

Trabalho de Banco de Dados

**Pedro Miguel Pinto Botelho, Ricardo Augusto Silva Bonfim,
Rômulo José Pereira Da Costa Junior**

¹Instituto de Computação – Universidade Federal do Amazonas(UFAM)
Av. Gen. Rodrigo Octávio,6200, Coroado I,Setor Norte do Campus–69080–900

{pedro.botelho, ricardo.bonfim, romulo.junior}@icomp.ufam.edu.br

1. Introdução

Neste projeto, seguimos o Modelo Relacional, uma estrutura que organiza os dados em tabelas relacionadas para facilitar a gestão e integridade das informações. Adotamos a abordagem ascendente bottom-up para construir o banco de dados, começando com elementos menores e integrando-os progressivamente em componentes maiores e mais complexos, permitindo uma evolução gradual e detalhada do sistema. Optamos pela Forma Normal de Boyce-Codd para o esquema de nossas tabelas. Utilizamos a linguagem Python para desenvolver os códigos, os quais têm acesso direto ao SGBD PostgreSQL através de comandos SQL. A figura 1 mostra um exemplo da tabela antes de ser normalizada:

ID	ASIN	TITLE	GROUP	SALESRANK	SIMILAR	CATEGORIES	REVIEWS
1	0827229534	Patterns of Preaching: A Sermon Sampler	Book	396585	0804215715 156101074X 0687023955 0687074231 082721619X	[Books[283155] Subjects[1000] Religion & Spirituality[22] Christianity[12290] Clergy[12360] Preaching[12368] [Books[283155] Subjects[1000] Religion & Spirituality[22] Christianity[12290] Clergy[12360] Sermons[12370]	2000-7-28 customer: A2JW67OY8U6HHK rating: 5 votes: 10 helpful: 9 2003-12-14 customer: A2VE83MZFP98ITY rating: 5 votes: 6 helpful: 5
2	0738700797	Candlemas: Feast of Flames	Book	168596	0738700827 1567184960 1567182836 0738700525 0738700940	[Books[283155] Subjects[1000] Religion & Spirituality[22] Earth-Based Religions[12472] Wicca[12484] [Books[283155] Subjects[1000] Religion & Spirituality[22] Earth-Based Religions[12472] Witchcraft[12486]	2001-12-16 customer: A11NCO6YTE4BTJ rating: 5 votes: 5 helpful: 4 2002-1-7 customer: A9CQ3PLRNIR83 rating: 4 votes: 5 helpful: 5

Figura 1. Exemplo de tabela sem normalização.

2. Diagrama

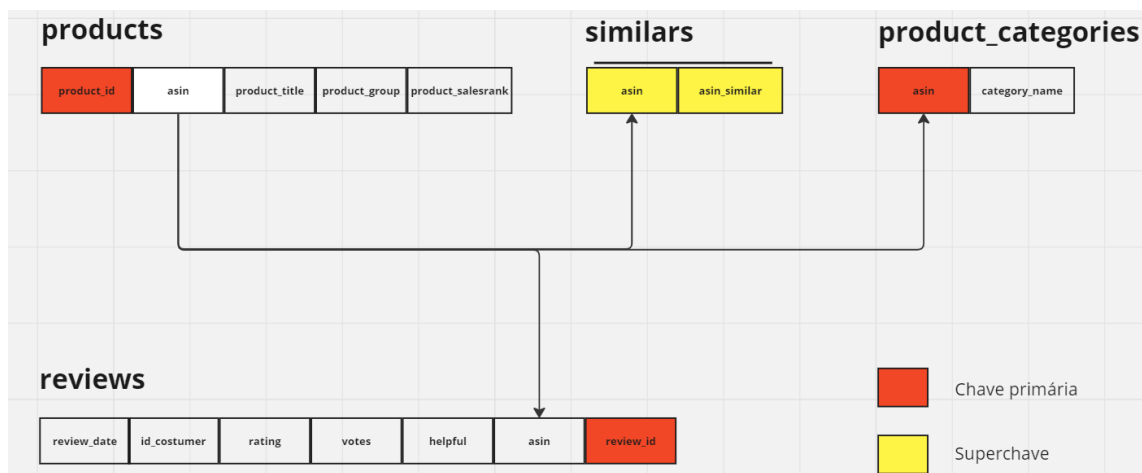


Figura 2. Diagrama do banco de dados.

3. Dicionário

3.1. Tabela 'products'

- **product_id** (inteiro, chave primária, não nulo): Identificador único e autoincrementado para cada produto na tabela.
- **asin** (char(10), chave única, não nulo): Código do produto no catálogo da Amazon, utilizado como identificador único.
- **product_title** (varchar(452)): Nome do produto com tamanho suficiente para armazenar o maior título da base de dados.
- **product_group** (varchar(13)): Grupo ao qual o produto pertence.
- **product_salesrank** (inteiro): Posição do produto no rank de vendas.

3.2. Tabela 'similars'

- **asin** (char(10), chave estrangeira referenciando products(asin), não nulo): Código do produto no catálogo da Amazon, servindo como chave estrangeira referenciando a tabela 'products'.
- **asin_similar** (char(10), não nulo): Código do produto similar ao produto referenciado pela coluna 'asin'.

3.3. Tabela 'reviews'

- **review_date** (data, não nulo): Data da review do produto.
- **id_customer** (varchar(18), não nulo): Identificador do cliente que fez a review.
- **rating** (inteiro, não nulo): Classificação atribuída à review.
- **votes** (inteiro, não nulo): Número de votos favoráveis para a review.
- **helpful** (inteiro, não nulo): Número de votos úteis para a review.
- **asin** (char(10), chave estrangeira referenciando products(asin), não nulo): Código do produto no catálogo da Amazon, servindo como chave estrangeira referenciando a tabela 'products'.
- **review_id** (inteiro, chave primária, auto-incremental, chave única, não nulo): Identificador único e autoincrementado para cada review na tabela.

3.4. Tabela 'product_categories'

- **asin** (char(10), chave estrangeira referenciando products(asin), não nulo): Código do produto no catálogo da Amazon, servindo como chave estrangeira referenciando a tabela 'products'.
- **category_name** (varchar(2500), não nulo): Nome da categoria à qual o produto pertence, a sequencia de caracteres separados por 'l' pode ser tratada como uma árvore e não possui valor exclusivo por produto.

4. Observação

A saída do código estará em um arquivo chamado "saida_tp1_3.3_DiaMesAno_HoraMinutoSegundo.txt".