

## Documentação do Banco de dados

Integrantes:

João Alfredo Holanda Bessa Neto

Raquel Folz Cavalcante

### 1. Descrição da base de dados

O arquivo de entrada de onde serão extraídos os dados de entrada será o "Amazon product co-purchasing network metadata" que faz parte do Stanford Network Analysis Project (SNAP). Os dados foram coletados em 2006 do site Amazon.com e contém informações sobre produtos e comentários de clientes sobre 548.552 produtos diferentes (livros, CDs de música, DVDs e fitas de vídeo VHS). Para cada produto, a seguinte informação está disponível:

- Título
- Posição no ranking de vendas (Salesrank)
- Lista de produtos ``similares`` (que foram adquiridos junto com o produto)
- Informação de categorização do produto - Categorias e subcategorias ao qual o produto pertence
- Comentários sobre os produtos:
  - ◆ Informação data, id do cliente, classificação, número de votos, o número de pessoas que acharam a avaliação útil

### 2. Esboço do esquema relacional

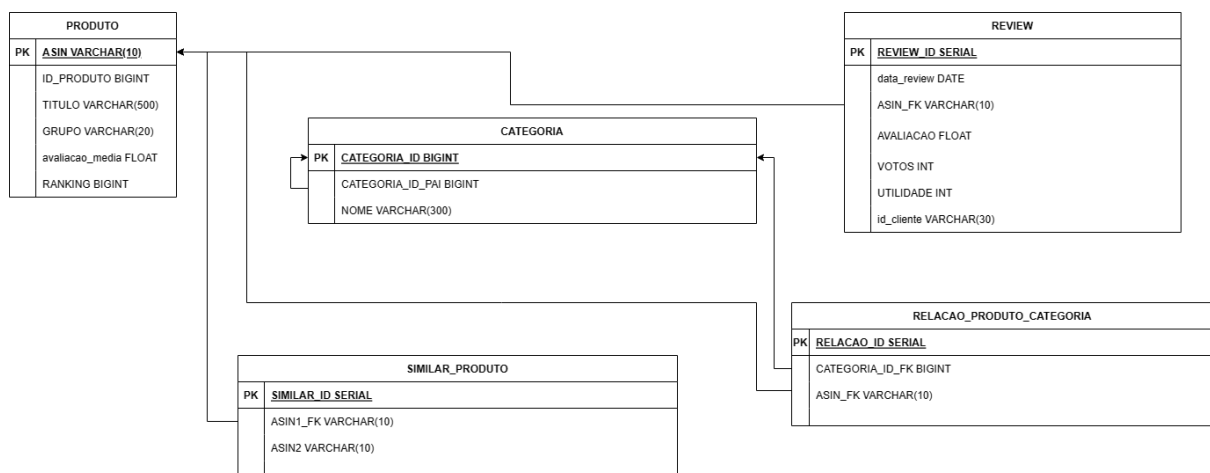
Com base nos requerimentos para a dashboard, o esboço do esquema foi gerado:

- PRODUTO
  - ASIN (VARCHAR(10), PRIMARY KEY)
  - ID\_PRODUTO (BIGINT)
  - TITULO (VARCHAR(500))
  - GRUPO (VARCHAR(20))
  - avaliacao\_media (FLOAT)
  - RANKING (BIGINT)
- SIMILAR\_PRODUTO
  - SIMILAR\_ID (SERIAL, PRIMARY KEY)
  - ASIN1\_FK (VARCHAR(10), FOREIGN KEY REFERENCES PRODUTO (ASIN))
  - ASIN2 (VARCHAR(10))
- CATEGORIA
  - CATEGORIA\_ID (BIGINT, PRIMARY KEY)
  - CATEGORIA\_ID\_PAI (BIGINT, FOREIGN KEY REFERENCES CATEGORIA (CATEGORIA\_ID))
  - NOME (VARCHAR(300))
- RELACAO\_PRODUTO\_CATEGORIA
  - RELACAO\_ID (SERIAL, PRIMARY KEY)

- CATEGORIA\_ID\_FK (BIGINT, FOREIGN KEY REFERENCES CATEGORIA(CATEGORIA\_ID))
  - ASIN\_FK (VARCHAR(10), FOREIGN KEY REFERENCES PRODUTO(ASIN))
- REVIEW
  - data\_review (DATE)
  - REVIEW\_ID (SERIAL, PRIMARY KEY)
  - ASIN\_FK (VARCHAR(10), FOREIGN KEY REFERENCES PRODUTO(ASIN))
  - AVALIACAO (FLOAT)
  - VOTOS (INT)
  - UTILIDADE (INT)
  - id\_cliente (VARCHAR(30))

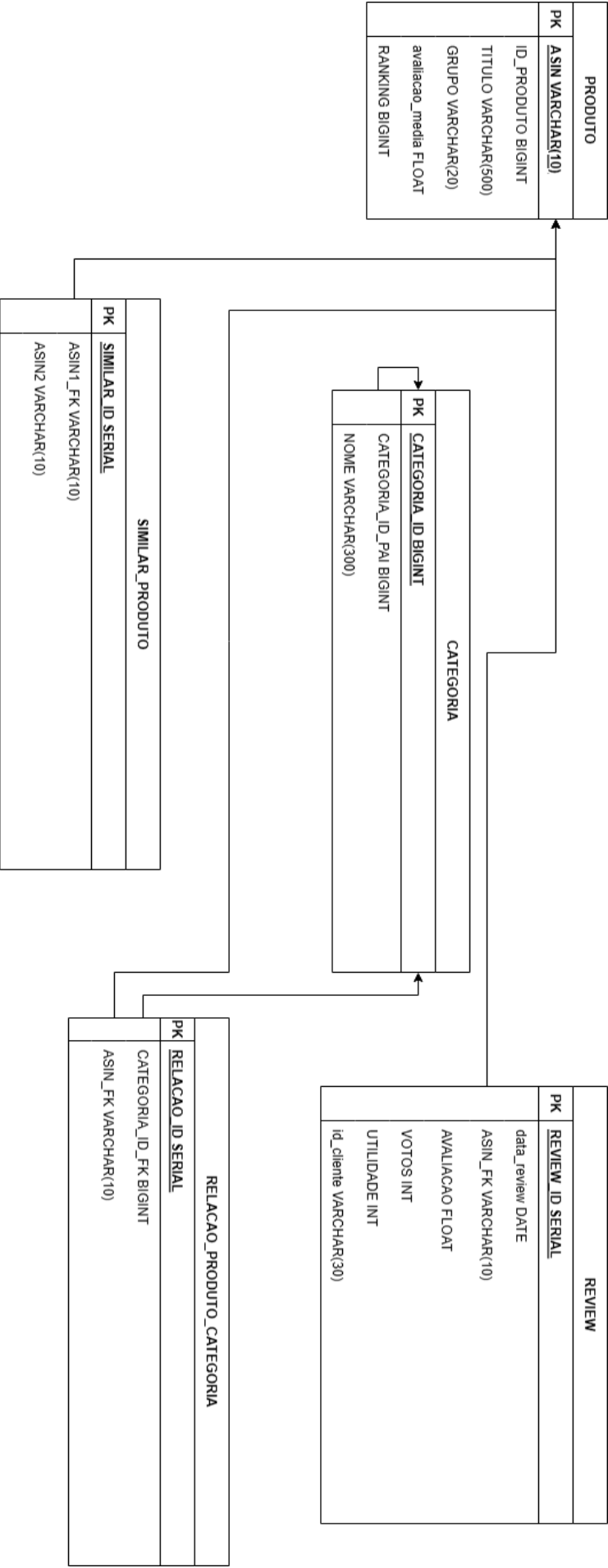
### 3. Esquema relacional

Com base no esboço, foi gerado o esquema relacional do banco de dados, este esquema representa o banco que foi implementado para suportar os requerimentos da dashboard:



A página a seguir mostra o esquema mais ampliado.

O esquema foi gerado com a ideia de trabalhar com 5 tabelas, com seus atributos respeitando a 3ª forma normal. O atributo ASIN de produto faz referência ao ID universal para livros e é utilizado em todas as outras tabelas para referenciar o livro relacionado à entidade. Não obstante, a tabela categoria possui um atributo chamado categoria pai que é chave estrangeira da chave primária da sua própria tabela, especificando relações de hierarquia de categoria.



#### 4. Dicionário de dados

Tabela:		PRODUTO	
Dado	Tipo do dado	Descrição	Restrições de domínio
ID_PRODUTO	INTEGER	Código de identificação do produto	Primary key
ASIN	VARCHAR	Identificação de produto interna da amazon	UNIQUE
TITULO	VARCHAR	Título do produto	
RANKING	INTEGER	Número no ranking de vendas	
AVALIACAO_MEDIA	FLOAT	Media de avaliações	
GRUPO	VARCHAR	Grupo que o produto pertence	
Tabela:		CATEGORIA	
CATEGORIA_ID	INTEGER	Código de identificação da categoria	Primary key
NOME	VARCHAR	Nome da categoria	
CATEGORIA_ID_PAI	INTEGER	Chave estrangeira que referencia outra categoria pai	NULL/FK
Tabela:		SIMILAR_PRODUTO	
RELAÇÃO_ID	SERIAL	Código de identificação da tabela SIMILAR_PRODUTO	Primary key
ASIN1_FK	INTEGER	Chave estrangeira que referencia o produto 1	NOT NULL/FK
ASIN2	INTEGER	ASIN que referencia o produto 2	
Tabela:		reviews	
REVIEW_ID	SERIAL	Código de identificação da tabela review	Primary key
ASIN_FK	INTEGER	Chave estrangeira que referencia o produto	FK
AVALIACAO	FLOAT	Avaliacao dada pelo cliente	
VOTOS	INTEGER	Quantidade de votos que a review recebeu	
UTILIDADE	INTEGER	Quantidade de votos positivos que a review recebeu	
DATA	DATE	Data que aconteceu o review	
Tabela:		RELACAO PRODUTO CATEGORIA	
RELAÇÃO_ID	SERIAL	Código de identificação da tabela RELACAO_PRODUTO_CATEGORIA	Primary key
ASIN_FK	VARCHAR	Chave estrangeira que referencia o produto	NOT NULL/FK
CATEGORIA_ID_FK	INTEGER	ASIN que referencia a categoria do produto	NOT NULL/FK