

Trabalho Prático I - Aplicação de Formas Normais na Modelagem de um Banco de Dados

Lucas Afonso Pereira Chagas - 22050316
Maria Vitória Costa do Nascimento - 22053592
Rodrigo Santos Corrêa - 22251139

¹Instituto de Ciências Exatas – Universidade Federal do Amazonas (UFAM)
Av. General Rodrigo Octavio Jordão Ramos, 1200 – Coroadó I, Manaus – AM – Brasil
CEP 69080-900

{lucas.chagas, vitoria.nascimento, rodrigo.correa}@icomp.ufam.edu.br

1. Necessidade de normalizar o banco de dados

Algumas incongruências podem se fazer presentes em uma tabela, por exemplo, quando vários dados são armazenados em um atributo do banco, surgem problemas de redundância e inconsistência, facilitando a ocorrência de erros em operações com o banco de dados.

Assim, surgem as regras de normalização do esquema do banco de dados foram criadas, sendo estas elaboradas para complementar a regra antecessora. Nesse trabalho, um dos objetivos é realizar a normalização do banco de dados de produtos da *Amazon*.

Asin	Title	Group	Salesrank	Similar	Categories	Reviews
1559362022	Wake Up and Smell the Coffee	Book	518927	1559360968 155936124 1559360828 1559361018 0743214552	!Books[283155]! !Subjects[1000]! !Literature & Fiction[17]! !Drama[2159]! !United States[2160]! !Arts & Photography[1]! !Performing Arts[521000]! !Theater[2154]! !General[2218]! !Bogorian, Eric[70116]! !(B)[70023]! !Authors, A-Z[70021]!	2002-5-13 cutomer: A2IGOA66Y6O8TQ rating: 5 votes: 3 helpful: 2 2002-6-17 cutomer: A2OIN4AUH84KNE rating: 5 votes: 2 helpful: 2 2003-1-2 cutomer: A2HN382JNT1CIU rating: 1 votes: 6 helpful: 1 2003-6-7 cutomer: A2FDJ79LDU4O18 rating: 4 votes: 1 helpful: 1 2003-6-27 cutomer: A39QMV9ZKRJXO5 rating: 4 votes: 1 helpful: 1 2004-2-17 cutomer: AUUVMSTQ1TXDI rating: 1 votes: 2 helpful: 0 2004-2-24 cutomer: A2C5K0QTL9UAT rating: 5 votes: 2 helpful: 2 2004-10-13 cutomer: A5XYF0Z3UH4HB rating: 5 votes: 1 helpful: 1
0688060676	Ten Black Dots	Book	144348	0140554955 0671792474 0440404959 0688147402 0689808925	!Books[283155]! Subjects[1000] Children's Books[4] Baby-3[2752] Basic Concepts[2753] General[2765] Ages 4-8[2785] !General[170062] !Authors & Illustrators, A-Z[170540] !(C)[170543] !Crews, Donald[170644] ![540744] !Specialty Stores[547930]! Books, Music & More[559958] !Books for Babies[692218] !Concept Books[694936] !General[694944]	2001-6-16 cutomer: A2MLCC7INJDUPI rating: 5 votes: 5 helpful: 5

Tabela 1. Amostra de dados antes de aplicar quaisquer regras de normalização ou decisões de projeto

2. Primeira Forma Normal

É a forma mais básica de normalização, sua propriedade diz respeito à atomicidade de um atributo, isto é, um atributo da tabela não pode ter valores repetidos ou em grupos, um atributo deve ser indivisível. Os dados múltiplos devem ser distribuídos em demais tabelas que forem necessárias.

A seguir, está a distribuição dos dados após a aplicação da Primeira Forma Normal (1NF).

	<u>asin</u>	<u>título</u>	<u>grupo</u>	<u>rank_vendas</u>
PRODUTO	1559362022	Wake Up and Smell the Coffee	Book	518927
	0688060676	Ten Black Dots	Book	144348

	<u>asin</u>	<u>asin_similar</u>
PRODUTO SIMILAR	1559362022	1559360968
	1559362022	1559361247
	1559362022	1559360828
	1559362022	1559361018
	1559362022	0743214552
	0688060676	0140554955
	0688060676	0671792474
	0688060676	0688147402
	0688060676	0440404959
	0688060676	0689808925

	<u>asin</u>	<u>id_usuario</u>	<u>data</u>	<u>nota</u>	<u>votos</u>	<u>votos_util</u>
AVALIAÇÕES	1559362022	A2IGOA66Y6O8TQ	2002-5-13	5	3	2
	1559362022	A2HN382JNT1CIU	2003-1-2	1	6	1
	0688060676	A2MLCC7INJDUPI	2001-6-16	5	5	5

	<u>asin</u>	<u>id_categoria</u>	<u>nome_categoria</u>
CATEGORIAS	0140445684	283155	Book
	0140445684	283155	Book
	0849943221	1000	Subjects
	0590464272	4	Children's Books
	0590464272	2966	Literature

Os atributos que possuíam tabelas internas deram origem a novas tabelas, PRODUTO SIMILAR, AVALIAÇÕES e CATEGORIAS. Os atributos que já eram atômicos permaneceram na tabela PRODUTO.

3. Segunda Forma Normal

Essa forma normal está relacionada com os atributos não-chave e a chave primária da tabela. Ou seja, qualquer atributo não classificado como chave deve, obrigatoriamente, ser dependente da chave primária correspondente. Caso contrário, deve-se gerar uma nova tabela com os dados devidamente relacionados.

Após a aplicação da 2NF, as seguintes modificações apareceram como resultado em nosso trabalho

Observou-se que a tabela CATEGORIAS gerada pela aplicação da 1FN não se enquadrava na 2FN, pois o atributo Nome Categoria não possuía relação direta com o

PRODUTO-CATEGORIA	asin	id_categoria
	0590464272	283155
	0590464272	1000
	0590464272	4
	0849943221	22
	0140445684	53

CATEGORIAS	id_categoria	nome_categoria
	283155	Book
	1000	Subjects
	297871	Authors, A-Z
	12489	Fiction
	735744	General

atributo de "Asin", dessa forma foi necessária a criação de uma nova tabela, denominada PRODUTO-CATEGORIA, que relaciona o ASIN do produto com o Id da Categoria. Na nova tabela CATEGORIAS foi mantido o Id que se relaciona com uma categoria.

4. Terceira Forma Normal

Nesta forma de normalização, tratam-se as dependências transitivas, isto é, quando uma coluna não-chave depende de outra ou mais colunas não-chave. Seguindo as outras formas, para aplicarmos a 3NF devemos primeiramente normalizarmos até a 2NF.

Caso essa dependência errônea seja encontrada, outra tabela deverá ser criada, com a sua *Primary Key* sendo a coluna não-chave determinante da relação. O atributo que virou *Primary Key* na nova tabela é mantido na tabela anterior, porém agora servirá como *Foreign Key* para associar as duas tabelas.

Durante o projeto, não encontramos a necessidade de aplicar a 3FN, uma vez que não identificamos algum atributo não-chave se relacionar com outro também não-chave, entretanto, notamos um problema na tabela AVALIAÇÕES, já que um usuário poderia fazer mais de uma avaliação sobre um mesmo produto. Ver seção 5.5 - A Tabela Avaliações.

5. Decisões de Projeto

Ao longo do desenvolvimento de toda a modelagem, algumas decisões tiveram de ser tomadas para que fizessem sentido e evitassem anomalias na implementação e nas consultas, para resultar nas respostas esperadas pelo dashboard. Essas decisões estão expostas nessa seção.

5.1. Padronizar o idioma

Os atributos identificados através da extração das informações do arquivo se encontravam em inglês, sendo esses, com exceção do Asin: Title, Group, Salesrank, Similar, Categories, Reviews. Para facilitar, estes foram traduzidos para o português: Título, Grupo, Rank Vendas, Categorias e Avaliações, respectivamente.

5.2. Parâmetros nas *Reviews*

No arquivo de entrada, o campo *Review* possuía 3 informações antes de exibir cada avaliação, são estas: *total*, *downloaded* e *avg rating*. Como essas informações não são solicitadas em qualquer enunciado do *dashboard*, foram desconsideradas para facilitar o processo de modelagem.

5.3. O Campo *Id*

Ainda no arquivo de entrada, o campo *Id* está presente na primeira linha de cada item novo, entretanto essa informação não contribui em nada para os requisitos do trabalho, sendo apenas referente à posição daquele item em relação ao arquivo de entrada. Assim, este campo foi desconsiderado.

5.4. Sobre produtos descontinuados

Os produtos registrados como descontinuados também foram desconsiderados por não agregarem no projeto

5.5. A Tabela Avaliações

Como visto, a tabela AVALIAÇÕES possuía uma combinação de *Asin* e *Customer* como chave primária, entretanto, foi observado que é possível que um usuário (*Customer*) fizesse várias avaliações sobre um mesmo produto, quebrando a regra de não repetição, fundamento das chaves primárias.

Sendo assim, foi criado o campo "*avaliação_id*", que servirá como a nova *primary key*, tornando cada avaliação única. Após essa mudança, verificamos todas as regras de normalização, todos os dados se mantiveram atômicos (1FN), assim como todos os atributos não-chave se relacionam com a chave primária (2FN). Já que não houve necessidade de normalização para a 3FN anteriormente, também não foi necessária depois da transformação.

<i>avaliação_id</i>	<i>asin</i>	<i>id_usuario</i>	<i>data</i>	<i>nota</i>	<i>votos</i>	<i>votos_util</i>
1	1559362022	A2IGOA66Y6O8TQ	2002-05-13	5	3	2
2	1559362022	A2OIN4AUH84KNE	2002-06-17	5	2	1
3	1559362022	A2HN382JNT1CIU	2003-01-02	1	6	1

Tabela 2. Amostra de revisões de um produto com *Asin*:1559362022

5.6. Sobre o *Dashboard*

5.6.1. Questão a) - Sobre as notas e avaliações úteis

Como devemos ordenar por notas e avaliações, decidimos o critério que $\text{nota} \geq 4$ são classificadas como bem avaliadas, $\text{nota} \leq 3$ como não bem avaliadas. Não adicionamos um critério que envolvesse os votos úteis, pois muitos produtos possuíam baixa quantidade de votos úteis, mas com boas notas.

5.6.2. Questão b) - Sobre o rank de vendas

Observamos que, quanto menor o rank de vendas, mais conhecido e avaliado era o produto, logo, deve ser mais vendido em comparação com os produtos que possuem um *ranking* maior. Assim, o resultado foi ordenado de forma decrescente.

5.6.3. Questão d) - Sobre os valores -1 e 0

Foi observado no arquivo de entrada que no atributo `rank_vendas` funciona do maior para o menor, ou seja, quanto menor o rank de vendas, mais vendido é o produto. Entretanto, percebemos que os valores -1 e 0 apareciam ao longo do arquivo de entrada e esses valores não possuíam mais vendas do que o produto com valor 1. Dessa forma, desconsideramos esses valores, pois viciavam os resultados quando ordenados.

6. Dicionário dos dados após a normalização

Tabela: Produto

- **Descrição da Tabela:** Armazena informações sobre um produto, cada produto é identificado por um "asin" único.
- **Atributos:**
 - **asin:** identificador único do produto.
 - **título:** Nome do produto
 - **grupo:** Grupo a que o produto pertence.
 - **rank_vendas:** posição que o produto ocupa em relação a um *ranking* de vendas.
- **Restrição de Integridade:**
 - **Chave Primária:** asin
 - **Chave Estrangeira:** não possui

Tabela: Produto Similar

- **Descrição da Tabela:** Armazena os produtos similares em relação a um determinado produto.
- **Atributos:**
 - **asin:** identificador de um produto
 - **asin_similar:** identificador de um produto similar ao produto referido
- **Restrição de Integridade:**
 - **Chave Primária:** combinação (asin, asin_similar)
 - **Chave Estrangeira:** asin (referencia PRODUTO.asin)

Tabela: Produto-Categoria

- **Descrição da Tabela:** Armazena a informação que ligam as categorias a zero ou mais produtos.
- **Atributos:**
 - **asin:** identificador de um produto.
 - **id_categoria:** identificador de uma categoria que o produto pertence.
- **Restrição de Integridade:**
 - **Chave Primária:** combinação (asin, id_categoria)

- **Chave Estrangeira:** asin (referencia PRODUTO.asin)

Tabela: Categorias

- **Descrição da Tabela:** Armazena informação sobre a divisão das categorias.
- **Atributos:**
 - **id_categoria:** identificador único de uma categoria.
 - **nome_categoria:** nome de uma categoria de produto.
- **Restrição de Integridade:**
 - **Chave Primária:** id_categoria
 - **Chave Estrangeira:** id_categoria (referencia PRODUTO_CATEGORIA.id_categoria)

Tabela: Avaliações

- **Descrição da Tabela:** Armazena informações sobre as avaliações feitas a cerca de um produto, cada avaliação é identificada por um id único em avaliacao_id.
- **Atributos:**
 - **avaliacao_id:** identificador único de uma avaliação
 - **asin:** identificador de um produto
 - **id_usuario:** identificador do usuário que realizou a avaliação
 - **data:** data em formato ano,mês,dia em que a avaliação foi realizada
 - **nota:** nota dada pelo usuário ao produto
 - **votos:** votos que a avaliação teve de outros usuários
 - **votos_util:** votos que a avaliação recebeu, considerando-a como útil, ou positiva.
- **Restrição de Integridade:**
 - **Chave Primária:** avaliacao_id
 - **Chave Estrangeira:** asin (referencia Produto.asin)

7. Diagrama em alto nível do banco de dados

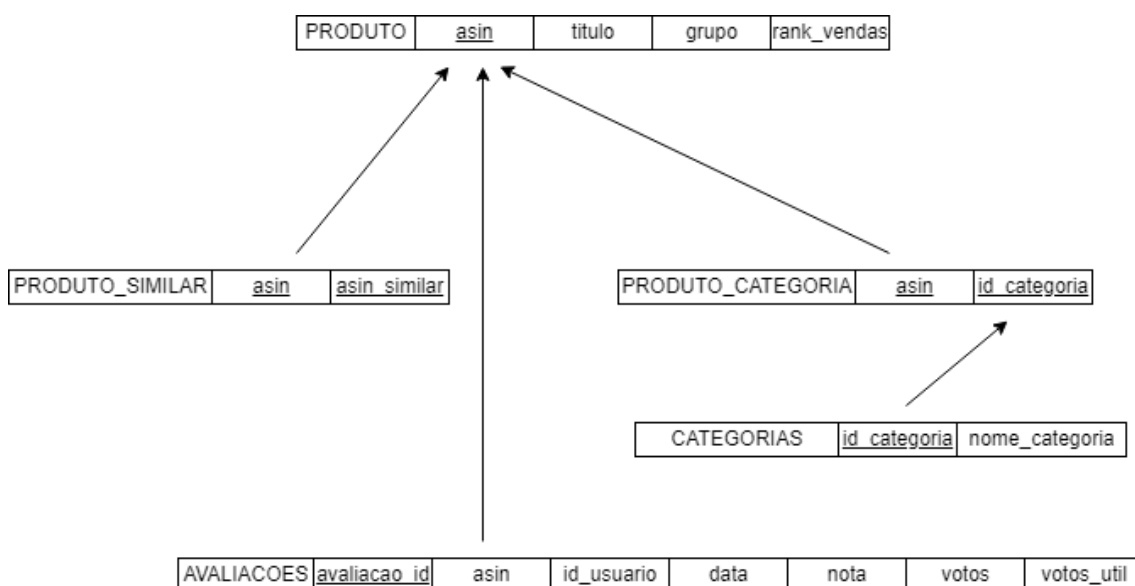


Figura 1. Representação com as relações entre as tabelas