

FORMA NORMAL

Equipe: Lucas Pereira, Roberta Brasil, Mikaelle Costa

Forma normal escolhida: 4FN

Justificativas da escolha:

- **Eliminação da redundância de dados:** A principal vantagem da 4FN é a sua capacidade de eliminar a repetição desnecessária de informações. Isso ocorre ao lidar com dependências multivaloradas, que acontecem quando um único registro em uma tabela pode estar associado a múltiplos valores de dois ou mais atributos independentes. A 4FN resolve isso decompondo a tabela em duas ou mais tabelas, cada uma armazenando um relacionamento multivalorado separado.
- **Prevenção de anomalias de atualização:** Como consequência direta da redução da redundância, a 4FN ajuda a prevenir anomalias de atualização, inserção e exclusão. Em uma tabela que não está na 4FN, a atualização de uma informação pode exigir a modificação de múltiplos registros, aumentando o risco de inconsistências caso alguma atualização falhe. Da mesma forma, a inserção de um novo dado sobre um atributo pode ser impossível sem a existência de outro, e a exclusão de um registro pode levar à perda de informações não relacionadas que estavam armazenadas na mesma linha.
- **Melhora na integridade e manutenção dos dados:** Ao garantir que cada tabela represente um único conceito bem definido, a 4FN torna o esquema do banco de dados mais lógico e fácil de entender. Isso simplifica a manutenção, pois as modificações e consultas se tornam mais diretas e menos propensas a erros. A integridade dos dados é aprimorada, pois as regras de dependência são aplicadas de forma mais rigorosa, garantindo que os dados armazenados sejam consistentes e precisos.
- **Simplificação das queries:** As tabelas representam uma única relação, o que torna o Banco de Dados mais fácil de ser consultado, ao simplificar as queries.

DICIONÁRIO DE DADOS

Tabela: Categories

- **Descrição:** Armazena as informações sobre as categorias de produtos. Cada linha representa uma categoria única.

- **Atributos:**
 - category_id (SERIAL): Identificador numérico único, gerado automaticamente para cada categoria para ficar mais simples de organizar. É a chave primária.
 - category_source_id (INT): O ID original da categoria, conforme consta no arquivo de origem. Usado para mapeamento durante o ETL.
 - category_name (TEXT): O nome da categoria (ex: "Books", "Drama").
 - **Chaves:**
 - **Chave Primária (PK):** category_id
 - **Chave Única (Unique):** category_source_id
 - **Restrições de Integridade:**
 - category_name **não pode ser nulo** (NOT NULL).
 - category_source_id **deve ser único**, garantindo que não haja IDs de origem duplicados.
-

Tabela: Category_Hierarchy

- **Descrição:** Tabela de associação que define a estrutura hierárquica (pai-filho) entre as categorias.
- **Atributos:**
 - parent_category_id (INT): Referencia o category_id da categoria "pai".
 - child_category_id (INT): Referencia o category_id da categoria "filho".
- **Chaves:**
 - **Chave Primária (PK):** (parent_category_id, child_category_id) - Chave composta.
 - **Chave Estrangeira (FK):** parent_category_id referencia Categories(category_id).
 - **Chave Estrangeira (FK):** child_category_id referencia Categories(category_id).
- **Restrições de Integridade:**

- **Referencial:** Uma categoria pai ou filho deve existir na tabela Categories. A exclusão de uma categoria em Categories pode afetar esta tabela (comportamento CASCADE se definido).
 - **CHECK:** parent_category_id <> child_category_id garante que uma categoria não pode ser sua própria pai.
 - **NOT NULL:** Ambos os campos são implicitamente NOT NULL por fazerem parte da chave primária.
-

Tabela: Products

- **Descrição:** Contém os dados principais de cada produto.
- **Atributos:**
 - source_id (INT): O ID original do produto, conforme consta no arquivo de dados de origem.
 - asin (VARCHAR(20)): Amazon Standard Identification Number. Identificador único do produto.
 - titulo (TEXT): O nome/título do produto.
 - group_name (TEXT): O grupo principal ao qual o produto pertence (ex: "Book", "DVD").
 - salesrank (INT): Ranking de vendas do produto. Pode ser nulo se não houver essa informação.
 - total_reviews (INT): O número total de avaliações que o produto possui, conforme informado no arquivo.
 - qntd_downloads (INT): O número de avaliações que foram efetivamente "baixadas" e estão presentes no arquivo.
 - average_rating (DECIMAL(3, 2)): A média de avaliação do produto, conforme informado no arquivo. Pode ser nulo.
 - similar_products_count (INT): A contagem de produtos similares, conforme informado no arquivo.
 - categories_count (INT): A contagem de categorias às quais o produto pertence, conforme informado no arquivo.
- **Chaves:**
 - **Chave Primária (PK):** asin
 - **Chave Única (Unique):** source_id

- **Nota:** tanto o source_id quanto o ASIN poderiam ser usados como primary key, mas o ASIN foi escolhido pelo fato de ser o identificador que a própria Amazon utiliza.
 - **Restrições de Integridade:**
 - source_id, titulo, group_name, similar_products_count, categories_count **não podem ser nulos** (NOT NULL).
 - **DEFAULT:** Os campos total_reviews, qntd_downloads, similar_products_count e categories_count têm valor padrão 0.
-

Tabela: Reviews

- **Descrição:** Armazena cada avaliação (review) individual feita por um cliente para um produto.
- **Atributos:**
 - review_id (SERIAL): Identificador numérico único, gerado automaticamente para cada review a fim de melhor organização.
 - product_asin (VARCHAR(20)): O ASIN do produto que foi avaliado.
 - customer_id (VARCHAR(20)): Identificador do cliente que escreveu a avaliação.
 - rating (SMALLINT): A nota (de 1 a 5) dada pelo cliente.
 - review_date (DATE): A data em que a avaliação foi feita.
 - votes (INT): O número total de votos que a avaliação recebeu.
 - helpful (INT): O número de votos que consideraram a avaliação como "útil".
- **Chaves:**
 - **Chave Primária (PK):** review_id
 - **Chave Estrangeira (FK):** product_asin referencia Products(asin).
- **Restrições de Integridade:**
 - **Referencial:** O product_asin deve corresponder a um produto existente na tabela Products.

- **CHECK:** rating >= 1 AND rating <= 5 garante que a nota esteja no intervalo válido.
 - product_asin, customer_id, rating, review_date, votes, helpful **não podem ser nulos** (NOT NULL).
 - **DEFAULT:** Os campos votes e helpful têm valor padrão 0.
-

Tabela: Related_products

- **Descrição:** Tabela de associação que registra a relação de "produtos similares" (muitos-para-muitos entre produtos).
 - **Atributos:**
 - product1_asin (VARCHAR(20)): O ASIN do primeiro produto na relação.
 - product2_asin (VARCHAR(20)): O ASIN do segundo produto na relação.
 - **Chaves:**
 - **Chave Primária (PK):** (product1_asin, product2_asin) - Chave composta.
 - **Chave Estrangeira (FK):** product1_asin referencia Products(asin).
 - **Chave Estrangeira (FK):** product2_asin referencia Products(asin).
 - **Restrições de Integridade:**
 - **Referencial:** Ambos os ASINs devem existir na tabela Products.
 - **CHECK:** product1_asin < product2_asin previne a duplicação de pares (ex: se (A, B) existe, (B, A) não pode existir) e impede que um produto seja relacionado a si mesmo.
 - **NOT NULL:** Ambos os campos são implicitamente NOT NULL por fazerem parte da chave primária.
-

Tabela: Product_category

- **Descrição:** Tabela de associação que implementa a relação muitos-para-muitos entre produtos e categorias.
- **Atributos:**
 - product_asin (VARCHAR(20)): O ASIN do produto.
 - category_id (INT): O ID da categoria.
- **Chaves:**
 - **Chave Primária (PK):** (product_asin, category_id) - Chave composta.
 - **Chave Estrangeira (FK):** product_asin referencia Products(asin).
 - **Chave Estrangeira (FK):** category_id referencia Categories(category_id).
- **Restrições de Integridade:**
 - **Referencial:** O product_asin deve existir na tabela Products e o category_id deve existir na tabela Categories.
 - **NOT NULL:** Ambos os campos são implicitamente NOT NULL por fazerem parte da chave primária.