

# Bases de Dados

## Normalização - Parte 2

Profa. Elaine Parros Machado de Sousa

# Normalização

- Formas Normais baseadas em dependências funcionais
  - baseadas em chave primária
    - 1ª FN
    - 2ª FN
    - 3ª FN
  - baseadas em chaves candidatas
    - definições genéricas de 2ª FN e 3ª FN
    - FN de Boyce-Codd (BCNF)
- Forma Normal baseada em dependências multivaloradas
  - 4ª FN

Parte 1

Parte 2

# Normalização

- Formas Normais baseadas em dependências funcionais
  - baseadas em chave primária
    - 1ª FN
    - 2ª FN
    - 3ª FN
  - baseadas em chaves candidatas
    - definições genéricas de 2ª FN e 3ª FN
    - FN de Boyce-Codd (BCNF)
- Forma Normal baseada em dependências multivaloradas
  - 4ª FN

# Normalização

- **Definições genéricas**  $\Rightarrow$  levam todas as chaves candidatas em consideração
  - **2ª FN genérica**  $\Rightarrow$  todos os atributos **não primários** possuem **dependência total**, transitiva ou não, de **todas as chaves** (primária, secundária, ...)
  - **3ª FN genérica**  $\Rightarrow$  todos os atributos **não primários** possuem **dependência total, não transitiva**, de **todas as chaves** (primária, secundária, ...)

# Exemplo

Matrícula → Municipio, NroLote, Área, Preço

Municipio, NroLote → Área, Matrícula, Preço

Área → Preço

**Lotes = {Matricula, Municipio, NroLote, Área, Preço }**

➤ Analise a relação **Lotes** considerando as definições genéricas de 2ª FN e 3ª FN

- 2ª FN genérica? **OK!!**
- 3ª FN genérica? **NÃO!!**

# Exemplo

Matrícula  $\rightarrow$  Municipio, NroLote, Área, Preço

Municipio, NroLote  $\rightarrow$  Área, Matrícula, Preço

Área  $\rightarrow$  Preço

Lotes = {Matricula, Municipio, NroLote, Área, Preço }

➤ **Normalizando para 3ª FN Genérica ....**

Lotes = {Matricula, Municipio, NroLote, Área}

Área = {Área, Preço}



# Exemplo

Estudante, Curso  $\rightarrow$  Instrutor

Instrutor  $\rightarrow$  Curso

**Treinamento = {Estudante, Curso, Instrutor}**

➤ Analisando a relação **Treinamento** considerando as definições genéricas de 2ª FN e 3ª FN...

- 2ª FN genérica? **OK!!**
- 3ª FN genérica? **OK!!**

**Mas... e a DF Instrutor  $\rightarrow$  Curso?**

# Forma Normal de Boyce - Codd (BCNF)

- **BCNF**  $\Rightarrow$  extensão da 3ª FN genérica
- uma relação R **está** na **BCNF** se:
  - estiver na 3ª FN genérica
  - **para toda DF não-trivial  $X \rightarrow A$**  válida para a relação R, **X** é uma **chave** em R



# Forma Normal de Boyce - Codd (BCNF)

- Na prática:
  - maioria das relações em 3FN genérica também está na BCNF
  - Exceção:
    - quando  $X \rightarrow A$  e:
      - X não é chave
      - A é atributo primário

# Exemplo

Estudante, Curso  $\rightarrow$  Instrutor

Instrutor  $\rightarrow$  Curso

**Treinamento = {Estudante, Curso, Instrutor}**

- A relação **Treinamento** está na BCNF?
- Alternativas de normalização por decomposição?

# Exemplo

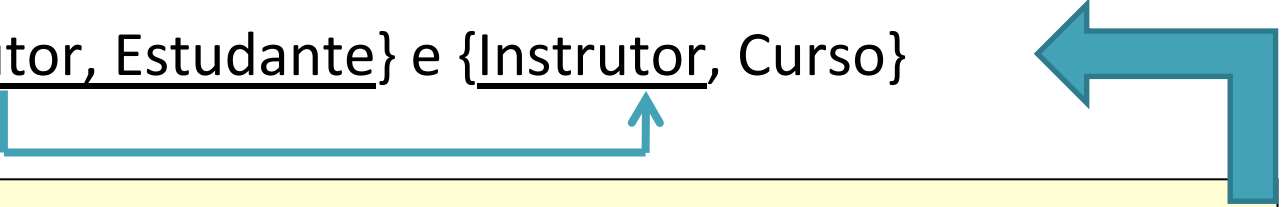
Estudante, Curso  $\rightarrow$  Instrutor

Instrutor  $\rightarrow$  Curso

Treinamento = {Estudante, Curso, Instrutor}

## ➤ Alternativas de decomposição:

- 1) {Estudante, Instrutor} e {Estudante, Curso}
- 2) {Curso, Instrutor} e {Curso, Estudante}
- 3) {Instrutor, Estudante} e {Instrutor, Curso}



**Quais os problemas de cada alternativa?  
Qual seria a melhor opção? Por que?**

# Exemplo

Estudante, Curso  $\rightarrow$  Instrutor

Instrutor  $\rightarrow$  Curso

Treinamento = {Instrutor, Estudante}



Instrutor = {Instrutor, Curso}

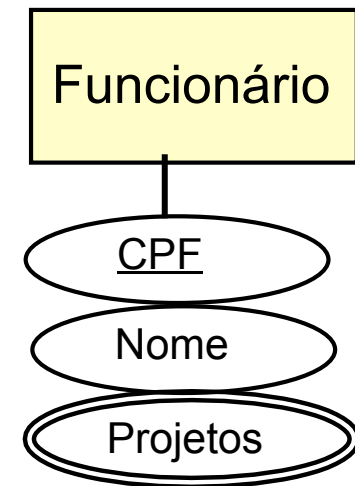
**Mas essa solução está realmente boa?  
Garantindo consistência dos dados  
(considerando as DFs)?**

# Normalização

- Formas Normais baseadas em dependências funcionais
  - baseadas em chave primária
    - 1ª FN
    - 2ª FN
    - 3ª FN
  - baseadas em chaves candidatas
    - definições genéricas de 2ª FN e 3ª FN
    - FN de Boyce-Codd (BCNF)
- Forma Normal baseada em **dependências multivaloradas**
  - 4ª FN

# Exemplo (simplificado...)

- Dados sobre funcionários na base de dados de uma empresa:
  - CPF e nome do funcionário
  - nomes dos projetos em que trabalha
- dependências funcionais
  - $\text{CPF} \rightarrow \text{Nome}$
  - Projeto?

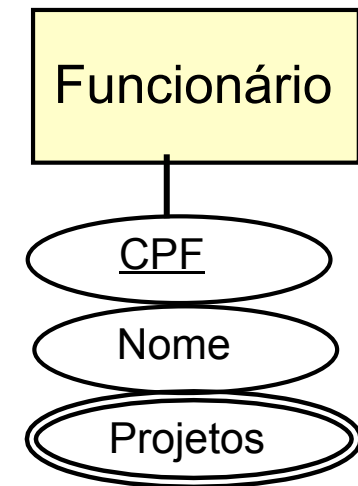


## Exemplo (cont.)

- Semanticamente:
  - um **conjunto** de valores de **Projeto** é determinado por **um** valor de **CPF** de funcionário, e somente por **CPF**



## Exemplo de Cenário de Dependência Multivalorada



# Dependência Multivalorada

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

$$\mathbf{A -\>> B}$$

- **A multidetermina B** (ou B é multidependente de A)  $\Rightarrow$  o **conjunto de valores** de B é determinado pelo valor de A, e **somente** pelo valor de A



# Dependência Multivalorada

- Dada uma DM  $X \twoheadrightarrow Y$  em  $R$

- se:

- (a)  $Y \subseteq X$  ou

- (b)  $X \cup Y = R$



**Dependência  
Multivalorada Trivial**

- caso contrário



**Dependência  
Multivalorada Não-Trivial**

# Exemplos

- DM Trivial (DMT)

CPF-» Projeto

Funcionario\_Projeto = {CPF, Projeto}

- DM Não-Trivial (DMNT)

CPF-» Projeto

Funcionario = {CPF, Projeto, nome}

# Dependência Multivalorada

- DM são semânticas (assim como as DF)
- **Simetria** na definição de DM:
  - se  $X \twoheadrightarrow Y$ , então

$$X \twoheadrightarrow Z, Z = R - (X \cup Y)$$

Funcionario = {CPF, Projeto, nome}

CPF  $\twoheadrightarrow$  Projeto



CPF  $\twoheadrightarrow$  Nome

# Exemplo

Funcionario = {CPF, Projeto, nome}

CPF -» Projeto

➤ Dados:

- Ana (CPF 123) trabalha nos projetos Museu Virtual e Cidadania;

➤ Como **armazenar** os dados na tabela Funcionário?

## Exemplo (cont.)

- Ana (CPF 123) trabalha nos projetos Museu Virtual e Cidadania;

Funcionario = {CPF, Projeto, Nome}  
{<123, Cidadania, Ana>,  
<123, Museu Virtual, Ana>}

**Anomalias?**

## 4ª Forma Normal

- Relação R está na 4ª Forma Normal se:
  - todas as dependências multivaloradas são **triviais** ou
  - para cada dependência multivalorada não-trivial  $A \twoheadrightarrow B$ , **A é uma chave** (completa) em R

# 4ª Forma Normal

## ■ Exemplos...

- Funcionario\_Projeto = {CPF, Projeto}

CPF -» Projeto

**OK!!**

- Funcionario = {CPF, Projeto, nome}

CPF -» Projeto

**Não!!**

# 4ª Forma Normal

- Intuição:

CPF -» Projeto

Funcionario = {CPF, ~~Projeto~~, nome}



Funcionario = {CPF, nome}

Projetos = {CPF, Projeto}



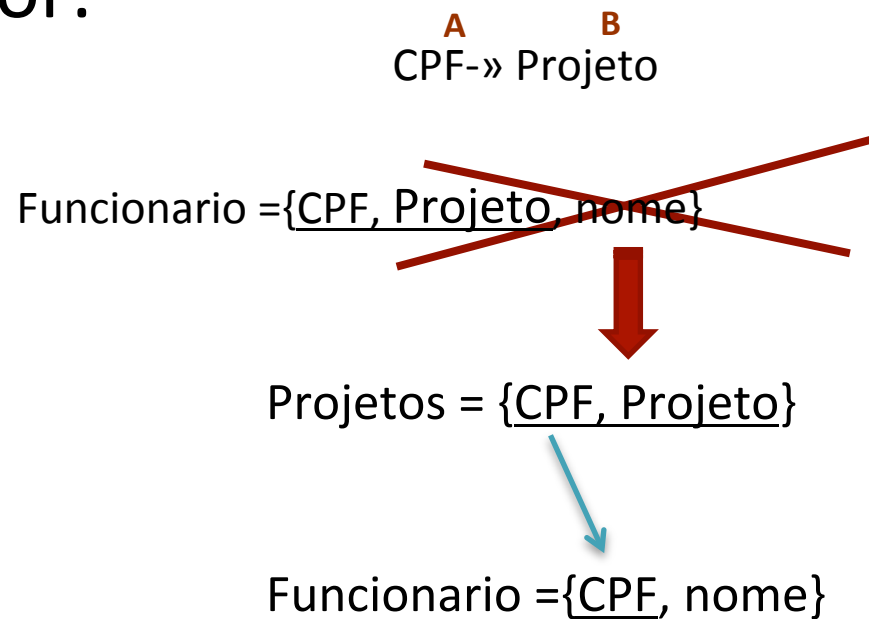
# 4ª Forma Normal

- Normalizando a relação para a 4ª FN....

➤ dada uma DMNT **A -» B** na relação **R**,  
substitui-se R por:

- **A ∪ B** e

- **R - B**

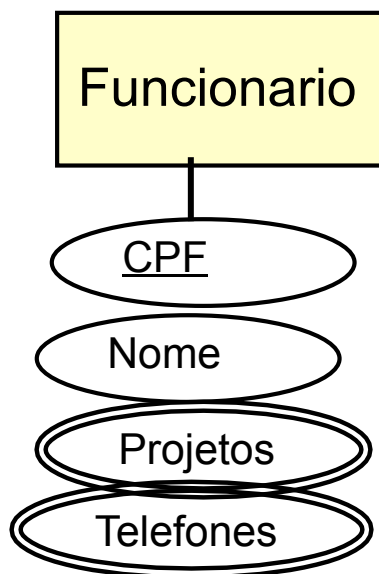


# 4ª Forma Normal

- Evita redundância de tuplas  $\Rightarrow$  evita anomalias de inclusão/remoção/alteração
- Cenários típicos de normalização para 4ªFN (ocorrências de DMNT):
  - atributos **multivalorados** ou **CR N:N** armazenados **de maneira incorreta** em múltiplas linhas
    - Ex: Funcionario = {CPF, projeto, nome}
  - atributos **multivalorados independentes** misturados em uma única tabela
    - Ex: Func\_Projetos\_Telefones = {CPF, Projeto, Telefone}

# Observação

- Lembrando o mapeamento ME-R → Modelo Relacional
  - multivalorados em geral definem novas relações
    - sem redundância e sem anomalias



Funcionario = {CPF, Nome}

Projetos = {CPF, Projeto}

Telefones = {CPF, Telefone}

# Considerações Finais (cont.)...

## ■ Normalização:

- ✓ uma relação por vez –
  - análise de 1ªFN, 2ª FN, 3ª FN, BCNF e 4ª FN
- ✓ **forma normal de uma relação**  $\Rightarrow$  forma normal mais restrita atendida
- ✓ **forma normal da base de dados**  $\Rightarrow$  forma normal mais restrita atendida por todas as relações

# Considerações Finais (cont.)...

## ■ Lembrando...

### ■ Propriedades desejáveis:

- 1) decomposição **sem perda de junção**  
(sem geração de tuplas ilegítimas)
- 2) decomposição com **preservação de dependências** (possibilidade de avaliar a DF  $\Rightarrow$  atributos na mesma tabela)

# Sugestão de Leitura

- **ELMASRI, R; NAVATHE, S.B. – *Sistemas de Banco de Dados*, Addison Wesley**
  - 4ª Edição
    - **Capítulos 10 e 11**
  - 6ª Edição
    - **Capítulo 15**

## EXERCÍCIO

Considere as seguintes relações e suas dependências funcionais.

1) Lotes = {ID-propriedade, municipio, nro\_lote, regiao, area, preço, imposto}

ID-propriedade  $\rightarrow$  municipio, nro\_lote, regiao, area, preço, imposto

municipio, nro\_lote  $\rightarrow$  ID-propriedade, regiao, area, preço, imposto

municipio  $\rightarrow$  imposto

area  $\rightarrow$  preço

regiao  $\rightarrow$  municipio

2) Carros\_Vendidos = {carro, data-venda, vendedor, %comissão, desconto-do-dia}

carro  $\rightarrow$  data-venda;

data-venda  $\rightarrow$  desconto-do-dia;

vendedor  $\rightarrow$  %comissão

carro  $\twoheadrightarrow$  vendedor

# EXERCÍCIO

- a) Indique se as formas normais 1FN, 2FN (baseada em PK e genérica), 3NF (baseada em PK e genérica) , BCNF e 4FN são atendidas.
- b) Normalize as relações para atender a cada uma dessas formas (faça todos os passos). Explique quais propriedades de decomposição são atendidas ou não, e o porquê.
- c) Discuta a influência dessa normalização no desempenho de consultas e dê exemplos.