

Trabalho 01 de Banco de Dados

Mateus Mota Nobrega – 21953021

Ricardo Oliveira Pinho de Lima – 22251127

20 de setembro de 2024

Arquivos e Execução

Os arquivos foram nomeados conforme as diretrizes estabelecidas pelo professor na especificação do trabalho. A seguir, apresentamos a descrição dos arquivos que compõem o sistema:

- **tp1_3.1.pdf**: Arquivo que contém a presente documentação, descrevendo o desenvolvimento do banco de dados, incluindo sua modelagem, criação das tabelas, inserção de dados, e as etapas realizadas ao longo do trabalho.
- **tp1_3.2.py**: Arquivo em Python responsável pela extração dos dados do conjunto "Amazon Product Co-Purchasing Network Metadata". Este script realiza a criação das tabelas em SQL (detalhadas na Seção *Dicionário do Esquema do Banco de Dados* deste relatório) e insere os dados extraídos no banco de dados, incluindo todas as suas respectivas relações.
- **tp1_3.3.py**: Script em Python que contém as principais consultas SQL requisitadas no trabalho, além de implementar a lógica por trás do dashboard em nível de terminal. O script apresenta uma interface para visualizar os resultados das consultas de forma interativa. Recomenda-se sua execução em uma tela grande para melhor visualização das informações apresentadas.

Arquivos de Dependências

Os arquivos `requirements.txt` e `install_dependencies.py` estão presentes na mesma pasta do projeto. O usuário poderá executar o arquivo `install_dependencies.py` para instalar todas as dependências necessárias de uma só vez, facilitando a configuração do ambiente.

Além disso, esses arquivos são componentes importantes do nosso trabalho, pois garantem que todas as bibliotecas requeridas sejam instaladas de maneira eficiente e organizada. O arquivo `requirements.txt` contém a lista de todas as bibliotecas necessárias, enquanto o script `install_dependencies.py` automatiza o processo de instalação.

Caso ocorra algum problema na execução do script instale as bibliotecas, execute os seguintes comandos:

```
pip install psycpg2
pip install tabulate
```

Para a execução correta dos arquivos `tp1_3.2.py` e `tp1_3.3.py`, como arquitetado por nós, é necessário que as essas bibliotecas citadas acima estejam instaladas no ambiente Python. Essas bibliotecas são essenciais para a conexão com o banco de dados PostgreSQL e para a exibição das informações no terminal de maneira organizada.

As bibliotecas necessárias são:

- **psycpg2**: Utilizada para realizar a conexão com o banco de dados PostgreSQL e executar as consultas SQL.
- **psycpg2.extras**: Fornece funções adicionais para facilitar a manipulação de dados no PostgreSQL, como a função `execute_values`, que acelera a inserção de dados em massa.
- **tabulate**: Biblioteca responsável por formatar os dados em tabelas organizadas no terminal, melhorando a visualização dos resultados.

Dicionário do Esquema do Banco de Dados

Tabela Categoria

- `CategoriaId` - Identificador da categoria.
- `Category_name` - Nome da categoria.
- `Parent_id` - Identificador da categoria pai.
- `CategoriaId` - Chave primária

Tabela Produto

- `Id` - Identificador do produto.
- `Asin` - Código ASIN do produto.
- `Titulo` - Nome do produto.

- Grupo - Grupo pertecente ao produto
- Sales_rank - Classificação no rank de vendas.
- Descontinuado - Diz se o produto foi descontinuado ou não.
- Id - Chave primária
- Asin - Chave única

Produtos Descontinuados

Adotamos o uso de um atributo chamado "descontinuado", que permite identificar se um produto foi descontinuado ou não. Durante a análise dos dados, constatamos que aproximadamente 5.000 produtos possuem este rótulo. Assim, ao solicitar um produto que foi descontinuado, o usuário receberá um feedback claro, informando que o produto solicitado está fora de circulação.

Com exceção dos campos ASIN e "descontinuado", os outros atributos na tabela poderão ter valores de domínio nulos. Essa abordagem garante que os dados relevantes sejam mantidos, enquanto permite flexibilidade para campos que podem não estar disponíveis para todos os produtos.

Tabela Categoria do Produto

- Asin - Código ASIN do produto.
- CategoriaId - Identificador da Categoria.
- Asin e CategoriaId - Chave composta (asin,categoriaId)
- Asin - Chave Estrangeira que referência asin na tabela produto
- CategoriaId - Chave Estrangeira que referência categoriaId na tabela produto

Tabela Similares

- Asin - Código ASIN do produto .
- Asin_similar - Código ASIN de produtos similares ao ASIN principal.
- Asin e Asin_similar - Chave composta (asin,categoriaId)
- Asin - Chave Estrangeira que referência asin na tabela produto

Tabela Review

- Id - Identificador único da review.
- Asin - Código ASIN do produto que foi avaliado.
- Cliente - Identificador do cliente que fez a review.
- Data_review - Data em que a review foi feita.
- Nota - Nota dada ao produto pelo cliente.
- Votos - Quantidade de votos que a review recebeu.
- Utilidade - Quantidade de votos marcando a review como útil.
- Id e Cliente - Chave composta (id, cliente).
- Asin - Chave Estrangeira que referência asin na tabela produto.

Esquema do Banco de Dados

A montagem dos esquemas foram feitas seguindo as regras das formas normais que implica em retirar produtos multivalorados, um dado por coluna nas linhas, atributos estão relacionados com suas respectivas chaves primárias e a organização dos atributos que dependem uns dos outros, porém que não são atributos chaves e usamos a quarta forma normal para eliminar as depências multivaloradas, abaixo o resultado final dessa modelagem:

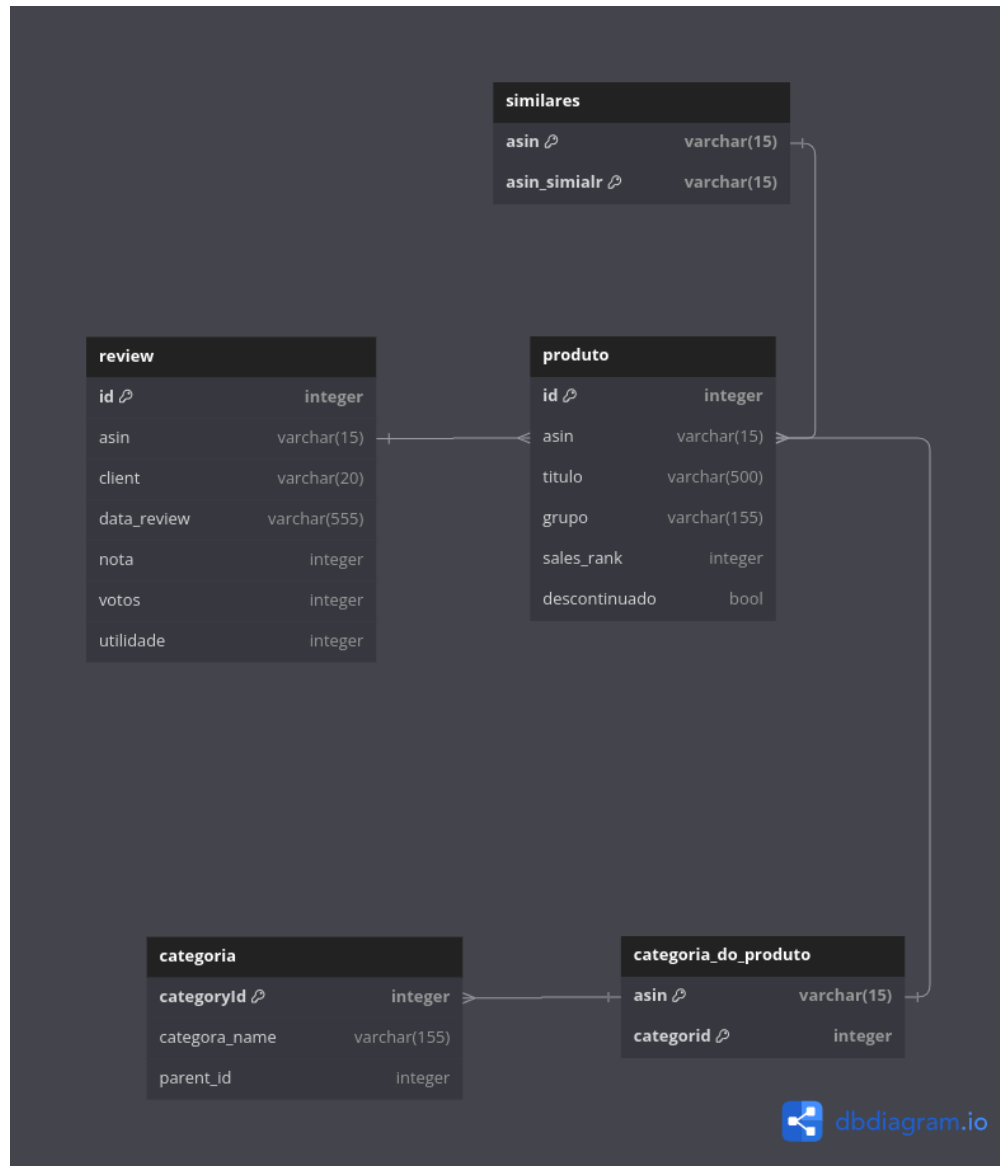


Figura 1: Esquema relacional do banco de dados.