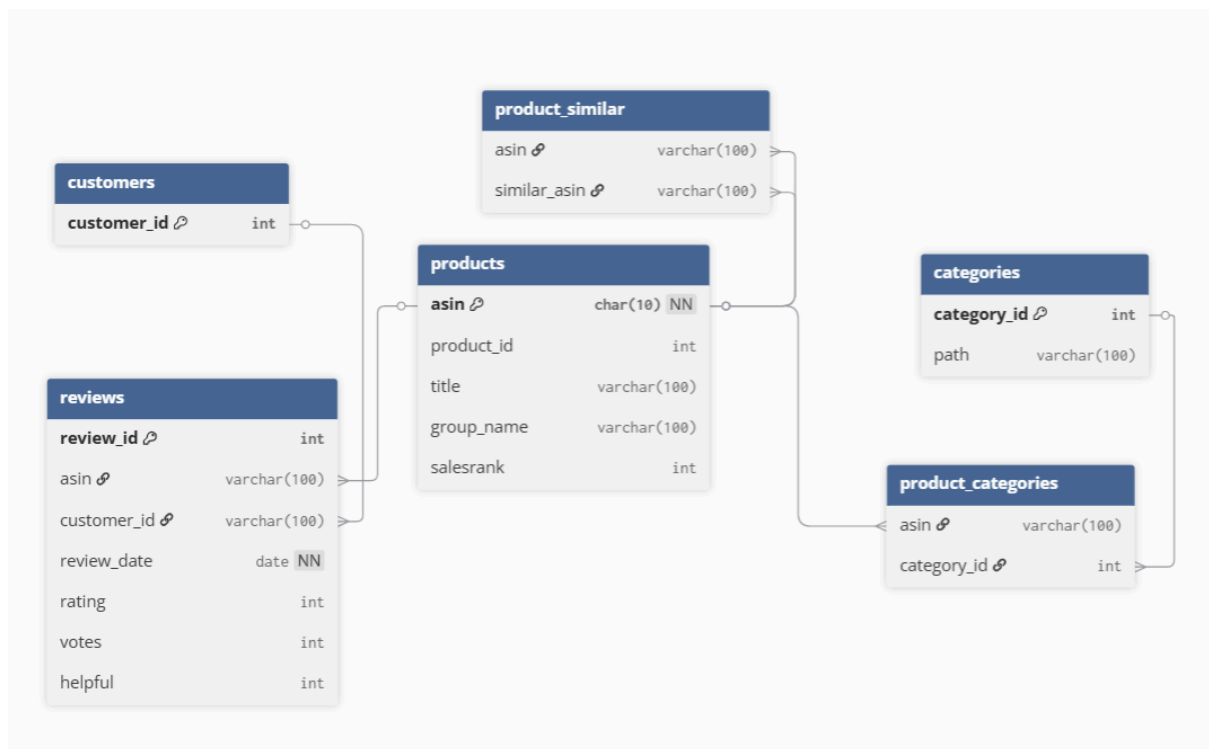


Documentação do Trabalho Prático 1 - Banco de Dados

Alunos: Guilherme Kaue, Kaio
Carvalho, Washington

1. Diagrama do banco de dados relacional

Diagrama correspondente ao esquema do banco de dados relacional, feito no dbdiagram.io



2. Dicionário de dados

Dicionário de dados feito para cada tabela do esquema, contendo descrição de cada relação, atributos, chaves e restrições de integridade.

Tabela	Atributo	Tipo	Descrição	Restrições
Products	product_id	int	Identificador do produto	
	ASIN	CHAR(10)	Amazon standard identification number	PK
	title	VARCHAR(100)	Título do produto	
	group_name	VARCHAR(100)	Nome do grupo do produto (Book, DVD, Video, Music)	
	salesrank	int	Rank da venda do produto	

Tabela	Atributo	Tipo	Descrição	Restrições
Product_similar	ASIN	CHAR(10)	Referência ao ASIN principal	PK, FK
	Similar_ASIN	CHAR(10)	Referência ao ASIN similar	PK, FK

Tabela	Atributo	Tipo	Descrição	Restrições
Categories	category_id	int	Identificador da categoria	PK
	path	VARCHAR(100)	Caminho de uma categoria	

Tabela	Atributo	Tipo	Descrição	Restrições
Product_categories	ASIN	CHAR(10)	Referência ao ASIN	PK, FK
	category_id	int	Referência à categoria do produto	PK, FK

Tabela	Atributo	Tipo	Descrição	Restrições
Customers	customer_id	int	Identificador do cliente	PK

Tabela	Atributo	Tipo	Descrição	Restrições
Reviews	review_id	int	Identificador do review	PK
	ASIN	CHAR(10)	Referência ao ASIN do produto	PK, FK
	customer_id	int	Referência ao id do cliente	FK
	review_date	DATE	Data da avaliação	
	rating	int	Nota dada pelo cliente (1 a 5)	
	votes	int	Votos da avaliação	
	helpful	int	Votos úteis da avaliação	

3. Observância das formas normais e justificativas

Neste banco de dados relacional, utilizamos a 3FN como forma normal de alto nível, pois assim evitamos redundância e inconsistência nos dados, além de melhorar a organização do banco.

Abaixo segue o conjunto de tabelas proposto. Depois de cada definição estão as justificativas das escolhas feitas.

Products

- **Atributos**

- asin CHAR(10) — **PK**
- product_id INTEGER — id numérico do arquivo (mantido como atributo informativo, pode ser NULL)
- title VARCHAR(100)
- group_name VARCHAR(100)
- salesrank INTEGER

- **Justificativas**

- O ASIN é o identificador externo usado em todas as linhas de similaridade e reviews no arquivo original; portanto é o identificador natural para joins com outras tabelas. O ASIN possui exatamente 10 caracteres.
- O product_id é mantido para rastreabilidade (é o "ID" do arquivo) mas não vira PK porque as relações externas referenciam ASIN.
- Adotamos o VARCHAR(100) para qualquer atributo que possua texto, pois é uma boa capacidade de caracteres para garantir que nunca falte espaço.

Product_similar

- **Atributos**

- asin CHAR(10) — PK, FK → products.asin
- similar_asin CHAR(10) — PK, FK → products.asin

- **Justificativas**

- Representa uma relação *produto* → *produto similar* (lista obtida na linha similar: do arquivo). Cada linha do arquivo lista varios ASINs pareados ao produto corrente; esse formato é naturalmente modelado como uma tabela de pares.
- A modelagem mantém o papel de cada coluna: asin = produto de referência; similar_asin = produto listado como similar.
- PK composta evita possível duplicidade do mesmo par.

Categories

- **Atributos**

- category_id INTEGER — PK
- path VARCHAR(100) — String representando o caminho de uma categoria

- **Justificativas**

- O arquivo apresenta *caminhos* de categorias (vários caminhos por produto). Cada caminho é tratado como uma entidade (um único path) para evitar repetir longas strings em product_categories.

Product_categories

- **Atributos**

- asin TEXT — PK, FK → products.asin
- category_id INTEGER — PK, FK → categories.category_id

- **Justificativa**

- Relação N:N entre produtos e caminhos de categorias. O arquivo mostra que um produto pode ter múltiplos caminhos; cada combinação vira uma linha nessa tabela.

Customers

- **Atributos**

- customer_id TEXT — PK

- **Justificativa**

- Os reviews referenciam clientes por um identificador textual. Separar customers evita repetição de strings de cliente e permite agregações por autor.

Reviews

- **Atributos**

- review_id INTEGER — PK
- asin CHAR(10) — PK, FK → products.asin
- customer_id TEXT — FK → customers.customer_id
- review_date DATE
- rating INTEGER
- votes INTEGER
- helpful INTEGER

- **Justificativas**

- Não existe um identificador numérico para cada review no arquivo, então usamos review_id serial como chave artificial para simplificar manipulação e permitir múltiplas reviews do mesmo cliente para o mesmo produto em datas diferentes.
- Usamos uma PK composta para garantir que cada review seja separada uma da outra.
- O rating deve variar de 1 a 5.