
Banco de Dados I

Normalização

Profa. Fabrícia Damando

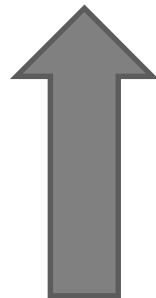
fdamando@yahoo.com.br

Introdução

- Projeto de BD
 - Nível conceitual
 - Nível lógico
 - Nível físico

Nível lógico

- a) Vasculhar a fundo o modelo conceitual identificando o conjunto completo de entidades que definem o sistema;
- b) Definir o conjunto completo de atributos para cada entidade;
- c) Identificar aqueles atributos candidatos a identificarem uma tabela (chaves primárias, chaves estrangeiras, etc...);
- d) Definir relacionamentos e cardinalidades;
- e) Identificar o domínio apropriado de cada atributo;
- f) Aplicar regras de normalização.



Normalização

- Processo através do qual busca-se eliminar, ou pelo menos reduzir, a redundância, a presença de valores nulos e as anomalias de atualização nas bases de dados.
- **Dependência Funcional**

Dependência funcional

- A Dependência Funcional pode ser descrita como uma restrição entre dois conjuntos de atributos de uma relação presente em um banco de dados

-
- Sendo X e Y subconjuntos de R a dependência funcional pode ser representada por $X \rightarrow Y$ onde X determina Y em R ;
 - A funcionalidade de X determina Y em um esquema de relação R se e somente se, sempre que duas tuplas de $r(R)$ coincidirem em seus valores de X , elas tiverem necessariamente que coincidir nos seus valores de Y
 - Exemplo: Matrícula aluno \rightarrow nome aluno

-
- Se X é uma chave candidata de R , $X \rightarrow Y$ para qualquer subconjunto de atributos Y de R . Se $X \rightarrow Y$ em R , isso não significa que $Y \rightarrow X$ em R .
 - Professor(Mat, Nome, Sexo)
 - A matrícula 49 sempre retornará o sexo 'M', mas o valor do campo sexo = 'M' nem sempre retornará a matrícula 49;

Tipos de dependência funcional

- Parcial
- Transitiva
- Multivalorada

Dependência funcional parcial

- Ocorre quando temos *mais de um atributo identificador* para uma relação e alguns dos atributos não identificadores dependem de somente um desses.
- Exemplo:
 - item_nota_fiscal(Nro_NF, Cod_prod, qtd, valor_unitário, data_emissao, descricao_produto)
- Os atributos data de emissão e descrição de produto dependem de somente parte da chave;

Dependência funcional transitiva

- Ocorre quando temos **atributos** dentro da relação que **dependem de um outro atributo não chave** que, por sua vez, depende do atributo chave da relação.
- Exemplo:
 - Funcionario(Matricula, Nome, cod_cargo, periculosidade_cargo)
- O atributo periculosidade cargo depende do código do cargo que, por sua vez, depende do atributo matrícula para a relação funcionário;

Dependência funcional multivalorada

- Ocorre quando para cada valor de um atributo A temos vários valores registrados para os atributos B e C, sendo que B e C independem um do outro – caso clássico para tabelas com três atributos chave;
- Projeto(Projeto, Engenheiro, Revisor)
- Haverá um conjunto de linhas para a combinação projeto * engenheiro + revisor e outro conjunto para projeto * revisor + engenheiro;

Formas normais

- Como determinar se determinado esquema está correto em relação a suas dependências funcionais?
- Qual técnica poderia ser utilizada para nos garantir de que nossas relações estão dentro de um padrão que evite problemas futuros com os dados?
- Para resolver essas perguntas foi criado um conjunto de técnicas que permitem a geração de modelos “livres” de redundância, chamadas formas normais.

-
- As formas normais são um processo para analisar os supostos esquemas de relações com base em suas Dependências Funcionais e chaves primárias.
 - Ao final do processo de normalização é desejável que cada tabela presente no banco de dados represente um único elemento.
 - As formas normais foram apresentadas por *E.F. Codd, 1970 - "A Relational Model of Data for Large Shared Banks"* inicialmente como três (1FN, 2FN e 3FN) e baseadas no conceito de dependência funcional, mais adiante Codd e R. Boyce refinaram essas regras na *Boyce-Codd Normal Form (BCNF)*.
 - A quarta forma normal foi criada para lidar com dependências multivaloradas e a quinta forma normal para lidar com dependências de junção
-

Objetivos - normalização

- O objetivo é auxiliar a eliminar redundância de dados e proteger a integridade dos dados quebrando entidades em formas mais simples;
- Aumentar o número de entidades pode significar aumentar a complexidade de lidar com todas elas, mas simplifica trabalhar com cada uma individualmente;
- Quanto maior o nível de normalização, maior o número de tabelas e menor a largura (número de colunas) dessas;
- Normalização tende a otimizar a eficiência do armazenamento e atualização do banco de dados, mas penalizar o processo de consulta;
- Eliminar dados duplicados
- Evitar o uso de código desnecessário para manter a integridade dos dados
- Manter as tabelas "estreitas" - aumentar o número de registros por página
- Reduzir o número de índices por tabela

Primeira forma normal

- O domínio de qualquer atributo de uma relação somente deve permitir valores atômicos.
- Não permitir atributos compostos ou multivalorados.
 - Exemplos:
 - Compostos:
 - CNPJCPF, endereço.rua endereço.nro endereço.cidade.
 - Multivalorados:
 - telefone(1- 5);
- Todas as instâncias da entidade devem possuir o mesmo número de valores
- Todas as instâncias da entidade devem ser diferentes umas das outras

- Procedimentos

- a) Identificar a chave primária da entidade;
- b) Identificar o grupo repetitivo e removê-lo da entidade;
- c) Criar uma nova entidade com a chave primária da entidade anterior e o grupo repetitivo.

A chave primária da nova entidade será obtida pela concatenação da chave primária da entidade inicial e a do grupo repetitivo

-
- Em outras palavras podemos definir que a **primeira forma normal** não admite repetições ou campos que tenha mais que um valor

→ Catálogo de CDs (fora da 1FN)

<u>codCD</u>	Nome CD	Gravadora	Preço	<u>NFaixa</u>	Musica	Autor	Tempo	<u>Tempo Tot</u>	CD Indicado
1	Mais do Mesmo	EMI	15,00	1	Será	Renato Russo	2:28	28:42	2
1	Mais do Mesmo	EMI	15,00	2	Ainda é cedo	Renato Russo	3:35	28:42	2
1	Mais do Mesmo	EMI	15,00	3	Geração Coca COLA	Renato Russo	2:20	28:42	2
1	Mais do Mesmo	EMI	15,00	4	Eduardo e Mônica	Renato Russo	4:32	28:42	2
1	Mais do Mesmo	EMI	15,00	5	Tempo Perdido	Renato Russo	5:00	28:42	2
1	Mais do Mesmo	EMI	15,00	6	Índios	Renato Russo	4:23	28:42	2
1	Mais do Mesmo	EMI	15,00	7	Que País é esse	Renato Russo	2:64	28:42	2
2	Bate Boca	Polygram	12,00	1	Meninos eu vi	Tom Jobim	3:20	21:00	1
2	Bate Boca	Polygram	12,00	2	Eu te Amo	Tom Jobim	2:54	21:00	1
2	Bate Boca	Polygram	12,00	3	Biscate	Chico Buarque	2:45	21:00	1
2	Bate Boca	Polygram	12,00	4	Noite dos Mascarados	Chico Buarque	3:20	21:00	1
2	Bate Boca	Polygram	12,00	5	Anos Dourados	Tom Jobim	4:00	21:00	1
2	Bate Boca	Polygram	12,00	6	BateBoca	Tom Jobim	4:41	21:00	1

→ Entidade CDs (na 1FN)

<u>codCD</u>	Nome CD	Gravadora	Preço	<u>TempoTot</u>	Indicado
1	Mais do Mesmo	EMI	15,00	28:42	2
2	Bate Boca	Polygram	12,00	21:00	1

Chave

Entidade Itens CD (na 1FN)

<u>codCD</u>	<u>NFaixa</u>	Musica	Autor	Tempo
1	1	Será	Renato Russo	2:28
1	2	Ainda é cedo	Renato Russo	3:35
1	3	Geração Coca Cola	Renato Russo	2:20
1	4	Eduardo e Mônica	Renato Russo	4:32
1	5	Tempo Perdido	Renato Russo	5:00
1	6	Indios	Renato Russo	4:23
1	7	Que País é esse	Renato Russo	2:64
2	1	Meninos eu vi	Tom Jobim	3:20
2	2	Eu te Amo	Tom Jobim	2:54
2	3	Biscate	Chico Buarque	2:45
2	4	Noite dos Mascarados	Chico Buarque	3:20
2	5	Anos Dourados	Tom Jobim	4:00
2	6	<u>BateBoca</u>	Tom Jobim	4:41

Chave Concatenada

Segunda forma normal

- Um esquema de relação R está na 2FN se todo atributo A em R que não seja parte da chave primária for completamente dependente em termos funcionais da chave primária de R (não existir dependência funcional parcial)
- Se a chave primária for composta de um único atributo esse teste não necessita ser realizado
- “Normalmente os projetistas de dados tendem a eliminar chaves compostas substituindo-as por uma chave simples do tipo contador e colocando os atributos identificadores sob uma regra de unicidade”.

-
- Uma tabela está na **Segunda Forma Normal** 2FN se ela estiver na 1FN e todos os atributos não chave forem totalmente dependentes da chave primária (dependente de toda a chave e não apenas de parte dela).

- Procedimentos

- a) Identificar os atributos que não são funcionalmente dependentes de toda a chave primária;
 - b) Remover da entidade todos esses atributos identificados e criar uma nova entidade com eles.

A chave primária da nova entidade será o atributo do qual os atributos do qual os atributos removidos são funcionalmente dependentes.

Entidade Autores (na 2FN)

<u>codAutor</u>	Autor
1	Renato Russo
2	Tom Jobim
3	Chico Buarque

+ Entidade Gravadoras (na 2FN)

<u>codGravadora</u>	Nome Gravadora	Endereço	e-mail
1	EMI	Rod Dutra, Km 229	contato@emi.com.br
2	Polygram	Av. Paulista, 220	contato@polygram.com.br

Entidade CDs (na 2FN)

<u>codCD</u>	Nome CD	<u>CodGravadora</u>	Preço	<u>TempoTot</u>	Indicado
1	Mais do Mesmo	1	15,00	28:42	2
2	Bate Boca	2	12,00	21:00	1

Entidade Música (na 2FN)

<u>Cod Musica</u>	Musica	<u>CodAutor</u>	Tempo
1	Será	1	2:28
2	Ainda é cedo	1	3:35
3	Geração Coca Cola	1	2:20
4	Eduardo e Mônica	1	4:32
5	Tempo Perdido	1	5:00
6	Indios	1	4:23
7	Que País é esse	1	2:64
8	Meninos eu vi	2	3:20
9	Eu te Amo	2	2:54
10	Biscate	3	2:45
11	Noite dos Mascarados	3	3:20

Entidade Itens CD (na 2FN)

<u>codCD</u>	<u>NFaixa</u>	<u>Cod Musica</u>
1	1	1
1	2	2
1	3	3
1	4	4
1	5	5
1	6	6
1	7	7
2	1	8
2	2	9
2	3	10
2	4	11
2	5	12
2	6	13

Terceira forma normal

- A relação não deve ter um atributo não-chave funcionalmente determinado por um outro atributo não-chave (ou por um conjunto de atributos não-chave). Ou seja, não deve haver dependência transitiva de um atributo não-chave na chave primária;

Terceira forma normal

- **FNBC - Boyce-Codd**
- Uma relação R está na BCFN se, e somente se, está na 1FN e se todo determinante for chave candidata de R. Cada tabela deve descrever um único objeto (eventualmente relacionamento entre objetos) e cada chave candidata é um valor (simples ou composto) escolhido para representar um objeto.
- Se há violação da FNBC, é porque há simultaneamente em uma tabela identificadores e dados relativos a distintos objetos ou relacionamentos entre estes objetos.

Ou seja...

- Uma tabela está na **Terceira Forma Normal** 3FN se ela estiver na 2FN e se Todos seus Atributos **NÃO CHAVES NÃO DEPENDEM de NENHUM OUTRO NÃO CHAVE.**
- Na **terceira forma normal** temos de eliminar aqueles campos que podem ser obtidos por alguma equação matemática, de outros campos da mesma tabela ou de outra tabela

- Procedimentos

- a) Identificar todos os atributos que são funcionalmente dependentes de outros atributos não chave;

- b) Removê-los

Entidade CDs (na 3FN)

<u>codCD</u>	Nome CD	<u>CodGravadora</u>	Preço	Indicado
1	Mais do Mesmo	1	15,00	2
2	Bate Boca	2	12,00	1

Excluída a coluna
Tempo total

Quarta forma normal

- A quarta forma normal trata das dependências multivaloradas: a cada valor de um atributo correspondem diversos valores em outro atributo.
 - Uma relação contendo uma dependência multivalorada que determina o atributo B a partir de um atributo A está na 4ªFN se, e somente se, todos os atributos da relação forem funcionalmente dependentes de A.
 - É violada quando uma relação tem dependências multivaloradas indesejáveis, e, portanto devem ser decompostas.
-
- Exemplo: (disciplinas, professores, alunos)

- Eliminação da Multiplicidade de Fatos Multivalorados.
- Um paciente pode necessitar de vários exames e dispor de vários planos de saúde.
- Portanto, plano de saúde e exame são fatos multivalorados em relação ao paciente.
- O paciente João dispõe de dois planos de saúde e necessita fazer dois exames.

Paciente	Plano de Saúde	Exame
João	Amil	Teor alcoólico
João	Blue-Life	Teor alcoólico
João	Amil	Sangue
João	Blue-Life	Sangue

- A tabela está normalizada na 3FN/FNBC (a chave é composta pela união de três atributos).
-
- Ainda assim, está difícil manipular mais de um fato multivalorado na mesma tabela.
- Resolve-se à situação fazendo a projeção dos subconjuntos que constituem cada fato multivalorado

Solicitação de Exame

<u>Paciente</u>	<u>Exame</u>
João	Teor alcoólico
João	Sangue

Suporte de Seguridade

<u>Paciente</u>	<u>Plano de Saúde</u>
João	Amil
João	Blue-Life

Revisando...

-
- Primeira Forma Normal (1FN) - Nenhum dos atributos possui repetições.
- Segunda Forma Normal (2FN) - Dependência da Chave Primária
- Terceira Forma Normal (3FN) - Não Dependência de Outros Atributos
- Quarta Forma Normal (4FN) - Dependência Multivalorada
- Quinta Forma Norma (5FN) -

Quinta forma normal

- Definição: uma relação de 4FN estará em 5FN, quando seu conteúdo não puder ser reconstruído (existir perda de informação) a partir das diversas relações menores que não possuam a mesma chave primária.
-
- Esta forma normal trata especificamente dos casos de perda de informação, quando da decomposição de relacionamentos múltiplos.

-
- A 5FN trata de casos bastante particulares, que ocorrem na modelagem de dados, que são os relacionamentos múltiplos (ternários, quaternários e n-ários).
 -
 - Ela fala que um registro está na sua 5FN, quando o conteúdo deste mesmo registro não puder ser reconstruído (junção) a partir de outros registros menores, extraídos deste registro principal.
 - Ou seja, se ao particionar um registro, e sua junção posterior não conseguir recuperar as informações contidas no registro original, então este registro está na 5FN.