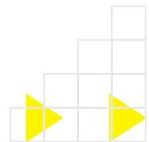




Curso 3: Administração de Dados Complexos em Larga Escala -- Online Analytical Processing --

Prof. Jose Fernando Rodrigues Junior

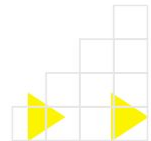
Objetivo: detalhes sobre o conceito de OLAP





OLAP

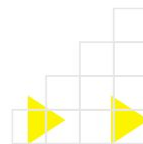
- **Definição:** uma categoria de software que visa à compreensão de dados por meio de agregação: **SUM, MAX, MIN, COUNT, AVG, VAR, e outras;**
- **Provê:** acesso rápido, consistente e interativo;
- **DW e OLAP,** em muitos casos, conceitos indissociáveis.



OLAP - FASMI

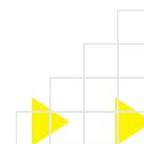


- **FASMI**: Fast Analysis of Multidimensional Information
 - **F**ast: agilidade em responder consultas
 - **A**nalysis: versatilidade analítica
 - **S**hared: dados/analistas múltiplos
 - **M**ultidimensional: orientado a dimensões de dados
 - **I**nformation: propósito fim



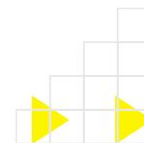
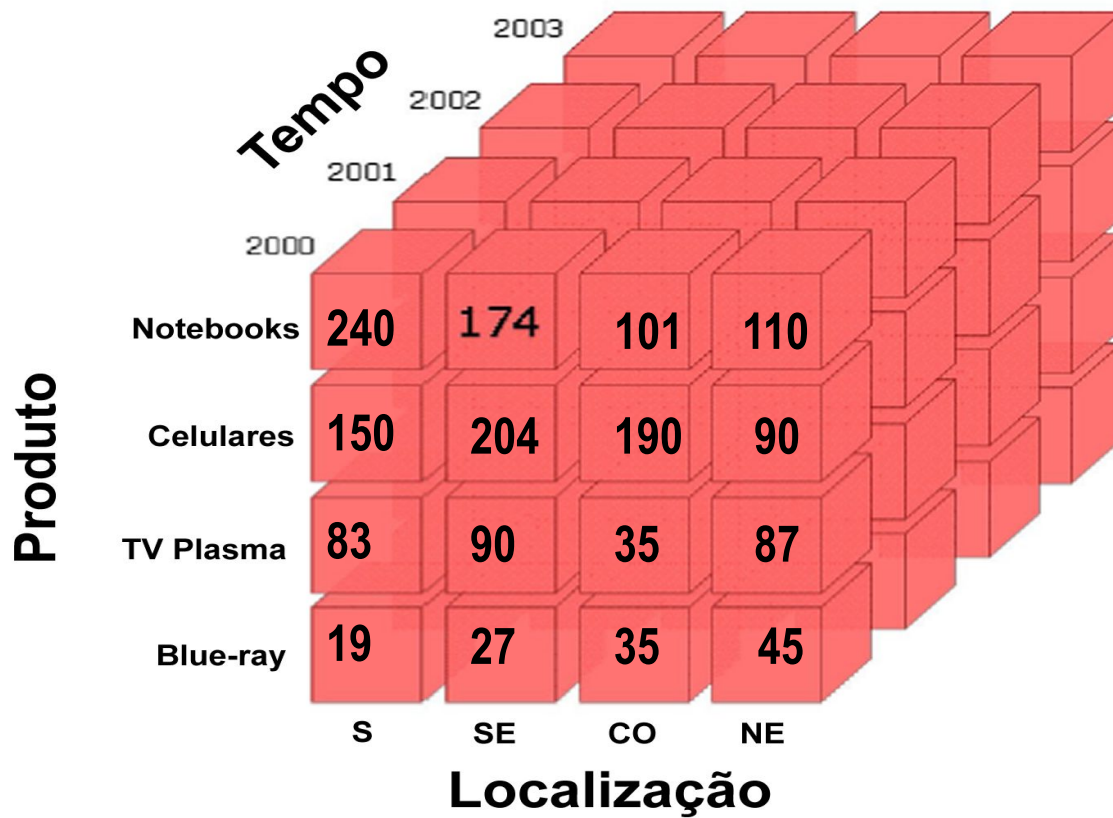


Cubo de dados OLAP





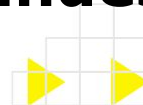
Cubo de dados





Cubo de dados

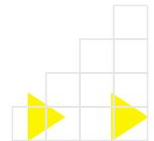
- Estrutura básica da prática de OLAP
 - Observam-se:
 - As **dimensões** dos dados
 - As **medidas** sobre os dados
 - O cubo é orientado a planos (faces)
 - Apesar da complexidade dos sistemas OLAP, seus **objetivos analíticos são básicos**: contagem, média, máximo, mínimo, soma, ...
- ☐ **Agregação, rápida e flexível, sobre grandes volumes de dados**





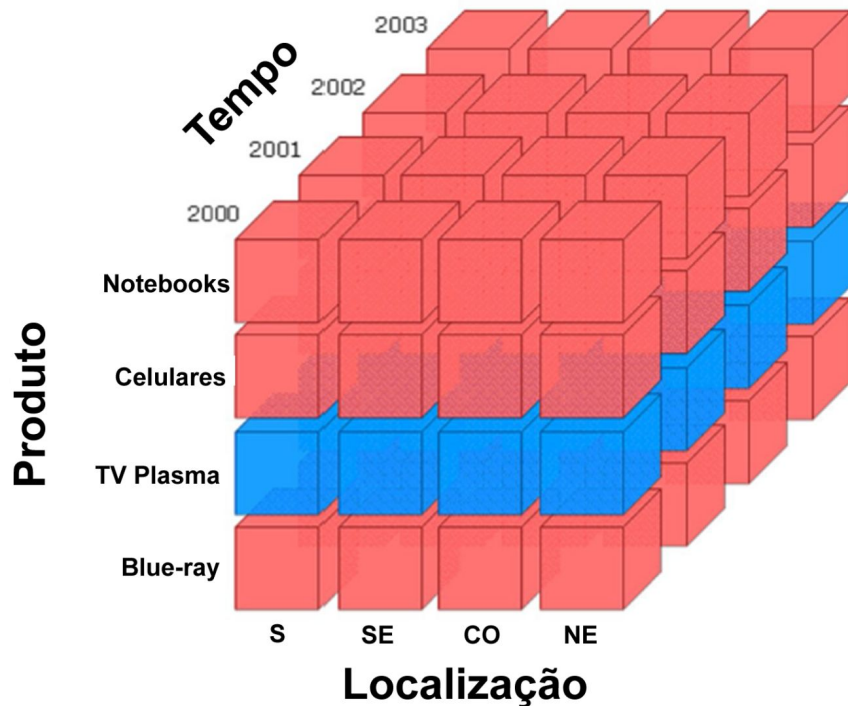
Cubo de dados - slicing

- Pode ser interessante ver o cubo a partir de diferentes perspectivas (planos)
- Operações sobre o cubo: **slicing, dicing e rotating (pivoting)**

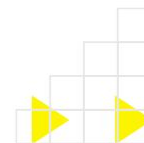




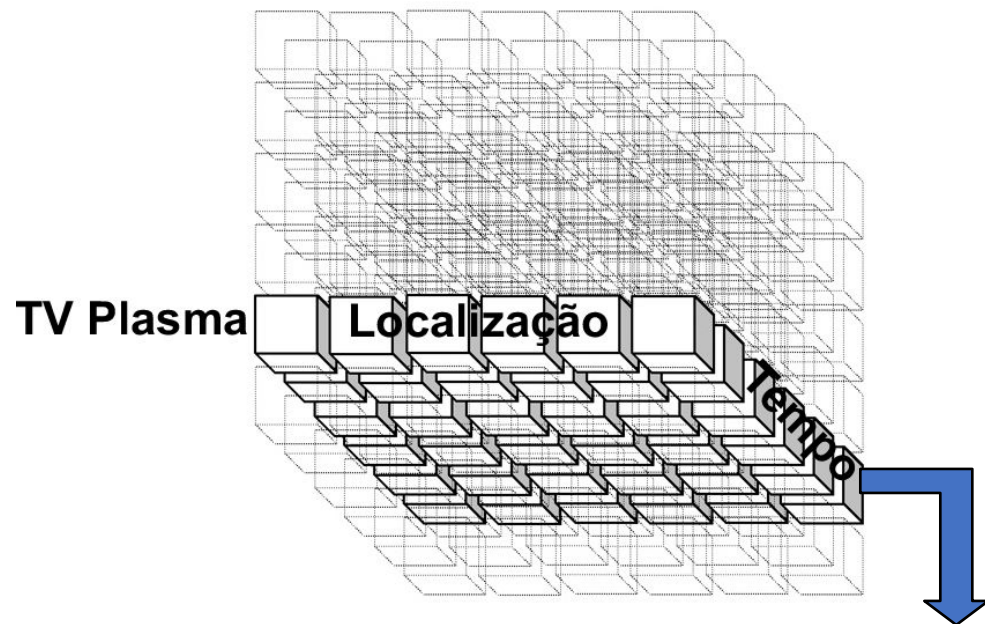
Cubo de dados - slicing



- A operação de slicing equivale a fatiar o cubo, definindo um novo plano de apreciação dos dados
- A operação geométrica é apenas uma analogia, o slicing dispara o processamento OLAP para calcular o novo plano

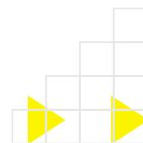


Cubo de dados - slicing

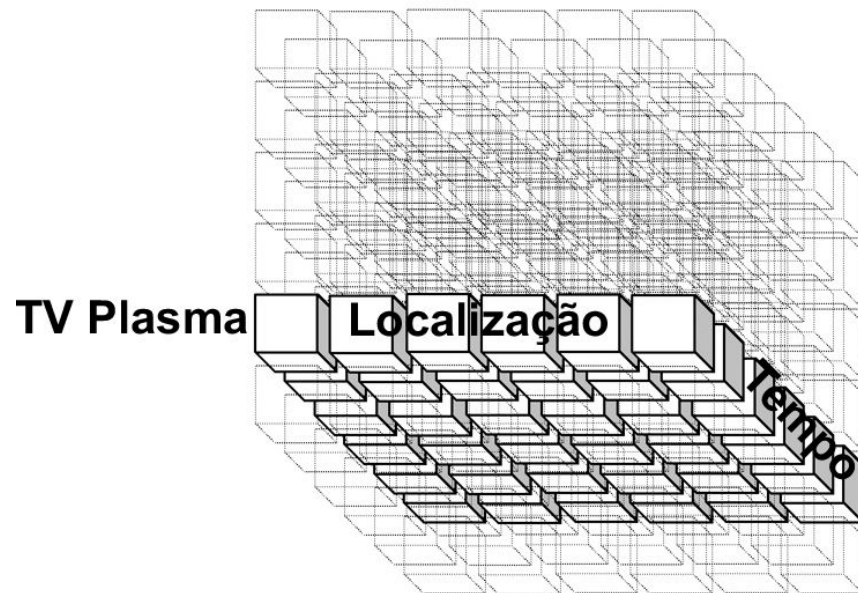


Dados calculados e acessíveis ao analista.

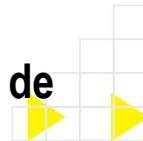
- A operação de slicing equivale a fatiar o cubo, definindo um novo plano de apreciação dos dados
- A operação geométrica é apenas uma analogia, o slicing dispara o processamento OLAP para calcular o novo plano



Cubo de dados - slicing

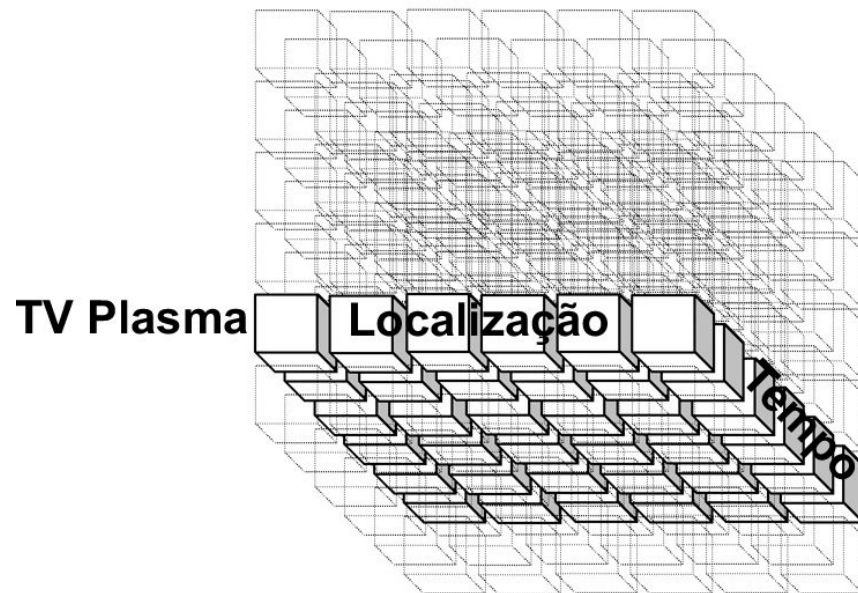


Exemplo de dado calculado: quantas TVs de plasma foram vendidas por região e por ano?





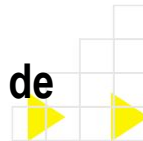
Cubo de dados - slicing



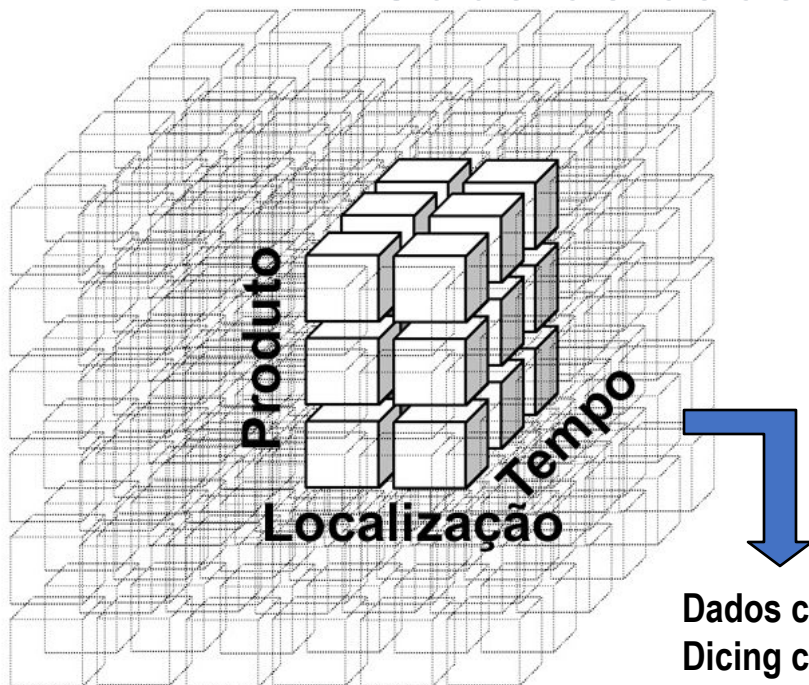
Via interação com o cubo, dispara-se o seguinte processamento:

```
SELECT regioao, ano, SUM(Nro_Unidades)
FROM VENDAS
WHERE produto = "TV Plasma"
GROUP BY regioao, ano
```

Exemplo de dado calculado: quantas TVs de plasma foram vendidas por região e por ano?

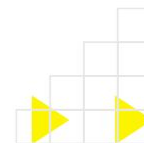


Cubo de dados - slicing

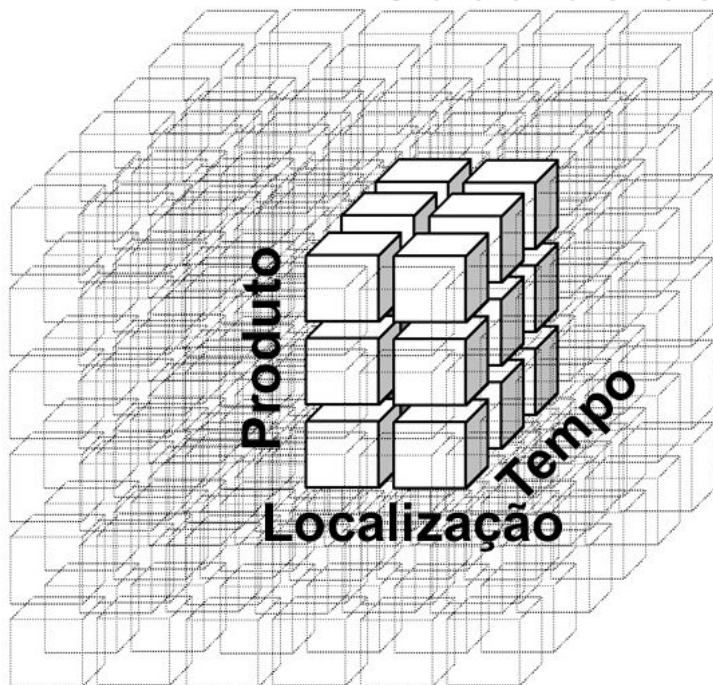


- A operação de dicing é semelhante ao slicing, mas usa dois, ou mais, planos de corte

Dados calculados e acessíveis ao analista.
Dicing conseguido com 5 planos de corte.



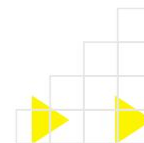
Cubo de dados - slicing



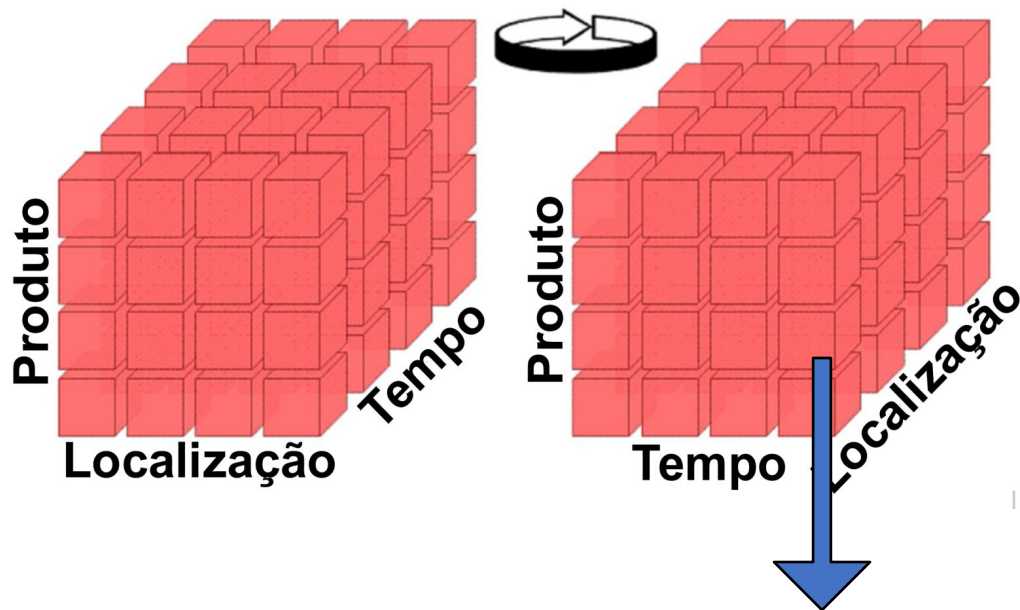
Via interação com o cubo, dispara-se o seguinte processamento:

```
SELECT regioao, ano, SUM(Nro_Unidades)
FROM VENDAS
WHERE (produto = "TV Plasma" OR
       produto = "Notebooks" OR
       produto = "Celulares")
AND
      (ano BETWEEN 2001 AND 2003)
AND
      (regiao = "SE" OR regiao = "CO")
GROUP BY regioao, ano
```

Exemplo de dado calculado: quantos Notebooks, Celulares, e TVs foram vendidos entre 2001 e 2003 nas regiões SE e CO?

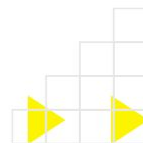


Cubo de dados - rotating



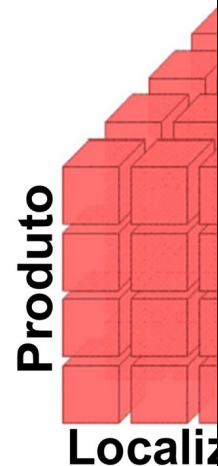
Novo plano: Produto x Tempo
Dados calculados e acessíveis ao analista.

- A operação de rotating muda a perspectiva do analista, permitindo diferentes consultas.





Cubo de dados - rotating

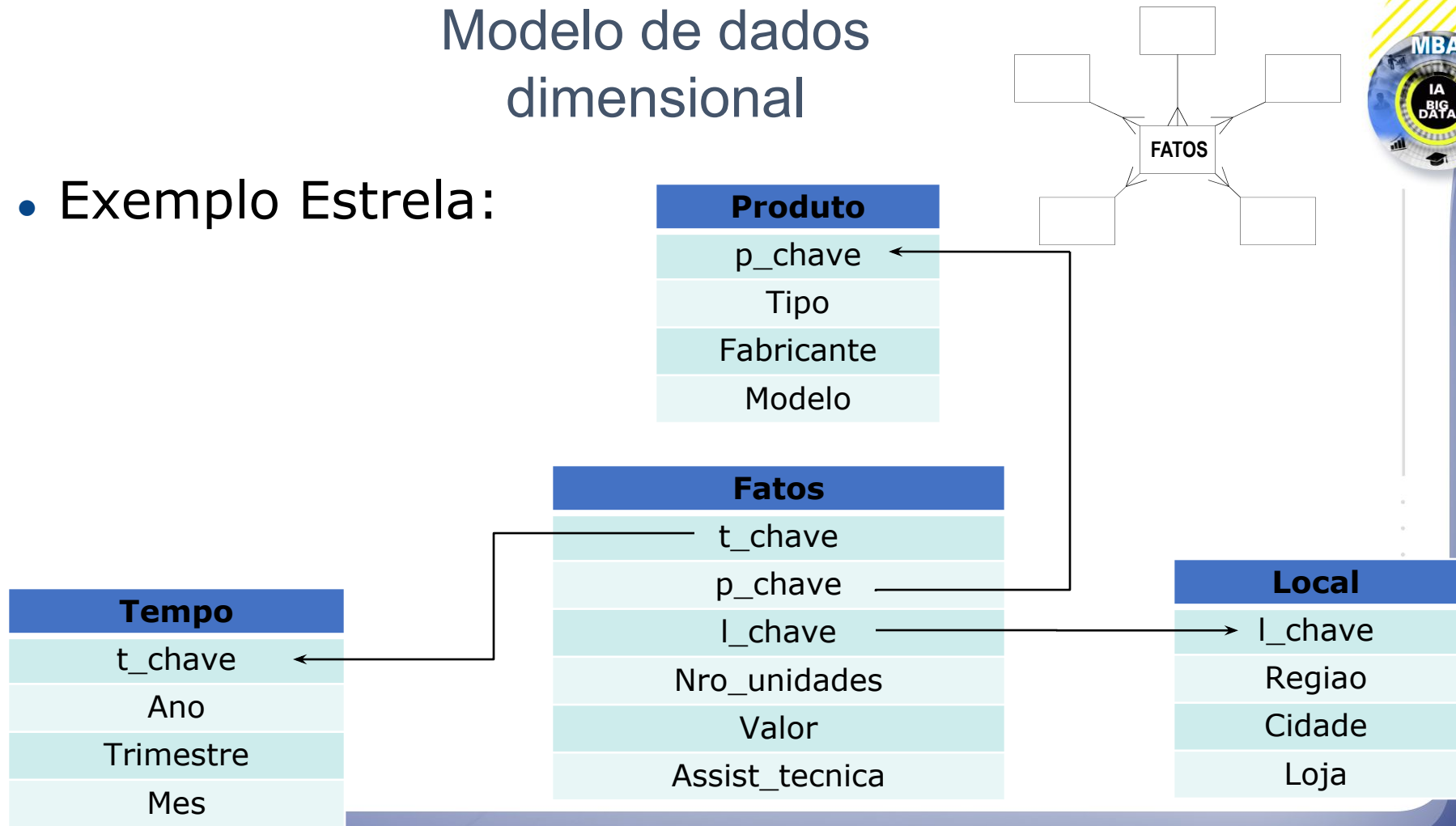


O termo Cubo de Dados é usado apenas para fins didáticos – aplicações de DW/OLAP geralmente envolvem mais do que 3 dimensões, definindo **hipercubos**.

Modelo de dados dimensional



- Exemplo Estrela:



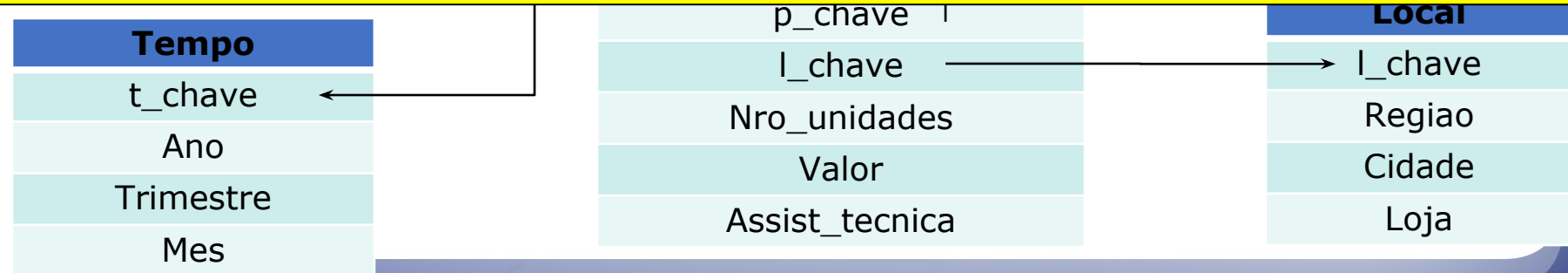
Modelo de dados dimensional



Exemplo:

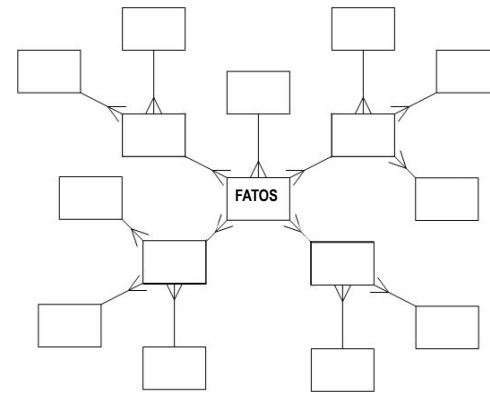
Quantas vendas e qual valor de vendas ocorreram considerando Ano, Tipo de Produto e Região?

```
SELECT Tempo.Ano, Produto.Tipo, Local.Regiao, Sum(Nro_unidades), Sum(valor)
FROM Fatos, Tempo, Produto, Local
WHERE Fatos.t_chave = Tempo.t_chave AND
      Fatos.p_chave = Produto.p_chave AND
      Fatos.l_chave = Local.l_chave
GROUP BY Tempo.Ano, Produto.Tipo, Local.Regiao
```

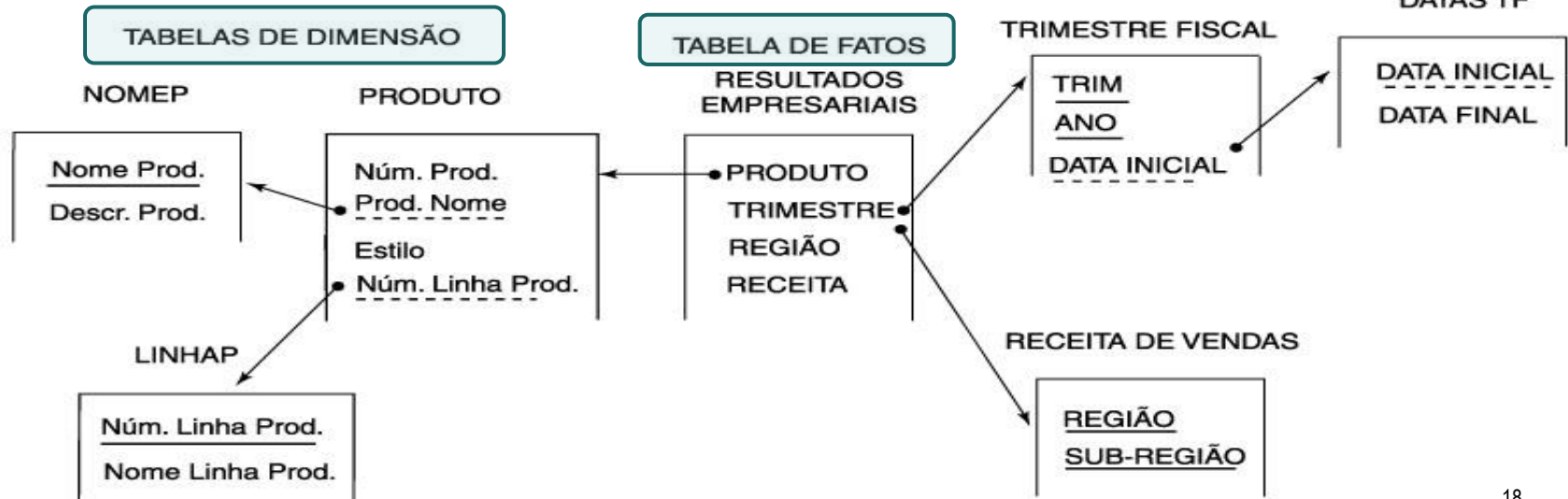


Modelo de dados dimensional

○ Exemplo *Snowflake*



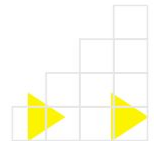
TABELAS DE DIMENSÃO



Modelo de dados dimensional



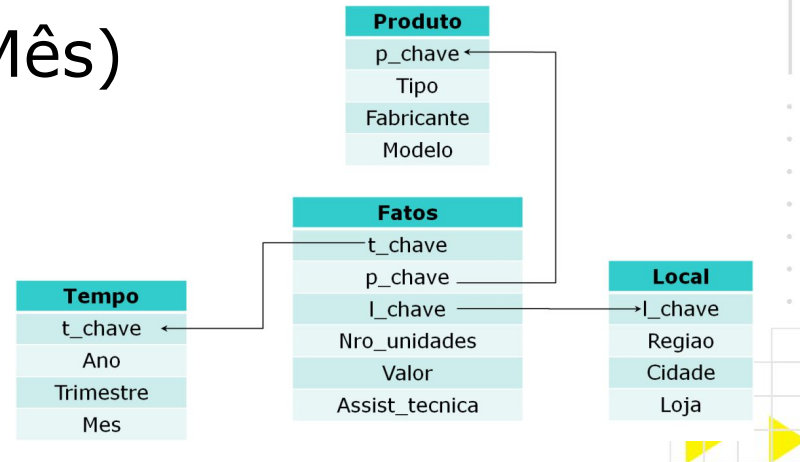
- Cubo de dados: análise dimensional das medidas (dados)
- **DW:** modelo de dados dimensional
 - **Dimensões:** contexto aos fatos
 - **Fatos:** números transacionais



Modelo de dados dimensional



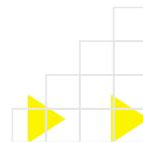
- Observe que as dimensões dos dados possuem uma **hierarquia granular**
- Por exemplo:
 - Tempo(Ano, Trimestre, Mês)





Drill down / Roll up

- Hierarquia das dimensões: apreciação dos dados **em diferentes granularidades**. Exemplo:
 - Itens_vendidos(ano) > Itens_vendidos(Trimestre)
 - Itens_vendidos(Trimestre) > Itens_vendidos(Mês)
- Duas outras operações muito importantes
 - Drill down
 - Roll up

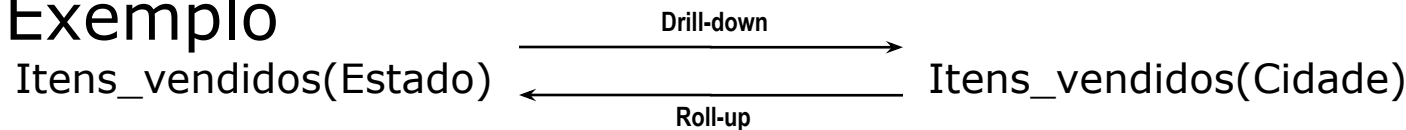




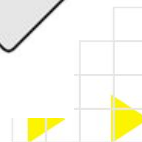
Drill down / Roll up

- **Drill-down e roll-up:** navegação ao longo dos níveis hierárquicos das dimensões

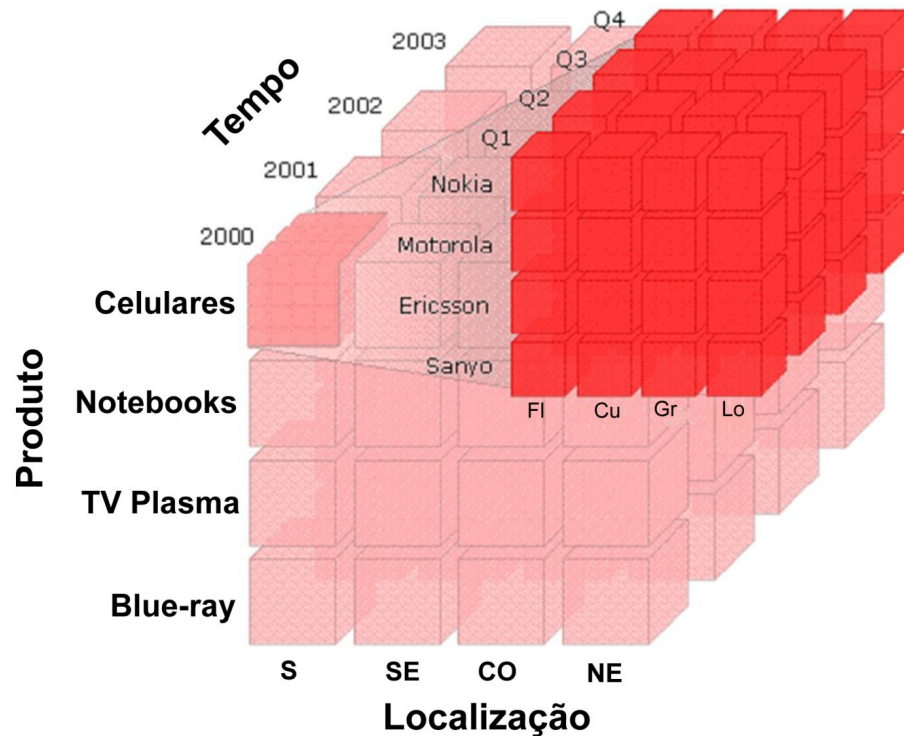
- **Exemplo**



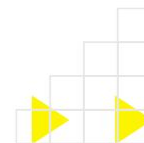
- Drill-down \Rightarrow mais detalhamento
- Roll-up \Rightarrow menos detalhamento



Drill down / Roll up



Drill-down sobre as três dimensões simultaneamente.





Drill down / Roll up

Exemplo:

Quantas vendas e qual valor de vendas ocorreram considerando Ano, Tipo de Produto e Região?

Drill Down em todas as dimensões

Quantas vendas e qual valor de vendas ocorreram considerando Trimestre, Fabricante e Cidade?

```
SELECT Tempo.Ano, Tempo.Trimestre, Produto.Tipo, Produto.Fabricante, Local.Regiao, Local.Cidade, Sum(Nro_unidades),  
Sum(valor)
```

```
FROM Fatos, Tempo, Produto, Local
```

```
WHERE Fatos.t_chave = Tempo.t_chave AND
```

```
      Fatos.p_chave = Produto.p_chave AND
```

```
      Fatos.l_chave = Local.l_chave
```

```
GROUP BY Tempo.Ano, Tempo.Trimestre, Produto.Tipo, Produto.Fabricante, Local.Regiao, Local.Cidade
```

Blue-ray

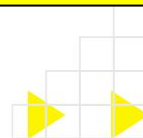
S

SE

CO

NE

Localização





Drill down / Roll up

Exemplo:

Quantas vendas e qual valor de vendas ocorreram considerando Ano, Tipo de Produto e Região?

Drill Down em todas as dimensões + slicing

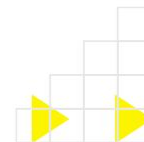
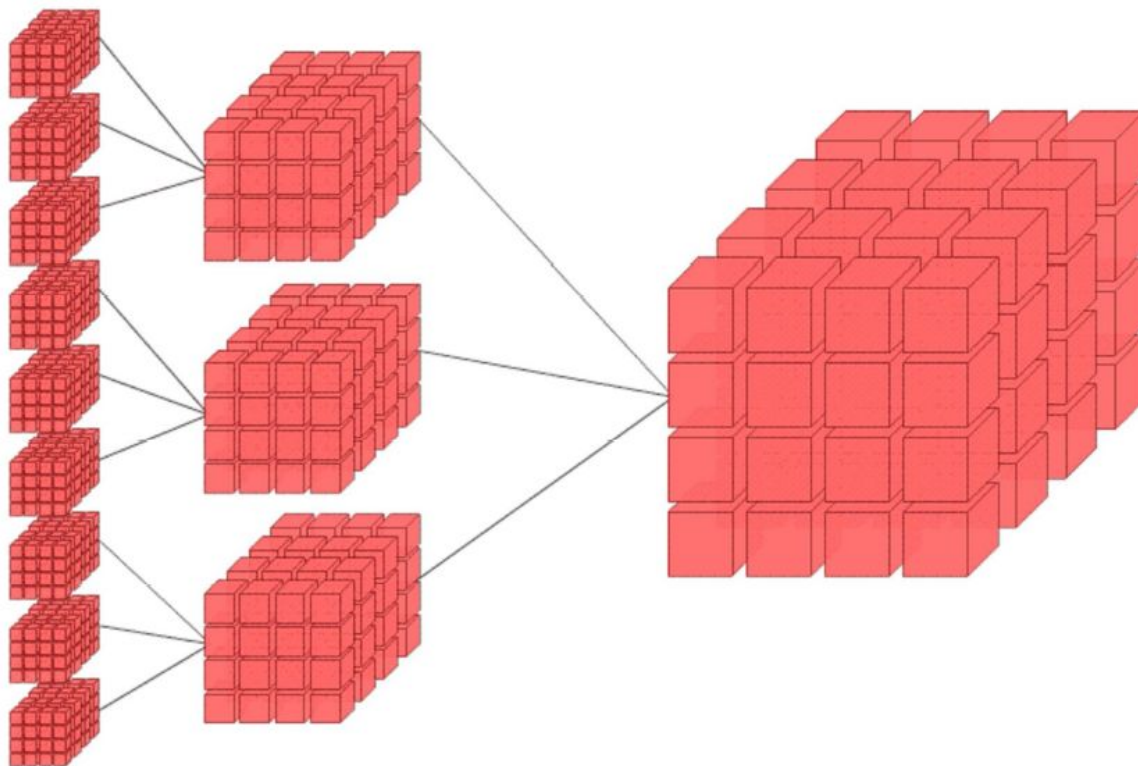
Quantas vendas e qual valor de vendas ocorreram considerando Trimestre, Fabricante e Cidade e considerando um slicing de ano entre 2001 e 2002?

```
SELECT Tempo.Ano, Tempo.Trimestre, Produto.Fabricante, Local.Cidade, Sum(Nro_unidades), Sum(valor)
FROM Fatos, Tempo, Produto, Local
WHERE Fatos.t_chave = Tempo.t_chave AND
      Fatos.p_chave = Produto.p_chave AND
      Fatos.l_chave = Local.l_chave AND
      Tempo.Ano between 2001 AND 2002
GROUP BY Tempo.Ano, Tempo.Trimestre, Produto.Fabricante, Local.Cidade
```

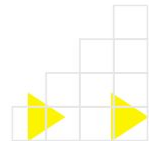
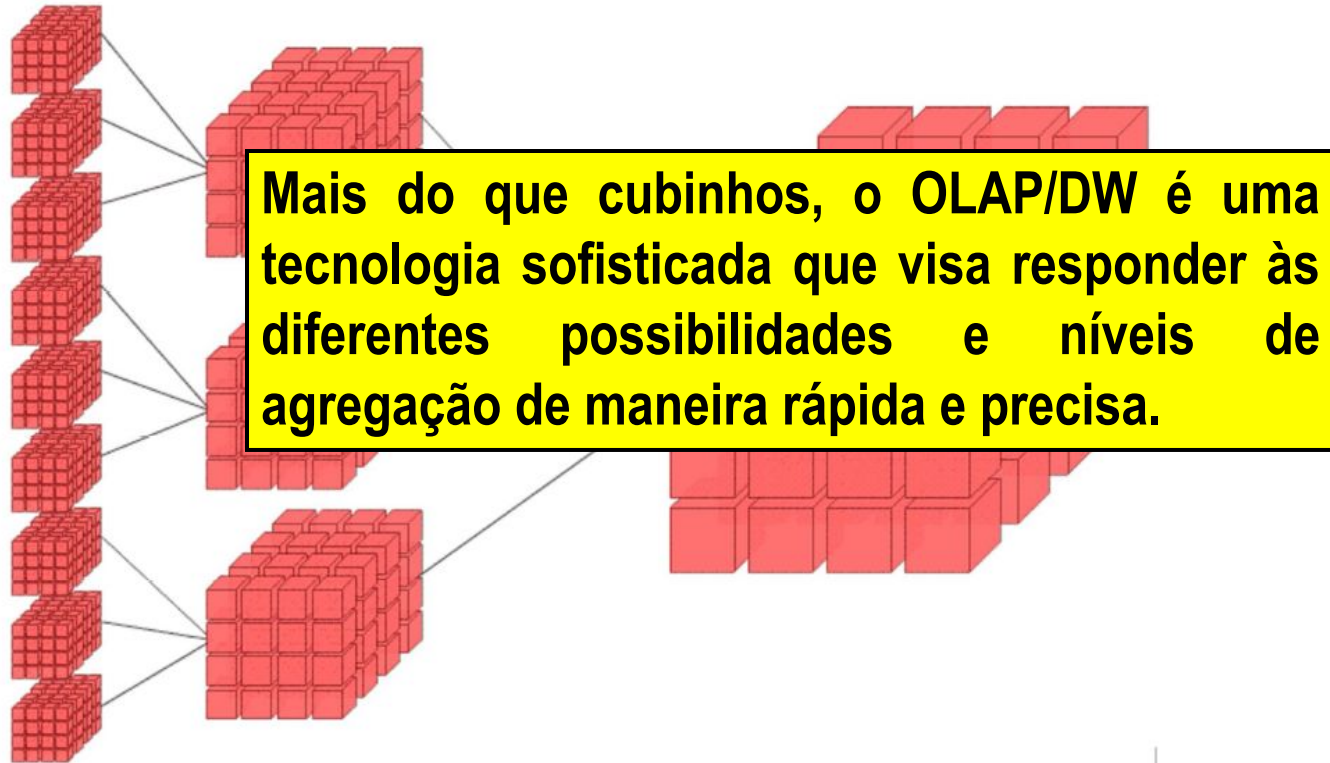
Localização



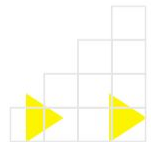
OLAP - Agregação



OLAP - Agregação



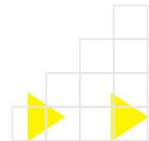
SQL OLAP



OLAP convencional single-node



- 📌 São **menos escaláveis**;
- 📌 Se baseiam em **processamento não distribuído**;
- 📌 Dependem de um hardware robusto para processamento ☐ **difícil de escalar**
- 📌 **Bilhões de tuplas** ☐ **latência muito grande**





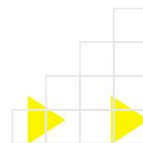
Relational OLAP (ROLAP)

- Recursos OLAP sobre SGBDs □ disparam todas as agregações possíveis sobre os atributos de interesse;
- SQL:1999
 - **GROUP BY CUBE**

```
SELECT ...  
FROM ...  
WHERE ...  
GROUP BY CUBE (trim, região)
```

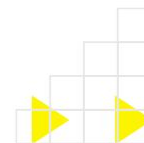
ex: selecionar

- ✓ total geral de vendas
- ✓ total de vendas por região
- ✓ total de vendas por trim.
- ✓ total de vendas por trim. por região



Relational OLAP (ROLAP)

- **GROUP BY CUBE** aplicado sobre **k** atributos
 - é equivalente ao agrupamento sobre cada um dos 2^k subconjuntos de atributos
 - ex: **k=2** \Rightarrow ... **GROUP BY CUBE** (a,b)
 - agrupamentos sobre: (a,b), (a), (b), (none)
 - agrupamento em (null) é o total geral de vendas
 - ex: **k=3** \Rightarrow ... **GROUP BY CUBE** (a,b,c)
 - agrupamentos sobre: (a,b,c), (a,b), (a,c), (b,c) (a), (b),(c), (none)



Relational OLAP (ROLAP)

○ Exemplo GROUP BY CUBE

```
SELECT Dept, Funcao,
        COUNT(*), SUM(Salario)
FROM Empregados
GROUP BY CUBE(Dept, Funcao);
```

GROUP BY CUBE(Dept, Funcao) =

- GROUP BY Dept, Funcao;
- GROUP BY Dept;
- GROUP BY Funcao;
- GROUP BY none; (ie, sem cláusula)

Dept	Funcao	COUNT(*)	SUM(Salario)
10	Secretario	1	100
10	Gerente	1	500
10	Presidente	1	900
10		3	1500
20	Analista	2	700
20	Secretario	2	240
20	Gerente	1	800
20		5	1740
	Secretario	3	340
	Gerente	2	1300
	Presidente	1	900
	Analista	2	700
		8	3240

