

# Normalização de BDs Relacionais



Melissa Lemos

melissa@inf.puc-rio.br

1

## Introdução

- Processo de normalização
  - Aplicar uma série de testes para se certificar se o esquema definido satisfaz a forma normal.
  - Inicialmente Codd propôs 1FN, 2FN e 3FN
  - Depois uma definição mais forte da 3FN foi proposta por Boyce-Codd e chamada de BCNF (Boyce-Codd Normal Form)
  - Todas estas formas normais são baseadas nas dependências funcionais entre os atributos de uma relação.
  - Depois ainda foram propostas a 4FN e a 5FN.

2

## Introdução

- O processo de normalização tem o objetivo de atender as propriedades
  - Minimizar redundância
  - Minimizar anomalias de inserção, remoção e atualização
- As relações que não satisfizerem as condições são decompostas em esquemas de relações menores que estão de acordo com as condições e possuem as propriedades anteriores.

3

## Definição

- **Super-chave** de uma relação
$$R = \{A_1, A_2, \dots, A_n\}$$
é um conjunto de atributos  $S \subseteq R$  com a seguinte propriedade
  - Não há duas tuplas  $t_1$  e  $t_2$  em um estado da relação  $r$  de  $R$  tal que  $t_1[S] = t_2[S]$

4

## Definição

- **Chave K** é uma super-chave com a propriedade adicional de que a remoção de qualquer atributo de  $K$  fará com que  $K$  não seja mais uma super-chave.
- A diferença é que a chave precisa ser mínima, ou seja, se  $K = \{A_1, A_2, \dots, A_k\}$  é chave de  $R$ , então  $K - \{A_i\}$  não é mais chave de  $R$  para qualquer  $i$ , sendo  $1 \leq i \leq k$ .

5

## Exemplo

- `EMPLOYEE(ENAME, SSN, BDATE, ADDRESS, DNUMBER*)`
- `{SSN}` é chave
- `{SSN}`, `{SSN, ENAME}`, `{SSN, ENAME, BDATE}`, etc são super-chaves.

6

## Definição

- Quando uma relação possui mais de uma chave primária, cada uma é chamada de **chave candidata**.
- Neste caso uma delas é arbitrariamente escolhida para ser a **chave primária**, e as outras são chamadas de **chaves secundárias**.

7

## Definição

- Um atributo de uma relação R é chamado de **primo** de R se for membro de uma chave candidata de R.
- Em caso contrário, o atributo é não primo.
- Ex.:
  - WORKS\_ON (SSN\*, PNUMBER\*, HOURS)
    - SSN e PNUMBER são primos
    - HOURS é não primo

8

## Primeira Forma Normal

- Uma relação não deve ter atributos não atômicos ou relações aninhadas.

9

## Exemplo – 1FN

- DEPARTMENT  
(DNAME, DNUMBER, DMGRSSN\*, DLOCATIONS)

DNAME	DNUMBER	DMGRSSN	DLOCATIONS
Pesquisa	1	33445566	{Rio, São Paulo, Salvador}
Administração	2	55667788	{Rio}
Rec Humanos	3	77889900	{Cabo Frio}

↑  
~~1FN~~

10

## Primeira Forma Normal

- Normalização
  - Forme novas relações para cada atributo não atômico ou relação aninhada

11

## Exemplo – 1FN

- DEPARTMENT  
(DNAME, DNUMBER, DMGRSSN\*, DLOCATIONS)



- DEPARTMENT  
(DNAME, DNUMBER, DMGRSSN\*), **Ok – 1FN**
- DEPT\_LOCATIONS  
(DNUMBER\*, DLOCATION)

**X – 1FN**

**Ok – 1FN**

12

## Exemplo – 1FN

- DEPARTMENT  
(DNAME, DNUMBER, DMGRSSN\*, DLOCATIONS) **X – 1FN**
- Outras saídas ...
- DEPARTMENT  
(DNAME, DNUMBER, DMGRSSN\*, DLOCATION) **Ok – 1FN**
  - Redundância
- DEPARTMENT  
(DNAME, DNUMBER, DMGRSSN\*, DLOCATION1, DLOCATION2, DLOCATION3) **Ok – 1FN**
  - Muitos nulos
  - Possível somente em alguns casos

13

## Segunda Forma Normal

- Para relações onde as chaves primárias contêm múltiplos atributos, nenhum atributo não chave deve ser dependente funcional de parte da chave primária.
- De forma geral: Uma relação R está em 2FN se todo atributo não primo A de R não for dependente parcialmente de qualquer chave de R.

14

## Segunda Forma Normal

- Uma relação está em 2FN se estiver em 1FN e todo atributo não primo não depender funcionalmente de uma parte da chave.

15

## Exemplo – 2FN

- EMP\_PROJ  
(SSN, PNUMBER, HOURS, ENAME, PNAME, PLOCATION)
- {SSN, PNUMBER} → HOURS
  - Dependência completa – OK!
    - Não: SSN → HOURS
    - Não: PNUMBER → HOURS
- {SSN, PNUMBER} → ENAME
  - Dependência parcial – Não 2FN!
  - SSN → ENAME, **X – 2FN**

16

## Segunda Forma Normal

- Normalização
  - Decomponha e crie uma nova relação para cada chave parcial e seus atributos dependentes.
  - Tenha certeza que foi mantida uma relação com a chave primária original e todos os atributos que são completamente dependentes dela.

17

## Exemplo – 2FN

- EMP\_PROJ  
(SSN, PNUMBER, HOURS, ENAME, PNAME, PLOCATION) **X – 2FN**
- Diagrama de dependência funcional:
- EP1(SSN, PNUMBER, HOURS) **Ok – 2FN**
- EP2(SSN, ENAME)
- EP3(PNUMBER, PNAME, PLOCATION)

18

## Terceira Forma Normal

- Uma relação não deve ter um atributo que não é chave determinado funcionalmente por outro atributo não chave (ou por um conjunto de atributos não chave).
- Ou seja, não deve haver dependência transitiva.
  - uma dependência funcional  $X \rightarrow Y$  em uma relação  $R$  é uma **dependência transitiva** se existe um conj. de atributos  $Z$  que não é chave candidata e nem um sub-conjunto de alguma chave de  $R$ , e é válido que  $X \rightarrow Z$  e  $Z \rightarrow Y$ .

19

## Terceira Forma Normal

- Uma relação está em 3FN se estiver em 2FN e todo atributo não primo depender apenas de um atributo primo.
- Ou ainda
  - If  $R$  is a relation scheme and  $F$  is a set of FD's on  $R$
  - $R$  is in 3NF if
    - whenever  $X \rightarrow A$  which is an element of  $F^+$
    - And when  $A$  is not in  $X$
  - Then
    - either  $X$  is a superkey for  $R$
    - Or  $A$  is prime (i.e. a member of a candidate key for  $R$ )

Uma relação em 3FN garante que não haverá anomalias de atualização.

20

## Exemplo – 3FN

X – 3FN

- EMP\_DEPT (ENAME, SSN, BDATE, ADDRESS, DNUMBER, DNAME, DMGRSSN)
- $SSN \rightarrow DMGRSSN$  existe através da transitividade por DNUMBER
  - $SSN \rightarrow DNUMBER$
  - $DNUMBER \rightarrow DMGRSSN$
  - DNUMBER não é chave e nem parte da chave de EMP\_DEPT
  - Intuitivamente nós vemos que a dependência entre DNUMBER e DMGRSSN não é desejada em EMP\_DEPT já que DNUMBER não é chave de EMP\_DEPT.

21

## Terceira Forma Normal

- Normalização
  - Decomponha e crie uma relação que inclua os atributos não chave que determinem funcionalmente os outros atributos não chave.

22

## Exemplo – 3FN

- EMP\_DEPT (ENAME, SSN, BDATE, ADDRESS, DNUMBER, DNAME, DMGRSSN) X – 3FN
- ↓
- ED1 (ENAME, SSN, BDATE, ADDRESS, DNUMBER) Ok – 3FN
- ED2 (DNUMBER, DNAME, DMGRSSN)

23

## Boyce-Codd Normal Form

- Uma relação em BCNF é mais restrita que a 3FN pois uma relação em BCNF está de acordo com a 3FN, mas o contrário não é verdadeiro.
- Uma relação está em FNBC se todas as dependências funcionais não triviais são do tipo atributo chave determina um ou mais atributos.

24

## Boyce-Codd Normal Form

- Ou, de forma equivalente
  - Uma relação está em FNBC quando R satisfizer  $X \rightarrow A$ ,  $A \notin X$  e  $X$  for super-chave de R.
- Ou ainda
  - If  $R$  is a relation scheme and  $F$  is a set of FD's on  $R$
  - $R$  is in BCNF if
    - whenever  $X \rightarrow A$  which is an element of  $F^+$
    - and when  $A$  is not in  $X$
  - Then  $X$  is a superkey for  $R$

FNBC elimina todos os problemas de anomalias referente a DFs.

25

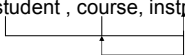
## Exemplo - FNBC

- $R(\text{cidade, end, cep})$ 
  - Chaves candidatas
    - $(\text{cidade, end})$
    - $(\text{end, cep})$
  - DFs
    - $(\text{cidade, end}) \rightarrow \text{cep}$
    - $\text{cep} \rightarrow \text{cidade}$
  - R está em 3FN
    - todo atributo não primo depende apenas de um atributo primo
  - R não está em FNBC
    - $\text{cep} \rightarrow \text{cidade}$  e  $\text{cep}$  não é chave.
    - Esta dependência funciona não é do tipo atributo chave determina um ou mais atributos

X –FNBC

26

## Exemplo - FNBC

- Teach(student, course, instructor)
- Chave candidata
  - $(\text{student, course})$
- DFs
  - $(\text{student, course}) \rightarrow \text{instructor}$
  - $\text{instructor} \rightarrow \text{course}$
- R está em 3FN
  - todo atributo não primo depende apenas de um atributo primo
- R não está em FNBC
  - $\text{instructor} \rightarrow \text{course}$  e  $\text{instructor}$  não é chave.

X –FNBC

27