

Normalização de Bancos de Dados

Formas Normais, Dependência Funcional, Dependência Transitiva, Dependência Multivalorada, Dependência de Junção

Prof. Leandro Colevati

Qualidade do Projeto Lógico

Como avaliar a qualidade do esquema da relação?

- Semântica
- Implementação/desempenho

Análise informal:

- Princípios para um bom projeto

Análise formal:

- Dependência funcional
- Normalização

Qualidade do Projeto Lógico

Análise Informal (princípios):

- Semântica de atributos
- Redução de redundância em tuplas:
 - prevenção de anomalias de inserção
 - prevenção de anomalias de remoção
 - prevenção de anomalias de alteração
- Redução de valores nulos
- Prevenção de geração de registros(tuplas) espúrios (ilegítimos)

Exemplo:

Combina informações de tipos diferentes de entidades

- Problema semântico

Redundância em relação às informações armazenadas

- Dados do departamento (Nome_Depto e CPF_Gerente_Depto)

Inserção

- Para inserir um empregado, é necessário cadastrar informações sobre o departamento (ou nulls)
- Tais informações podem gerar dados inconsistentes sobre o departamento

Exclusão:

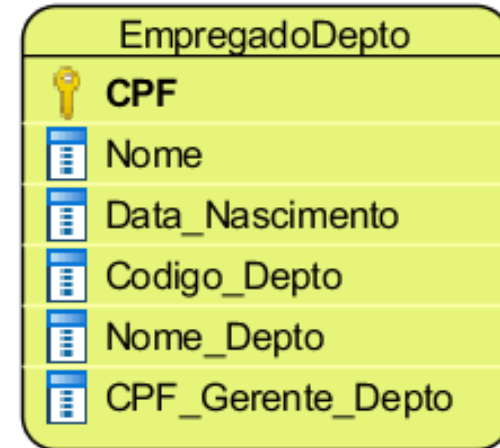
- Apagar um empregado pode significar apagar as informações do departamento

Atualização:

- Mudar o valor de um atributo de um registro de EmpregadoDepto pode implicar em ter de alterar outros valores correspondentes
 - Ex.: mudar Codigo_Depto

Valores *null*:

- Se muitos atributos não se aplicarem a muitos registros da relação, poderemos desperdiçar espaço de armazenamento.
 - Ex: Incluir no escritório na relação “empregados”, sendo que somente 10% destes possuem de fato um escritório



Qualidade do Projeto Lógico

Análise Formal:

- Dependências Funcionais:
 - Restrições entre atributos:
 - Avaliação da qualidade dos esquemas de relação
 - Garantia de consistência da base de dados

Dependência Funcional

É uma restrição entre dois subconjuntos de atributos (A e B) de \mathcal{R} , sendo denotada por $A \rightarrow B$

Especifica uma restrição nos possíveis registros de $R(\mathcal{R})$:

- Se $t_i[A] = t_j[A]$ então $t_i[B] = t_j[B]$ para quaisquer i, j

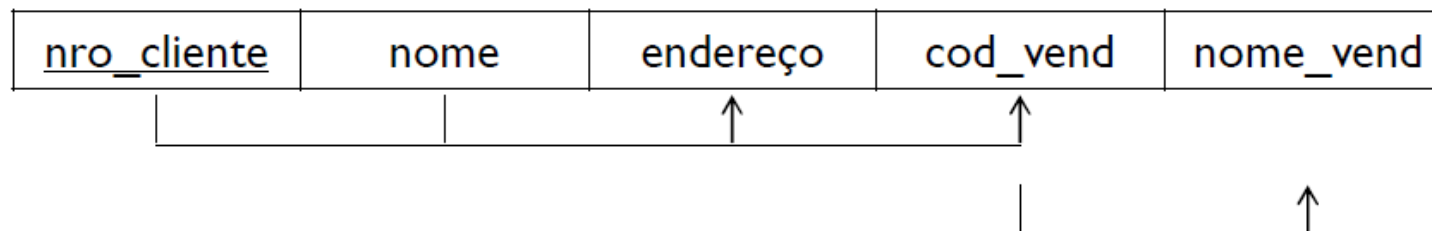
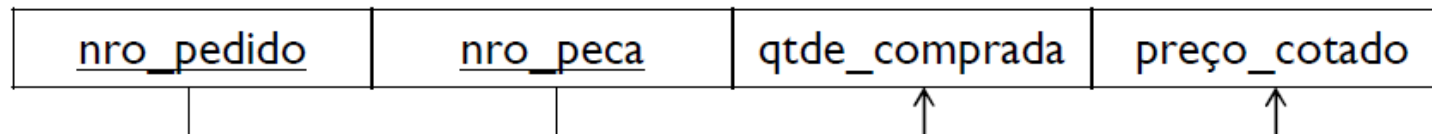
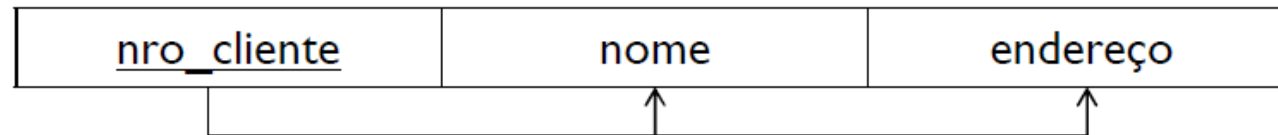
Neste caso diz-se que A determina funcionalmente B (ou alternativamente que B depende funcionalmente de A)

Exemplos:

$\{RA\} \rightarrow \{\text{Nome_Aluno}, \text{Curso_Aluno}\}$

$\{\text{Codigo_Cliente}, \text{Codigo_Vendedor}\} \rightarrow \{\text{quantidade}, \text{valor_total}\}$

Notação para Dependência Funcional



Dependência Funcional

Propriedade semântica, identificada pelo projetista do BD

Pode ser verificada na instância do BD mas não é definida a partir dela

Exemplo: Considerando a relação:

Produto {ID, Nome, Valor}

t (INT, VARCHAR(50), DECIMAL(4,2))

r1<1, 'Mouse', 59.99>

r2<2, 'Teclado', 99.99>

r3<3, 'Monitor', 999.99>

As DFs são atendidas ?

ID → Nome

ID → Valor

Nome → Valor

Valor → Nome

Consistência

Na construção de um SGBD baseado no modelo relacional:

- Definição das relações baseada na análise de DFs;
- Formas normais;
- Uma relação está em uma determinada forma normal quando satisfaz certas propriedades baseadas nas DFs;
- Colocar uma relação em uma forma normal
 - **Normalização**

Normalização

Normalização de Relações:

- Baseada nas DFs;
- Garante consistência na construção do sistema:
 - redução de anomalias;
 - redução de redundância;
- Formas Normais (FNs) baseadas em DFs:
 - baseadas em chave primária: 2a FN, 3a FN;
 - baseadas em chaves candidatas: FN de Boyce-Codd (FNBC ou, em Inglês, BCNF).
- FN baseada em dependências multivaloradas:
 - 4a FN
- FN baseada em dependências de junção:
 - 5a FN
- Para atender uma Forma Normal, a Relação deve atender a regra definida e sua anterior

1ª Forma Normal (1FN)

\mathcal{R} está na 1FN se:

- todo valor em \mathcal{R} for atômico
- \mathcal{R} não contém grupos de repetição

Considerações:

- geralmente considerada parte da definição de \mathcal{R}
- não permite atributos multivalorados, compostos ou suas combinações

1ª Forma Normal (1FN)

Exemplo:

Cliente (nro_cli, nome, {end_entrega})

nro_cli	nome	end_entrega
124	João dos Santos	Rua 10, 1024 Rua 24, 1356
311	José Ferreira Neves	Rua 46, 1344 Rua 98, 4456

1ª Forma Normal (1FN)

Exemplo: Cliente (nro_cli, nome, {end_entrega})

Adequação:

1. Substituir o grupo de repetição pelo máximo de valores estabelecido para o grupo (genérico e gera muitos *null*)
2. Expandir a chave primária (redundância)
3. Gerar nova relação para o grupo de repetição definindo nova PK (genérica, sem redundância)

1. Cliente (nro_cli, nome, rua1, numero1, rua2, numero2)
2. Cliente (nro_cli, nome, rua, numero)
3. Cliente (nro_cli, nome)
Cliente_Entrega (nro_cli, rua, numero)

2ª Forma Normal (2FN)

Definição: O esquema de relação \mathcal{R} está na 2FN se todo atributo não primário* A em \mathcal{R} tem dependência funcional total da chave primária de \mathcal{R}

Deve estar na 1FN

Exemplo: Pedido (codigo_pedido, data_pedido, codigo_peca, descricao_peca, quantidade_comprada, preco_cotado)

DFs: codigo_pedido \rightarrow data_pedido

codigo_peca \rightarrow descricao_peca

{codigo_pedido, codigo_peca} \rightarrow {quantidade_comprada, preco_cotado}

2ª Forma Normal (2FN)

Adequação: Para cada sub-conjunto de atributos da chave primária, gerar uma relação com esse sub-conjunto como sua chave primária;

- Incluir os atributos da relação original na relação correspondente à chave primária apropriada:
 - colocar cada atributo junto com a coleção mínima da qual ele depende, atribuindo um nome a cada relação.

Exemplo:

Pedido (codigo_pedido, data)

Peca (codigo_peca, descricao_peca)

Pedido_Peca (codigo_pedido, codigo_peca, quantidade_comprada, preco_cotado)

3ª Forma Normal (3FN)

1. Está na 2FN;
2. Nenhum atributo não primário de R for transitivamente dependente da chave primária.

Dependência transitiva: Atributo não chave faz dependência funcional com atributo não chave

Exemplo: Cliente (codigo_cli, nome_cli, telefone_cli, codigo_vendedor, nome_vendedor)

DFs: codigo_cli → nome_cli, telefone_cli
 codigo_vendedor → nome_vendedor

3ª Forma Normal (3FN)

Adequação:

- Para cada determinante que não é uma chave candidata, remover da relação os atributos que dependem desse determinante
- Criar uma nova relação contendo todos os atributos da relação original que dependem desse determinante
- Tornar o determinante a chave primária da nova relação

Exemplo:

Cliente (codigo_cli, nome_cli, telefone_cli, codigo_vendedor (FK ?))

Vendedor (codigo_vendedor, nome_vendedor)

Assim como a 2FN, a 3FN evita:

- Inconsistência e anomalias causadas por redundância de informações;
- Perda de informação em operações de remoção/alterações na relação.

3ª Forma Normal (3FN)

- Considera-se que uma Relação (\mathcal{R}) já está normalizado se atende à 3FN

Forma Normal de Boyce-Codd (FNBC)

\mathcal{R} está na FNBC se para cada dependência funcional $X \rightarrow A$, X é uma superchave de \mathcal{R}

Diferença entre FNBC e 3FN:

- 3FN permite A primário – não se aplica à FNBC
- Se \mathcal{R} está na FNBC \rightarrow \mathcal{R} está na 3FN
- Se \mathcal{R} está na 3FN, não necessariamente \mathcal{R} está na FNBC.

Na prática, a maioria dos esquemas de relação que está na 3FN também está na FNBC.

Forma Normal de Boyce-Codd (FNBC)

Exemplo: Filme (nome, ano, duração, nome_estudio, endereço_estudio)

nome	ano	duração	nome_estudio	endereço_estudio
Star Wars	1977	124	Fox	Elm St.
Empire Strikes Back	1980	143	Fox	Elm St.
Gone With the Wind	1939	181	Paramount	Pine St.
Lion King	1994	124	Disney	Oak St.
Return of the Jedi	1983	165	Fox	Elm St.
Pocahontas	1995	115	Disney	Oak St.

Problemas:

Redundância;

Anomalia de Updates;

Anomalia de Deletes;

Forma Normal de Boyce-Codd (FNBC)

Adequação: Aplicar a 3FN (Gera decomposição em Relações)

- Não deve ser feita qualquer decomposição

Solução errada:

Filme (nome, ano, duração (FK))

Estudio (duração, nome_estudio, endereço_estudio)

nome	ano	duração
Star Wars	1977	124
Empire Strikes Back	1980	143
Gone With the Wind	1939	181
Lion King	1994	124
Return of the Jedi	1983	165
Pocahontas	1995	115

duração	nome_estudio	endereço_estudio
124	Fox	Elm St.
143	Fox	Elm St.
181	Paramount	Pine St.
124	Disney	Oak St.
165	Fox	Elm St.
115	Disney	Oak St.

Registros
Espúrios

Viola FN



Forma Normal de Boyce-Codd (FNBC)

Solução:

Filme (nome, ano, duração, nome_estudio(FK))

Estudio (nome_estudio, endereço_estudio)

nome	ano	duração	nome_estudio
Star Wars	1977	124	Fox
Empire Strikes Back	1980	143	Fox
Gone With the Wind	1939	181	Paramount
Lion King	1994	124	Disney
Return of the Jedi	1983	165	Fox
Pocahontas	1995	115	Disney

nome_estudio	endereço_estudio
Fox	Elm St.
Paramount	Pine St.
Disney	Oak St.

4ª Forma Normal (4FN)

Um esquema de relação está na 4FN se:

- Atende a FNBC
- todas as Dependências Multivaloradas são triviais ou;
- para cada DM não-trivial $A \twoheadrightarrow B$, A é uma superchave em R

Dependência Multivalorada (DM): restrição entre dois conjuntos de atributos:

- A multidetermina B (ou B é multidependente de A)

Atributos Multivalorados de uma Relação devem ser decompostos para uma nova relação, definindo o correto relacionamento entre relações e a chave primária (fundamentalmente composta) – Uma das possíveis adequações da 1FN.

4ª Forma Normal (4FN)

- Exemplo:

Pessoa (codigo, nome, {email})

- Adequação:

Pessoa (codigo, nome)

Pessoa_Email (codigo_pessoa (FK), email)

Evita redundância nos registros

Evita inconsistências causadas por

- inclusão/remoção/alteração de tuplas;

5ª Forma Normal (5FN)

Estar na 4FN

Não deve haver dependência de junção

- Evitar a criação de relação PK/FK quando não for estritamente necessário

- Exemplo:

Produto (idProduto, descrição, idFornecedor(FK), quantidade)

Fornecedor (idFornecedor, nome)

Nota (idNota, idVendedor(FK), idProduto(FK), idFornecedor(FK))

- Adequação:

Produto (idProduto, descrição, idFornecedor(FK), quantidade)

Fornecedor (idFornecedor, nome)

Nota (idNota, idVendedor(FK), idProduto(FK))

Exemplo Engenharia Reversa

| FICHA TÉCNICA PARA RESTAURANTE

Nome da receita Brigadeiro Cremoso Tamanho da receita 100 porções Custo da receita 12.00 Custo da porção 0.12							FOTO
Ingredientes	Quantidade	Unidade	Forma de Compra	Preço	Rendimento %	Preço limpo	Custo
Leito Condensado	1	Lata	Kg	20.00	100	20.00	8.00
Custo Total							

Exemplo Engenharia Reversa

Solução:

Receita(Codigo, Nome, Tamanho, Custo_Total, Custo_Porcao, Foto, Custo_Receita)

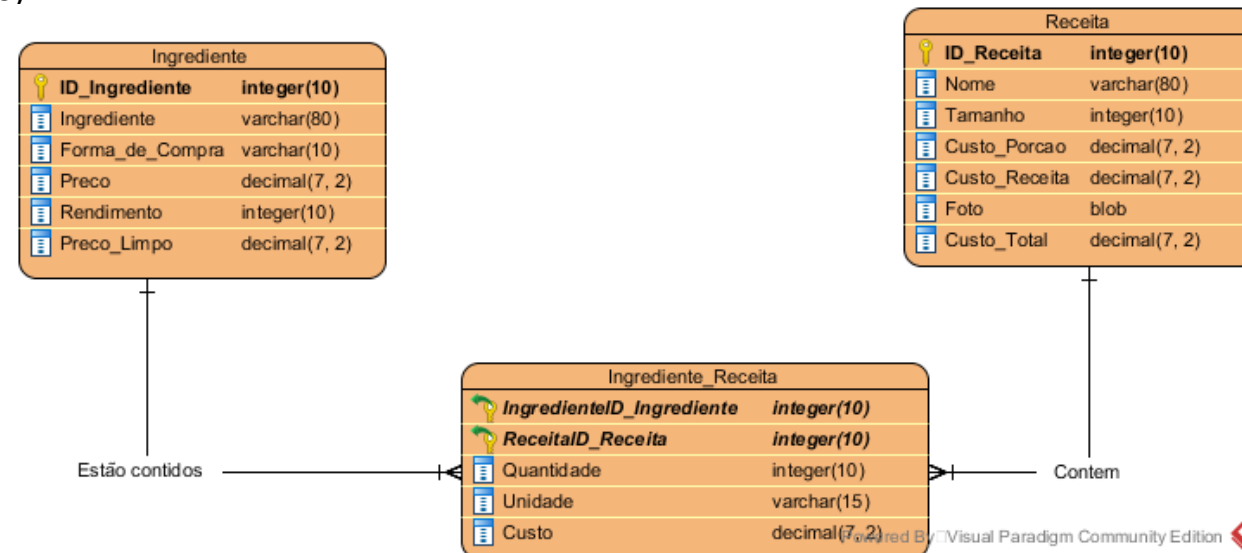
t<Int, Varchar(60), Int, Decimal(7,2), Decimal(7,2), Blob, Decimal(7,2)>

Ingrediente(ID_Ingrediente, Nome, Forma_Compra, Preco, Rendimento, Preco_Limpo)

t<Int, Varchar(50), Varchar(15), Decimal(7,2), Int, Decimal(7,2)>

Receita_Ingrediente(ID_Ingrediente, ReceitaCodigo, Quantidade, Unidade, Custo)

t<Int, Int, Int, Varchar(10), Decimal(7,2)>



Exercícios

1. vendedor (nro_vend, nome_vend, {cliente (nro_cli, nome_cli)})

As seguintes dependências funcionais devem ser garantidas na normalização:

nro_vend \rightarrow nome_vend;

nro_cli \rightarrow nome_cli.

Observação: considere que um vendedor

- pode atender diversos clientes, e um cliente
- pode ser atendido por diversos vendedores.

Exercícios

2. aluno (nro_aluno, cod_depto, nome_depto, sigla_depto, cod_orient, nome_orient, {fone_orient}, cod_curso)

As seguintes dependências funcionais devem ser garantidas na normalização:

cod_depto → {nome_depto, sigla_depto};

cod_orient → {nome_orient, fone_orient};

nro_aluno → {cod_depto, cod_orient, cod_curso};

Observações adicionais:

- um aluno somente pode estar associado a um departamento;
- um aluno cursa apenas um único curso;
- um aluno somente pode ser orientado por um único orientador

Exercícios

3. aluno (nro_aluno, nome_aluno, {curso (nro_curso, descrição_curso, ano_ingresso, nro_depto, nome_depto)})

As seguintes dependências funcionais devem ser garantidas na normalização:

nro_aluno → nome_aluno;

nro_curso → descrição_curso;

nro_depto → nome_depto;

{nro_aluno, nro_curso} → ano_ingresso;

nro_curso → nro_depto.

Observações adicionais:

- um aluno pode cursar mais do que um curso;
- um curso somente pode ser oferecido por um único departamento.

Exercícios

4. Aplicar as formas normais ao formulário abaixo:

The image shows a screenshot of a web form titled "Cadastro de médicos". The form is organized into several sections:

- Identificação:** Includes a text field for "Nome" (filled with "AARAO BURLAMAQUI BENCHIMOL"), a dropdown for "Tipo" (set to "CRM"), a text field for "Número" (filled with "6379"), and a dropdown for "UF" (set to "RJ").
- Especialidades:** A text field containing "CLÍNICA MÉDICA".
- Endereço:** Includes text fields for "Logradouro", "Bairro", and "Cidade", along with a dropdown for "UF" and a text field for "CEP".
- Telefones:** Three empty text input fields.
- Internet:** A text field for "E-mail" and a button labeled "enviar".
- Outras informações:** Includes text fields for "Secretária", "Última receita", and "Comissão" (filled with "10"), and a dropdown for "Representante" (set to "CESAR ROBERTO") and a text field for "Última visita".

At the top of the form, there is a toolbar with navigation arrows and icons for file operations (new, open, save, print, delete, search, refresh).