

Relatório BD

Aline de Souza Sahdo e Silva, Carolina Falabelo Maycá,
Fernando Lucas Almeida Nascimento

¹Instituto de Computação (ICOMP) – Universidade Federal do Amazonas (UFAM)

{aline.silva, carolina.mayca, fernando.nascimento}@icomp.ufam.edu.br

1. Modelo Relacional e Justificativas de Projeto

Por já existir um arquivo com os dados prontos, o processo de modelagem seguiu uma abordagem *bottom-up*, na qual foi analisado o formato e a estrutura do arquivo de entrada para identificar atributos e relações. A partir disso, foram definidas as tabelas que compõem o esquema relacional, garantindo que todas as dependências fossem representadas sem redundâncias e que as restrições de integridade estivessem ativas.

As relações obtidas foram as seguintes:

- **Products:** Relação central do banco, contendo todos os produtos do e-commerce. O atributo `ASIN` foi definido como chave primária por ser único no arquivo de entrada. O atributo `Id` foi mantido como identificador numérico interno (único) para fins de possíveis otimizações ou integrações futuras. Atributos como `title`, `group`, `salesrank`, `numSimilar`, `numCategories`, `numReviews`, `downloaded` e `avg_rating` mantêm os domínios de acordo com os tipos presentes no dataset (texto, inteiro ou numérico decimal).
- **Reviews:** Relação que armazena as avaliações feitas pelos clientes. Foi criada como tabela separada para evitar repetições de dados, já que cada produto pode ter diversas avaliações. A relação apresenta uma chave estrangeira (`ASIN`) referenciando `Products(ASIN)`, garantindo que não existam reviews sem produto associado. Há também atributos como `date`, `customer`, `rating`, `votes` e `helpful`.
- **Categories:** Representa a hierarquia de categorias de produtos. O atributo `id_categorie` foi definido como chave primária. Para modelar a hierarquia, o atributo `id_father` é uma chave estrangeira autorreferente para `Categories(id_categorie)`, permitindo a formação de árvores de categorias.
- **Categorie_product:** Tabela de associação que resolve o relacionamento N:M entre produtos e categorias. A chave primária é composta pelo par (`asin`, `id_category`), garantindo que um produto não seja repetido mais de uma vez na mesma categoria. Ambas as colunas são chaves estrangeiras que referenciam, respectivamente, `Products(ASIN)` e `Categories(id_categorie)`.
- **Similar:** Tabela responsável por armazenar a lista de produtos similares de cada produto. Assim como `Categorie_product`, resolve um relacionamento N:M, desta vez entre produtos e seus similares. A chave primária é composta por (`asin_product`, `asin_similar`) e ambas as colunas são chaves estrangeiras que referenciam `Products(ASIN)`.

Na tabela de `Products`, foi utilizado o `ASIN` como chave primária, visto que ele já é o identificador único fornecido no arquivo de entrada e é referenciado em outras tabelas. Em `Reviews`, foi criada uma chave substituta (`id SERIAL`), pois as avaliações não possuem identificador próprio. Para as tabelas `Categorie_product` e `Similar`, foram utilizadas chaves compostas para evitar duplicação e representar corretamente relações N:M.

Foi usado também o `ON DELETE CASCADE` nas chaves estrangeiras para sempre garantir a consistência dos dados, para que quando um produto seja removido, todas as informações dele também sejam removidas.

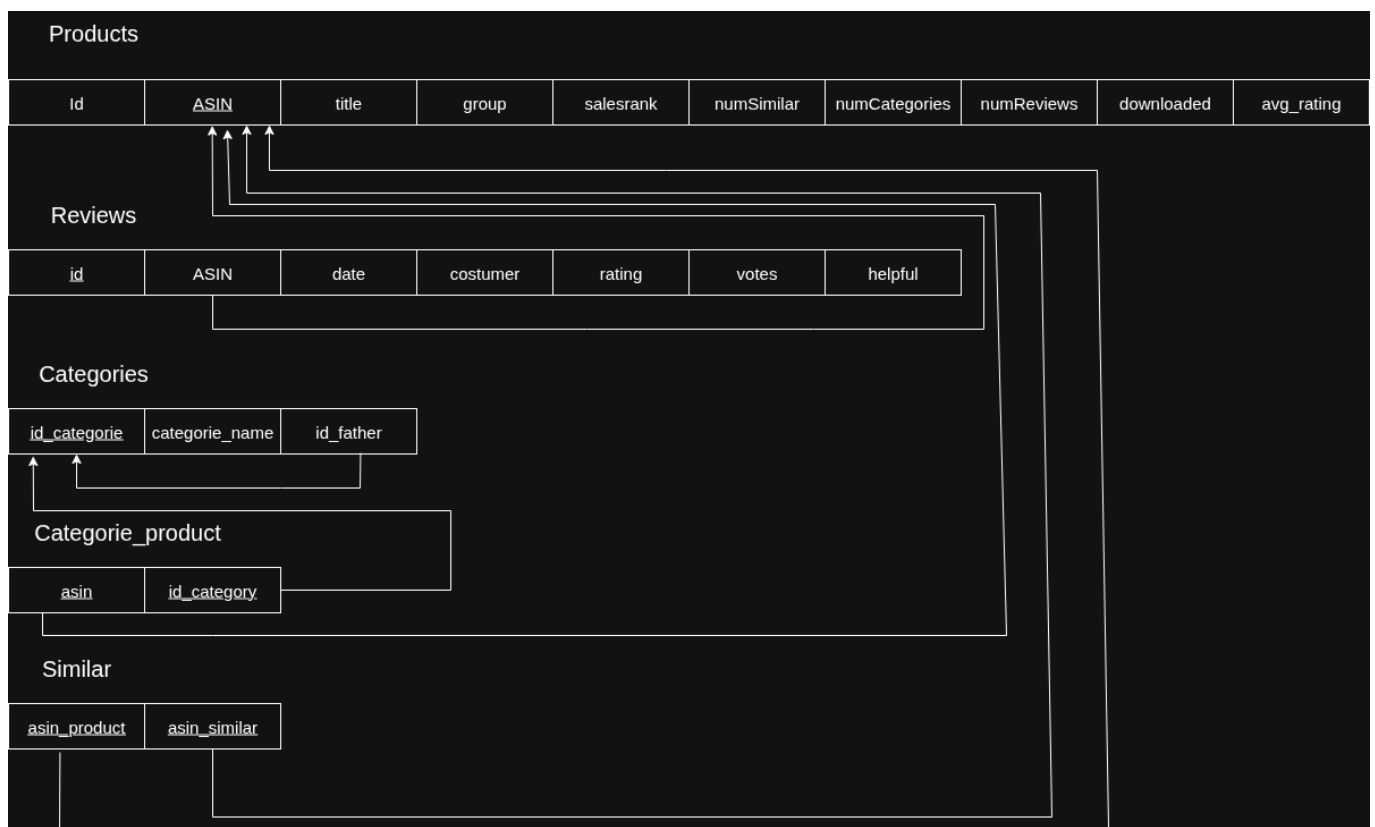
1.1. Normalização

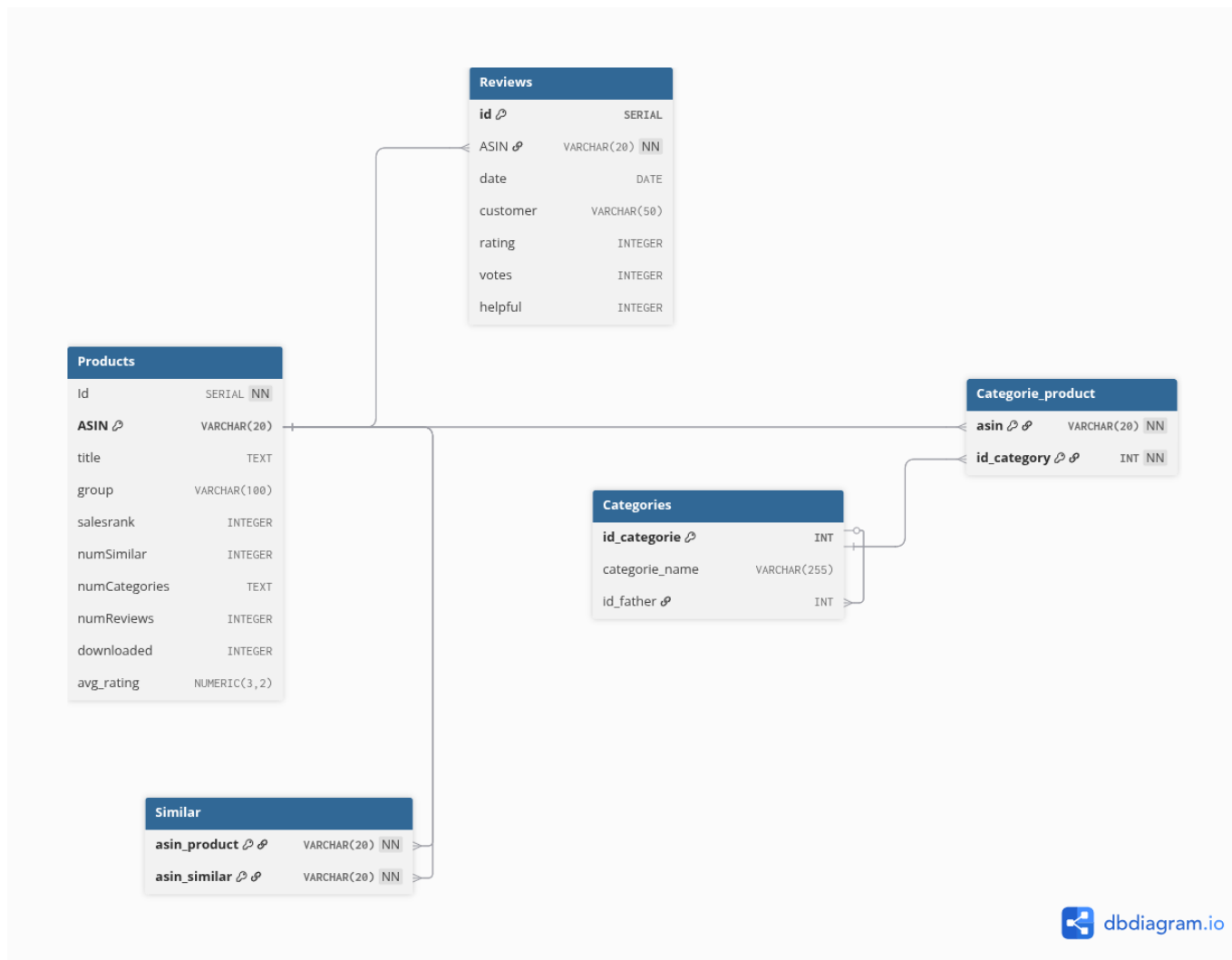
As relações do banco foram feitas observando uma das formas normais de alto nível, que é a **Terceira Forma Normal (3FN)**. Para fazer isso:

- Todas as dependências parciais foram eliminadas (2FN). Ou seja, cada atributo não-chave depende de toda a chave primária e não apenas de parte dela. Um exemplo seria na tabela `Categorie_product`, o atributo de categoria depende de toda a combinação de produto e categoria, evitando redundâncias.
- Todas as dependências transitivas foram eliminadas (3FN). Significando que nenhum atributo não-chave depende de outro atributo não-chave. Por exemplo, atributos derivados ou calculáveis, como `avg_rating`, foram mantidos de forma consistente e isolada para evitar inconsistências.
- Cada tabela representa um único conceito (produto, review, categoria), permitindo fazer as operações de tabela sem afetar dados de outras entidades ou causar inconsistências.

Todas as relações estão na Terceira Forma Normal, uma vez que cada atributo não-chave depende unicamente da chave primária e não existem dependências transitivas.

2. Diagrama do Esquema Relacional





3. Dicionário de Dados

3.1. Tabela Products

Atributo	Tipo	Chave/Restrição	Descrição
ASIN	VARCHAR(20)	PK, UNIQUE	Identificador único do produto, utilizado como chave primária e nas relações com outras tabelas.
Id	SERIAL	UNIQUE, NOT NULL	Identificador numérico interno gerado automaticamente, não usado como chave primária.
title	TEXT	-	Nome ou título do produto.
group	VARCHAR(100)	-	Grupo ou categoria principal do produto (por exemplo: Book, DVD, Music).
salesrank	INTEGER	-	Posição do produto no ranking de vendas.
numSimilar	INTEGER	-	Quantidade de produtos similares.
numCategories	TEXT	-	Número de categorias nas quais o produto aparece.
numReviews	INTEGER	-	Total de reviews recebidas.

downloaded	INTEGER	-	Número de reviews que foram baixadas.
avg_rating	NUMERIC(3,2)	-	Média das avaliações recebidas para o produto.

3.2. Tabela Reviews

Atributo	Tipo	Chave/Restrição	Descrição
id	SERIAL	PK	Identificador único de cada review, criado para diferenciá-los
ASIN	VARCHAR(20)	FK: Products(ASIN) ON DELETE CASCADE	Produto ao qual a review está associada.
date	DATE	-	Data em que a review foi feita.
customer	VARCHAR(50)	-	Cliente que fez a review.
rating	INTEGER	-	Nota atribuída pelo cliente (ex.: 1 a 5).
votes	INTEGER	-	Número de votos recebidos pela review.
helpful	INTEGER	-	Quantidade de votos marcando a review como útil.

3.3. Tabela Categories

Atributo	Tipo	Chave/Restrição	Descrição
id_categorie	INT	PK	Identificador único de cada categoria.
categorie_name	VARCHAR(255)	-	Nome da categoria.
id_father	INT	FK: Categories(id_categorie)	Categoria pai (auto-referência), permitindo hierarquia de categorias.

3.4. Tabela Categorie_product

Atributo	Tipo	Chave/Restrição	Descrição
asin	VARCHAR(20)	PK (composta), FK: Products(ASIN) ON DELETE CASCADE	Produto associado à categoria.
id_category	INT	PK (composta), FK: Categories(id_categorie) ON DELETE CASCADE	Categoria à qual o produto pertence.

3.5. Tabela Similar

Atributo	Tipo	Chave/Restrição	Descrição
asin_product	VARCHAR(20)	PK (com-posta), FK: Products(ASIN) ON DELETE CASCADE	Produto de origem.
asin_similar	VARCHAR(20)	PK (com-posta), FK: Products(ASIN) ON DELETE CASCADE	Produto considerado similar ao produto de origem.