

Dicionário de Dados: *Amazon product co-purchasing network metadata* -
Stanford Network Analysis Project (SNAP)

Tabela: Groups

Descrição da Tabela: Armazena informações dos grupos de produtos. Cada grupo é composto por um identificador único (id) e um nome, podendo ser: “Toy”, “DVD”, “Video Games”, “Software”, “Baby Product”, “Sports”, “Video”, “Music”, “CE”, “Book”.

Atributos:

- **group_id:** Identificador único do grupo.
- **name:** Nome do grupo.

Restrições de Integridade:

- **Chave Primária:** group_id.

Tabela: Products

Descrição da Tabela: Armazena informações relevantes sobre os produtos, e é considerada a tabela principal do sistema.

Atributos:

- **product_id:** Identificador único do produto.
- **asin:** Amazon Standard Identification Number, um identificador único alfanumérico de 10 caracteres criado pela Amazon.com para identificação interna.
- **title:** Título do produto.
- **salesrank:** A posição do produto no ranking de vendas da Amazon.com.
- **total_reviews:** Quantidade total de avaliações do produto.
- **group_id_fk:** Identificador único do grupo ao qual o produto pertence.

Restrições de Integridade:

- **Chave Primária:** product_id.
- **Chave estrangeira:** group_id_fk (Referencia Groups.group_id)

Tabela: Reviews

- **Descrição da Tabela:** Armazena as avaliações dos usuários feita a um determinado produto.

Atributos:

- **review_id:** Identificador único da avaliação.
- **date:** Data que a avaliação foi publicada.
- **rating:** Nota de 1 a 5, de acordo com o padrão da Amazon.com, sobre o produto em questão.
- **votes:** Total de votos na avaliação.
- **helpful:** Total de votos úteis da avaliação (quantas pessoas acharam determinada avaliação útil).
- **customer_id:** Identificador único de 15 caracteres do usuário autor da avaliação.
- **product_id_fk:** Identificador único do produto do qual a avaliação foi feita.

Restrições de Integridade:

- **Chave Primária:** review_id.
- **Chaves Estrangeiras:** product_id_fk (Referencia Products.product_id)

Tabela: Category

- **Descrição da Tabela:** Armazena informações sobre as categorias de determinado produto. A categoria completa é em

formato de grafo, havendo uma hierarquia de subcategorias que formam uma categoria completa. Esta tabela armazena as subcategorias.

Atributos:

- **category_id**: Identificador único da subcategoria.
- **name**: Nome da categoria.
- **parent_id**: Identificador único de uma subcategoria pai dada uma subcategoria. Quando nulo, indica uma subcategoria raiz.

Restrições de Integridade:

- **Chave Primária**: category_id.
- **Chaves Estrangeiras**: parent_id (Referencia Category.category_id).

Tabela: ProductsCategories

- **Descrição da Tabela**: Trata-se de uma tabela intermediária para garantir a relação N:N de produtos e categorias. Relaciona um produto à uma subcategoria folha (sem filhos) do qual este produto pertence.

Atributos:

- **category_id_fk**: Identificador único de subcategoria folha.
- **product_id_fk**: Identificador único do produto.

Restrições de Integridade:

- **Chave Primária**: category_id_fk e product_id_fk (formam uma única chave, com dois elementos).

- **Chaves Estrangeiras:** category_id_fk (Referencia Categories.category_id) e product_id_fk (Referencia Products.product_id).

Tabela: ProductProduct

- **Descrição da Tabela:** Trata-se de uma tabela intermediária para garantir a relação N:N de produtos e produtos (similares). Relaciona um produto a alguns produtos similares.

Atributos:

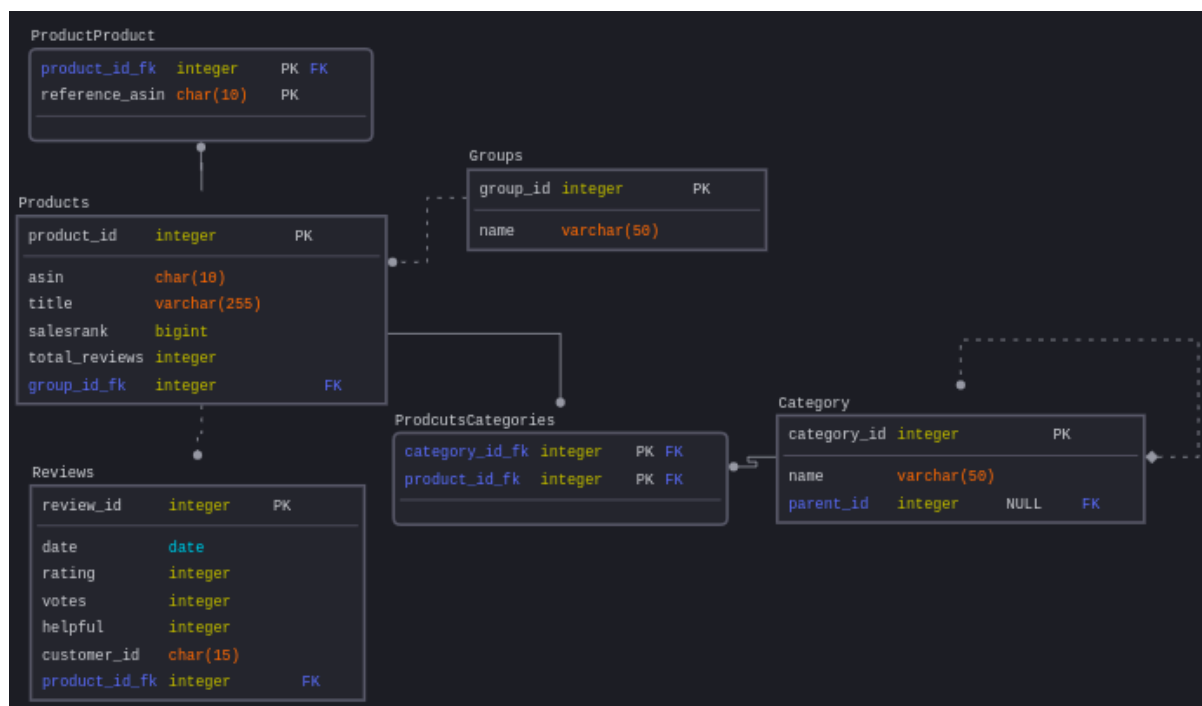
- **product_id_fk:** Identificador único do produto.
- **referenc_asin:** Identificador ASIN do produto similar.

Restrições de Integridade:

- **Chave Primária:** product_id_fk e reference_asin (formam uma única chave, com dois elementos).
- **Chaves Estrangeiras:** product_id_fk (Referencia Products.product_id).

Observações Gerais

- **Chaves Primárias:** Usadas para garantir a unicidade dos registros em cada tabela.
- **Chaves Estrangeiras:** Estabelecem ligações entre tabelas, garantindo a integridade referencial.
- **Integridade de Dados:** Restrições e tipos de dados ajudam a manter a consistência e a precisão dos dados no banco de dados.



Esquema do banco de dados e suas relações*

*No software utilizado para a modelagem, não é possível indicar uma chave primária com dois ou mais valores. Por isso, as tabelas “ProductsCategories” e “ProductProduct” aparecem com duas chaves primárias na figura. Contudo, trata-se de uma chave primária com dois valores.