

Aplicando Formas Normais em banco de dados e Esquema do TP1

Vinicius L. N. Fonseca - 22050031,
Matheus Silva do Santos - 22052573

¹Instituto de Computação – Universidade Federal do Amazonas (UFAM)
Caixa Postal 69077-470 – Manaus – AM – Brazil

{vinicius.fonseca, matheus.silva{@icomp.ufam.edu.br}}

1. Normalizando o Banco de Dados

A normalização de banco de dados é um processo de organização das tabelas de um banco de dados relacional para reduzir a redundância e garantir a integridade dos dados. A normalização é importante porque ajuda a evitar problemas como inconsistências nos dados e redundâncias desnecessárias, o que pode levar a erros e perda de eficiência em operações de busca e atualização.

Geralmente dividida em diferentes níveis, chamados de Formas Normais (FN). A primeira forma normal (1FN) requer que cada coluna de uma tabela contenha apenas valores atômicos (indivisíveis), não permitindo valores compostos ou repetidos. A segunda forma normal (2FN) requer que cada coluna da tabela dependa totalmente da chave primária da tabela, eliminando dados duplicados ou redundantes. A terceira forma normal (3FN) exige que cada coluna da tabela dependa apenas da chave primária e não de outras colunas, o que elimina dependências transitivas e ajuda a garantir que cada tabela seja mantida de forma independente. A figura 1, mostra uma instanciação simplificada da tabela antes de ser normalizada:

PRODUCT						
ASIN	TITLE	GROUP	SALES RANK	SIMILAR	CATEGORIES	REVIEWS
827229534	Patterns of Preaching: A Sermon Sampler	Book	396585	0804215715 156101074X 0687023955 0687074231 082721619X	[Books[283155]]Subjects[1000] Religion & Spirituality[22] Christianity[12290] Clergy[12360] Preaching[12368]	2000-7-28 customer: A2JW67OY8U6HHK rating: 5 votes: 10 helpful: 9 2003-12-14 customer: A2VE83MZF98ITY rating: 5 votes: 6 helpful: 5
738700797	Candlemas: Feast of Flames	Book	168596	0738700827 1567184960 1567182836 0738700525 0738700940	[Books[283155]]Subjects[1000] Religion & Spirituality[22] Earth-Based Religions[12472] Wicca[12484]	2003-12-14 customer: A2VE83MZF98ITY rating: 5 votes: 6 helpful: 5
486287785	World War II Allied Fighter Planes Trading Cards	Book	1270652	NULL	[Books[283155]]Subjects[1000] Home & Garden[48] Crafts & Hobbies[5126] General[5144]	2002-1-7 customer: A9CQ3PLRNIR83 rating: 4 votes: 5 helpful: 5

Figura 1. Exemplo de Tabela sem Normalização

1.1. Forma Normal 1

A primeira forma normal (1NF) é uma propriedade de uma relação numa base de dados relacional. Requer que cada atributo de uma relação tenha valores atômicos (indivisíveis). Isto significa que um atributo não deve conter valores múltiplos ou grupos repetitivos. Se um atributo tiver valores múltiplos, deve ser dividido em atributos separados numa nova

relação. Isto assegura que os dados são armazenados de forma consistente e estruturada, permitindo uma consulta mais eficiente e reduzindo o risco de anomalias de dados.

Com base na figura 1, é possível observar que os atributos Similar, Categories e Review não estão em conformidade com a 1NF do modelo relacional de banco de dados, pois são atributos multivalorados ou compostos. Portanto, é necessário realizar a modificação desses atributos para adequá-los ao padrão da 1NF, o que resulta na obtenção das seguintes tabelas:

PRODUCT			
ASIN	TITLE	GROUP	SALES RANK
827229534	Patterns of Preaching: A Sermon Sampler	Book	396585
738700797	Candlemas: Feast of Flames	Book	168596
486287785	World War II Allied Fighter Planes Trading Cards	Book	1270652

Figura 2. Tabela PRODUCT, após a ajustes quanto a atributos multivalorados e compostos (SIMILAR, CATEGORIES e REVIEW)

SIMILAR_PRODUCTS		PRODUCT_CATEGORY							
ASIN	PRODUCT	ASIN	SIMILAR		ASIN	PRODUCT	ID	CATEGORY	NAME
827229534		804215715			827229534		283155		Books
827229534		156101074X			827229534		1000		Subjects
827229534		687023955			827229534		22		Religion & Spirituality
827229534		687074231			827229534		12290		Christianity
827229534		082721619X			827229534		12360		Clergy
738700797		738700827			827229534		12368		Preaching
738700797		1567184960			738700797		283155		Books
738700797		1567182836			738700797		1000		Subjects
738700797		738700525			738700797		22		Religion & Spirituality
738700797		738700940			738700797		12472		Earth-Based Religions
486287785		NULL			738700797		12484		Wicca
					486287785		283155		Books
					486287785		1000		Subjects
					486287785		48		Home & Garden
					486287785		5126		Crafts & Hobbies
					486287785		5144		General

Figura 3. Tabelas criadas: SIMILAR_PRODUCTS e PRODUCT_CATEGORY, após a separação de seus atributos compostos e multivalorados de PRODUCT

PRODUCT REVIEW						
ASIN	PRODUCT	COSTUMER	DATE	RATE	VOTES	HELPFUL
827229534		A2JW67OY8U6HHK	28/07/2000	5	10	9
827229534		A2VE83MZF98ITY	14/12/2003	5	6	5
738700797		A2VE83MZF98ITY	14/12/2003	5	6	5
486287785		A9CQ3PLRNIR83	07/01/2002	4	5	5

Figura 4. Tabela PRODUCT_REVIEW criada após a separação de seus atributos compostos e multivalorados de PRODUCT

As tabelas da figura 3 e 4 (PRODUCT_CATEGORY e PRODUCT_REVIEW) foram criadas para resolver o problema de atributos compostos e multivalorados. Embora

seja comum adicionar colunas extras para valores compostos, a criação dessas tabelas foi necessária devido à natureza multivalorada desses atributos.

1.2. Forma Normal 2

A Segunda Forma Normal (2NF) é um nível de normalização de base de dados que se baseia na 1NF, eliminando dependências parciais de uma chave primária composta. Por outras palavras, 2NF assegura que cada atributo não-chave de uma tabela depende funcionalmente de toda a chave primária, e não apenas de uma parte dela. Em essência, 2NF ajuda a assegurar que os dados sejam logicamente organizados e evita a redundância de dados e anomalias que podem ocorrer por ter uma tabela com múltiplas dependências funcionais apenas de uma parte da chave primária.

A partir das tabelas geradas na 1FN, é possível observar que a tabela **PRODUCT_CATEGORY** na figura 3 não está na 2FN. Isso ocorre porque o nome de uma categoria não depende do **ASIN_PRODUCT** para ser definido. Portanto, é necessário utilizar apenas uma parte da chave composta, resultando na seguinte tabela:

PRODUCT_CATEGORY		CATEGORY_NAME	
ASIN_PRODUCT	ID_CATEGORY	ID_CATEGORY	NAME
827229534	283155	283155	Books
827229534	1000	1000	Subjects
827229534	22	22	Religion & Spirituality
827229534	12290	12290	Christianity
827229534	12360	12360	Clergy
827229534	12368	12368	Preaching
738700797	283155	12472	Earth-Based Religions
738700797	1000	12484	Wicca
738700797	22	48	Home & Garden
738700797	12472	5126	Crafts & Hobbies
738700797	12484	5144	General
486287785	283155		
486287785	1000		
486287785	48		
486287785	5126		
486287785	5144		

Figura 5. Tabela alterada: **PRODUCT_CATEGORY. Tabela criada: **CATEGORY_NAME**, após normalização 2FN em **PRODUCT_CATEGORY****

Observa-se que a redução de informações redundantes ocorreu, mesmo que o exemplo seja pequeno. É possível perceber o impacto que essa mudança trará ao ser aplicado na base de dados completa.

Embora possa ser inicialmente questionável se **PRODUCT_REVIEW**, figura 4, satisfaz a segunda forma normal (2FN), uma análise mais minuciosa revela que todos os atributos não chave dependem da chave composta como um todo. Para acessar informações como a nota de avaliação (**RATE**), a data da avaliação (**DATE**), a utilidade da avaliação (**HELPFUL**) e os votos positivos recebidos (**VOTE**), é necessário conhecer tanto o código do produto (**ASIN_PRODUCT**) quanto o autor do comentário (**CUSTOMER**).

1.3. Forma Normal 3

A Terceira Forma Normal (3NF) é um processo de normalização de bases de dados que visa eliminar dados redundantes através da remoção de dependências transitivas. Uma dependência transitiva ocorre quando um atributo não-chave de uma tabela depende de outro atributo não-chave, que, por sua vez, depende da chave primária. Para levar uma tabela a 3NF, todos os atributos não-chave devem ser dependentes apenas da chave primária, e não de qualquer outro atributo não-chave. Isto significa que se existirem atributos dependentes de outros atributos não-chave, estes devem ser movidos para a sua própria tabela.

Analizando as tabelas geradas a partir da forma 1FN e 2FN na figura 6, podemos facilmente notar que não há atributos não chave que dependam de outros atributos não chave

PRODUCT			
ASIN	TITLE	GROUP	SALESRANK
827229534	Patterns of Preaching: A Sermon Sampler	Book	396585
738700797	Candlemas: Feast of Flames	Book	168596
486287785	World War II Allied Fighter Planes Trading Cards	Book	1270652

PRODUCT_CATEGORY	
ASIN_PRODUCT	ID_CATEGORY
827229534	283155
827229534	1000
827229534	22
827229534	12290
827229534	12360
827229534	12368
738700797	283155
738700797	1000
738700797	22
738700797	12472
738700797	12484
486287785	283155
486287785	1000
486287785	48
486287785	5126
486287785	5144

CATEGORY_NAME	
ID_CATEGORY	NAME
283155	Books
1000	Subjects
22	Religion & Spirituality
12290	Christianity
12360	Clergy
12368	Preaching
12472	Earth-Based Religions
12484	Wicca
48	Home & Garden
5126	Crafts & Hobbies
5144	General

SIMILAR_PRODUCTS	
ASIN_PRODUCT	ASIN_SIMILAR
827229534	804215715
827229534	156101074X
827229534	687023955
827229534	687074231
827229534	082721619X
738700797	738700827
738700797	1567184960
738700797	1567182836
738700797	738700525
738700797	738700940
486287785	NULL

PRODUCT_REVIEW					
ASIN_PRODUCT	COSTUMER	DATE	RATE	VOTES	HELPFUL
827229534	A2JW67OY8U6HHK	28/07/2000	5	10	9
827229534	A2VE83MZF98ITY	14/12/2003	5	6	5
738700797	A2VE83MZF98ITY	14/12/2003	5	6	5
486287785	A9CQ3PLRNIR83	07/01/2002	4	5	5

Figura 6. Tabelas geradas até a 2FN

1.4. Forma Normal de Boyce Codd

A FNBC (Forma Normal de Boyce-Codd) é uma forma normal de banco de dados que visa eliminar anomalias de dependência funcional. Ela é uma forma mais restritiva que a 3NF e garante que não haja dependências funcionais não triviais em atributos que não façam parte da chave primária. Isso garante a integridade dos dados e evita redundância desnecessária de informações.

Após realizar testes de inserção de dados em um trabalho, foi constatado que na tabela `PRODUCT_REVIEW`, um mesmo `CUSTOMER` pode fazer mais de uma avaliação para um mesmo produto. Dessa forma, a criação de uma chave composta por `ASIN_PRODUCT`, `CUSTOMER` e `DATE` causaria conflitos, já que essas informações estariam duplicadas e apresentam valores diferentes para `RATE`, `VOTES` e `HELPFUL`. Para solucionar o problema de duplicidade de informações em `PRODUCT_REVIEW`, foi criado um novo atributo chamado `ID` e, em seguida, foi estabelecida uma chave composta por `ID`, `ASIN_PRODUCT` e `CUSTOMER`.

PRODUCT_REVIEW						
ID	ASIN_PRODUCT	COSTUMER	DATE	RATE	VOTES	HELPFUL
1	827229534	A2JW67OY8U6HHK	28/07/2000	5	10	9
2	827229534	A2VE83MZF98ITY	14/12/2003	5	6	5
3	738700797	A2VE83MZF98ITY	14/12/2003	5	6	5
4	486287785	A9CQ3PLRNIR83	07/01/2002	4	5	5

Figura 7. Alteração em `PRODUCT_REVIEW`

2. Visão geral do esquema

