

BANCO DE DADOS

Normalização

PROJETO DE BANCO DE DADOS RELACIONAL

○ Objetivos implícitos no projeto:

○ Preservação da Informação

"A informação é muito difícil de se quantificar - logo, consideramos a preservação de informação em matéria de manutenção de todos os conceitos, incluindo tipos de atributos, tipos de entidade e tipos de relacionamentos, bem como generalização / especialização descritos usando modelos ER."

○ Redundância mínima

"Implica diminuir o armazenamento redundante da mesma informação e reduzir a necessidade de múltiplas atualizações para manter a consistência entre diversas cópias da mesma informação, em resposta a eventos do mundo real que exijam fazer uma atualização."

PROJETO DE BANCO DE DADOS RELACIONAL

DIRETRIZES INFORMAIS DE PROJETO (ELMASRI, R. NAVATHE, S. B. 2011)

- **Garantir que a semântica dos atributos seja clara no esquema.**
 - Projete um esquema de relação de modo que seja fácil explicar seu significado.
 - Não combine atributos de vários tipos de entidade e de relacionamento em uma única relação.
 - “A facilidade com que o significado dos atributos de uma relação pode ser explicado é uma medida informal de quão bem a relação está projetada.”
- **Reduzir a informação redundante nas tuplas.**
 - Projete os esquemas de relação da base de modo que nenhuma anomalia de inserção, exclusão ou modificação esteja presente nas relações.

PROJETO DE BANCO DE DADOS RELACIONAL

DIRETRIZES INFORMAIS DE PROJETO (ELMASRI, R. NAVATHE, S. B. 2011)

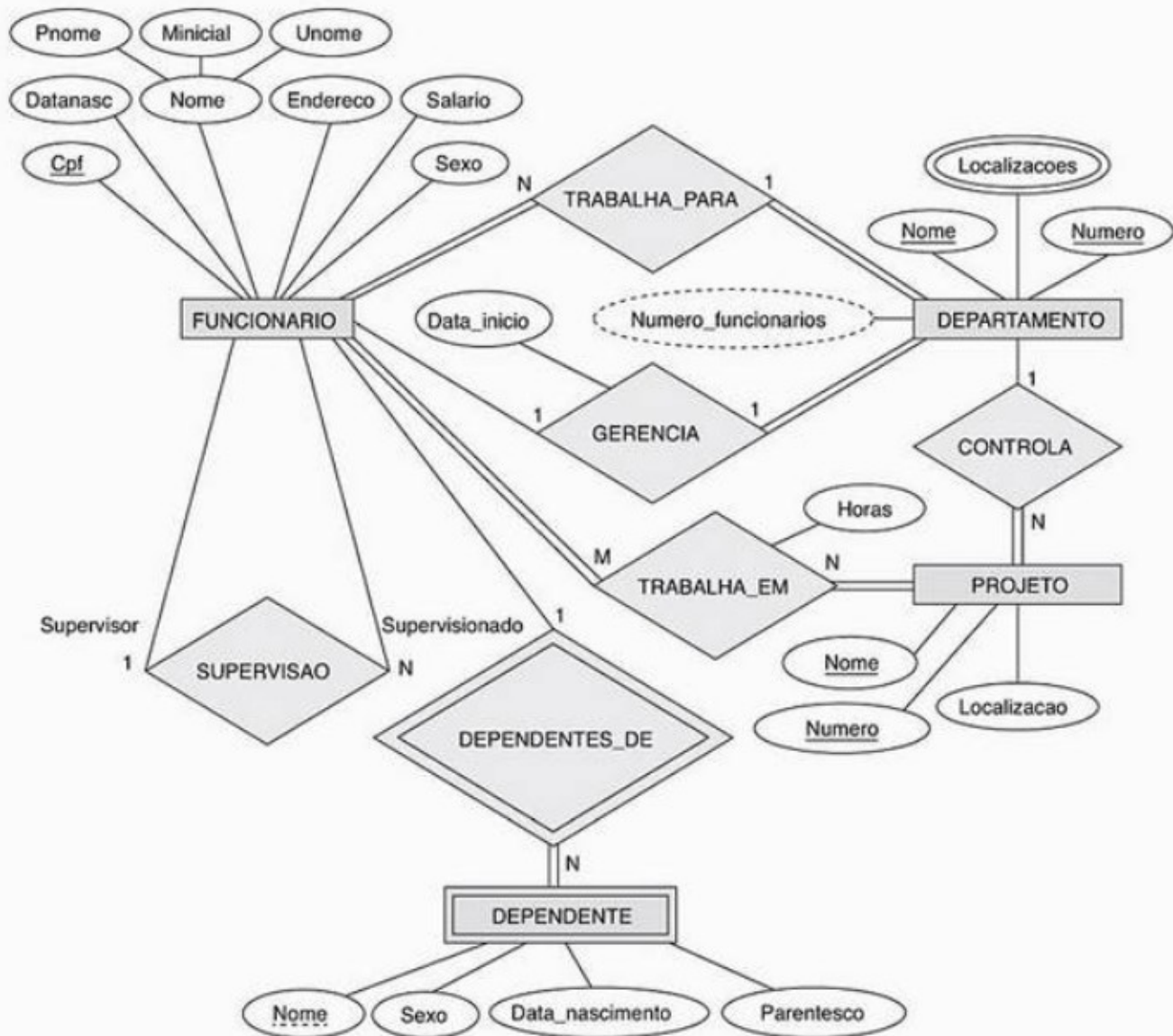
- **Reduzir os valores NULL nas tuplas.**

- Evite colocar atributos em uma relação cujos valores podem ser NULL com muita frequência.

- **Anular a possibilidade de gerar tuplas falsas.**

- Projete esquemas de relação de modo que possam ser relacionados com condições de igualdade (chave estrangeira -> chave primária).
- Evite relações com atributos correspondentes que não sejam combinações (chave estrangeira -> chave primária), pois a junção sobre esses atributos pode produzir resultados falsos.

DIAGRAMA DE ESQUEMA CONCEITUAL ER PARA BANCO DE DADOS EMPRESA. FONTE: ELMASRI,R. NAVATHE, S. B. 2011



FUNC_DEP

FNOME	CPF	DATANASC	ENDERECO	DNUMERO	DNOME	CPF_GERENTE
Silva, João B.	123456789	09-01-1965	Rua das oligarquias, 832, Natal, RN	5	Pesquisa	123456789
Soares, Zezo	098765432	08-12-1975	Rua dos políticos, 432, Natal, RN	4	Aministração	123456789
Lima, Ronaldo K.	321456098	19-01-1985	Rua da honestidade, 12, Natal, RN	4	Aministração	321456987
Leite, Joice A.	321456987	20-06-1989	Rua da República, 2018, Natal, RN	5	Pesquisa	321456987

FUNC_PROJ

CPF	PROJNUMERO	HORAS	FNOME	PROJNOME	PROJLOCAL
123456789	1	32,5	Silva, João B.	ProdutoX	Mossoró
098765432	2	7,5	Silva, João B.	ProdutoY	Natal
321456098	3	40,0	Lima, Ronaldo K.	ProdutoZ	Currais Novos
321456987	1	20,0	Leite, Joice A.	ProdutoX	Mossoró

PROJETO DE BANCO DE DADOS RELACIONAL

EXEMPLOS DE VIOLAÇÃO

1) **Garantir que a semântica dos atributos seja clara no esquema.**

Uma tupla no esquema Func_Dep inclui informações de duas outras entidades.

Uma tupla no esquema Func_Proj relaciona um funcionário (CPF) a um projeto (projnumero), mas também inclui nome do funcionário, nome e local do projeto (Relacionamento, projeto e trabalha_em).

2) **Reduzir a informação redundante nas tuplas.**

3) **Reduzir os valores NULL nas tuplas.**

4) **Anular a possibilidade de gerar tuplas falsas.**

PROJETO DE BANCO DE DADOS RELACIONAL

EXEMPLOS DE VIOLAÇÃO

- 1) Garantir que a semântica dos atributos seja clara no esquema.
- 2) **Reduzir a informação redundante nas tuplas.**
 - 1) Compare a relação Func_Dep com as relações Funcionário e Departamento
- 3) Reduzir os valores NULL nas tuplas.
- 4) Anular a possibilidade de gerar tuplas falsas.

FUNC_DEP

FNOME	CPF	DATANASC	ENDERECO	DNUMERO	DNOME	CPF_GERENTE
Silva, João B.	123456789	09-01-1965	Rua das oligarquias, 832, Natal, RN	5	Pesquisa	123456789
Soares, Zezo	098765432	08-12-1975	Rua dos políticos, 432, Natal, RN	4	Aministração	321456987
Lima, Ronaldo K.	321456098	19-01-1985	Rua da honestidade, 12, Natal, RN	4	Aministração	321456987
Leite, Joice A.	321456987	20-06-1989	Rua da República, 2018, Natal, RN	5	Pesquisa	123456789

FNOME	CPF	DATANASC	ENDERECO	DNUMERO
Silva, João B.	123456789	09-01-1965	Rua das oligarquias, 832, Natal, RN	5
Soares, Zezo	098765432	08-12-1975	Rua dos políticos, 432, Natal, RN	4
Lima, Ronaldo K.	321456098	19-01-1985	Rua da honestidade, 12, Natal, RN	4
Leite, Joice A.	321456987	20-06-1989	Rua da República, 2018, Natal, RN	5

FUNCIONARIO

DNUMERO	DNOME	CPF_GERENTE
5	Pesquisa	123456789
4	Aministração	123456789

DEPARTAMENTO

REDUNDÂNCIA E ANOMALIAS DE ATUALIZAÇÃO

O armazenamento de junções naturais de relações leva a um problema adicional: as **ANOMALIAS DE ATUALIZAÇÃO**:

○ Anomalias de Inserção

- Para se inserir uma nova tupla na Tabela 1, é preciso inserir também um novo Departamento, pois assim como CPF é chave, DNumero também é, então é possível efetuar a inserção apenas se ambas existirem;
- Como inserir um departamento que ainda não possui funcionários?

○ Anomalias de Remoção

- Se os dados de algum empregado for apagado da tabela, também se perde a informação sobre o departamento que esse empregado está vinculado.

○ Anomalias de Modificação

- Imagine que o departamento 5 muda de nome. Para alterar essa informação na base de dados seria necessário alterar as tuplas dos funcionários Leite e Silva.

PROJETO DE BANCO DE DADOS RELACIONAL

DIRETRIZES INFORMAIS DE PROJETO

- A relação FUNC_DEP está mal projetada. Ela poderia ser dividida em duas relações: Funcionario e Departamento.

FUNC_DEP

FNOME	CPF	DATANASC	ENDERECO	DNUMERO	DNOME	CPF_GERENTE
Silva, João B.	123456789	09-01-1965	Rua das oligarquias, 832, Natal, RN	5	Pesquisa	123456789
Soares, Zezo	098765432	08-12-1975	Rua dos políticos, 432, Natal, RN	4	Aministração	123456789
Lima, Ronaldo K.	321456098	19-01-1985	Rua da honestidade, 12, Natal, RN	4	Aministração	321456987
Leite, Joice A.	321456987	20-06-1989	Rua da República, 2018, Natal, RN	5	Pesquisa	321456987

PROJETO DE BANCO DE DADOS RELACIONAL

DEPENDÊNCIA FUNCIONAL

- Uma dependência funcional é uma restrição entre dois conjuntos de atributos em uma relação.

Suponha que um esquema de banco de dados relacional possua n atributos A_1, A_2, \dots, A_n .

Suponha que o banco de dados inteiro é descrito por uma única relação *universal* $R = \{A_1, A_2, \dots, A_n\}$.

(Não, não armazenaremos tudo em uma única tabela, apenas vamos representá-lo, para essa definição)

Uma **dependência funcional**, indicada por $X \rightarrow Y$, entre dois conjuntos de atributos X e Y que são subconjuntos de R , especifica uma restrição sobre possíveis tuplas que podem formar um estado de relação $r(R)$.

A restrição é que, para quaisquer duas tuplas t_1 e t_2 em r que tenham $t_1[X] = t_2[X]$, elas também devem ter $t_1[Y] = t_2[Y]$.

PROJETO DE BANCO DE DADOS RELACIONAL

DEPENDÊNCIA FUNCIONAL

- Uma dependência funcional $X \rightarrow Y$ pode ser lida como:
 - "X determina Y";
 - "O valor de Y em uma tupla depende do valor de X";
 - "Y depende de X";
- Pode-se dizer que existe uma dependência funcional de X para Y.
- Considere a relação FUNCPROJ abaixo.
 - CPF \rightarrow FNOOME
 - PROJNUMERO \rightarrow {PROJNOME, PROJLOCAL}
 - {CPF, PROJNUMERO} \rightarrow Horas

FUNCPROJ	<u>CPF</u>	<u>PROJNUMERO</u>	HORAS	FNOOME	PROJNOME	PROJLOCAL
----------	------------	-------------------	-------	--------	----------	-----------

PROJETO DE BANCO DE DADOS RELACIONAL

NORMALIZAÇÃO

- O processo de NORMALIZAÇÃO, leva o esquema projetado a uma série de testes comparando-o com as **formas normais** definidas.
 - Inicialmente foram definidas por Codd (1972) 3 formas normais;
 - Um aprofundamento da terceira forma normal foi proposto por Boyce e Codd;
 - Posteriormente foram propostas mais duas formas normais;
- Dessa forma, tem-se:
 - Primeira Forma Normal - 1FN;
 - Segunda Forma Normal - 2FN;
 - Terceira Forma Normal - 3FN;
 - Forma Normal de Boyce-Codd - FNBC;

 - Quarta Forma Normal - 4FN;
 - Quinta Forma Normal - 5FN;

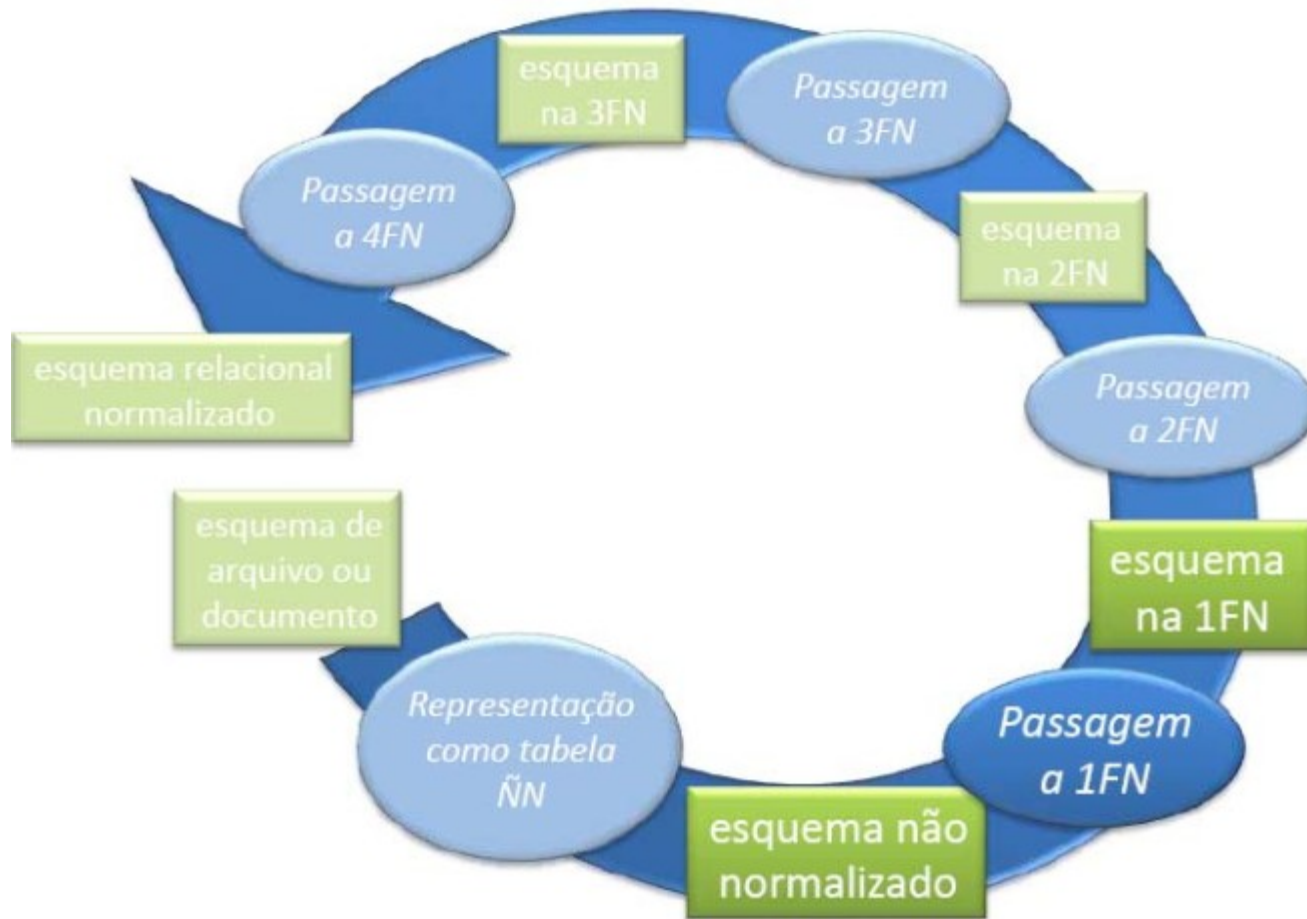
PROJETO DE BANCO DE DADOS RELACIONAL

NORMALIZAÇÃO

- O processo de NORMALIZAÇÃO, fornece aos projetistas de banco de dados:
 - Uma estrutura formal para analisar os esquemas de relação com base nas chaves e nas dependências funcionais.
 - Uma série de testes de forma normal que podem ser executados em relações individuais.

PROJETO DE BANCO DE DADOS RELACIONAL

NORMALIZAÇÃO



FORMAS NORMAIS

PRIMEIRA FORMA NORMAL - 1FN

A 1FN afirma que o domínio de um atributo deve incluir apenas valores atômicos (simples, indivisível).

- O valor de qualquer atributo em uma tupla deve ser um único valor do domínio desse atributo.

DNOME	DNUMERO	CPF_GERENTE	DLOCAL
Pesquisa	5	123456789	Natal, Mossoró
TI	4	123456789	Mossoró
RH	5	321456987	Natal, Caicó



DNOME	DNUMERO	CPF_GERENTE	DLOCAL
Pesquisa	5	123456789	Natal
TI	4	123456789	Mossoró
Pesquisa	5	321456987	Mossoró

FORMAS NORMAIS

SEGUNDA FORMA NORMAL - 2FN

Um esquema de relação R está em 2FN se cada atributo não principal A em R for total e funcionalmente dependente da chave primária de R .

- DEPENDÊNCIA TOTAL DA CHAVE
- O teste para a 2FN envolve testar as dependências funcionais $X \rightarrow Y$, cujos atributos do lado esquerdo fazem parte da chave primária.
- Se a chave primária tiver um único atributo, o teste não precisa ser aplicado.

FORMAS NORMAIS

SEGUNDA FORMA NORMAL - 2FN

- Uma dependência funcional $X \rightarrow Y$ é uma **dependência funcional total** se a remoção de qualquer atributo A de X significar que a dependência não se mantém mais.
 - Para qualquer atributo $A \in X$, $(X - \{A\})$ não determina Y funcionalmente.
- Uma dependência funcional $X \rightarrow Y$ é uma **dependência parcial** se algum atributo $A \in X$ puder ser removido de X e a dependência ainda se mantiver.
 - Para qualquer atributo $A \in X$, $(X - \{A\}) \rightarrow Y$.

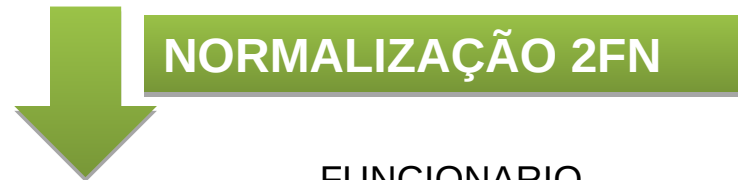
FUNCPROJ	<u>CPF</u>	<u>PROJNUMERO</u>	HORAS	FNOME	PROJNOME	PROJLOCAL	DEPENDÊNCIA FUNCIONAL
							TOTAL
							PARCIAL
							PARCIAL

FORMAS NORMAIS

SEGUNDA FORMA NORMAL - 2FN

<u>CPE</u>	<u>PROJNUMERO</u>	HORAS	FNOME	PROJNOME	PROJLOCAL	DEPENDÊNCIA FUNCIONAL
						TOTAL
						PARCIAL
						PARCIAL

FUNCPROJ



PROJETO

<u>PROJNUMERO</u>	PROJNOME	PROJLOCAL

FUNCIONARIO

<u>CPE</u>	FNOME

FUNCPROJ

<u>CPE</u>	<u>PROJNUMERO</u>	HORAS

FORMAS NORMAIS

TERCEIRA FORMA NORMAL - 3FN

Um esquema de relação R está na 3FN se ele satisfaz a 2FN e nenhum atributo não principal de R for transitivamente dependente da chave primária.

- ELIMINAR DEPENDÊNCIAS FUNCIONAIS TRANSITIVAS
- Uma **dependência é transitiva** se em $X \rightarrow Y$ houver um conjunto de atributos Z em R que nem sejam chave candidata nem um subconjunto de qualquer chave de R e:
 - $X \rightarrow Z$
 - $Z \rightarrow Y$

FNOME	<u>CPF</u>	DATAN ASC	ENDERECO	DNUMERO	DNOME	CPF_GERENTE	DEPENDENCIA FUNCIONAL
							TOTAL
							TRANSITIVA

FUNCDEP

FORMAS NORMAIS

TERCEIRA FORMA NORMAL - 3FN

FUNCDEP

FNOME	<u>CPF</u>	DATAN ASC	ENDERECO	DNUMERO	DNOME	CPF_GERENTE	DEPENDENCIA FUNCIONAL
							TOTAL
							TRANSITIVA



FUNCIONARIO

FNOME	<u>CPF</u>	DATAN ASC	ENDERECO	DNUMERO

DEPARTAMENTO

<u>DNUMERO</u>	DNOME	CPF_GERENTE

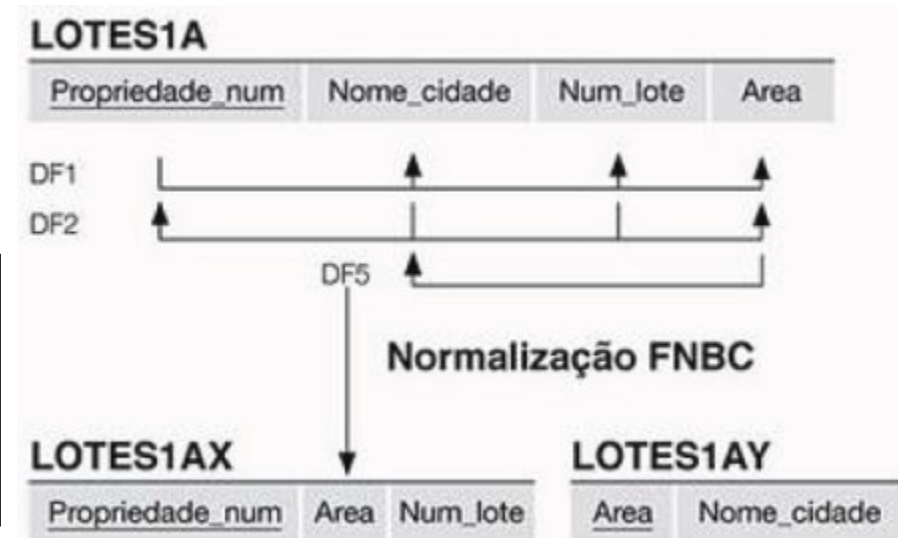
FORMAS NORMAIS

TERCEIRA FORMA NORMAL DE BOYCE-CODD - FNBC

Um esquema de relação R está na FNBC se toda vez que uma dependência funcional $X \rightarrow A$ se mantiver em R, X for uma super-chave de R.

- Toda relação na FNBC está na 3FN, mas uma relação na 3FN não necessariamente está na FNBC.
 - A FNBC é mais rigorosa.
- Não pode haver dependência funcional com um atributo não chave.

“Na prática, a maioria dos esquemas de relação que estão na 3FN também estão na FNBC. Somente se $X \rightarrow A$ se mantiver em um esquema de relação R com X não sendo uma superchave e A sendo um atributo principal é que R estará na 3FN, mas não na FNBC.”



PROJETO DE BANCO DE DADOS RELACIONAL

○ Observe a relação abaixo:

<u>Enome</u>	<u>Projnome</u>	<u>Dnome</u>
Silva	X	João
Silva	Y	Ana
Silva	X	Ana
Silva	Y	João

- A relação está em qual FN?
- Ainda restam problemas?

PROJETO DE BANCO DE DADOS RELACIONAL

DEPENDÊNCIA MULTIVALORADA

- As dependências multivaloradas são uma consequência da primeira forma normal (1FN).

Uma dependência multivalorada $X \twoheadrightarrow Y$ especificada sobre o esquema de relação R, onde X e Y são subconjuntos de R, determina a seguinte restrição sobre qualquer estado de relação r de R:

Se duas tuplas t1 e t2 existirem em r tais que $t1[X] = t2[X]$, então duas tuplas t3 e t4 também deverão existir em r com as seguintes propriedades, nas quais usamos Z para indicar $(R - (X \cup Y))$:

- $t3[X] = t4[X] = t1[X] = t2[X]$
- $t3[Y] = t1[Y]$ e $t4[Y] = t2[Y]$
- $t3[Z] = t2[Z]$ e $t4[Z] = t1[Z]$

Fnome	Projnome	Dnome
Silva	X	João
Silva	Y	Ana
Silva	X	Ana
Silva	Y	João

PROJETO DE BANCO DE DADOS RELACIONAL

DEPENDÊNCIA MULTIVALORADA

- Uma MVD $X \twoheadrightarrow Y$ em R é chamada de **MVD trivial** se:
 - Y for um subconjunto de X ; ou
 - $X \cup Y = R$

<u>Enome</u>	<u>Projnome</u>
Silva	X
Silva	Y

- Uma MVD que não satisfaz ambas as condições é, portanto, chamada **MVD NÃO TRIVIAL**.

- Relações com MVD não triviais tendem a ser relações cuja chave é composta por todos os atributos juntos (relações de todas as chaves).

<u>Enome</u>	<u>Projnome</u>	<u>Dnome</u>
Silva	X	João
Silva	Y	Ana
Silva	X	Ana
Silva	Y	João

FORMAS NORMAIS

QUARTA FORMA NORMAL - 4FN

Um esquema de relação R está na 4FN com relação a um conjunto de dependências F (que inclui dependências funcionais e dependências multivaloradas) se, para cada dependência multivalorada não trivial $X \twoheadrightarrow Y$ em F , X é uma superchave para R .

- Uma relação de todas as chaves está sempre na FNBC, pois não tem DFs.
- Uma relação de todas as chaves não está na 4FN.
- Uma relação que não está na 4FN devido a uma MVD não trivial precisa ser decomposta para convertê-la em um conjunto de relações na 4FN.
- A decomposição remove a redundância causada pela MVD.

FORMAS NORMAIS

QUARTA FORMA NORMAL - 4FN

O processo de normalização se dá decompondo a relação separando cada MVD em uma relação distinta, transformando as MVDs em triviais:

FUNC		
<u>Enome</u>	<u>Projnome</u>	<u>Dnome</u>
Silva	X	João
Silva	Y	Ana
Silva	X	Ana
Silva	Y	João

FUNC_PROJETOS	
<u>Enome</u>	<u>Projnome</u>
Silva	X
Silva	Y

FUNC_DEPENDETES	
<u>Enome</u>	<u>Nome_dependente</u>
Silva	João
Silva	Ana

PROJETO DE BANCO DE DADOS RELACIONAL

DEPENDÊNCIA DE JUNÇÃO

- A dependência de junção é semântica e, por isso, difícil e raramente detectada na prática.

Uma **dependência de Junção (DJ)**, indicada por $DJ(R_1, R_2, \dots, R_n)$, especificada no esquema de relação R , determina uma restrição sobre os estados r de R . A restrição indica que cada estado válido r de R deve ter uma decomposição de junção não aditiva para R_1, R_2, \dots, R_n .

$$*(\pi_{R_1}(r), \pi_{R_2}(r), \dots, \pi_{R_n}(r)) = r$$

- Ou seja, a junção das decomposições de $r(R)$ deve ser exatamente igual ao estado da relação r , sem tuplas falsas.

PROJETO DE BANCO DE DADOS RELACIONAL

DEPENDÊNCIA DE JUNÇÃO

- Observe a geração de tuplas falsas na junção Fornece:

<u>Nome_fornece</u>	<u>Nome_peca</u>	<u>Nome_proj</u>
Silva	Peneira	ProjX
Silva	Porca	ProjY
Adam	Peneira	ProjY
Walter	Porca	ProjZ
Adam	Prego	ProjX
Adam	Peneira	ProjX
Silva	Peneira	ProjY

PROJETO DE BANCO DE DADOS RELACIONAL

QUINTA FORMA NORMAL – PROJEÇÃO-JUNÇÃO

- Pode não haver decomposição de junção não aditiva de R em dois esquemas de relação, mas pode haver uma decomposição de junção não aditiva em mais de dois esquemas.
- Pode não haver dependência funcional em R que viole qualquer forma normal até a FNBC, e também pode não haver MVD não trivial que viole a 4FN.
- O que diz a 5FN:

Um esquema de relação R está na quinta forma normal (5FN) (ou forma normal projeção-junção – FNPJ) com relação a um conjunto F de dependências funcionais, multivaloradas e de junção se, para cada dependência de junção não trivial DJ (r_1, R_2, \dots, R_n) em F^+ , cada R_i é uma superchave de R.

FORNECE

<u>Nome_fornece</u>	<u>Nome_peca</u>	<u>Nome_proj</u>
Silva	Peneira	ProjX
Silva	Porca	ProjY
Adam	Peneira	ProjY
Walter	Porca	ProjZ
Adam	Prego	ProjX

Adam	Peneira	ProjX
Silva	Peneira	ProjY

R1

<u>Nome_fornece</u>	<u>Nome_peca</u>
Silva	Peneira
Silva	Porca
Adam	Peneira
Walter	Porca
Adam	Prego

R2

<u>Nome_fornece</u>	<u>OR_proj</u>
Silva	ProjX
Silva	ProjY
Adam	ProjY
Walter	ProjZ
Adam	ProjX

R3

<u>Nome_peca</u>	<u>OR_proj</u>
Peneira	ProjX
Porca	ProjY
Peneira	ProjY
Porca	ProjZ
Prego	ProjX

FORMAS NORMAIS

NORMALIZAÇÃO

- Com o processo de normalização:
 - Aumenta o número de relações do esquema;
 - Diminui a ocorrência de redundâncias;

DÚVIDAS?



REFERÊNCIAS BIBLIOGRÁFICAS

ELMASRI, Ramez. NAVATHE, Shamkant B. Sistemas de Banco de Dados. **Projeto de Banco de Dados**. 6ª Edição. São Paulo. Pearson Addison Wesley, 2011.