

OLAP e Data Warehousing

Sistemas de Apoio à Decisão

OLAP e Data Warehousing

- Sumário
 - *Data warehousing*
 - OLAP
 - Cubo de dados
 - ROLAP, MOLAP e HOLAP
 - OLTP vs. OLAP

Data Warehousing

- *Data warehouse*
 - Definição: “A Data Warehouse is a **subject oriented, integrated, time variant, non volatile** collection of data in support of management’s decision making process”.
 - Uma base de dados que é mantida em **separado** das bases de dados operacionais
 - Suporta o processamento de informação através de uma **plataforma sólida** com **dados históricos** consolidados para **análise**

Data Warehousing

- O que é o *data warehousing*?
 - Antes de mais é um processo
 - Um **processo** é uma sequência de **atividades** executadas de forma **sistemática e organizada**, que são realizadas por **intervenientes** com **responsabilidades** bem definidas
 - *Data warehousing* é o **processo** de **construção** e **utilização** de *data warehouses*

OLAP e Data Warehousing

- Sumário
 - *Data warehousing*
 - OLAP
 - Cubo de dados
 - ROLAP, MOLAP e HOLAP
 - OLTP vs. OLAP

OLAP

- O que é o OLAP?
 - OLAP - *OnLine Analytical Processing*
 - Termo introduzido em 1993 num “white paper” por Edgar Codd
 - Processo **interativo de criação, análise e relato** a partir de dados
 - **Análise de grandes quantidades de dados em tempo-real** para suporte à **tomada de decisão**

OLAP

- OLAP - *OnLine Analytical Processing*
 - Os dados são percebidos e manipulados como se estivessem armazenados num **array multidimensional**
 - Estruturas de armazenamento **multidimensionais**
 - Cubos de dados multidimensionais
 - Envolve um conjunto de **interfaces** e **aplicações**

OLAP

- OLAP - *OnLine Analytical Processing*
 - Agregação de dados de muitas e diferentes maneiras
 - O número de agrupamentos possível é muito grande
 - Todos os agrupamentos têm de ser considerados
 - Problema de processamento analítico

OLAP

- OLAP - *OnLine Analytical Processing*
 - Exemplo: Armazém

Fornecedor	Produto	Quantidade
F1	P1	300
F1	P2	200
F2	P1	300
F2	P2	400
F3	P2	200
F4	P2	200

OLAP

- OLAP - *OnLine Analytical Processing*
 - Exemplo: Armazém
 - Quantidade total em stock
 - Quantidade total em stock por fornecedor
 - Quantidade total em stock por produto
 - Quantidade total em stock por fornecedor e por produto
 - Necessário efetuar 4 consultas
 - Solução com limitações

OLAP e Data Warehousing

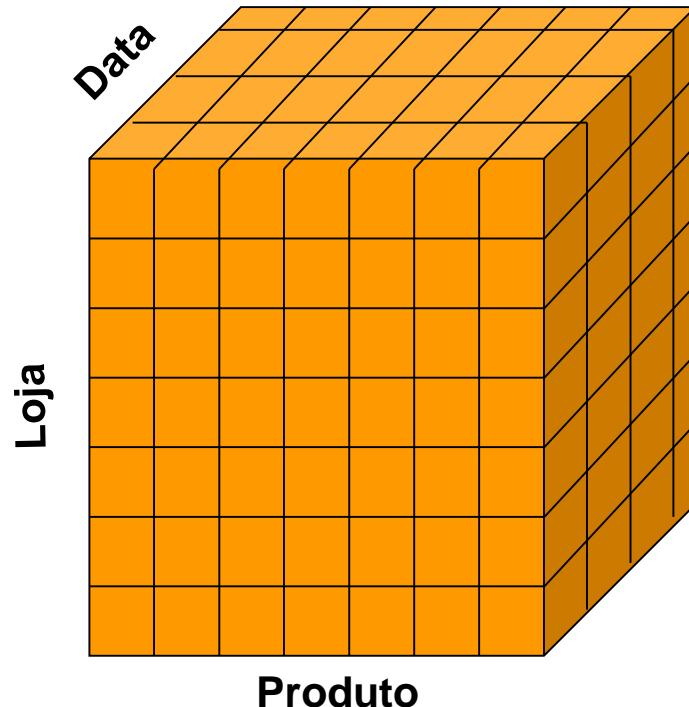
- Sumário
 - *Data warehousing*
 - OLAP
 - **Cubo de dados**
 - ROLAP, MOLAP e HOLAP
 - OLTP vs. OLAP

Cubo de Dados

- O que é um cubo de dados?
 - Estrutura de dados que armazena os dados em formato multidimensional, tornando-os mais fáceis e rápidos de analisar
 - Quando existem mais de três dimensões designa-se por hipercubo

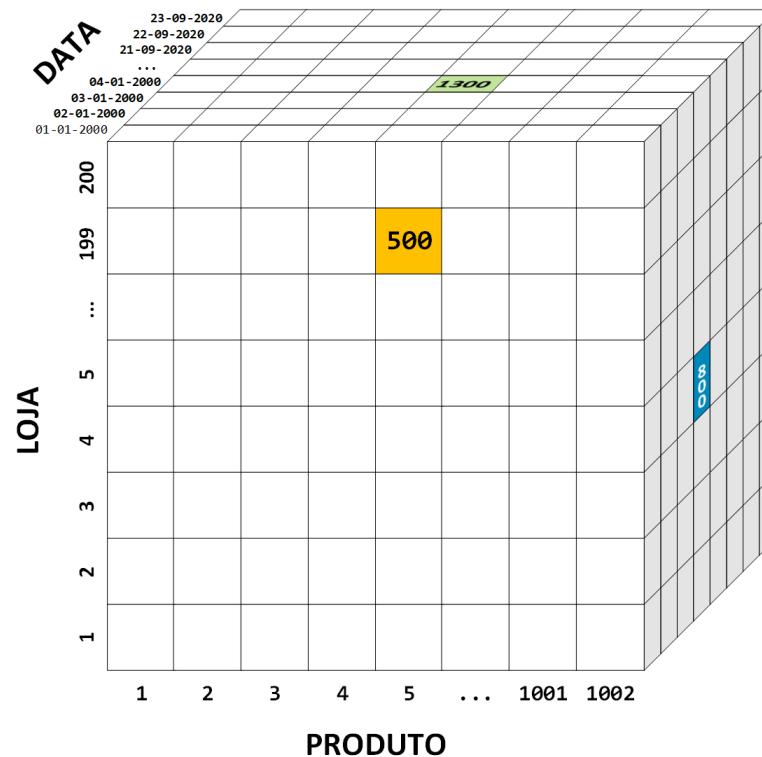
Cubo de Dados

- Cubo de dados multidimensional
 - Exemplo: Quantidade vendida



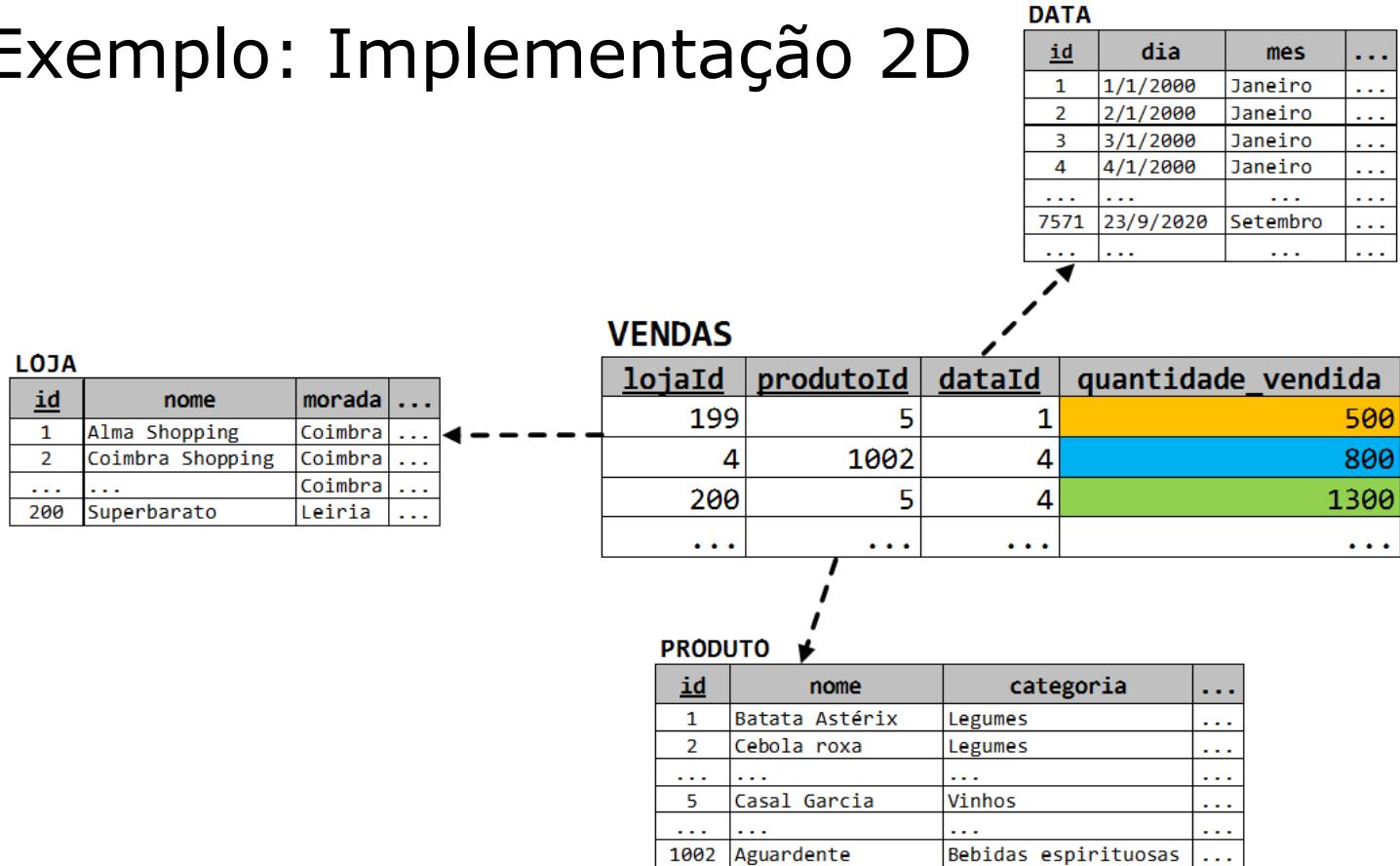
Cubo de Dados

- Cubo de dados multidimensional
 - Exemplo: Quantidade vendida



Cubo de Dados

- Cubo de dados multidimensional
 - Exemplo: Implementação 2D



Cubo de Dados

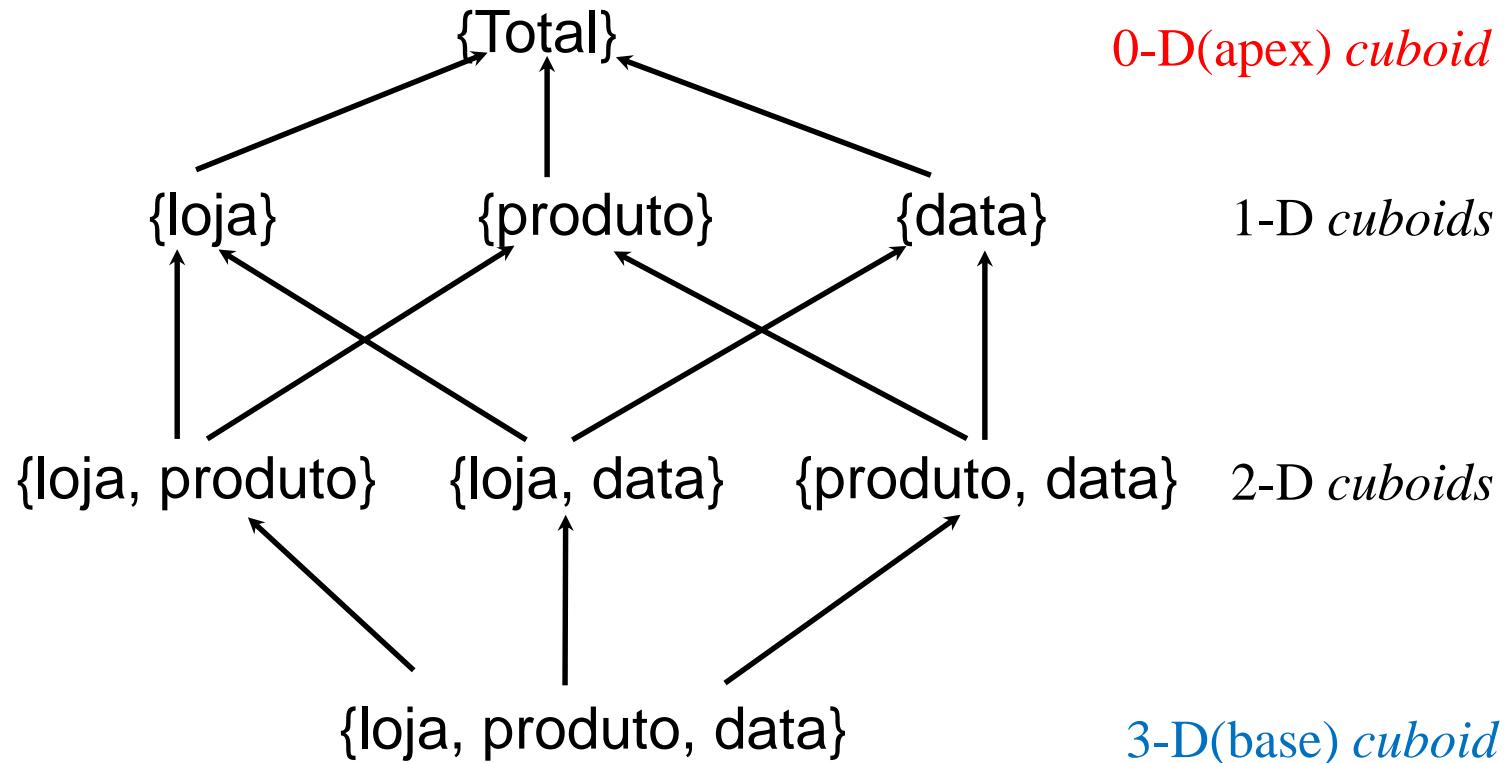
- Cubo de dados multidimensional
 - Exemplo: Unidades vendidas
 - **Variáveis independentes (dimensões)**
 - Data
 - Loja
 - Produto
 - **Variável dependente**
 - Quantidade vendida
 - Os valores da **variável dependente** estão armazenados nas **células do cubo de dados**

Cubo de Dados

- Conceito de *cuboid*
 - n-D *cuboid*
 - Contém o **nível mais baixo de agregação**, designado por *base cuboid*
 - 0-D *cuboid*
 - Contém o **nível mais elevado de agregação**, designado por *apex cuboid*
 - Grafo de *cuboids*
 - Forma o cubo de dados completo

Cubo de Dados

- Conceito de *cuboid*



Cubo de Dados

- Operações OLAP
 - Operações típicas num cubo de dados
 - *Drilling*
 - *Rolling*
 - *Slicing*
 - *Dicing*
 - *Pivoting*
 - Estas **operações** num cubo de dados são instantâneas
 - Numa **base de dados relacional** podem demorar (minutos/horas)

Cubo de Dados

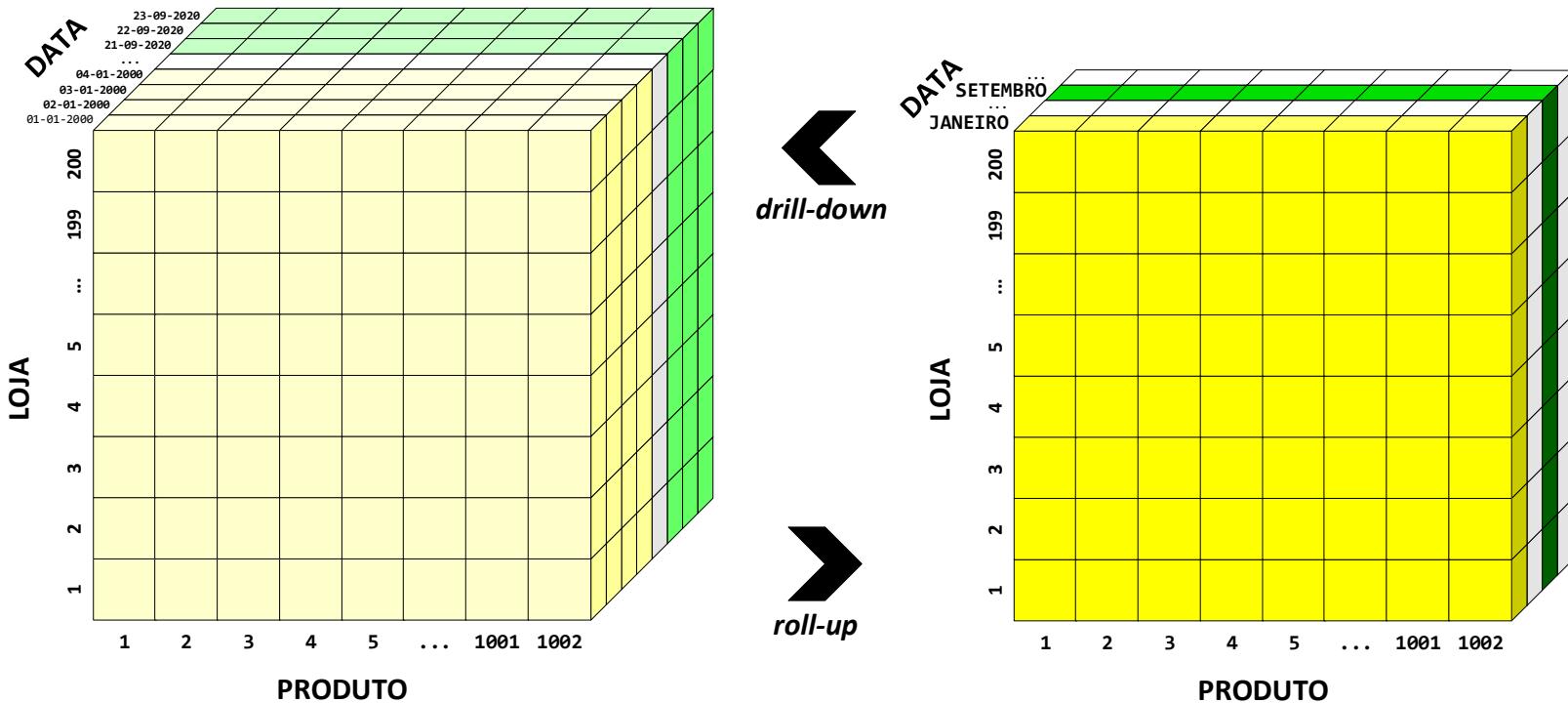
- Operações OLAP: *Drilling*
 - *Drill-down*: consiste em **aumentar o nível de detalhe** de uma consulta, ou seja, ir de um nível mais elevado para um nível mais baixo de agregação
 - As operações de *drill-down* podem ser feitas até se atingir o **grão mais fino** de informação, isto é o **nível atómico** dos dados
 - Introduzindo novas dimensões/atributos

Cubo de Dados

- Operações OLAP: *Rolling*
 - *Roll-up*: consiste em **diminuir o nível de detalhe** de uma consulta
 - Operação de agregação num **grão superior** de informação
 - Diminuindo o número de dimensões ou subindo numa hierarquia
 - Inverso do *drill-down*

Cubo de Dados

- Operações OLAP: *Drill-down / Roll-up*



Cubo de Dados

- Operações OLAP: *Drill-down / Roll-up*

```
SELECT p.nome, l.nome, d.mes, total
FROM vendas
    JOIN produto p ON ...
    JOIN loja l ON ...
    JOIN
        (SELECT SUM(quantidade_vendida) AS total, data.mes
         FROM vendas JOIN data ON ...
          GROUP BY data.mes);
```

drill-down



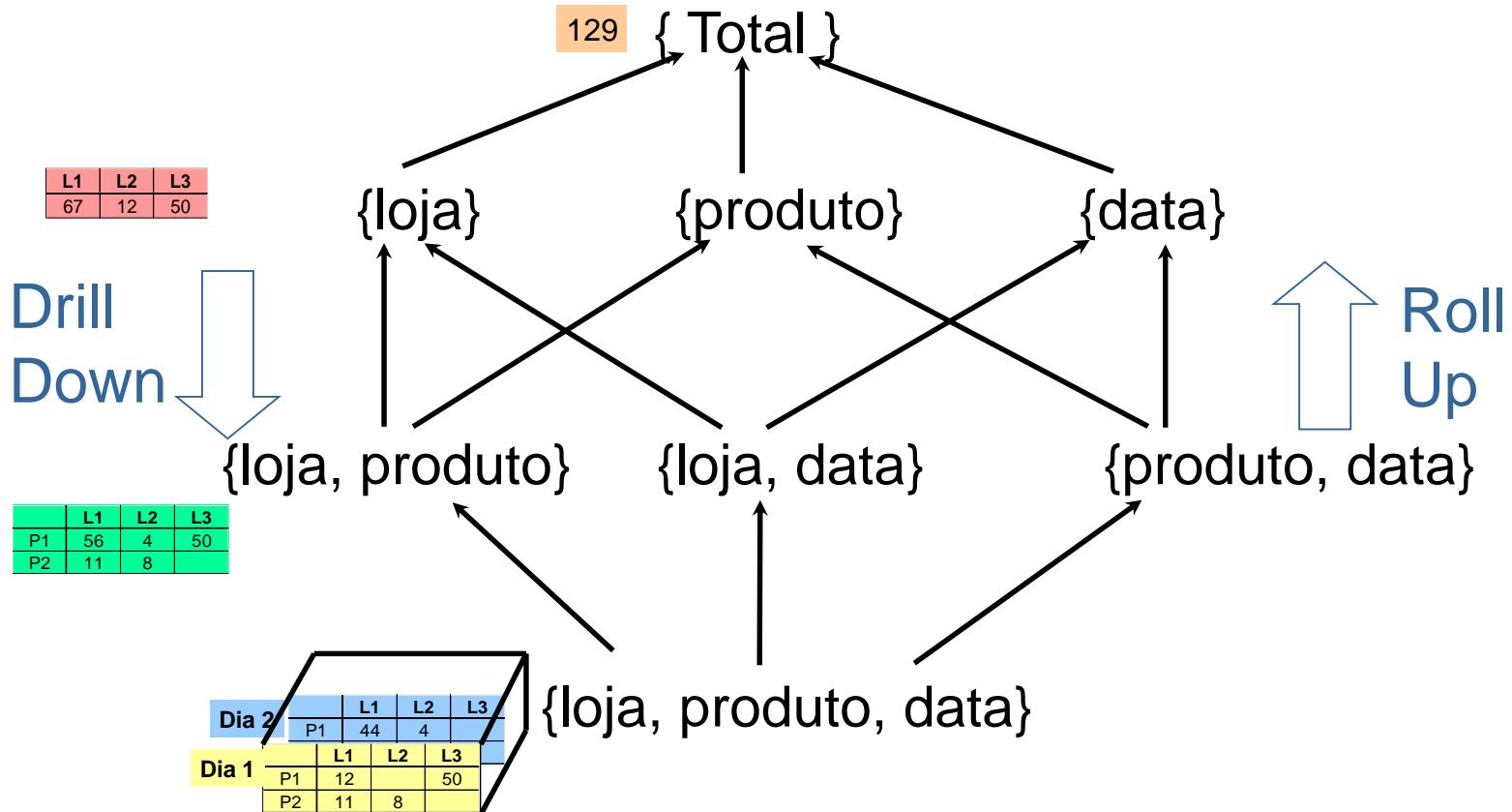
roll-up



```
SELECT p.nome, l.nome, d.dia, quantidade_vendida
FROM vendas
    JOIN produto p ON...
    JOIN loja l ON ...
    JOIN data d ON ...;
```

Cubo de Dados

- Operações OLAP: *Drill-down* e *Roll-up*

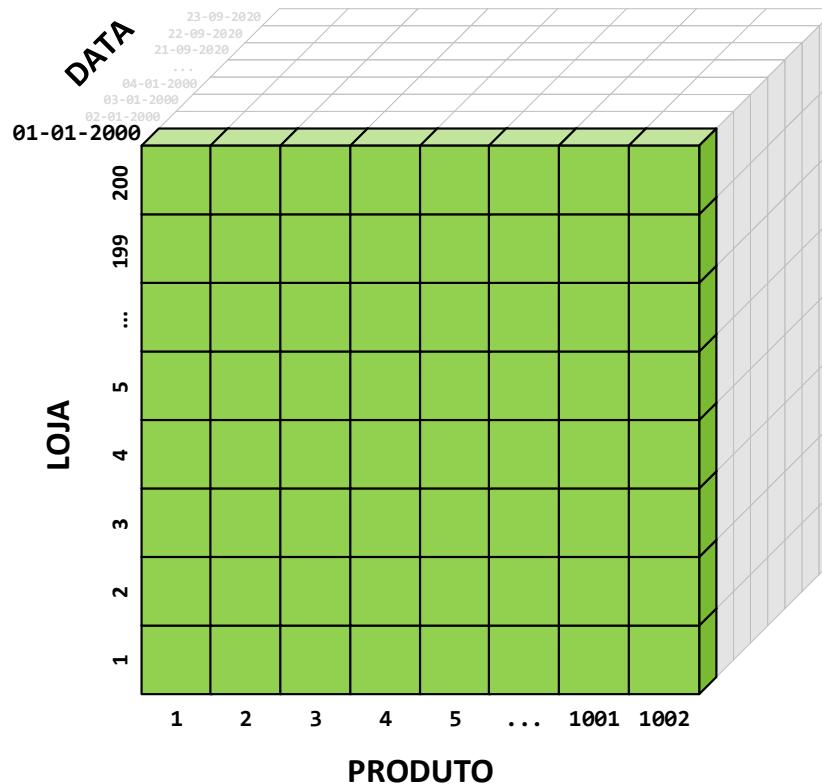


Cubo de Dados

- Operações OLAP: *Slicing*
 - Consiste em **cortar o cubo** ao longo de uma ou mais dimensões
 - **Fixa-se um valor** numa das **dimensões** de análise
 - Nas outras dimensões continua a ser possível realizar operações de *drill-down* e *roll-up*

Cubo de Dados

- Operações OLAP: *Slicing*

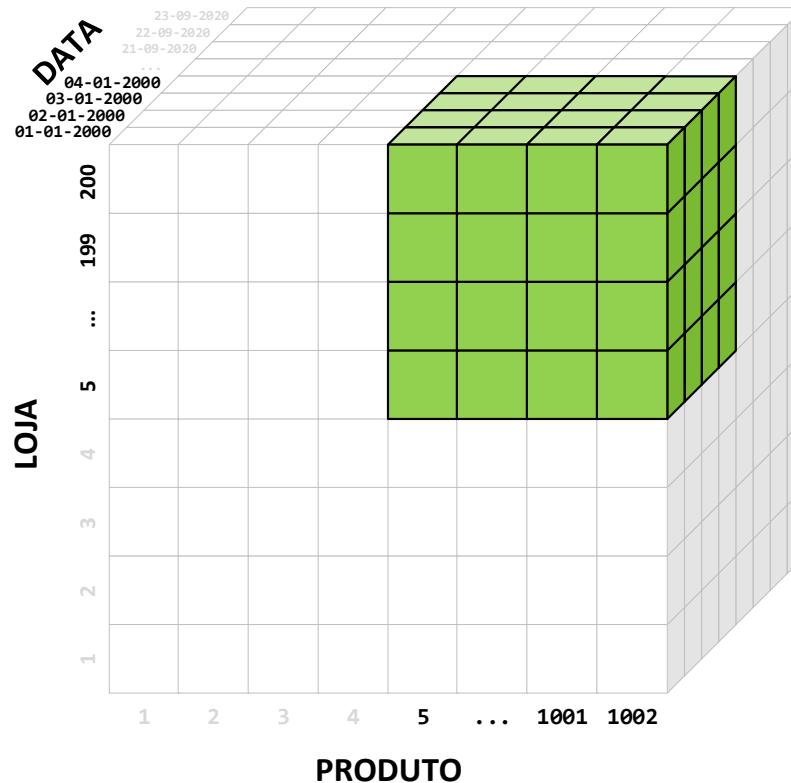


Cubo de Dados

- Operações OLAP: *Dicing*
 - Consiste em **definir um subcubo** limitando os valores de uma ou mais dimensões
 - **Concentração da análise**
 - Colocam-se em segundo plano dimensões que de momento não interessam
 - Permite trazer para primeiro plano as restantes dimensões

Cubo de Dados

- Operações OLAP: *Dicing*

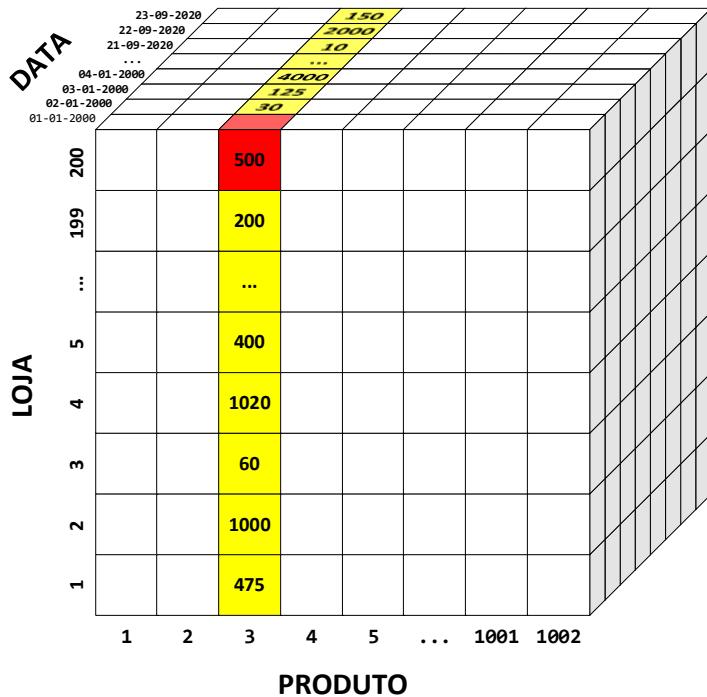


Cubo de Dados

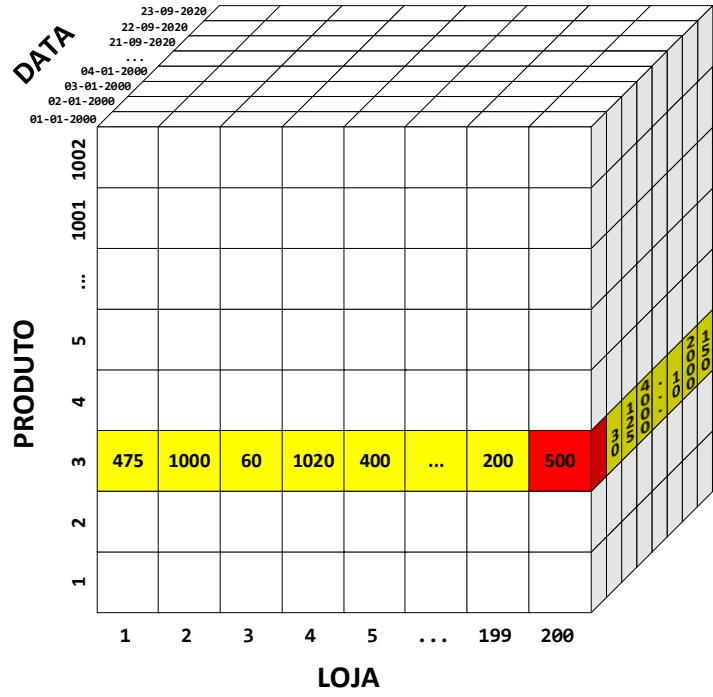
- Operações OLAP: *Pivoting*
 - Permite a **rotação do cubo** de dados
 - Útil em termos de **visualização**
 - Permite transformar uma visualização 3D numa série de planos 2D

Cubo de Dados

- Operações OLAP: *Pivoting*



PIVOT



Cubo de Dados

- Outras operações OLAP
 - *Drill-through*
 - Detalhe além do cubo (ao nível dos registos)
 - *Drill-across*
 - Consultas que envolvem mais do que um data mart
 - *Ranking*
 - Ordenação de resultados
 - *Filtering*
 - Aplicação de filtros nos dados

Cubo de Dados

- No mundo real
 - Os **dados** são armazenados em **bases de dados relacionais**
 - É possível representar um problema com n dimensões utilizando **tabelas 2D?**
 - É possível combinar o OLAP com pesquisas em SQL?
 - Extensões à linguagem SQL
 - CUBE
 - ROLLUP, etc.

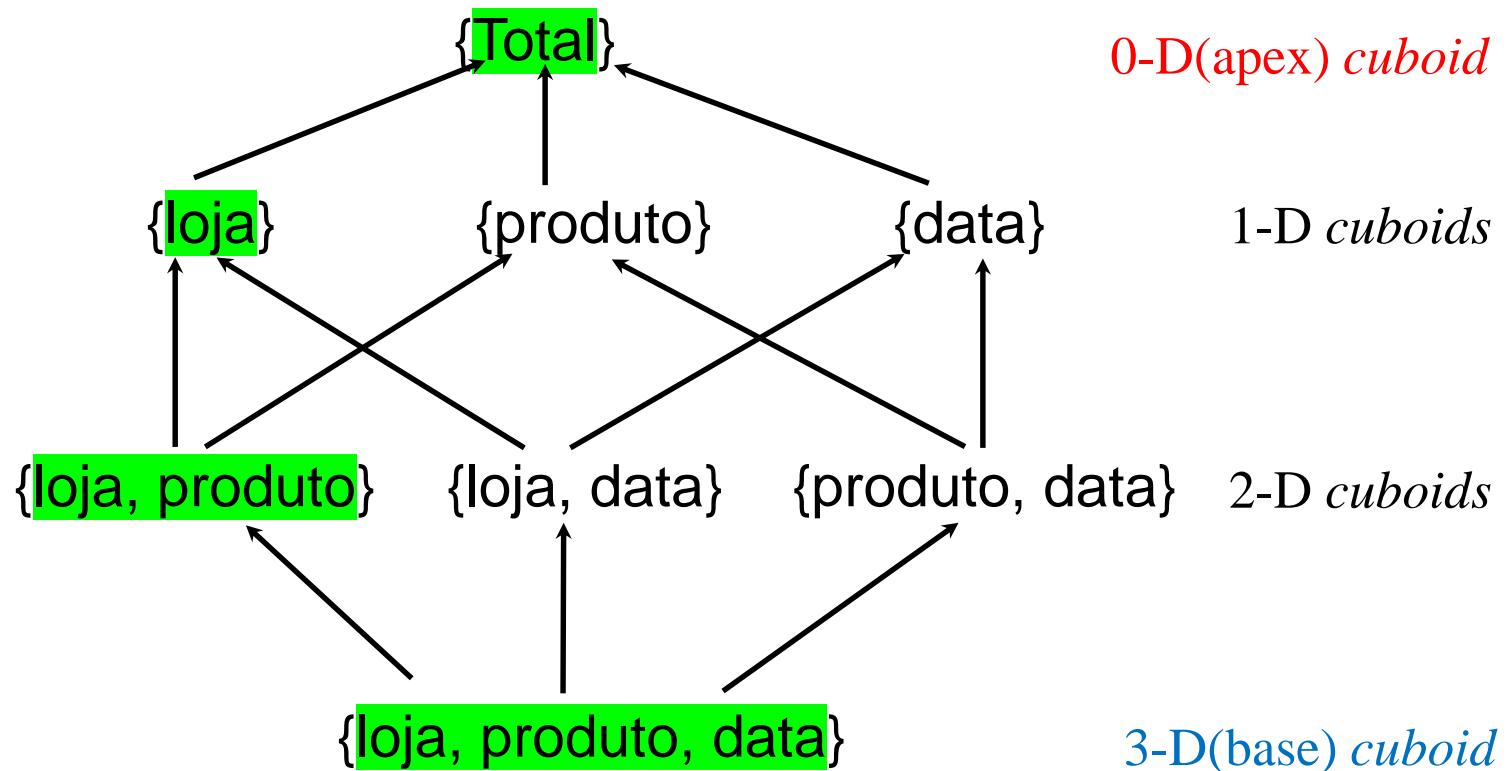
Cubo de Dados

- ROLLUP
 - Calcula um **caminho** através do **grafo de cubóides**
 - A ordem no GROUP BY interessa

```
SELECT loja, produto, data, SUM(quantidade)
FROM SALES
GROUP BY ROLLUP(loja, produto, data)
```
 - Agregações calculadas
 - {loja,produto,data}
 - {loja,produto}
 - {loja}
 - {Total}

Cubo de Dados

- ROLLUP



Cubo de Dados

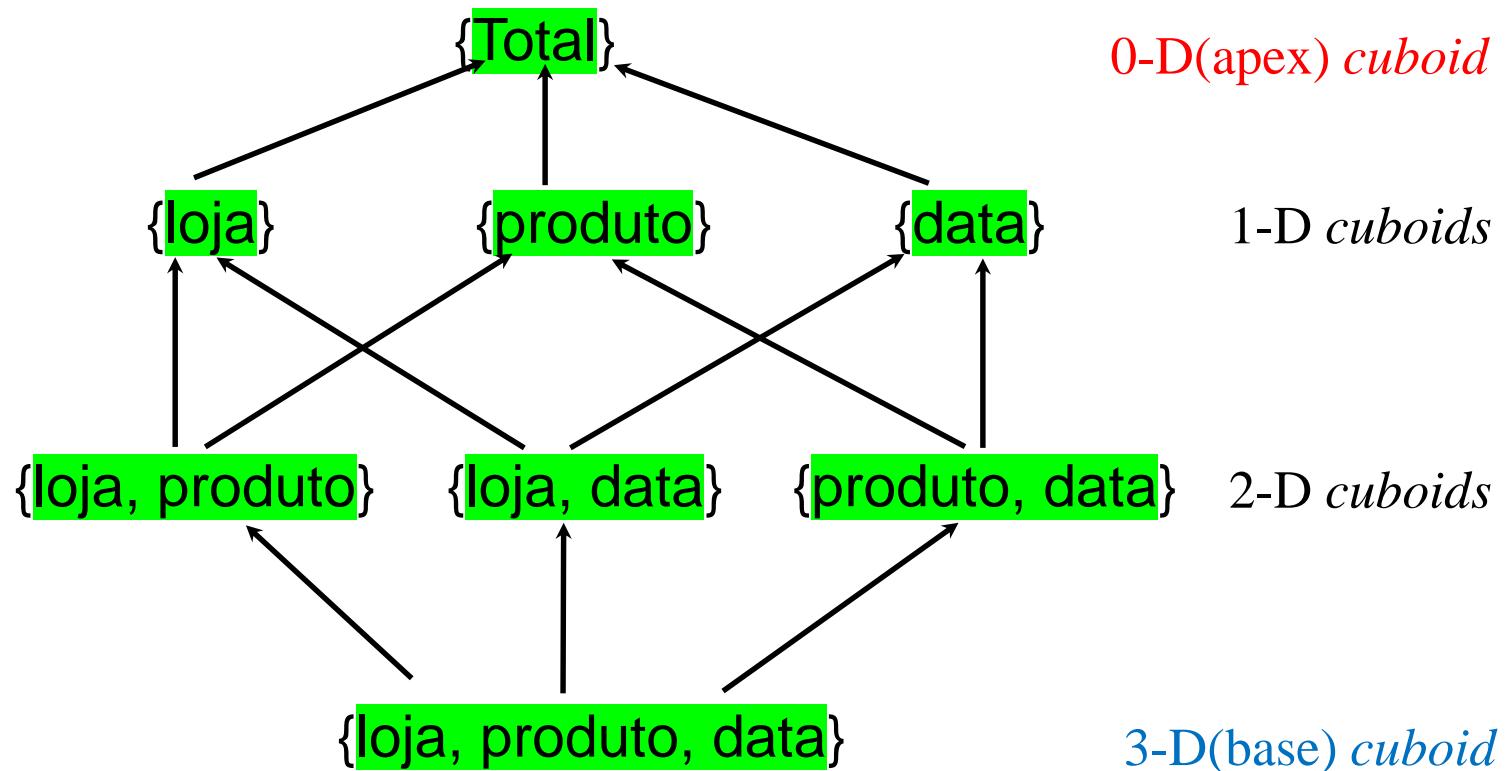
- CUBE
 - Calcula o **grafo de cubóides completo**
 - A **ordem no GROUP BY não é relevante**

```
SELECT loja, produto, data, SUM(quantidade)  
FROM SALES  
GROUP BY CUBE(loja, produto, data)
```

- Agregações calculadas
 - $\{\text{loja}, \text{produto}, \text{data}\}$
 - $\{\text{loja}, \text{produto}\}$, $\{\text{loja}, \text{data}\}$, $\{\text{produto}, \text{data}\}$
 - $\{\text{loja}\}$, $\{\text{produto}\}$, $\{\text{data}\}$
 - $\{\text{Total}\}$

Cubo de Dados

- CUBE



OLAP e Data Warehousing

- Sumário
 - *Data warehousing*
 - OLAP
 - Cubo de dados
 - ROLAP, MOLAP e HOLAP
 - OLTP vs. OLAP

ROLAP, MOLAP e HOLAP

- O que é o **ROLAP**, **MOLAP** e **HOLAP**?
 - São **modelos físicos** de **armazenamento de dados**
 - Definem a **arquitetura OLAP** para a construção dos **cubos de dados**
 - Na prática são servidores OLAP

ROLAP, MOLAP e HOLAP

- ROLAP - *Relational OLAP*
 - OLAP sobre motores de bases de dados relacionais
 - Permite a **escalabilidade** dos dados
 - Extensões do SQL
 - Modelo utilizado nesta unidade curricular

ROLAP, MOLAP e HOLAP

- MOLAP - *Multidimensional OLAP*
 - OLAP sobre motores de bases de dados multidimensionais
 - Forma clássica de OLAP
 - Fornece um processamento mais rápido para as consultas
 - Suporta menos dados devido à sua esparsidão (células vazias)

ROLAP, MOLAP e HOLAP

- HOLAP - *Hybrid OLAP*
 - OLAP sobre motores de bases de dados multidimensionais e relacionais
 - Compromisso entre o ROLAP e MOLAP
 - Razoavelmente rápido em termos de pesquisas
 - Relativamente poupado em termos de espaço de armazenamento

OLAP e Data Warehousing

- Sumário
 - *Data warehousing*
 - OLAP
 - Cubo de dados
 - ROLAP, MOLAP e HOLAP
 - OLTP vs. OLAP

OLTP vs. OLAP

	OLTP	OLAP
Objetivo chave	Disponibilidade	Performance
Tamanho comum	Mega a Gigabyte	Giga a Tera/Petabyte
Granularidade	Atómico	Atómico e agregado
Origem dos dados	Interna	Interna e externa
Atualizações	Em tempo real	<i>Batch</i>
Período dos dados	Sobretudo atual	Histórico
Consultas	Pré-definidas	Pré-definidas e <i>ad hoc</i>
Atividade	Operacional	Analítica
Utilizadores	Muitos	Poucos
Modelo de dados	Normalizado	Desnormalizado

OLAP e Data Warehousing

- Referências
 - The Data Warehouse Toolkit, R. Kimball, John Wiley & Sons, 2013 (3^a edição)
 - Building the Data Warehouse, W. Inmon, John Wiley & Sons, 2005 (4^a edição)