

Trabalho 1 de Banco de Dados

Estefany Licinha Mendes

Luiz Henrique Ribeiro

Samuel Davi Chagas

1. Modelagem do Projeto e Normalização

O processo de modelagem seguiu uma abordagem *bottom-up*, partindo da análise do dataset de entrada e das relações implícitas entre os atributos. O objetivo foi garantir que todas as tabelas estivessem na Terceira Forma Normal (3FN), eliminando redundâncias e assegurando a integridade dos dados.

As principais etapas da normalização foram:

- **1FN (Primeira Forma Normal):**

Todos os atributos são atômicos, ou seja, não existem atributos multivalorados ou compostos em uma mesma célula. Por exemplo, o campo multivalorado similar do dataset original foi decomposto em uma tabela própria (Similar), evitando listas dentro de atributos.

- **2FN (Segunda Forma Normal):**

Foram eliminadas dependências parciais. Cada atributo não-chave depende da chave primária completa, e não apenas de parte dela. Um exemplo é a tabela Produto_categoria, o relacionamento N:N é representado por chave composta (id_produto, id_categoria), de forma que cada associação depende do par completo e não de apenas uma das colunas.

- **3FN (Terceira Forma Normal):**

Foram eliminadas dependências transitivas, garantindo que atributos não-chave não dependam de outros atributos não-chave. Como exemplo temos o atributo grupo, que foi mantido em Produto porque depende apenas de id_produto e não de outro atributo derivado.

Assim, cada tabela representa um conceito único:

- **Produto:** informações centrais dos produtos.
- **Similar:** relações de similaridade entre produtos.
- **Categoria:** lista única de categorias.
- **Produto_categoria:** associação N:N entre produtos e categorias.
- **Avaliacao:** informações sobre reviews de clientes.
- **Cliente:** identificação dos avaliadores.

2. Dicionário de dados

2.1 Tabela Produto

Descrição: Armazena os produtos da empresa, incluindo os identificadores internos, os identificadores da Amazon(ASIN) e outras informações.

Atributo	Domínio	Descrição	Restrições
id_produto	INT	Identificador do produto	UNIQUE, NOT NULL
asin	CHAR(10)	Código de identificação padrão único fornecido pela Amazon	PK
nome_produto	TEXT	Nome do produto	NOT NULL
grupo	VARCHAR(50)	Grupo que o produto pertence	NOT NULL
posicao_ranking	INTEGER	Posição do produto no ranking da Amazon	-

Quadro 1 - Atributos da tabela Produto e suas respectivas informações.

Justificativa da modelagem:

- O atributo grupo não foi decomposto em uma tabela separada porque depende apenas de id_produto e não há violação das formas normais.

2.2 Tabela Similar

Descrição: Modela a relação de similaridade entre produtos, de acordo com o ASIN.

Atributo	Domínio	Descrição	Restrições
id_asin	CHAR(10)	ID do produto de referência	PK (composta) , FK, NOT NULL
id_asin_similar	CHAR(10)	ID do produto similar	PK (composta) , FK, NOT NULL

Quadro 2 - Atributos da tabela Similar e suas respectivas informações.

Justificativa da modelagem:

- A escolha de usar id_asin em vez de id_produto preserva a lógica do dataset de entrada, que relaciona produtos por ASIN;
- Esta tabela foi criada para normalizar o campo multivalorado “similar” do dataset;
- A PK composta vai evitar duplicação de pares de similaridade.

Regras de integridade:

- FOREIGN KEY (id_asin) → Produto(asin);
- FOREIGN KEY (id_asin_similar) → Produto(asin);
- Em ambos atributos foi usado ON DELETE CASCADE, garantindo exclusão consistente.

2.3 Tabela Categoria

Descrição: Contém a lista de categorias de produtos.

Atributo	Domínio	Descrição	Restrições
id_categoria	SERIAL	ID da categoria	PK
nome_categoria	TEXT	Nome da categoria	UNIQUE, NOT NULL

Quadro 3 - Atributos da tabela Categoria e suas respectivas informações.

Justificativa de modelagem:

- De acordo com o dataset, um produto pode pertencer a várias categorias. Assim, a decomposição foi necessária para evitar redundância;
- nome_categoria foi definido como atributo único, garantindo integridade sem duplicatas.

2.4 Tabela Produto_categoria

Descrição: Relaciona produtos com categorias (N:N).

Atributo	Domínio	Descrição	Restrições
id_produto	CHAR(10)	ID do produto	PK (composta), FK, NOT NULL
id_categoria	INT	ID da categoria	PK (composta), FK, NOT NULL

Quadro 4 - Atributos da tabela Produto_categoria e suas respectivas informações.

Justificativa de modelagem:

- A relação de muitos-para-muitos entre produtos e categorias exigiu a criação desta tabela;
- A PK composta vai garantir que um produto não seja relacionado repetidamente à mesma categoria.

Regras de integridade:

- FOREIGN KEY (id_produto) → Produto(asin);
- FOREIGN KEY (id_categoria) → Categoria(id_categoria);
- Em ambos atributos foi usado ON DELETE CASCADE, garantindo exclusão consistente.

2.5 Tabela Avaliacao

Descrição: Armazena as avaliações feitas por clientes sobre produtos.

Atributo	Domínio	Descrição	Restrições
id_review	SERIAL	ID da avaliação	PK
id_produto	CHAR(10)	ID do produto avaliado	FK, NOT NULL
id_cliente	VARCHAR(70)	ID do cliente que realizou a avaliação	FK, NOT NULL
data	DATE	Data da avaliação	NOT NULL
hora	TIME	Hora da avaliação	NOT NULL
rating	INTEGER	Nota dada ao produto	CHECK (rating BETWEEN 1 AND 5), NOT NULL
votos	INTEGER	Votos totais da avaliação	NOT NULL
helpful	INTEGER	Votos de utilidade	NOT NULL

Quadro 5 - Atributos da tabela Avaliacao e suas respectivas informações.

Justificativa de modelagem:

- A criação desta tabela ocorreu para evitar que informações de avaliações ficassem dentro de Produto, pois isso causaria multivalores.

Regras de integridade:

- FOREIGN KEY (id_produto) → Produto(asin);

- FOREIGN KEY (id_cliente) → Cliente(id_cliente);
- Em ambos atributos foi usado ON DELETE CASCADE, garantindo exclusão consistente.
- Avaliação só existe se houver produto e cliente válidos.

2.6 Tabela Cliente

Descrição: Armazena clientes que fazem avaliações.

Atributo	Domínio	Descrição	Restrições
id_cliente	VARCHAR(70)	ID do cliente	PK

Quadro 6 - Atributos da tabela Cliente e suas respectivas informações.

Justificativa da decomposição:

- O arquivo de entrada registra avaliações associadas a identificadores de clientes. Assim, criamos a tabela Cliente para manter a consistência.

3. Diagrama do esquema relacional

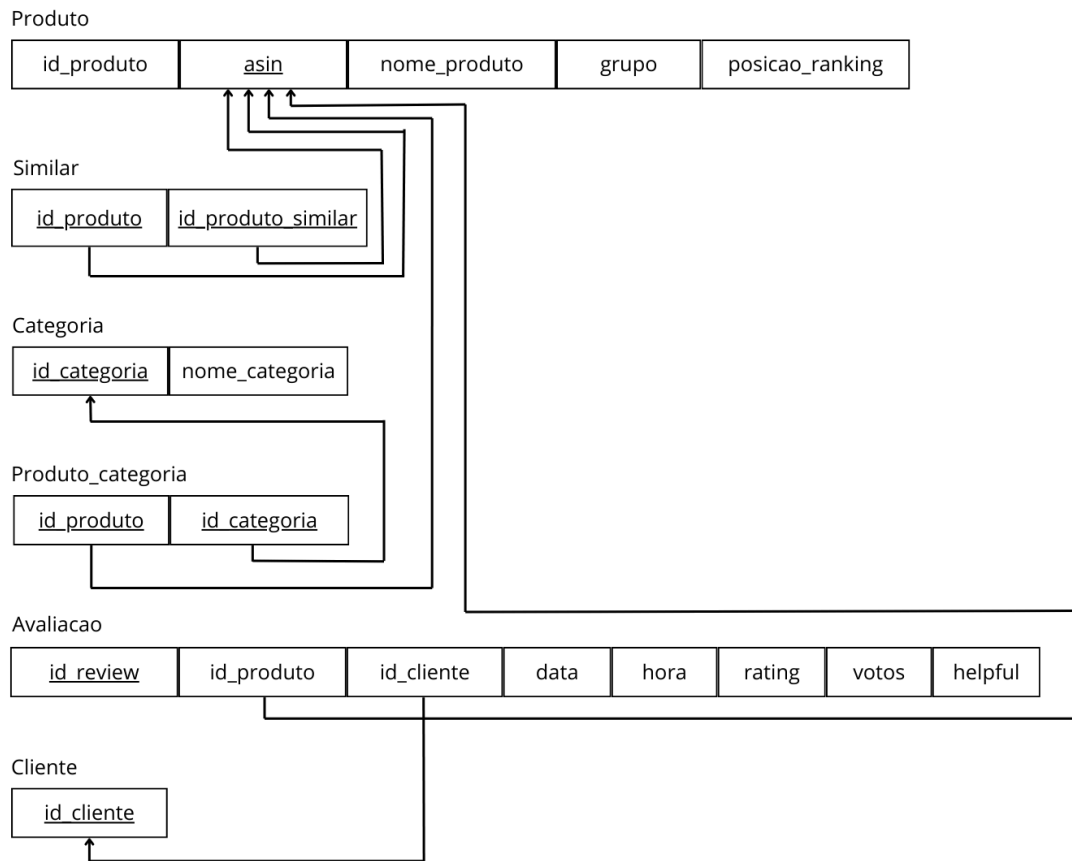


Figura 1 - Esquema Relacional