$$

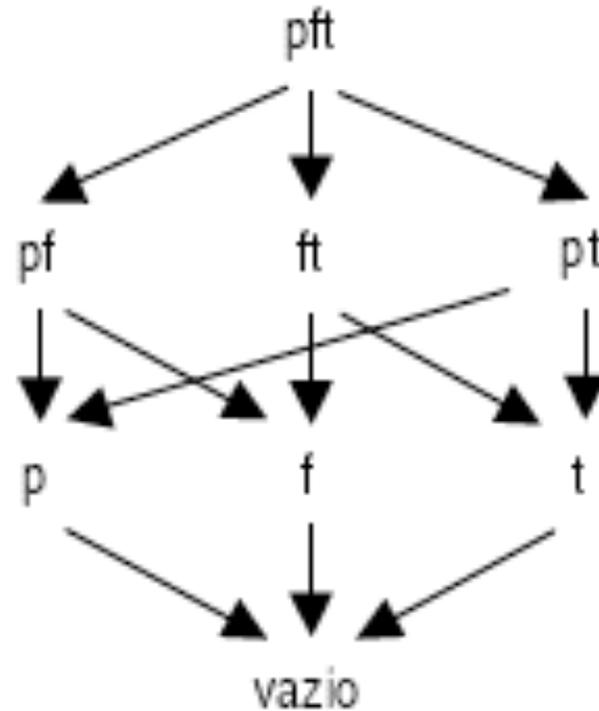
$$\text{descendentes}(v) = \{ w \mid w \preceq v \} .$$

$$\text{ancestrais_diretos}(v) = \text{pais}(v) = \{ w \mid v \prec w, \exists u, v \prec u, u \prec w \} .$$

$$\text{descendentes_diretos}(v) = \text{filhos}(v) = \{ w \mid w \prec v, \exists u, w \prec u, u \prec v \} .$$

$$\text{sendo que } v \prec w \Rightarrow v \preceq w \wedge v \neq w .$$

Exemplo



níveis de agregação

inferior

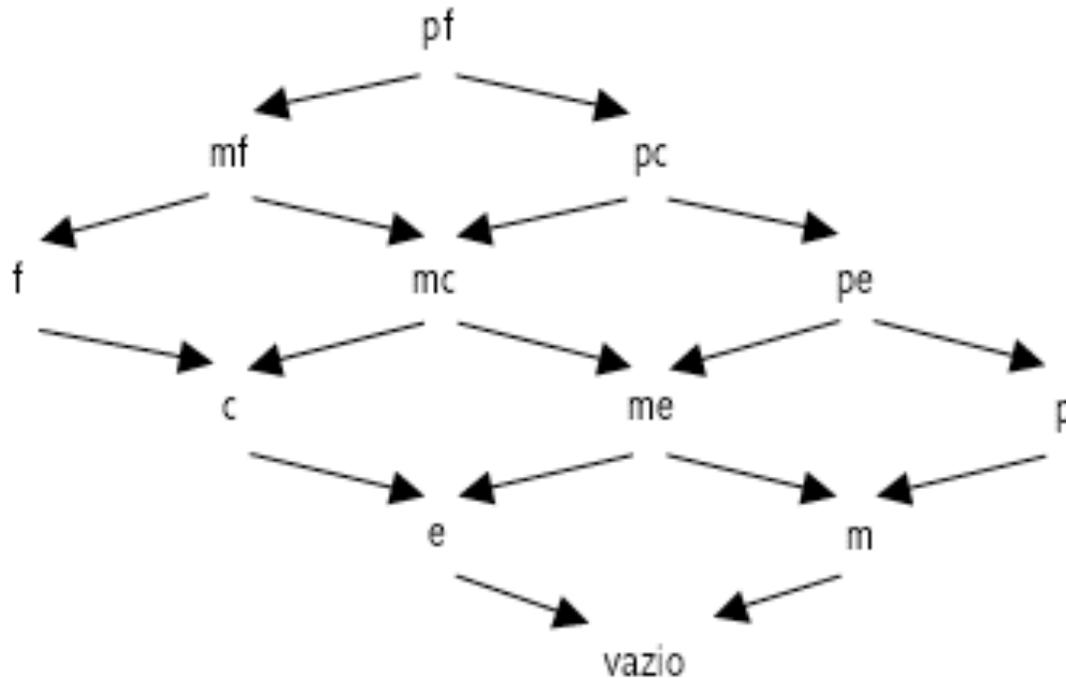
superior

dimensões produto (p), filial (f), tempo (t)

Exemplo

níveis de agregação

inferior



superior

dimensões produto (p), filial (f)

hierarquias de atributos: (i) p → marca (m);

(ii) f → cidade (c) → estado (e)

Problemas Relacionados

- Três grandes linhas de pesquisa
 1. Identificação de quais visões devem ser materializadas
 2. Manutenção incremental das visões
 3. Reformulação transparente de consultas dos usuários de SSD usando visões materializadas

Identificando Visões

- Problema
 - requisito **processamento de consultas eficiente** é conflitante tanto com o **tamanho** do DW quanto com o **tempo gasto** para manter a consistência dos dados
- Trabalhos existentes
 - *entradas*: restrição de espaço, consultas frequentes dos usuários, uso de índices, custo de manutenção
 - *saída*: quais visões devem ser materializadas

Mantendo Visões

- Problema
 - visões materializadas tornam-se **inconsistentes** sempre que as **relações base** são **alteradas**
- Passos
 - detecção e propagação de alterações relevantes nos provedores
 - atualização **incremental** das visões materializadas tanto de nível inferior quanto dos demais níveis de agregação

Reformulando Visões

- Problema
 - a existência de diversas visões correlacionadas permite que uma **mesma consulta** seja respondida usando-se **diferentes visões** materializadas
- Trabalhos existentes
 - dado uma consulta Q e um conjunto de visões materializadas, encontrar uma reescrita de Q , chamada de Q' , de forma que Q' seja **equivalente** a Q