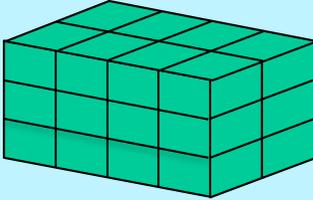
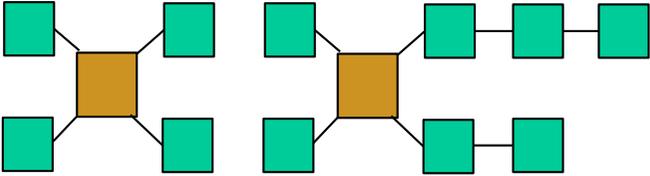
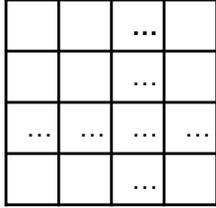
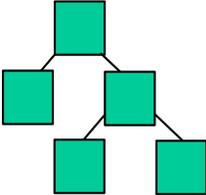
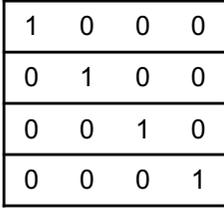


Modelagem Multidimensional

Processamento Analítico de Dados

Profa. Dra. Cristina Dutra de Aguiar Ciferri

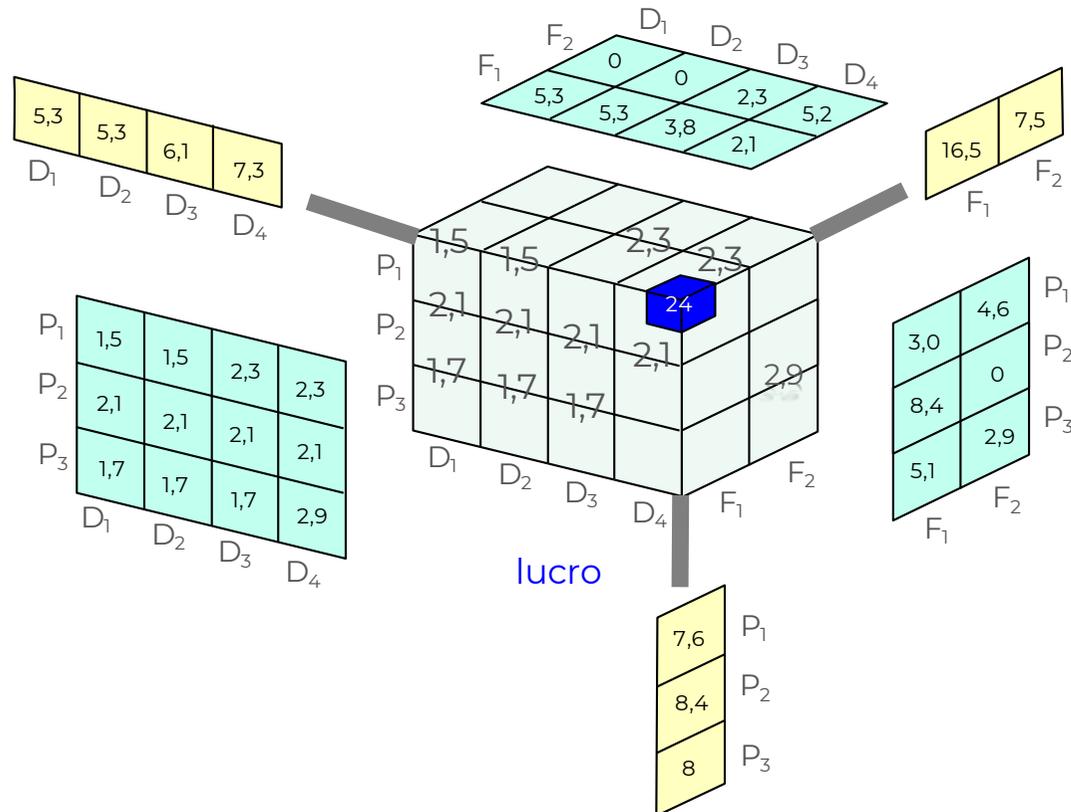
Arquitetura de 3 Camadas

	esquema	operações
conceitual	 <p>metáfora do cubo de dados</p>	<p>Cube Álgebra</p>
lógico	 <p><i>esquemas estrela e floco de neve</i></p> <p>ROLAP</p>	 <p><i>estruturas matriciais</i></p> <p>MOLAP</p> <p>...</p>
físico	 <p><i>índices: árvores</i></p>  <p><i>índices bitmap</i></p> <p>ROLAP</p>	<p><i>estruturas e algoritmos proprietários</i></p> <p>MOLAP</p>
		<p>processamento e otimização de consultas</p>

Cubo de Dados Multidimensional

Data	Produto	Filial	Lucro
D1	P1	F1	1,5
D1	P2	F1	2,1
D1	P3	F1	1,7
D2	P1	F1	1,5
D2	P2	F1	2,1
D2	P3	F1	1,7
D3	P1	F2	2,3
D3	P2	F1	2,1
D3	P3	F1	1,7
D4	P1	F2	2,3
D4	P2	F1	2,1
D4	P3	F2	2,9

□ nível inferior



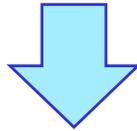
Visões multidimensionais

- nível superior
- nível intermediário 2
- nível intermediário 1
- nível inferior

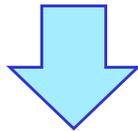
função de agregação: SOMA

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 ...

Dimensão

- Características
 - perspectiva de análise dos usuários de SSD
 - composta por atributos
- Exemplos
 - produto: chaveProduto, nomeProduto, marca, categoria, departamento
 - filial: chaveFilial, nomeFilial, cidade, estado, região, país
 - data: chaveData, dia, mês, trimestre, semestre, ano

Medida Numérica Aditiva

- Pode ser **somada** considerando **todas as dimensões**
- Exemplos
 - unidades vendidas
 - lucro
- São agregadas usando
 - função de agregação **SOMA**

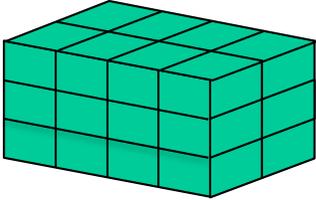
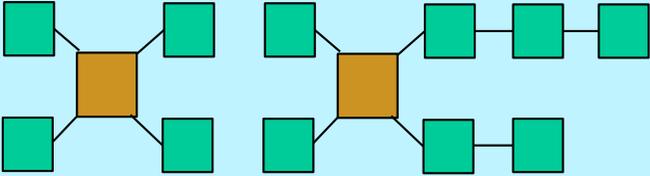
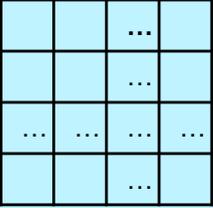
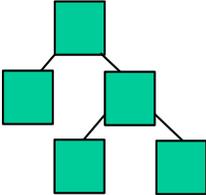
Medida Numérica Não Aditiva

- Não pode ser **somada**
- Exemplos
 - **preço**, com a semântica de preço do produto
- São agregadas usando
 - Funções de agregação **AGV**, **MAX**, **MIN**, **COUNT**
 - Outra função complexa

Medida Numérica Semiaditiva

- Podem ser **somadas** considerando somente **algumas dimensões**
- Exemplo
 - **Número de clientes**, com a semântica de vendas de produtos
 - **Não aditiva**: para dois produtos P1 e P2 vendidos pela mesma filial no mesmo dia, não é possível somar o número de clientes do produto P1 com o número de clientes do produto P2, desde que o mesmo cliente pode estar sendo contabilizado duas vezes
 - **Aditiva**: número de clientes do produto P1 por dia pode ser agregado para se obter o número de clientes do produto P1 por mês

Arquitetura de 3 Camadas

	esquema	operações																
conceitual	 <p>metáfora do cubo de dados</p>	<p>Cube Álgebra</p>																
lógico	<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;">  <p><i>esquemas estrela e floco de neve</i></p> <p>ROLAP</p> </div> <div style="text-align: center;">  <p><i>estruturas matriciais</i></p> <p>MOLAP</p> </div> </div>	<p>SQL MDX ...</p>																
físico	<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;">  <p><i>índices: árvores</i></p> <p>ROLAP</p> </div> <div style="text-align: center;"> <table border="1" data-bbox="853 1150 1077 1358"> <tr><td>1</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>1</td></tr> </table> <p><i>índices bitmap</i></p> <p>MOLAP</p> </div> </div>	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	<p><i>estruturas e algoritmos proprietários</i></p> <p>processamento e otimização de consultas</p>
1	0	0	0															
0	1	0	0															
0	0	1	0															
0	0	0	1															

Sistemas ROLAP (OLAP Relacional)

- Baseado no uso do **modelo relacionais**
- *Data Warehouse*
 - representado como uma **coleção de esquemas de relação**
- Sistema gerenciador de banco de dados (SGBD)
 - estendido para oferecer suporte às operações analíticas

Máquina Relacional

- SGBD adaptado para processamento analítico
- Armazenamento dos dados
 - organizado especialmente para refletir a visão multidimensional dos dados
- Funcionalidades adicionais
 - novos operadores SQL
 - estruturas de indexação especializadas
 - otimização de consultas complexas
 - execução paralela de consultas complexas

Máquina ROLAP

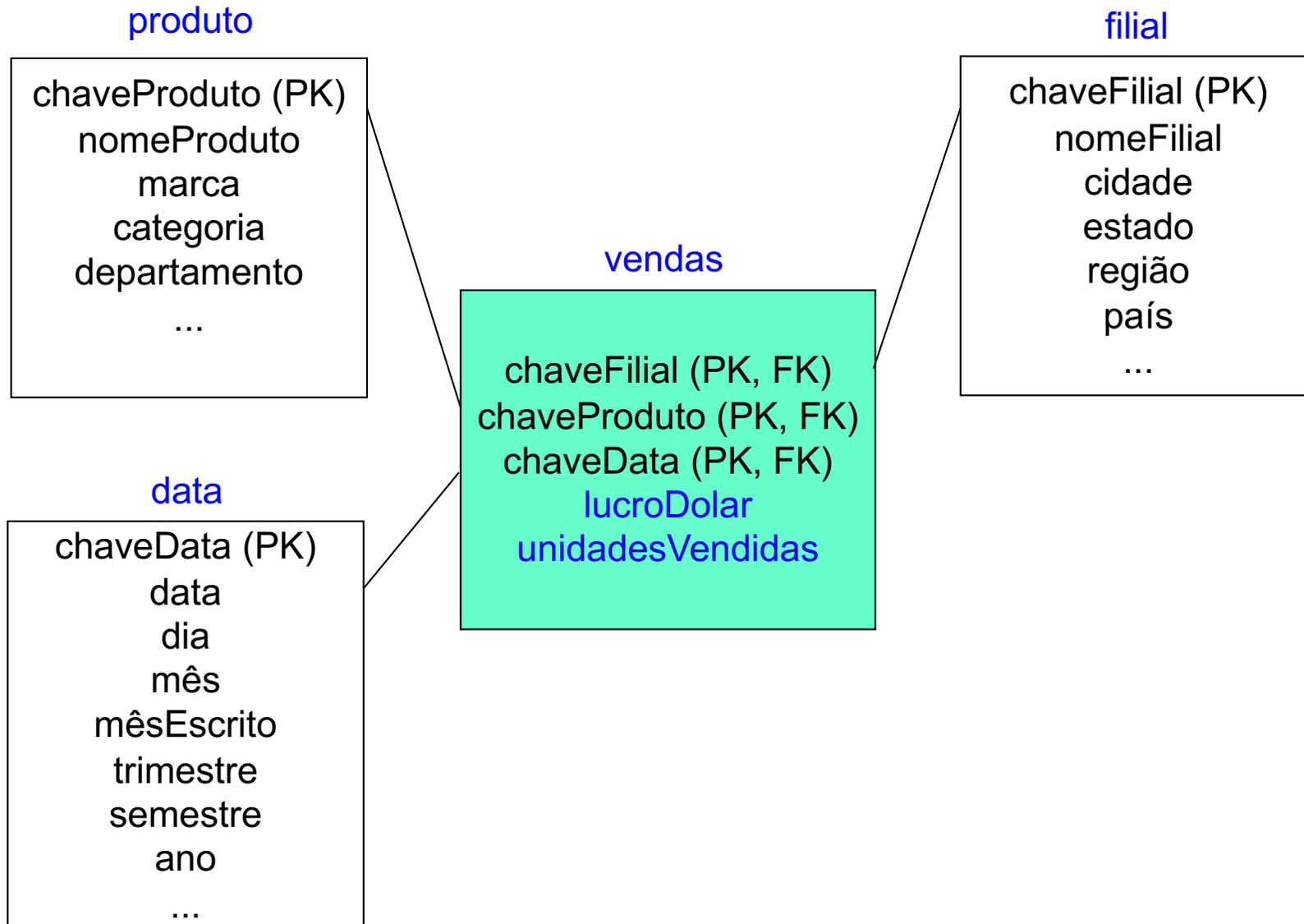
- Camada de *software* que complementa a máquina relacional
- Objetivos
 - suprir as limitações da máquina relacional
 - exibir os dados multidimensionalmente
- Funcionalidade adicional
 - materialização de visões

Tipos de Esquema

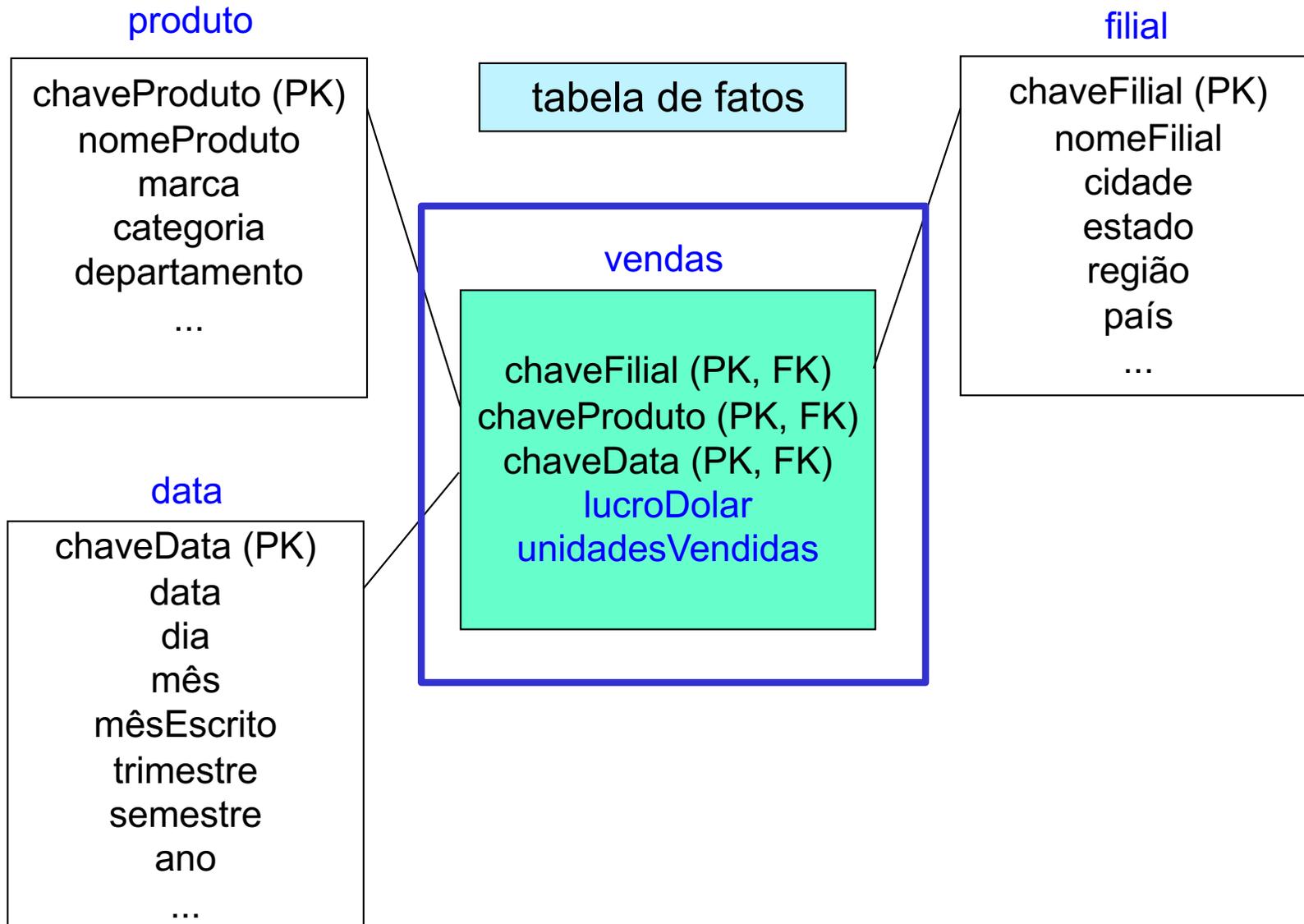
- Organizam o armazenamento dos dados do DW
 - multidimensionalidade dos dados
 - orientados a assunto
 - históricos
- Esquemas
 - **estrela** (star)
 - **floco de neve** (snowflake)
 - **estrela-floco** (starflake)

Tipos de Tabela
Tabela de Fatos
Tabela de Dimensão

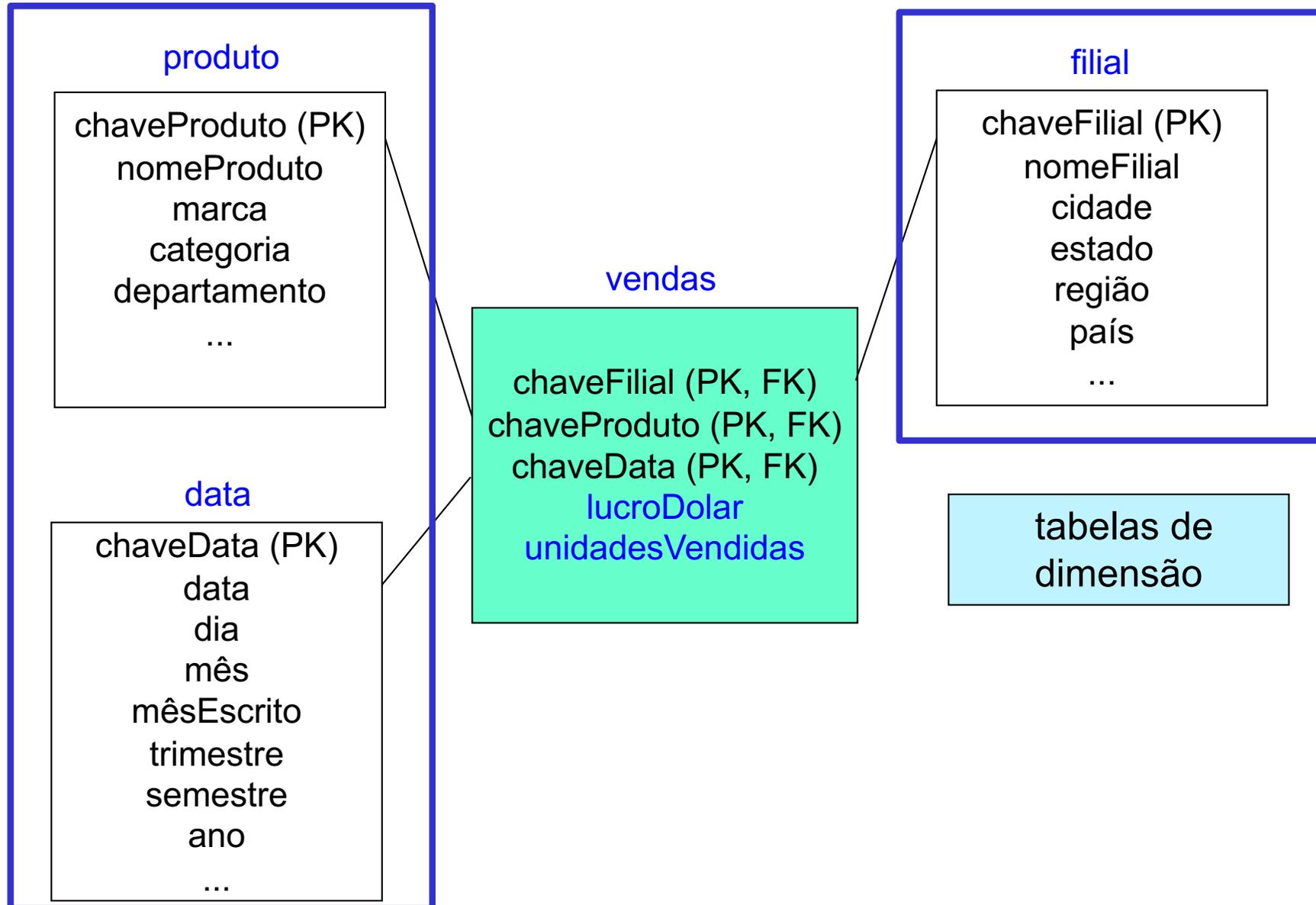
Esquema Estrela



Esquema Estrela



Esquema Estrela



Esquema Estrela

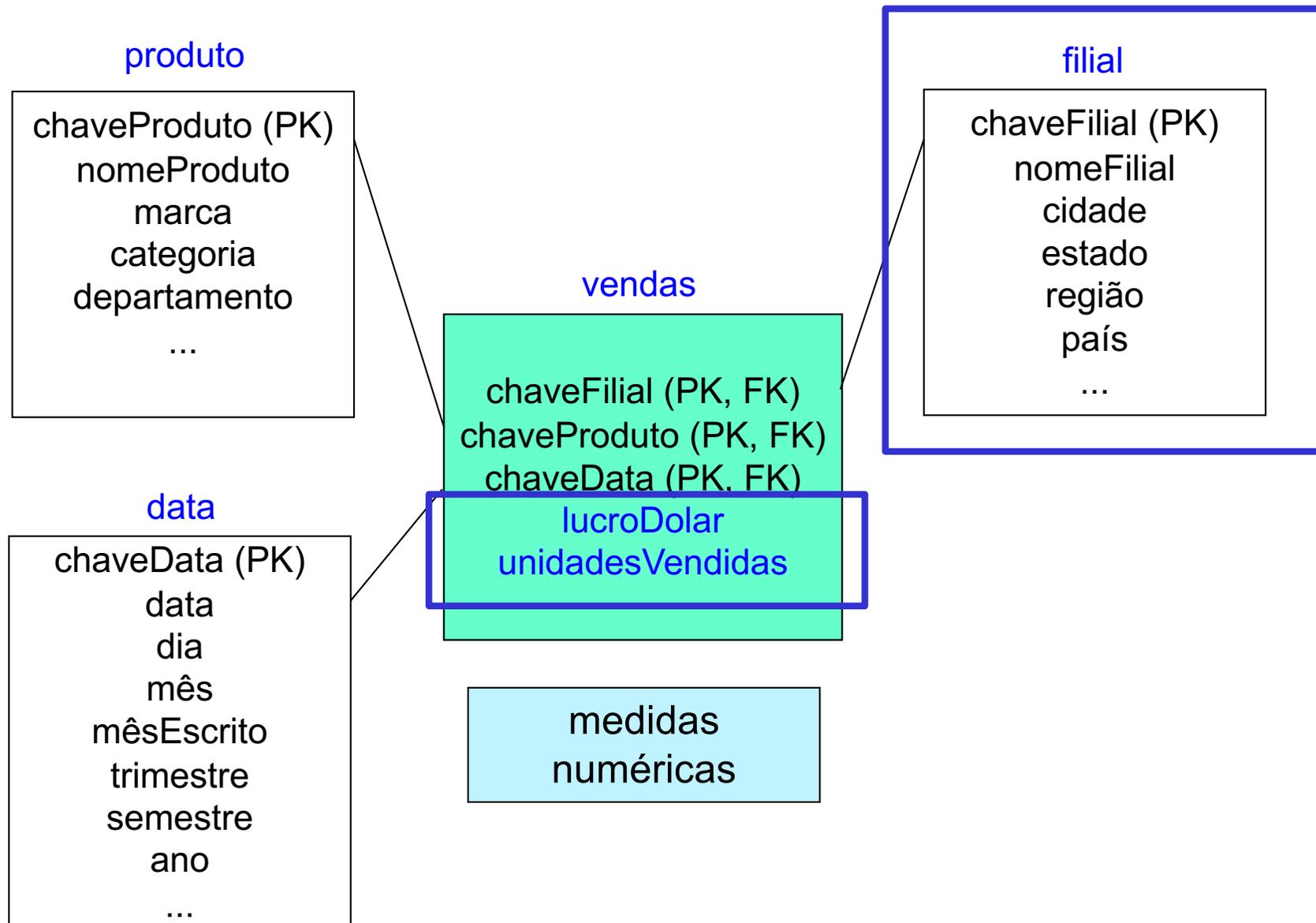


Tabela de Dimensão

- Localizada visualmente na **extremidade** da estrela
- Armazena
 - uma chave primária (chave artificial)
 - atributos da dimensão
- Características
 - usualmente larga e curta
 - com redundância
 - dados desmoralizados

Tabela de Dimensão **Filial**

chaveFilial	nomeFilial	cidade	estado	região	país	...
F1	Vende Barato	São Carlos	São Paulo	Sudeste	Brasil	...
F2	Para Poucos	São Carlos	São Paulo	Sudeste	Brasil	...
F3	Atende Você	Araraquara	São Paulo	Sudeste	Brasil	...
F4	Mega Vendão	Campinas	São Paulo	Sudeste	Brasil	...
F5	Compra Fácil	Limeira	São Paulo	Sudeste	Brasil	...
F6	Atende Você II	Araraquara	São Paulo	Sudeste	Brasil	...
F7	Tudo Aqui	Belo Horizonte	Minas Gerais	Sudeste	Brasil	...
...

- Representa o aspecto geográfico (WHERE)

Tabela de Dimensão **Data**

chaveData	data	dia	mês	mêsEscrito	trimestre	semestre	ano	...
D1	01/01/2020	1	1	janeiro	1	1	2020	...
D2	02/01/2020	2	1	janeiro	1	1	2020	...
D3	03/01/2020	3	1	janeiro	1	1	2020	...
D4	04/01/2020	4	1	janeiro	1	1	2020	...
D5	05/01/2020	5	1	janeiro	1	1	2020	...
D6	06/01/2020	6	1	janeiro	1	1	2020	...
D7	07/01/2020	7	1	janeiro	1	1	2020	...
...

- Representa o aspecto temporal (WHEN)
- Pode ser povoada antes das demais tabelas

Tabela de Dimensão **Produto**

chaveProduto	nomeProduto	marca	categoria	departamento	...
P1	carne moída	Carnes e Cia	alimento	açougue	...
P2	frango	Frangote	alimento	açougue	...
P3	bolacha salgada	Bolachinhas	alimento	biscoitos	...
P4	ovo	Ovo Dourado	alimento	hortifruti	...
P5	bucha	Esfregando	limpeza	higiene do lar	...
P6	sabão em pó	Bolhas e Bolhas	limpeza	higiene do lar	...
...

- Representa o aspecto sendo investigado (WHAT)
- Praticamente povoada com dados do ambiente operacional

Tabela de Fatos

- Localizada visualmente no **centro** da estrela
- Armazena
 - as medidas numéricas relevantes ao negócio (**fatos**)
 - uma chave estrangeira para cada tabela de dimensão
 - **chave primária: combinação das chaves estrangeiras**
- Características
 - usualmente fina e longa
 - sem redundância
 - sem dados esparsos

Tabela de Fatos **Vendas**

data	produto	filial	lucro	unidadesVendidas
D1	P1	F1	1,5	5
D1	P2	F1	2,1	6
D1	P3	F1	1,7	7
D2	P1	F1	1,5	5
D2	P2	F1	2,1	5
D2	P3	F1	1,7	7
D3	P1	F2	2,3	4
D3	P2	F1	2,1	5
D3	P3	F1	1,7	7
D4	P1	F2	2,3	4
D4	P2	F1	2,1	5
D4	P3	F2	2,9	10
...

Esquema Relacional

filial (chaveFilial, nomeFilial, cidade, estado, região, país, ...)

produto (chaveProduto, nomeProduto, marca, categoria, departamento, ...)

data (chaveData, dia, mês, mêsEscrito, trimestre, semestre, ano, ...)

vendas (chaveData, chaveProduto, chaveFilial,
lucroDolar, unidadesVendidas)

Dimensões Clássicas

- **When**
 - ex.: tempo, data, hora
- **What**
 - ex.: produto, funcionário, cliente
- **Where**
 - ex.: filial, departamento, setor
- **Why**
 - ex.: promoção, cargo

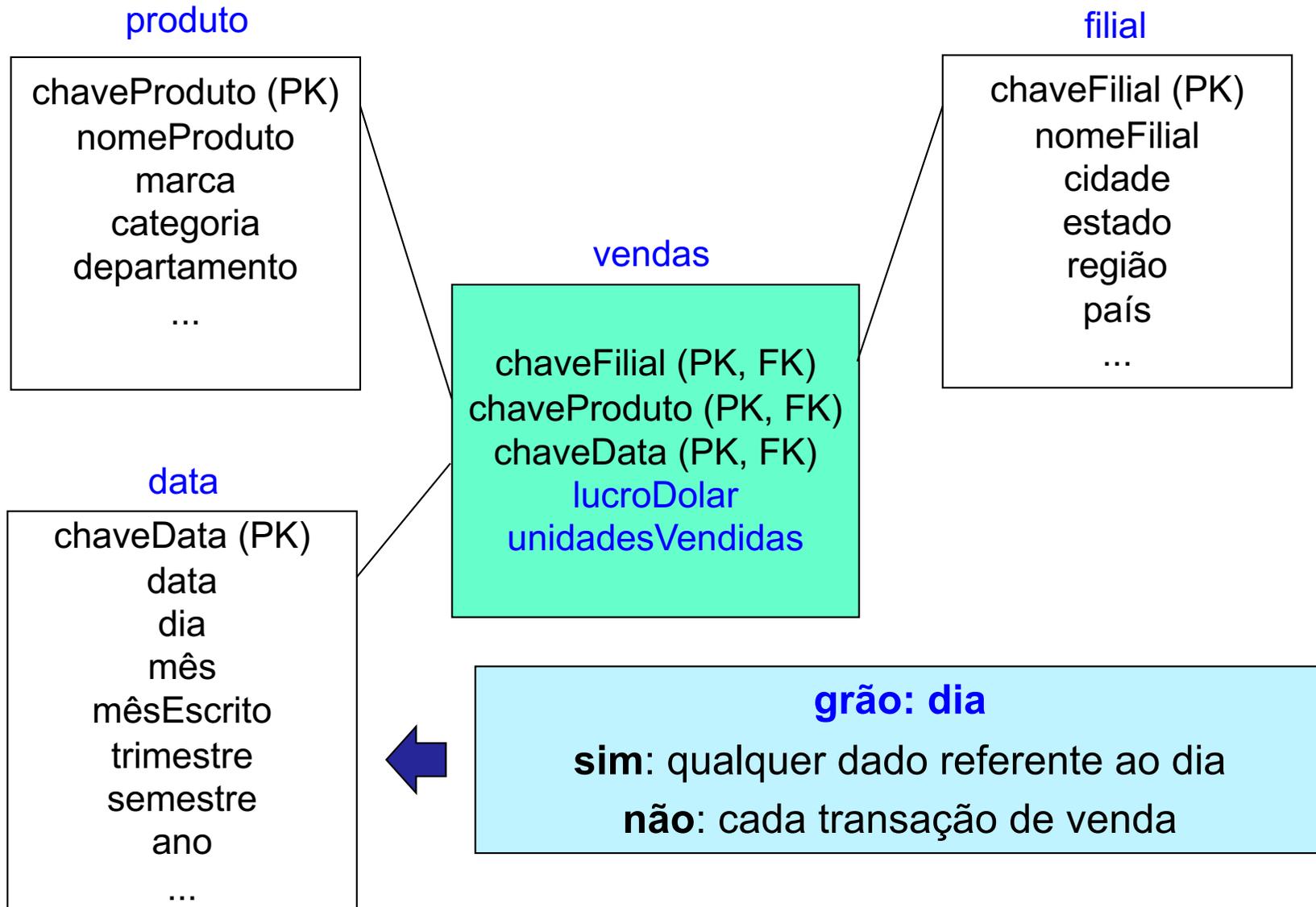
Granularidade

- Grau de detalhamento em que os dados são armazenados
 - Dados **mais detalhados** (ou **menos agregados**)
 - Dados **menos detalhados** (ou **mais agregados**)
- Aspecto de projeto muito importante
 - **Volume** de dados do *data warehouse*
 - **Consultas** que podem ser respondidas

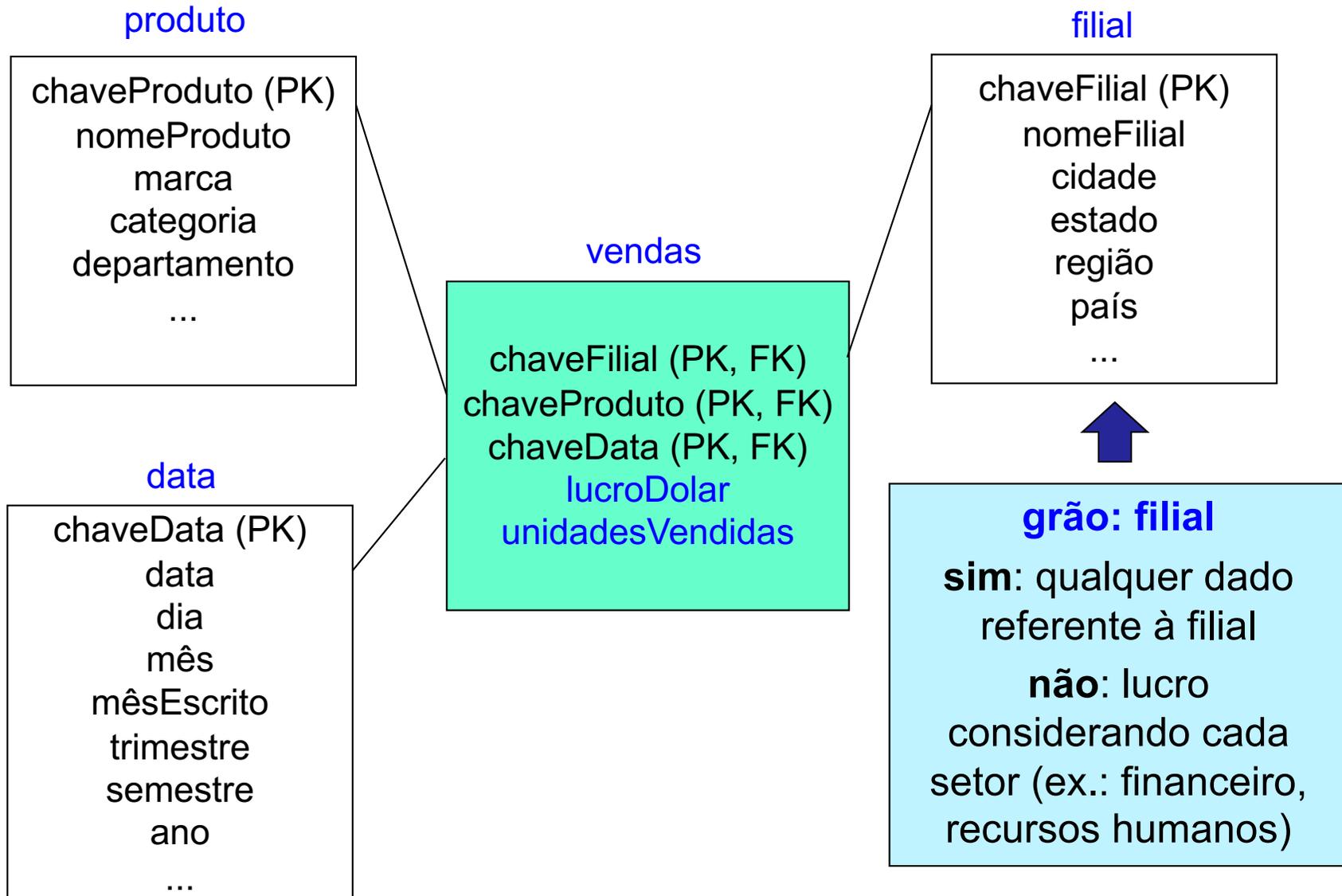
Tamanho do Grão

- Grão muito pequeno
 - Dados mais detalhados
 - *Data warehouse* é muito mais volumoso
 - Quantidade de consultas que podem ser respondidas é maior
- Grão muito grande
 - Dados menos detalhados
 - *Data warehouse* é menos volumoso
 - Quantidade de consultas que podem ser respondidas é menor

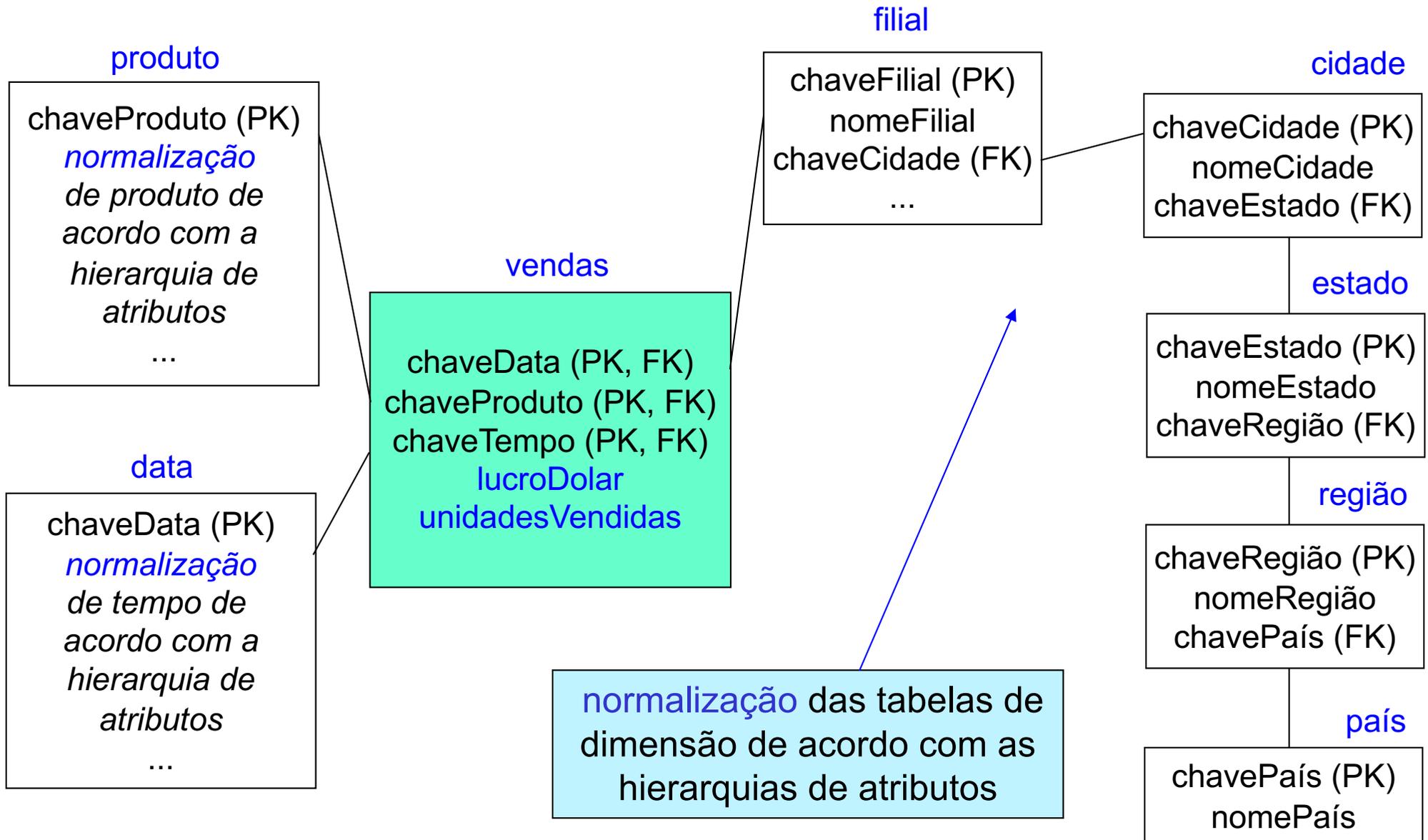
Grão da Tabela Data



Grão da Tabela Filial



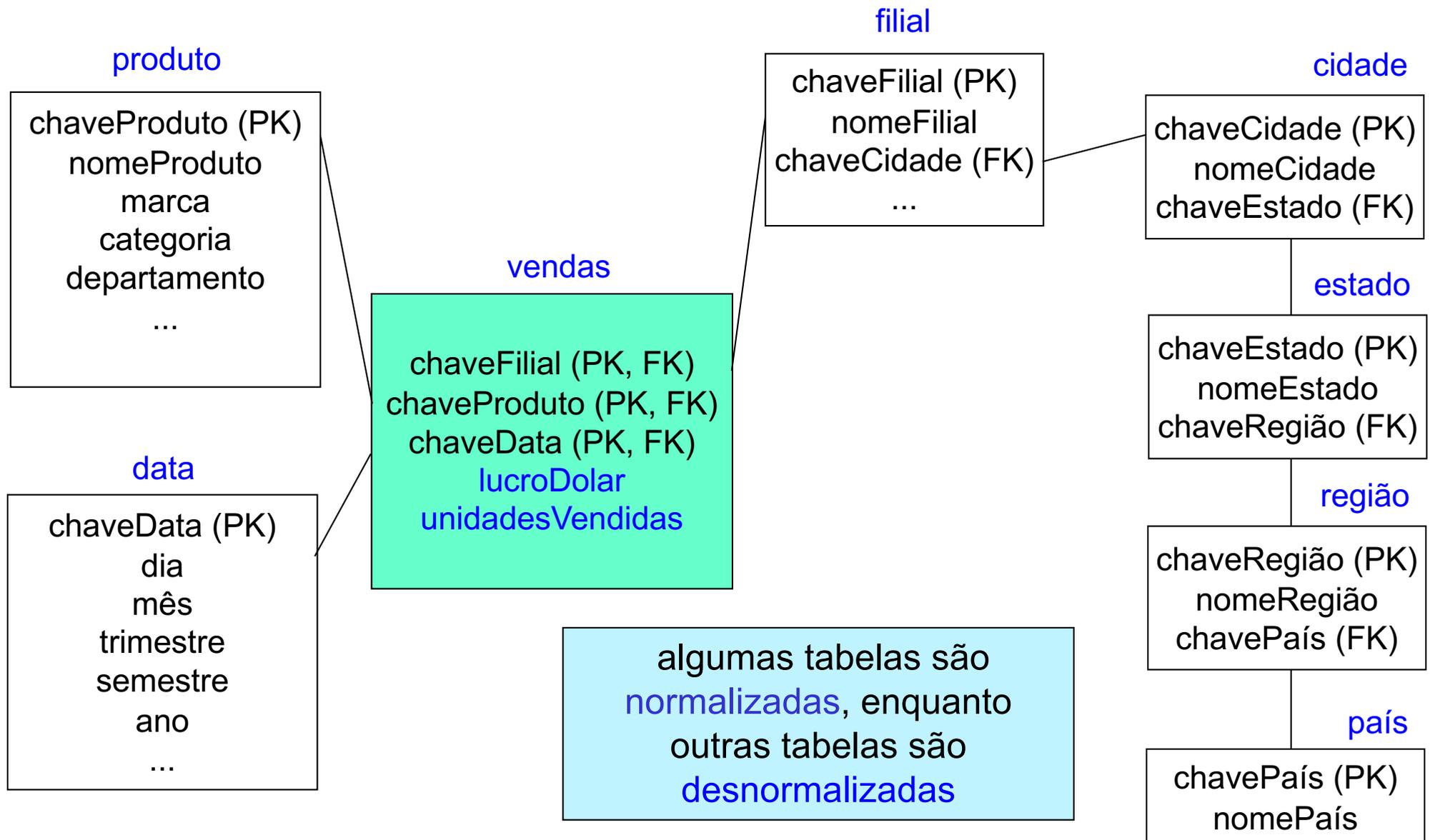
Esquema Floco de Neve



Estrela *versus* Floco de Neve

Característica	Estrela	Floco de Neve
redundância dos dados nas tabelas de dimensão	redundantes dados desnormalizados	sem redundância dados normalizados
processamento de junção-estrela	mais simples junção da(s) tabela(s) de dimensão necessária(s) com a tabela de fatos	mais complexa junção da(s) tabela(s) de dimensão necessária(s) com a tabela de fatos + junção da(s) tabela(s) de dimensão normalizada(s)

Esquema Estrela-Floco (Starflake)



Esquema Estrela-Floco

país (chavePaís, nomePaís)

região (chaveRegião, nomeRegião, chavePaís)

estado (chaveEstado, nomeEstado, chaveRegião)

cidade (chaveCidade, nomeCidade, chaveEstado)

filial (chaveFilial, nomeFilial, chaveCidade, ...)

produto (chaveProduto, nomeProduto, marca, categoria, departamento, ...)

data (chaveData, dia, mês, mêsEscrito, trimestre, semestre, ano, ...)

vendas (chaveData, chaveProduto, chaveFilial,
lucroDolar, unidadesVendidas)

