

Megadados

Aula 10 – Dependências Funcionais,
Normalização

2020 – Engenharia

Fábio Ayres <fabioja@insper.edu.br>

Aula de hoje

- Dependências funcionais
- Formas normais

Dependências Funcionais

Dependências funcionais

São relacionamentos do tipo “muitos para um” entre atributos de uma relação (não confundir com relacionamentos entre relações)

Formalmente: Para dois conjuntos de atributos X e Y de uma relação, dizemos que “ Y é funcionalmente dependente de X ” ou que “ X determina funcionalmente Y ” (e denotamos $X \rightarrow Y$) se, e somente se, qualquer tupla em X tem precisamente uma tupla equivalente em Y .

Ou seja:

$X \rightarrow Y \Leftrightarrow$ duas linhas da tabela que concordam em X deverão concordar em Y

Exemplos:

Liste todas as dependências funcionais do *schema*

FilmesAtores(id_filme, filme, estúdio, data, id_ator, ator, salario)

Considere que um mesmo estúdio jamais lança filmes com mesmo título na mesma data

Use as abreviações a seguir:

Atributo	Abreviação
id_filme	#F
filme	F
estúdio	E
data	D
id_ator	#A
ator	A
salario	S

Solução

As dependências funcionais mais importantes neste problema são:

$$\{\#F\} \rightarrow \{F, E, D\}, \{\#A\} \rightarrow \{A\}, \{\#F, \#A\} \rightarrow \{S\} \text{ e } \overbrace{\{F, E, D\} \rightarrow \{\#F\}}^{\text{bizarro!}}$$

Porém existem várias outras dependências funcionais que podem ser derivadas destas através da aplicação de algumas regras:

1. Reflexão: $B \subset A \Rightarrow A \rightarrow B$
2. Aumento: $A \rightarrow B \Rightarrow AC \rightarrow BC$
3. Transitividade: $(A \rightarrow B) \wedge (B \rightarrow C) \Rightarrow A \rightarrow C$

Exercício: ache algumas dependências funcionais resultantes da aplicação destas regras

Fecho de um conjunto de atributos

Para uma relação r , um conjunto S de dependências funcionais e um conjunto Z de atributos, o **fecho** Z^+ de Z é o conjunto todos os atributos da relação r que são funcionalmente dependentes de Z . Para obter Z^+ podemos aplicar o seguinte algoritmo:

```
algoritmo FECHO( $Z, S$ ):  
     $Z^+ = Z$  (por reflexão)  
    mudou = Verdadeiro  
    enquanto mudou:  
        mudou = Falso  
        para  $X \rightarrow Y$  em  $S$ :  
            se  $(X \subset Z^+) \wedge (Y \notin Z^+)$ :  
                 $Z^+ := Z^+ \cup Y$  (por transitividade)  
                mudou = Verdadeiro  
    retorna  $Z^+$ 
```

Exercício

Suponha que tenhamos uma relação $R(A, B, C, D, E, F)$ e as seguintes dependências funcionais:

$$S = \{A \rightarrow BC, E \rightarrow CF, B \rightarrow E, CD \rightarrow EF\}$$

(abreviamos A no lugar de $\{A\}$, BC no lugar de $\{B, C\}$, etc.)

Calcule o fecho Z^+ de $Z = \{A, B\}$

Chaves

Superchave: um conjunto de atributos K é uma *superchave* da relação se K^+ é o conjunto completo de atributos desta relação

Chave candidata: uma superchave irredutível, ou seja, nenhum atributo pode ser removido da superchave sem que seu *status* de superchave seja cancelado

Chave primária: a chave candidata escolhida como identificador da relação

Formas normais

Material de estudo

Molina, H. C., Ullman, J. D., Widom, J. "Database Systems: The Complete Book", cap 3

Silberschatz, A., Korth, H. F., Sudarshan, S. "Sistema de Banco de Dados", cap 8

Date, C. J. "Introdução a Sistemas de Bancos de Dados", caps. 11 a 13

CUIDADO: Livro do Poderoso: cap. 3

- **Existem alguns erros no livro na parte da segunda forma normal!**

Formas normais

São regras para ajudar a **diminuir redundância** e **aumentar integridade** no projeto de um banco de dados.

Inventadas por **Edgar F. Codd**, criador do modelo relacional, em 1970.

Formas normais

Chamamos de **normalização** o processo de rearranjar o banco de dados para que obedeça às várias **formas normais**. O processo contrário, de regredir o projeto do banco de dados para formas normais menores, chama-se **denormalização**.

Formas normais

- 1NF: Primeira forma normal (*first normal form*)
- 2NF: Segunda forma normal
- 3NF: Terceira forma normal
- Outras formas normais:
 - BCNF: Forma normal de Boyce-Codd (*Boyce-Codd Normal Form*)
 - 4NF, 5NF
 - São raramente usadas

1NF

Definição:

“Uma entidade está na primeira forma normal quando cada tupla tem apenas um valor por atributo”

Exemplo 1NF

Esta tabela não está na 1NF!

id_usuario (PK)	nome	id_endereço	endereço	telefone
1	Juca	100	Rua Quatá, 300, São Paulo/SP	111-1111, 333-3333
2	Maria	200	Rua Casa do Ator, 537, Ap 42, São Paulo/SP	222-2222
3	Roberto	300	Av. Bias Fortes 382, Belo Horizonte/MG	444-4444

Múltiplos valores
por célula!

Múltiplos valores
por célula!

Solução?

id_usuario (PK)	nome	id_endereço	endereço	cidade	estado	telefone
1	Juca	100	Rua Quatá, 300	São Paulo	SP	111-1111
1	Juca	100	Rua Quatá, 300	São Paulo	SP	333-3333
2	Maria	200	Rua Casa do Ator, 537	São Paulo	SP	222-2222
3	Roberto	300	Av. Bias Fortes 382	Belo Horizonte	MG	444-4444

Anomalias

- **Atualização** (*update*): quando a atualização da tabela pode gerar inconsistências em função da redundância de informação.
- **Inserção** (*insert*): quando a inserção de dados não pode ser realizada sempre, pelo risco de deixar a base inconsistente.
- **Remoção** (*delete*): quando a remoção de linhas da tabela acaba removendo informação totalmente não-relacionada.

Anomalia de atualização

id_usuario	nome	endereço	cidade	estado	telefone
1	Juca	Rua Quatá, 300	São Paulo	SP	111-1111
1	Juca	Rua Quatá, 300	São Paulo	SP	333-3333
2	Maria	Rua Casa do Ator, 537	São Paulo	SP	222-2222
3	Roberto	Av. Bias Fortes 382	Belo Horizonte	MG	444-4444



id_usuario	nome	endereço	cidade	estado	telefone
1	Juca	Rua Quatá, 300	São Paulo	SP	111-1111
1	Juca	Rua Quatá, 1000	São Paulo	SP	333-3333
2	Maria	Rua Casa do Ator, 537	São Paulo	SP	222-2222
3	Roberto	Av. Bias Fortes 382	Belo Horizonte	MG	444-4444

E agora, onde mora o Juca?

Anomalia de inserção

id_usuario	nome	endereço	cidade	estado	telefone
1	Juca	Rua Quatá, 300	São Paulo	SP	111-1111
1	Juca	Rua Quatá, 300	São Paulo	SP	333-3333
2	Maria	Rua Casa do Ator, 537	São Paulo	SP	222-2222
3	Roberto	Av. Bias Fortes 382	Belo Horizonte	MG	444-4444
???	???	???	Rio Branco	AC	???

Como inserir uma cidade sem ninguém morando lá ainda?

Anomalia de remoção

id_usuario	nome	endereço	cidade	estado	telefone
1	Juca	Rua Quatá, 300	São Paulo	SP	111-1111
1	Juca	Rua Quatá, 300	São Paulo	SP	333-3333
2	Maria	Rua Casa do Ator, 537	São Paulo	SP	222-2222
3	Roberto	Av. Dias Fortes 382	Belo Horizonte	MG	444-4444

Ao remover o Roberto, a cidade de Belo Horizonte/MG desapareceu da base de dados!

2NF

Definição:

“Uma entidade está na segunda forma normal quando obedece à 1NF e todos os seus atributos não-chave dependem funcionalmente da chave primária inteira”

Exemplo 2NF

id_pessoa (PK)	id_projeto (PK)	nome_pessoa	Nome_projeto	doação
1	33	Mario	Jacaré	150
1	44	Mario	Sabiá	240
3	33	Ana	Jacaré	350

Solução:



id_pessoa (PK)	id_projeto (PK)	doação
1	33	150
1	44	240
3	33	350

id_pessoa (PK)	nome_pessoa
1	Mario
3	Ana

id_projeto (PK)	Nome_projeto
33	Jacaré
44	Sabiá

3NF

Definição:

“Uma tabela está na 3FN quando estiver na 2FN e se todos os atributos que não pertencem à chave primária não dependerem funcionalmente de outro atributo que não a chave.”

3NF

Definição:

“Uma tabela está na 3FN quando estiver na 2FN e se não existem dependências transitivas para a chave primária.”

Exemplo 3NF

Estes atributos pertencem à endereço,
não diretamente à usuário!

id_usuario (PK)	nome	id_endereço	endereço	cidade	estado
1	Juca	100	Rua Quatá, 300	São Paulo	SP
2	Maria	200	Rua Casa do Ator, 537	São Paulo	SP
3	Roberto	300	Av. Bias Fortes 382	Belo Horizonte	MG

Solução 3NF

id_usuario (PK)	nome	id_endereço (FK)
1	Juca	100
2	Maria	200
3	Roberto	300

id_endereço (PK)	endereço	cidade	estado
100	Rua Quatá, 300	São Paulo	SP
200	Rua Casa do Ator, 537	São Paulo	SP
300	Av. Bias Fortes 382	Belo Horizonte	MG

Acabou?

Solução 3NF

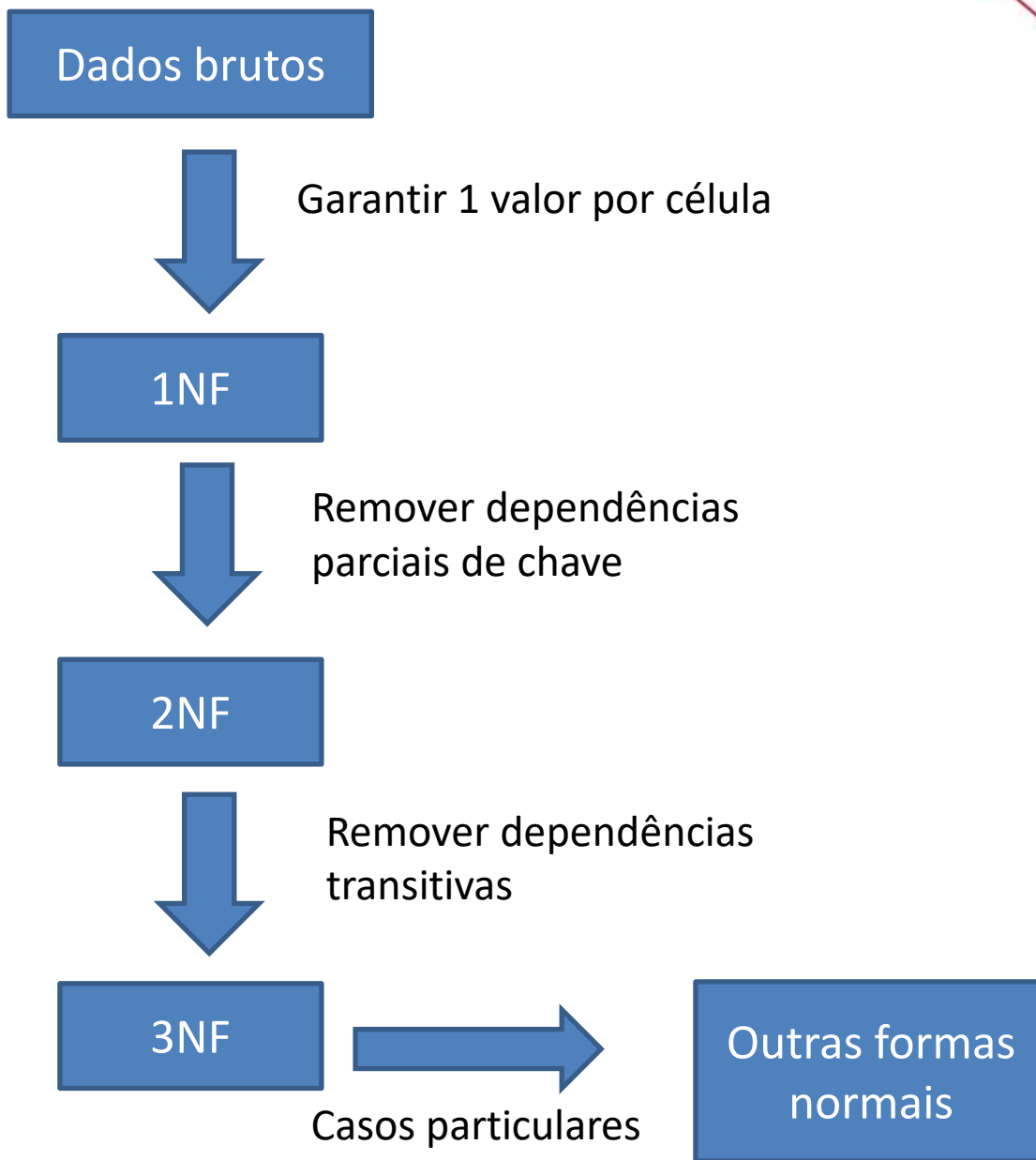
id_usuario (PK)	nome	id_endereço (FK)
1	Juca	100
2	Maria	200
3	Roberto	300

Analise do ponto de vista
das anomalias

id_endereço (PK)	endereço	id_cidade (FK)
100	Rua Quatá, 300	55
200	Rua Casa do Ator, 537	55
300	Av. Bias Fortes 382	66

id_cidade (PK)	cidade	estado
55	São Paulo	SP
66	Belo Horizonte	MG

Resumo



Exercício

Você recebeu um arquivo CSV com as informações do Clube dos Jogadores de Bolinha de Gude.

- Defina as dependências funcionais essenciais
- Passe para a 3NF

Nota: uma bolinha é identificável pela sua cor, tamanho e material.

Cor	Tamanho	RGB	Peso (g)	Diâmetro (cm)	Material	Valor (R\$)	Dono	Qtde
Azul	P	(0, 0, 255)	2	2	Vidro	3.00	Juca	3
Vermelha	G	(255, 0, 0)	15	5	Vidro	10.00	Ana	5
Azul	G	(0, 0, 255)	100	5	Aço	15.00	Juca	6
Vermelha	P	(255,0,0)	2	2	Vidro	4.00	Ana	2

Insper

www.insper.edu.br