

Arquitetura de Dados Analíticos

Modelagem Dimensional e o Esquema Estrela




Professor

Gustavo Dias

Mestre em Gestão e Desenvolvimento Regional

Consultor em Business Intelligence, com atuação em projetos na área da saúde e varejo.

Contatos:

 p.gustavodias@gmail.com

 <https://www.linkedin.com/in/pgustavodias/>

 pgustavodias



Agenda de aula

- Conceitos chave;
- Modelagem Dimensional;
- Esquema Estrela.

Conhecimentos desenvolvidos



Conceitos Chave

DW

Coleção de dados orientados por assuntos, integrada, variante no tempo e não volátil que tem por objetivo dar suporte ao processo de tomada de decisão.



Datawarehouse



Fundamental para uma estratégia eficiente:

- Entender do negócio;
- Processo ETL bem estruturado;
- Monitoramento constante.

Conceitos Chave

Arquiteturas:

OLTP: Sistemas do ambiente operacional cujo objetivo é executar o maior número de transações possíveis no menor tempo de processamento.

OLAP: Sistemas que permitem aos usuários de alto nível navegarem entre os dados da empresa, proporcionando uma visão multidimensional.



O papel dos SGDB's



OLTP: características que atendam as altas demandas transacionais;

OLAP: características que atendam a performance analítica.

Conceitos Chave

Modelo de Dados:

Modelo Relacional e Modelo Dimensional.

Medidas e Dimensões:

Visão quantitativa e qualitativa dos dados.



.modelagem dimensional

Modelagem Dimensional

Diferente do modelo relacional ao qual estamos habituados, o modelo dimensional surgiu para atender sistemas de processamento analítico.

Tendo em vista esse objetivo, os dados fragmentados armazenados em modelos relacionais trazem uma considerável complexidade à formulação de uma consulta por um usuário final.

O excesso de tabelas e de relacionamentos tornam as consultas mais lentas e inviabilizam uma abordagem analítica.

Modelagem Dimensional

A partir da modelagem dimensional é possível **modelar logicamente dados** para melhorar o desempenho de consultas e promover facilidade de utilização.

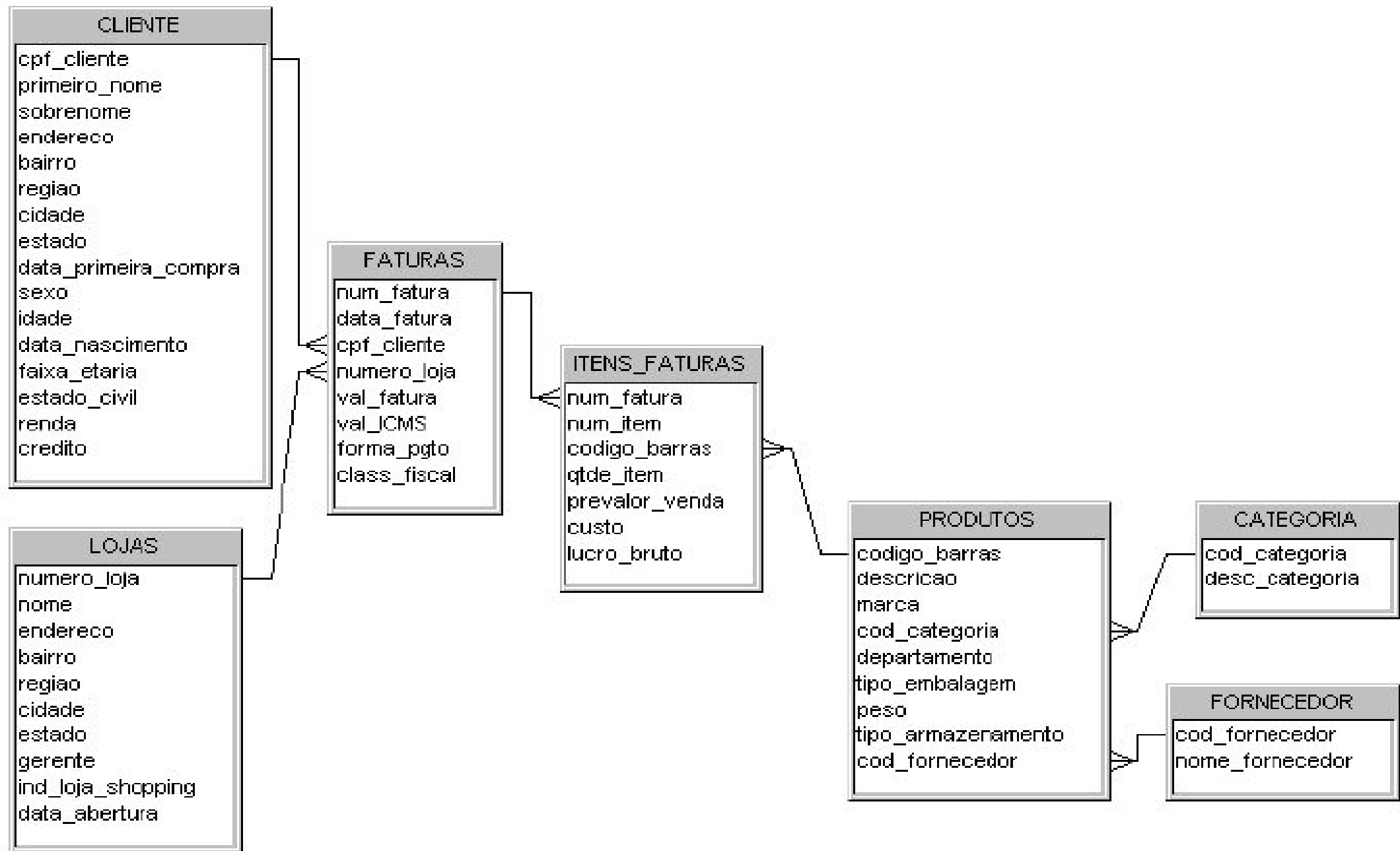
Eles são **compreensíveis, previsíveis, ampliáveis e resistentes** aos ataques de grupos de usuários de negócio.

Modelagem Dimensional

Exemplificando...

MR para o sistema de vendas de uma rede de lojas de departamentos.

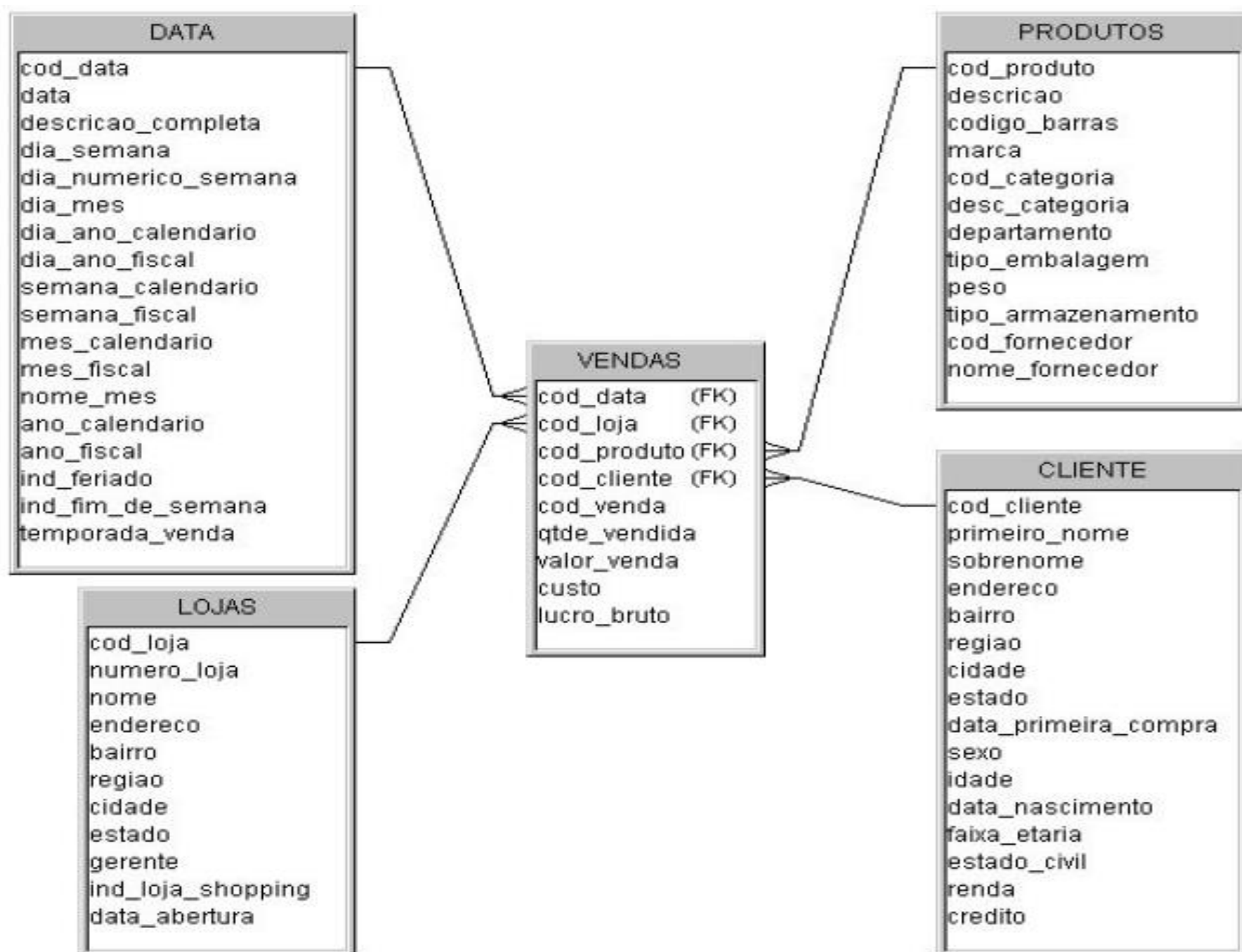
Cada venda ocorrida representa uma linha na tabela Faturas e n linhas na tabela Itens_Fatura.



Modelagem Dimensional

Exemplificando...

MD desse sistema, preparado para atender necessidades de análises das vendas, a partir do cruzamento de informações das lojas da rede, com produtos vendidos, clientes e data da venda.



.esquema estrela

Visão Geral do Modelo

O nome “estrela” se dá pela disposição das tabelas no modelo, que consiste em:

Tabela de Fatos: centro do modelo;

Tabela de Dimensões: aspectos do fato.

Características



Estrutura simples; poucas tabelas; relacionamentos bem definidos; mais próximo do negócio; leitura facilitada para usuários finais.

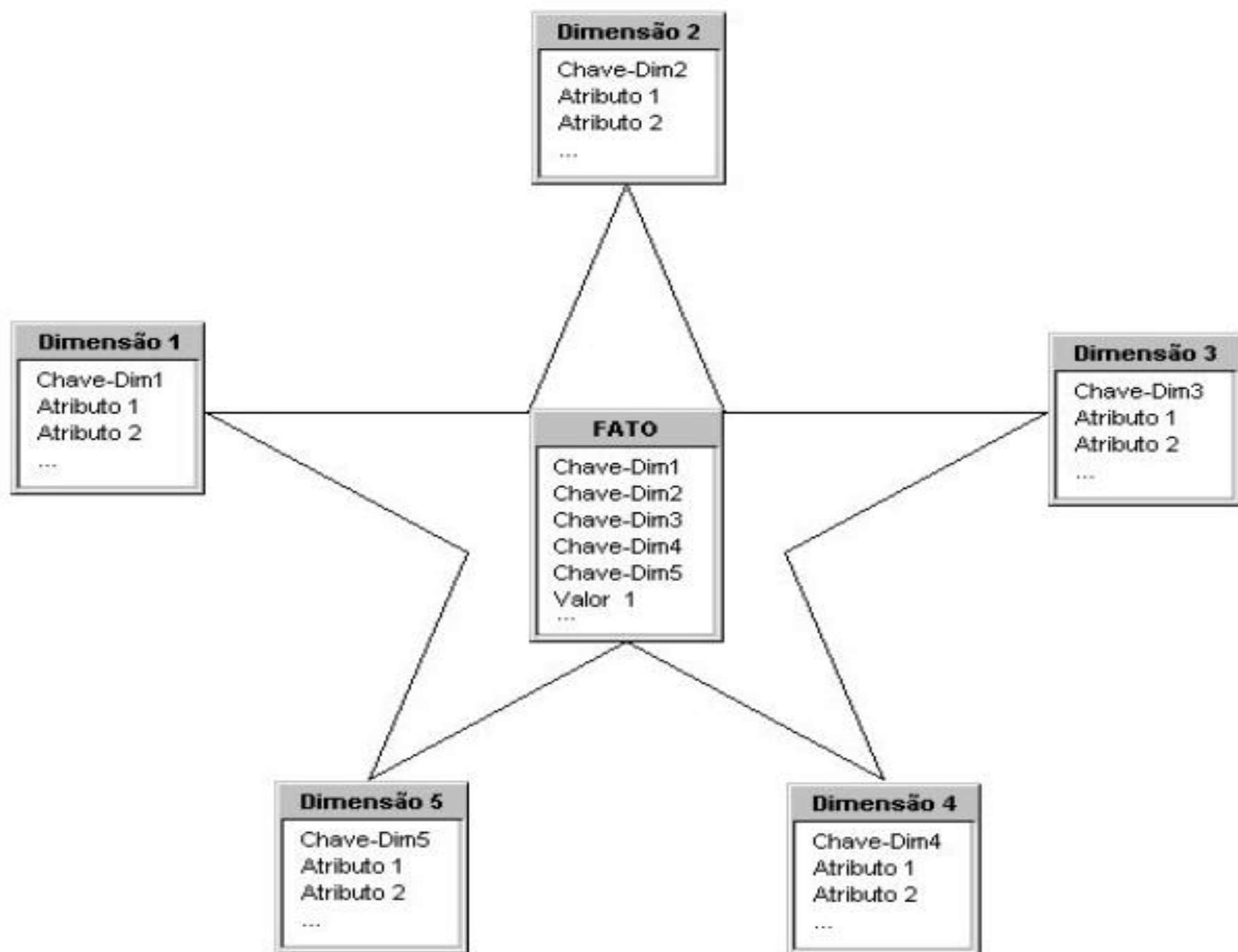


Tabela de Fatos

Principal tabela de um modelo dimensional, onde as medidas numéricas de interesse da empresa estão armazenadas.

Tem como objetivo registrar os fatos que estão sendo analisados, mas sem misturar assuntos em uma mesma tabela, ou seja, deve representar uma unidade do processo do negócio.

Tabela de Fatos

Questões que ajudam a modelar tabelas de fatos:

1. Que **processo** estamos modelando?
2. O que usamos para **medir** este processo?
3. Quais os **indicadores críticos de desempenho** desse processo?

VENDAS	
cod_data	(FK)
cod_loja	(FK)
cod_produto	(FK)
cod_cliente	(FK)
cod_venda	
qtde_vendida	
valor_venda	
custo	
lucro_bruto	

Tabela de Fatos

Existem tipos distintos de tabelas fato a se conhecer:

- a) **Transacional (Transaction Fact Table)**
 - Armazena registros detalhados das transações;
 - Cada linha representa uma transação ou evento.
- b) **Agregada (Aggregate Fact Table)**
 - Armazena dados agregados ou resumos das transações;
 - As métricas são agregadas por uma ou mais dimensões.
- c) **Consolidada (Consolidated Fact Table)**
 - Armazena dados consolidados de múltiplas fontes.
- d) **Fotografia (Snapshot Fact Table)**
 - Captura o estado dos dados em um determinado momento do tempo.

Dimensões

Contém **descrições textuais** do negócio e possui informações para análise qualitativa.

São **aspectos** pelos quais se pretende observar as métricas relativas ao processo que está sendo modelado.

A qualidade do banco é proporcional à qualidade dos atributos de dimensões.

Cada dimensão é definida com uma única **chave primária**.

Essa chave é a base da **integridade referencial** no relacionamento com a tabela fato.

Dimensões

Uma dimensão é estruturada em forma de **hierarquia**, permitindo uma visão escalável e organizada dos dados.

Exemplo:

Dimensão: Data

Hierarquia: Ano > Mês > Dia

Papel das Hierarquias:

1. **Organização estruturada e ordenada;**
2. **Agregação e Drill-Down;**
3. **Navegação facilitada;**
4. **Análise comparativa.**

.exercitando

Fato ou Dimensão

Contexto de uma rede de supermercados.

Tabela que guarda os dados das compras realizadas.



Fato ou Dimensão

Contexto de uma rede de supermercados.

Tabela que guarda os dados das movimentações de estoque.



Fato ou Dimensão

Contexto de uma rede de supermercados.

Tabela que guarda os dados dos
fornecedores.



Fato ou Dimensão

Contexto de uma rede de supermercados.

Tabela que guarda os dados dos produtos.



Fato ou Dimensão

Contexto de uma rede de supermercados.

Tabela que guarda os dados das
pagamentos dos funcionários.



Fato ou Dimensão

Contexto de uma rede de supermercados.

Tabela que guarda os dados dos funcionários.



Referências

INMON, William H. Building the Data Warehouse. 4. ed. Hoboken: John Wiley & Sons, 2005.

KIMBALL, Ralph; ROSS, Margy. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 3. ed. Hoboken: John Wiley & Sons, 2013.

LAUDON, Kenneth C.; LAUDON, Jane P. Sistemas de informação gerenciais. 11. ed. São Paulo: Pearson, 2014.

RIBEIRO, Fernanda; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Dados, informação e conhecimento: conceitos e relações. São Paulo: Cultura Acadêmica, 2020.

TORINO, Emanuelle. Arquitetura de dados no contexto da Ciência da Informação. 2023. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista (UNESP), Marília, 2023.

Referências

Aplicação:

- Inmon (2005) → fundamentação clássica de Data Warehouse corporativo, com foco em integração, orientação por assunto, variabilidade temporal e não volatilidade dos dados.
- Kimball & Ross (2013) → base conceitual e prática para modelagem dimensional (fatos e dimensões), amplamente utilizada em projetos de DW modernos e alinhada ao material anexo trabalhado.
- Ribeiro & Santos (2020) → sustentação teórica da evolução do dado para informação e conhecimento, contribuindo para a compreensão do valor analítico dos ambientes de DW, Data Lake e Lakehouse.
- Torino (2023) → fundamentação contemporânea sobre arquitetura de dados, metadados, governança e integração de repositórios analíticos, conectando DW, Data Lake e Lakehouse em uma visão arquitetural atual.

Obrigado!

