

BANCO DE DADOS II

Data Warehouse

DATA WAREHOUSE BIG DATA DATA SCIENCE

CHAUDHURI, Surajit; DAYAL, Umeshwar. **An overview of data warehousing and OLAP technology.** ACM Sigmod record, v. 26, n. 1, p. 65-74, 1997.

SAGIROGLU, Seref; SINANC, Duygu. **Big data: A review.** In: Collaboration Technologies and Systems (CTS), 2013 International Conference on. IEEE, 2013. p. 42-47.

- **Pesquisar 4 e 5 Vs do Big Data.**

DHAR, Vasant. **Data science and prediction.** Communications of the ACM, v. 56, n. 12, p. 64-73, 2013.

CONTEXTO

Os bancos de dados tradicionais equilibram o requisito de acesso aos dados com a exigência por integridade/consistência.

Em alguns contextos organizacionais complexos e em espaços de tomadas de decisão, muitas pessoas precisam de acesso de leitura aos dados, mas:

- Acesso a um volume maior do que o de aplicações convencionais;
- Frequentemente dados oriundos de diversos bancos de dados;
- Acesso rápido;

Para atender a demanda de análise rápida e precisa, em um nível de detalhe correto, tem-se:

- Data warehouse;
- OLAP - Processamento Analítico On-line;
- Data Mining;

ELMASRI, Ramez; NAVATHE, Shamkant. Database systems. Boston, MA: Pearson Education, 2011.

DATA WAREHOUSE



Data Warehouse é uma coleção de informações, bem como um sistema de suporte. Contudo há uma distinção importante para os bancos de dados tradicionais:

- Bancos de dados tradicionais são transacionais;
- Data Warehouses são otimizados para a recuperação de dados, não para processamento de transação de rotina.
 - Servem principalmente a aplicações de apoio a decisão.

"Uma coleção de dados orientada a assunto, integrada, não volátil, variável no tempo para o suporte às decisões da gerência."

INMON, W. H. EIS and the data warehouse: a simple approach to building an effective foundation for EIS. Database Programming & Design, v. 5, n. 11, p. 70-73, 1992.

"Em sentido amplo, o termo data warehouse é usado para se referir a um banco de dados que contém um armazenamento muito grande de dados históricos. O dado é armazenado como uma série de snapshots, nos quais cada registro representa o dado em um tempo específico."

IBM Informix 12.10. Overview of data warehousing.

DATA WAREHOUSE



Data Warehouses oferecem acesso a dados para análises complexas, descoberta de conhecimento (KDD - Knowledge Discovery in Databases) e tomada de decisão em vários tipos de aplicações:

- **DSS - Sistemas de Apoio à Decisão**
 - Também conhecidos como **EIS** - Sistemas de Informações executivas.
 - Auxilia os stakeholders na tomada de decisões estratégicas com dados de nível mais alto.
- **OLAP - Processamento Analítico On-Line**
 - Análise de dados complexos em data warehouse.
 - Demanda mais recursos para armazenamento e processamento.

DATA WAREHOUSE CARACTERÍSTICAS

- Visão conceitual específica:
 - **Modelo de dados Multidimensional.**
- Dados não voláteis (mudam em baixa frequência);
- Tempo **não** real:
 - Atualização periódica;
 - Atualizações incrementais;
 - Atualização tratadas por componente específico de aquisição que oferece o pré-processamento exigido;
- Dados previamente limpos e reformatados;

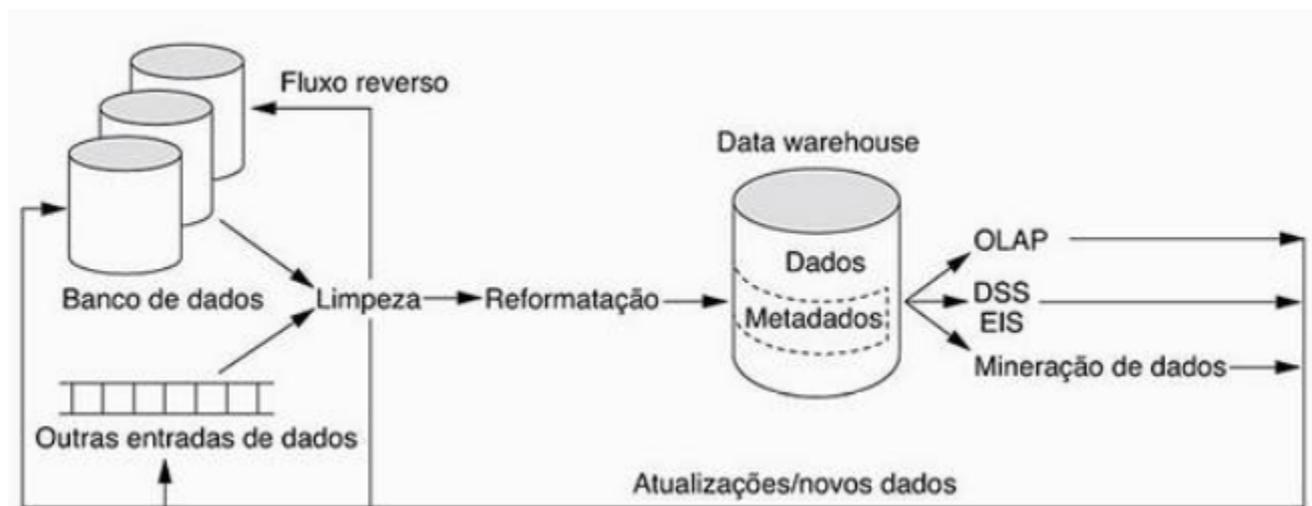
- Dimensionalidade genérica;
- Dimensões e níveis de agregação ilimitados;
- Operações irrestritas entre dimensões;
- Tratamento dinâmico de matriz esparsa;
- Arquitetura cliente-servidor;
- Suporte para múltiplos usuários;
- Acessibilidade;
- Transparência;
- Manipulação de dados intuitiva;
- Desempenho de relatório consistente;
- Recurso de relatório flexível;

ELMASRI, Ramez; NAVATHE, Shamkant. Database systems. Boston, MA: Pearson Education, 2011.

DATA WAREHOUSE ETL PROCESS

O processo de data warehousing inclui a possível limpeza e reformatação dos dados antes de serem carregados no **armazém**.

Ferramentas de **ETL** (Extract, Transform and Load) realizam esse processo.

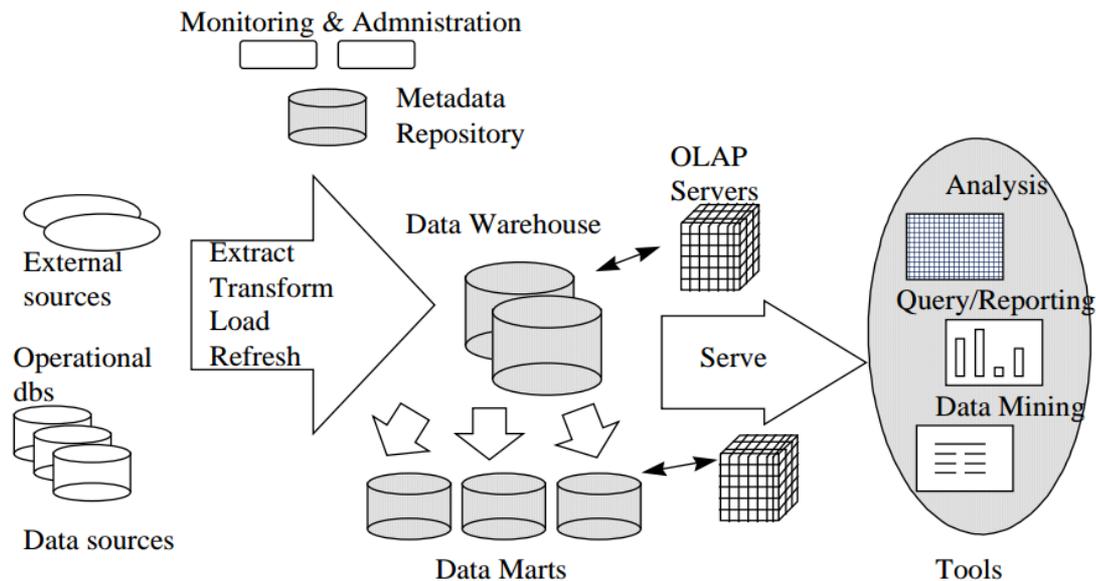


ELMASRI, Ramez; NAVATHE, Shamkant. Database systems. Boston, MA: Pearson Education, 2011.

DATA WAREHOUSE ARQUITETURA

Data warehouses, são uma ordem de magnitude (em alguns casos duas ordens de magnitude) maiores que os bancos de dados de origem.

Esse imenso volume de dados tem sido tratado em **Enterprise Data Warehouses**, **Data Warehouses virtuais** e **Data Marts**.



Data Warehousing Architecture

CHAUDHURI, Surajit; DAYAL, Umeshwar. An overview of data warehousing and OLAP technology. *ACM Sigmod record*, v. 26, n. 1, p. 65-74, 1997.

DATA WAREHOUSE MODELO MULTIDIMENSIONAL

Modelos multidimensionais extraem o máximo dos relacionamentos inerentes aos dados para dispô-los em matrizes multidimensionais - cubos de dados.

Uma planilha padrão, ou relação, seria bidimensional.

Exemplo:

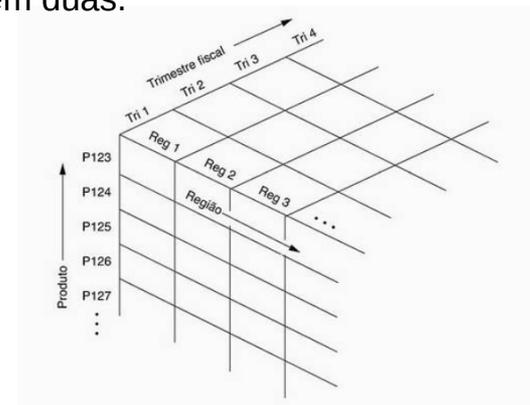
- Dimensão 1: Vendas regionais por período
- Dimensão 2: Produto

Diagrama de uma matriz bidimensional representando dados de vendas regionais por produto. O eixo vertical é rotulado 'Produto' e contém as entradas P123, P124, P125, P126 e um símbolo de ponto e vírgula. O eixo horizontal é rotulado 'Região' e contém as entradas Reg 1, Reg 2, Reg 3 e um símbolo de ponto e vírgula. A matriz é composta de células vazias que representam os pontos de dados.

Um cubo de dados multidimensionais decompõe a dimensão 1 em duas.

Exemplo:

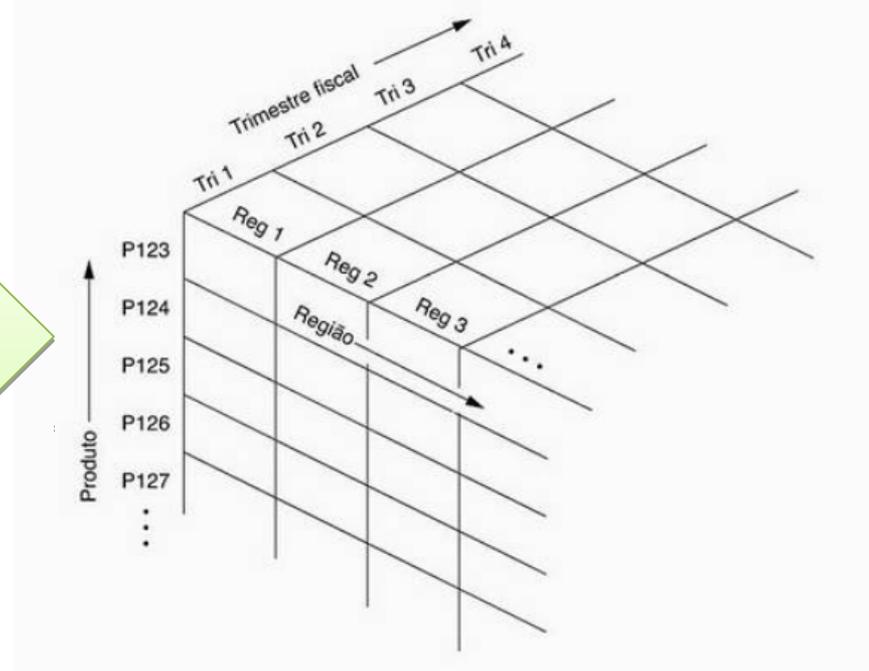
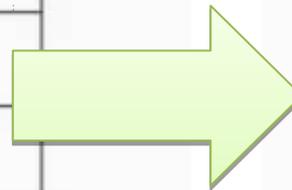
- Dimensão 1: Períodos Fiscais
- Dimensão 2: Produtos
- Dimensão 3: Regiões da empresa



ELMASRI, Ramez; NAVATHE, Shamkant. Database systems. Boston, MA: Pearson Education, 2011.

DATA WAREHOUSE MODELO MULTIDIMENSIONAL

Produto	Região			
	Reg 1	Reg 2	Reg 3	...
P123				
P124				
P125				
P126				
⋮				



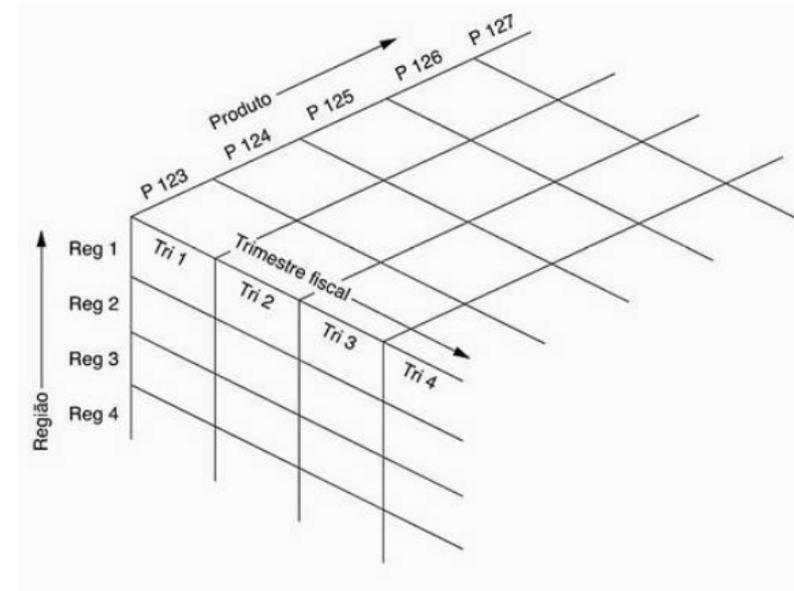
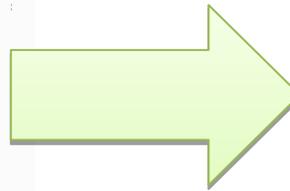
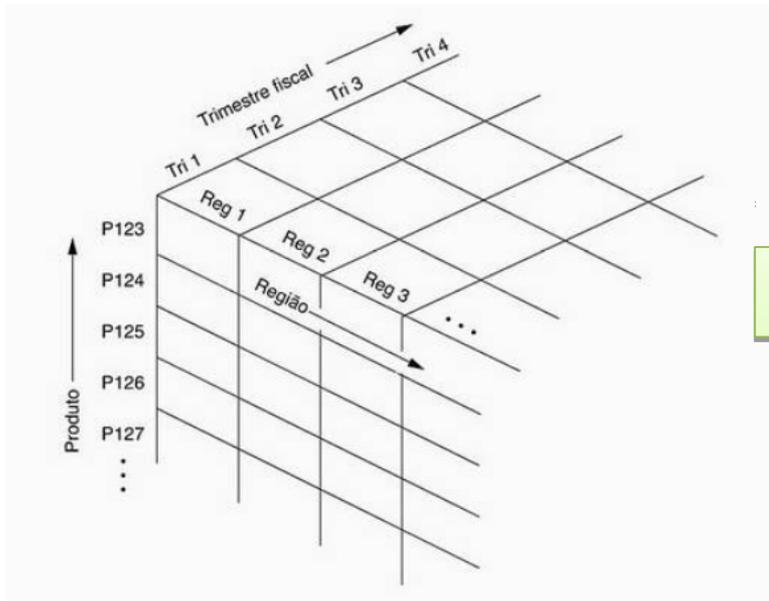
DATA WAREHOUSE OPERAÇÕES

Pivot / Giro / Reorientação:

- O cubo tem a disposição dos eixos reorientada (girando sobre um pivot).

Exemplo:

Mostrar as receitas de vendas regionais como linhas, os totais de receita por trimestre fiscal como colunas e os produtos da empresa na terceira dimensão.

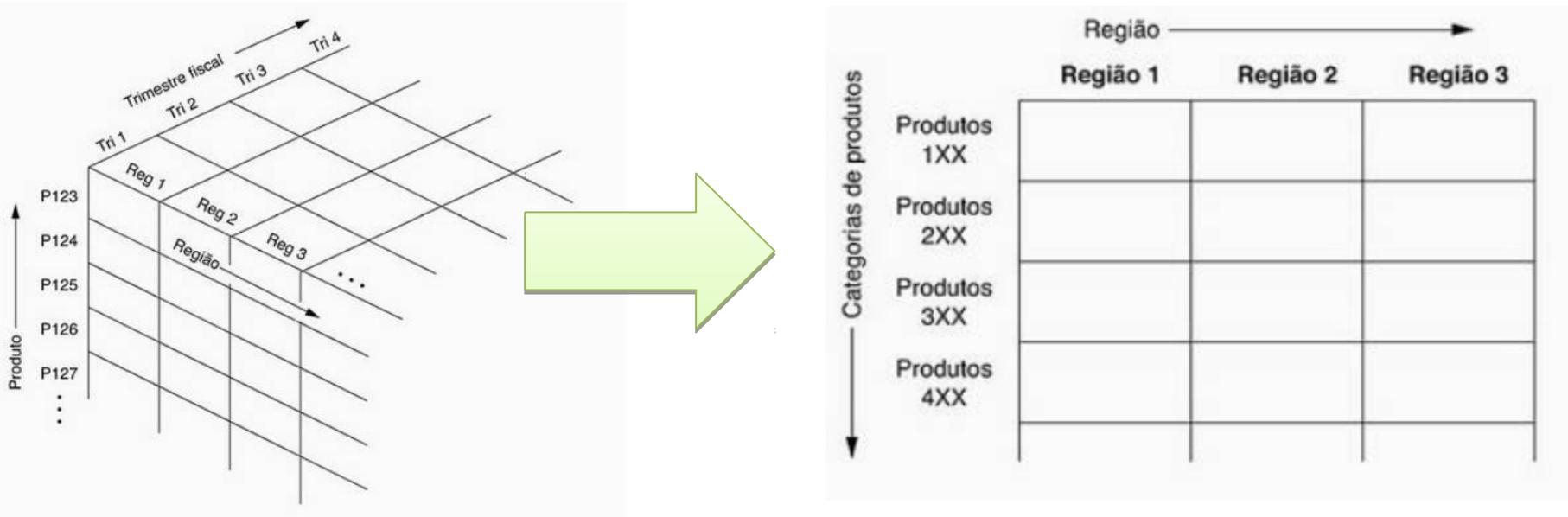


ELMASRI, Ramez; NAVATHE, Shamkant. Database systems. Boston, MA: Pearson Education, 2011.

DATA WAREHOUSE OPERAÇÕES

Rollup:

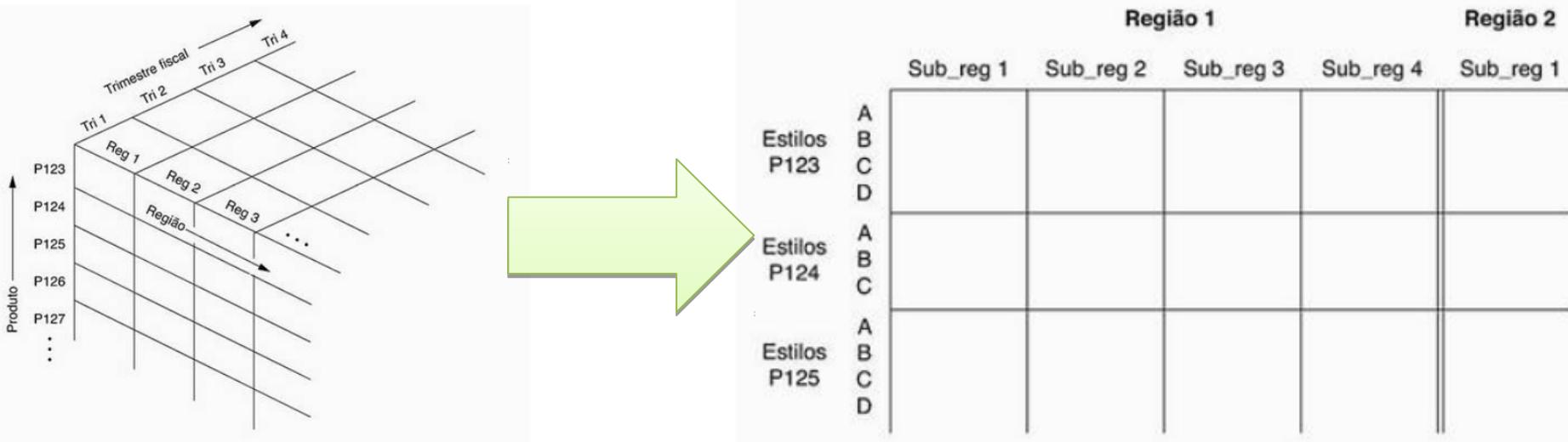
- Realiza a exibição dos dados subindo na hierarquia, agrupando os dados em unidades maiores ao longo de uma dimensão.
- Exemplo:
 - somar dados semanais em meses, trimestres ou ano.



DATA WAREHOUSE OPERAÇÕES

Drill-Down:

- Realiza a exibição dos dados descendo na hierarquia, desagregando os dados em unidades menores ao longo de uma dimensão.
- Apresenta uma visão mais detalhada.
- Exemplos:
 - Desagregar as vendas do país por região e sub-região;
 - Desagregar produtos por estilos;



ELMASRI, Ramez; NAVATHE, Shamkant. Database systems. Boston, MA: Pearson Education, 2011.

DATA WAREHOUSE OPERAÇÕES

Slice e Dice:

- Realiza operações de projeção e seleção nas N dimensões.

Ordenação:

- Ordena os dados por valor ordinal.

Atributos derivados:

- Atributos são calculados por operações sobre valores armazenados e derivados.

DATA WAREHOUSE

TABELAS DE FATOS E DE DIMENSÕES

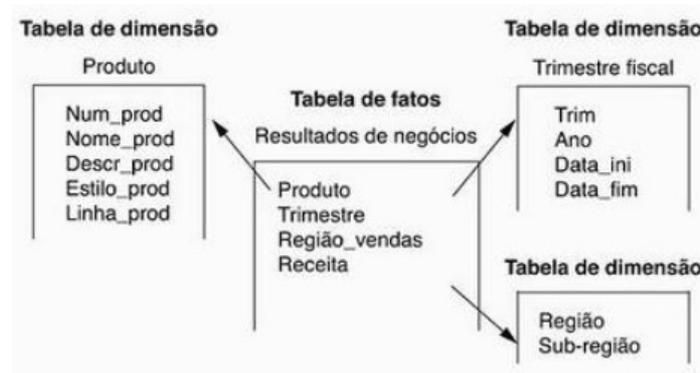
O modelo dimensional envolve dois tipos de relações (tabelas):

Tabela de dimensão:

- Armazena atributos nos quais uma dimensão se decompõe.
- **Ex.** Dimensão Produto: Número, Nome, Descrição, Estilo, Linha...

Tabela de Fatos:

- Armazena os fatos registrados em tuplas.
- Cada fato possui atributos e seus valores são ponteiros / referências para tabelas de dimensão.
- **Ex.** Produto1 vendido em 10/10/2018 em Mosoró por R\$10.000.000.000,00.



ELMASRI, Ramez; NAVATHE, Shamkant. Database systems. Boston, MA: Pearson Education, 2011.

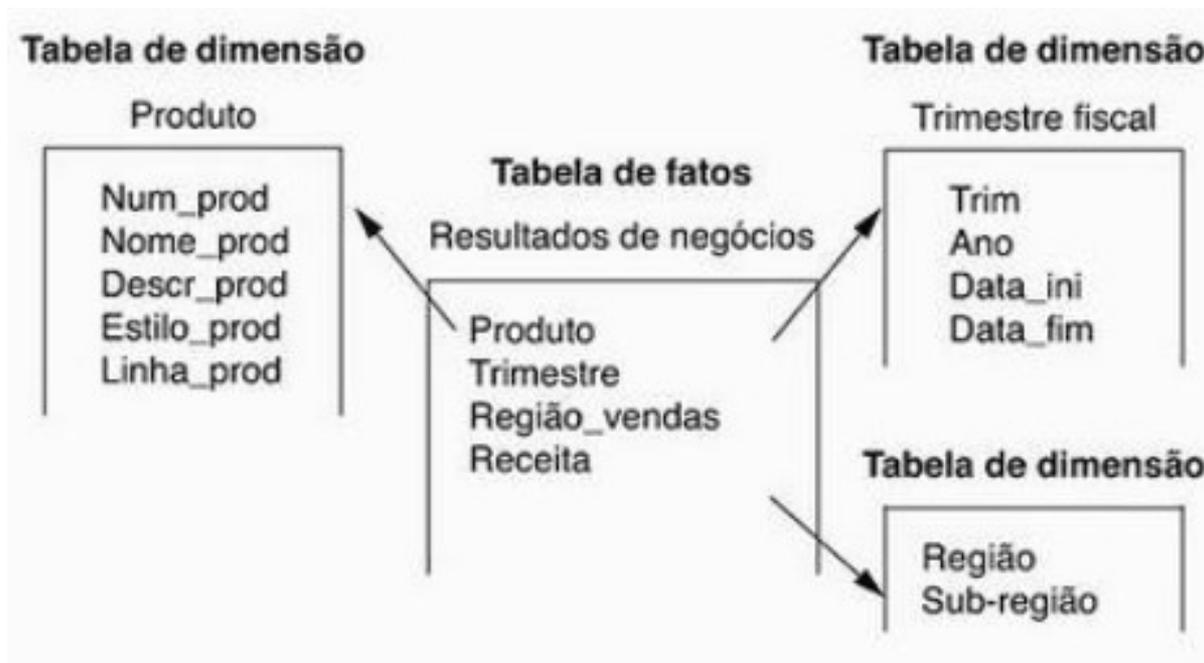
DATA WAREHOUSE

ESQUEMAS MULTIDIMENSIONAIS

Os esquemas multidimensionais comumente variam entre Esquema Estrela e Esquema Floco de Neve.

Esquema estrela:

- Uma tabela de fatos possui apenas uma tabela para cada dimensão.



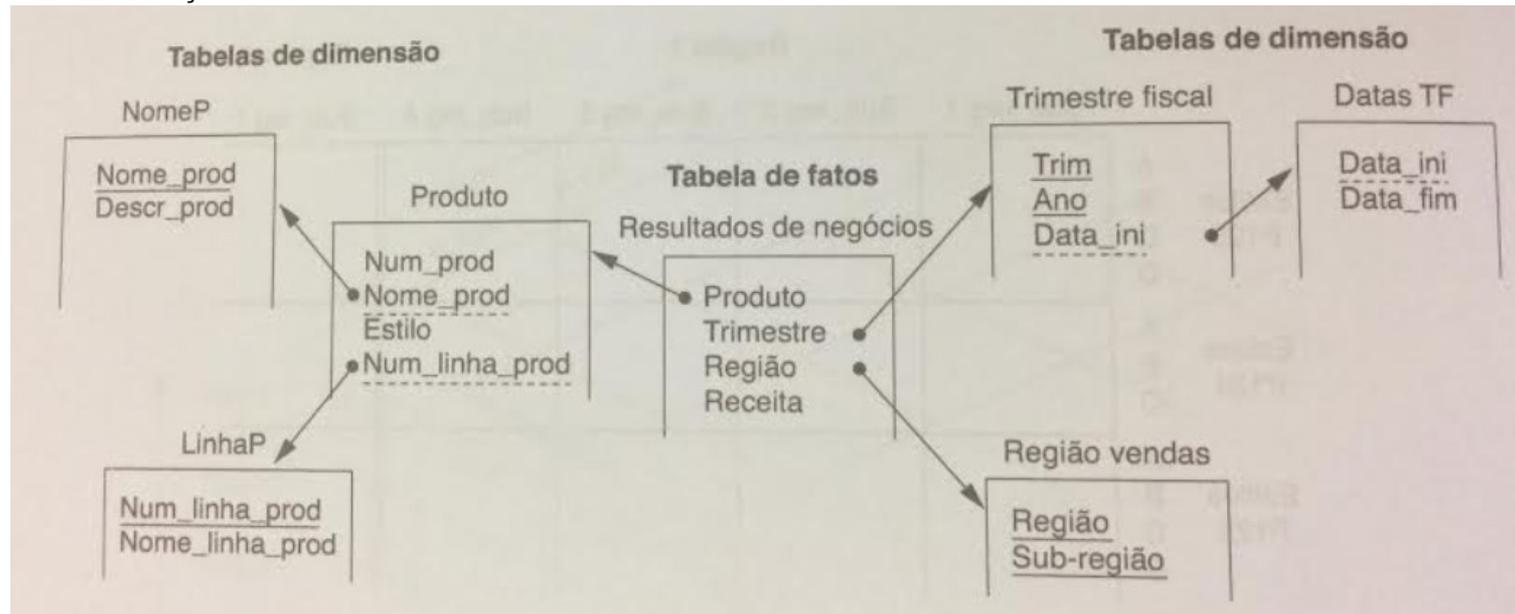
ELMASRI, Ramez; NAVATHE, Shamkant. Database systems. Boston, MA: Pearson Education, 2011.

DATA WAREHOUSE ESQUEMAS MULTIDIMENSIONAIS

Os esquemas multidimensionais comumente variam entre Esquema Estrela e Esquema Floco de Neve.

Esquema Floco de Neve:

- As tabelas de dimensões são normalizadas até a 3FN, gerando uma hierarquia de relações.



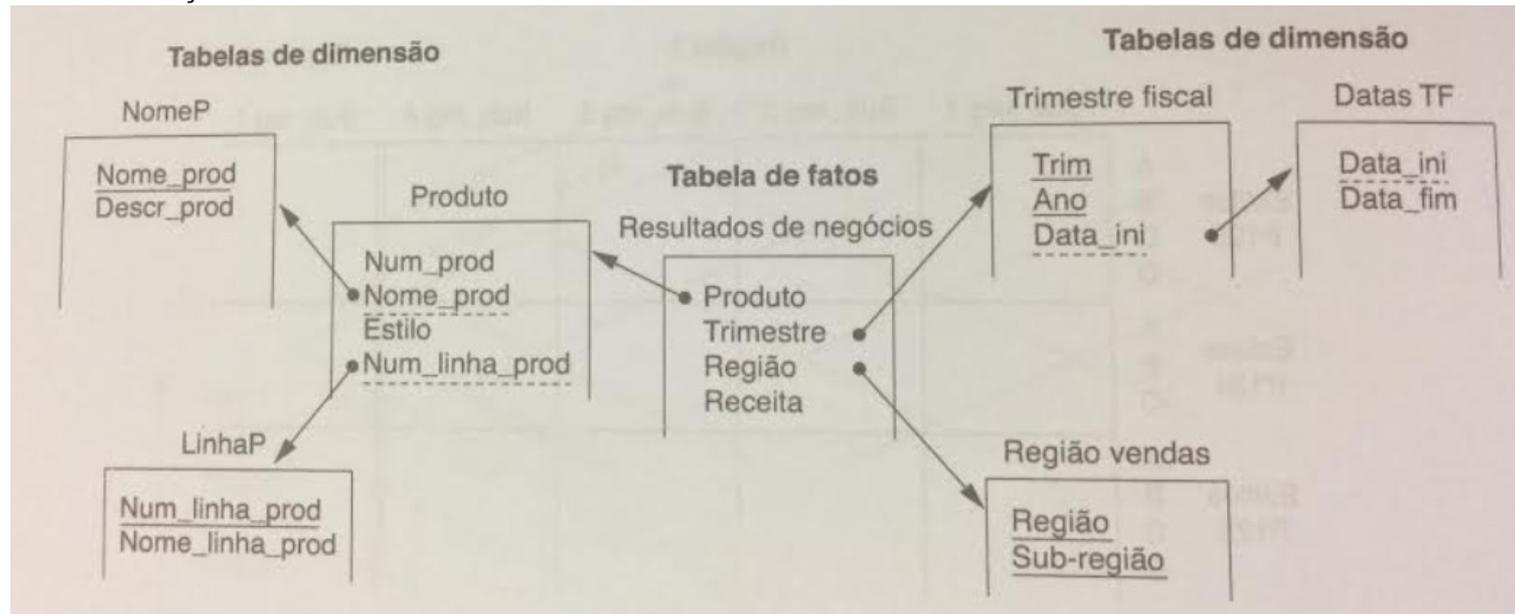
ELMASRI, Ramez; NAVATHE, Shamkant. Database systems. Boston, MA: Pearson Education, 2011.

DATA WAREHOUSE ESQUEMAS MULTIDIMENSIONAIS

Os esquemas multidimensionais comumente variam entre Esquema Estrela e Esquema Floco de Neve.

Esquema Floco de Neve:

- As tabelas de dimensões são normalizadas até a 3FN, gerando uma hierarquia de relações.

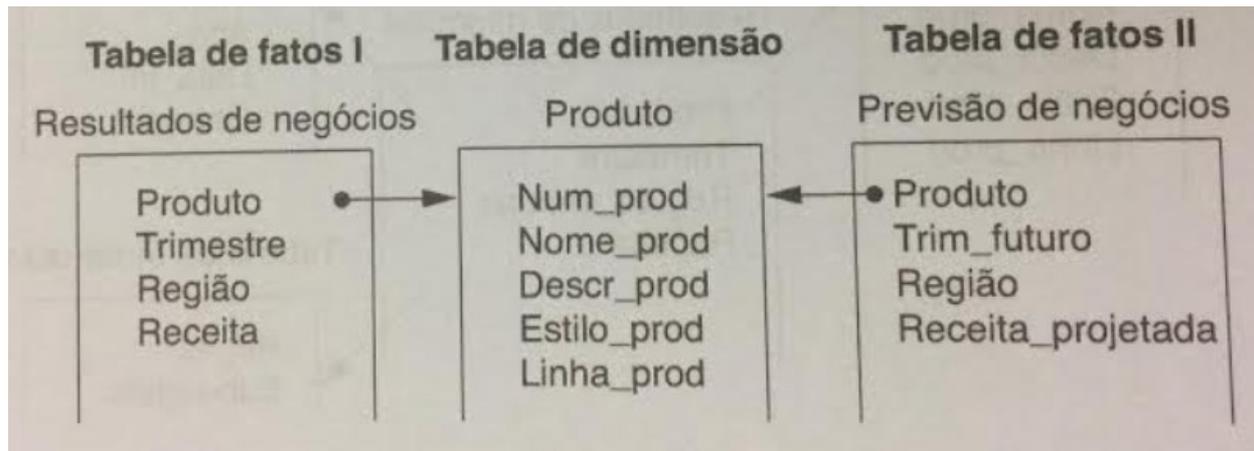


ELMASRI, Ramez; NAVATHE, Shamkant. Database systems. Boston, MA: Pearson Education, 2011.

DATA WAREHOUSE CONSTELAÇÃO DE FATOS

Ocorre quando mais de uma tabela de fato acessa uma mesma dimensão.

Uma constelação de fatos é um conjunto de tabelas de fatos que compartilham algumas tabelas de dimensão.



DÚVIDAS?

