

Normalização

Base de Dados - 2017/18 Carlos Costa

1

Introdução



- Já estudámos aspectos de desenho conceptual de base de dados e respectivo mapeamento para o modelo relacional.
- No entanto, nunca apresentámos um processo formal de analisar se determinado grupo de atributos de um esquema de relação é melhor do que outro.
- O desenho de uma base de dados relacional resulta num conjunto de relações. Existe um objectivo implícito nesse processo de desenho:
 - Preservação da informação
 - Todos os conceitos capturados pelo desenho conceptual que são mais tarde mapeados para o desenho lógico.
 - Minimizar a redundância dos dados
 - Minimizar o armazenamento duplicado de dados em relações distintas, reduzindo a necessidade de múltiplos updates e consequente problema 2 de consistência entre múltiplas cópias da mesma informação.



Desenho de BD - Esquemas de Relação

Análise de Qualidade:

- Critérios Informais
- Critérios Formais
 - Dependências Funcionais, Multivalor e Junção
- Processo de Normalização
 - Formas Normais
 - · Baseadas em critérios formais

3



Critérios Informais

- Clareza da semântica dos atributos da relação
- Redundância de informação no tuplo
- Redução dos NULLs nos tuplos
- Junção de relações baseada em PK e FK

deti

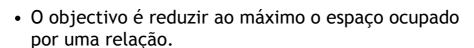
H deti

Semântica dos atributos da relação

- O desenho de um esquema de relação deve ser fácil de explicar.
- Verificar se existe uma semântica clara entre os atributos de uma relação.
 - Evitar que uma relação corresponda a uma mistura de atributos de diferentes entidades e relacionamentos.
 - Exemplos de mau desenho:



Redundância de Informação no Tuplo



• No mau exemplo anterior verificámos que também há duplicação desnecessária de informação.





Redução dos NULLs nos tuplos

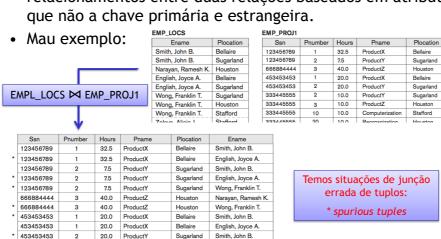
- Há situações em que temos uma grande quantidade de atributos numa relação:
 - Muitos dos atributos não se aplicam a todos os tuplos da relação.
- Consequência: existência de muitos NULLs nesses tuplos
 - · Desperdício de espaço
 - Difícil interpretação do seu sentido desses atributos (Null pode ter vários significados)
- Recomendação: Criar outra relação para esses atributos.
 Exemplo:
 - Imaginando que queremos incluir o número do gabinete na relação Employee mas só 15% dos funcionários têm esse número.
 - Solução: criar uma nova relação EMP_OFFICES(Essn, Office_number) só com tuplos de funcionários com gabinete.

7

deti

Junção de Relações baseada em PK e FK

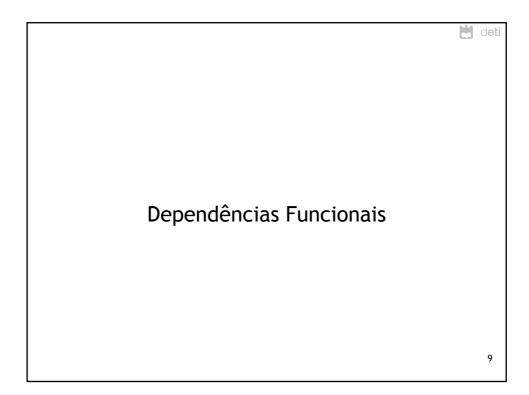
 Devemos evitar esquemas de relação que estabeleçam relacionamentos entre duas relações baseados em atributos que não a chave primária e estrangeira.



English, Joyce A

Sugarland

20.0 ProductY



Dependências Funcionais (DF)



- Considerando a relação:
 - R(A1, A2, ..., An)
 - Subconjunto de atributos X,Y ⊆ R
- Dependência Funcional: X→Y
 - tuplos: $t1, t2 \in R$
 - $t1[X] = t2[X] \Rightarrow t1[Y] = t2[Y]$

Restrição

- Formalismo de análise de esquemas relacionais.
 - Permite <u>descrever restrições</u> dos atributos que os tuplos devem respeitar em todo o momento (invariantes).
 - Permite detectar e descrever problemas com precisão. 10

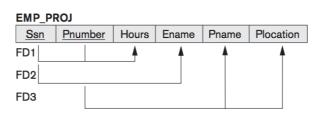
Dependências Funcionais



- X→Y ... por outras palavras:
 - Y é funcionalmente dependente de X.
 - Os valores da componente X do tuplo define de forma única a componente Y do respectivo tuplo.
- Uma DF é uma propriedade do esquema de relação R que <u>não pode ser inferido</u> de uma qualquer <u>instância</u> de R, i.e. r(R).
 - Deve ser definida por alguém que conhece a semântica dos atributos da relação.

11

Dependências Funcionais - Exemplo



- Pela semântica dos atributos da relação EMP_PROJ podemos inferir as seguintes DF:
 - Ssn → Ename
 - Pnumber → {Pname, Plocation}
 - {Ssn, Pnumber} → Hours

O Ssn determina de forma única o nome do funcionários.

O número do projecto determina de forma única o seu nome e localização.

O Ssn e o número do projecto determinam de forma única o número de horas que um funcionário trabalha para o projecto.

FD: Functional Dependency



DF - Algumas Propriedades

- X, Y, Z e W subconjuntos de atributos de R
- DP obedecem aos Axiomas de Armstrong (AA):
 - Reflexibilidade: $Y \subseteq X \Rightarrow X \rightarrow Y$
 - Aumento: X→Y então XZ→YZ
 - Transitividade: X→Y e Y→Z então X→Z
- AA dão origem às seguintes Regras de Inferência:
 - Decomposição: X→YZ então X→Y e X→Z
 - União: X→Y e X→Z então X→YZ
 - Pseudotransitividade: X→Y e YW→Z então XW→Z
- $X \rightarrow Y \not \Rightarrow Y \rightarrow X$
- Se X é chave candidata de R então: $X \rightarrow R$



deti

13

Tipos de Dependências Funcionais

- Dependência Parcial
 - atributo depende de parte dos atributos que compõem a chave da relação.
- Dependência Total
 - atributo depende de toda a chave da relação.
- Dependência Transitiva
 - atributo que não faz parte da chave da relação depende de um atributo que também não faz parte da chave da relação.



Introdução



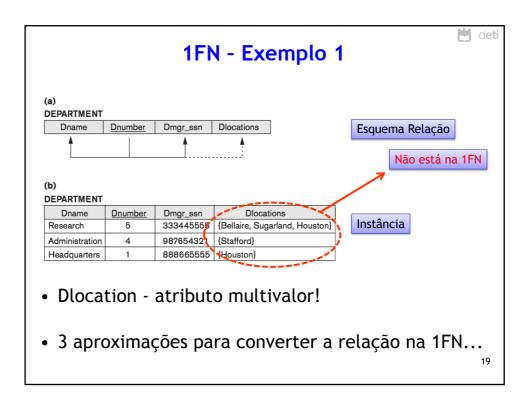
- Objectivo: Reduzir a Redundância
- Utilizámos DF para especificar alguns aspectos semânticos do esquema da relação.
 - Mas a redundância está associada a DF não desejadas!
- Vamos assumir que:
 - Existe um conjunto de DF associadas a cada esquema de relação;
 - Que cada relação tem uma chave primária definida;
- Processo de Normalização:
 - Formas Normais
 - · Conjunto de testes (condições) para validação de cada forma.
 - Cada forma superior tem menos DF que a anterior.

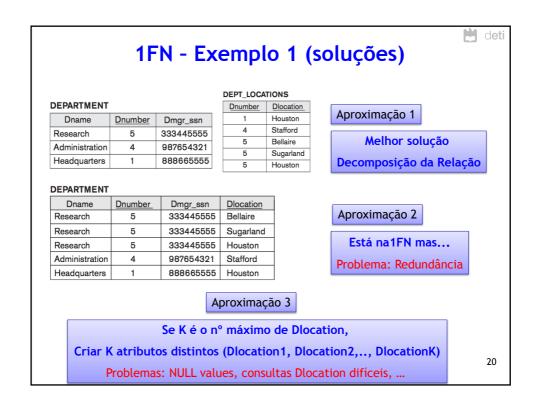


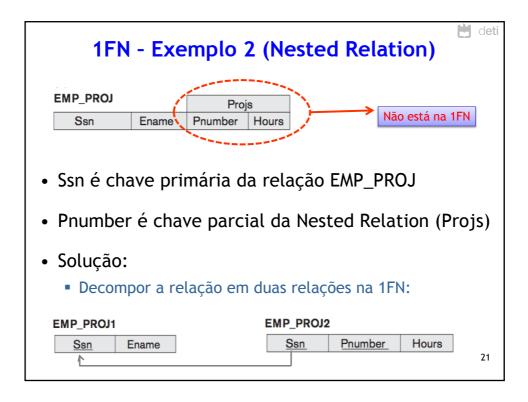
Formas Normais

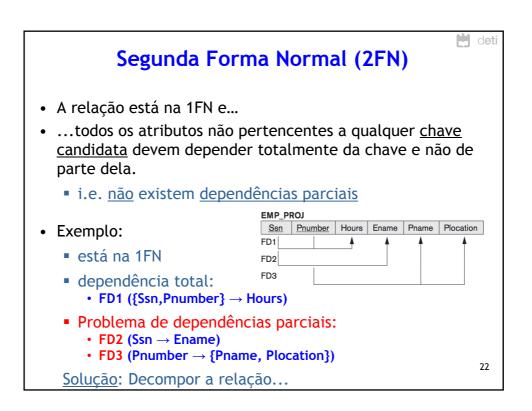
- O processo de normalização consiste em efetuar um conjunto de testes para certificar se um desenho de BD relacional satisfaz determinada Forma Normal (FN).
 - Relações que não satisfazem os testes de determinada forma normal são decompostas em relações menores.
- Codd propôs três FN baseadas em DF
 - Primeira (1FN), Segunda (2FN) e Terceira (3FN)
 - A 3FN satisfaz as condições da 2FN e esta as da 1FN
- Mais tarde Boyce e Codd propuseram uma definição mais restritiva da 3NF à qual se chamou:
 - Boyce-Codd Normal Form (BCNF)
- Foram ainda propostas a 4FN e 5FN baseadas respectivamente em dependências multivalor e de junção.

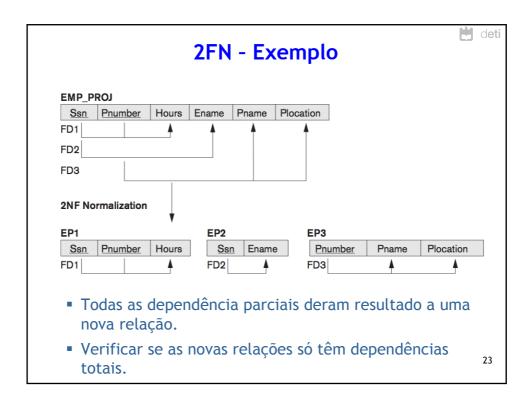
deti Primeira Forma Normal (1NF) • Definição formal de uma relação básica do modelo relacional: Atributos são atómicos (simples e indivisíveis) · Não permite atributos composto ou multivalor Não suporta relações dentro de relações (Nested Relation) • Não é possível utilizar uma relação como valor de um atributo de um tuplo. EMP_PROJ Projs Ssn Pnumber Hours EMP_PROJ Ssn Hours **Nested Relation** Ename 123456789 Smith, John B. 32.5 7.5 666884444 Narayan, Ramesh K 40.0 20.0 453453453 English, Joyce A. 18 EMP_PROJ(Ssn, Ename, {PROJS(Pnumber, Hours)}) { } significa que o atributo PROJS é multivalo

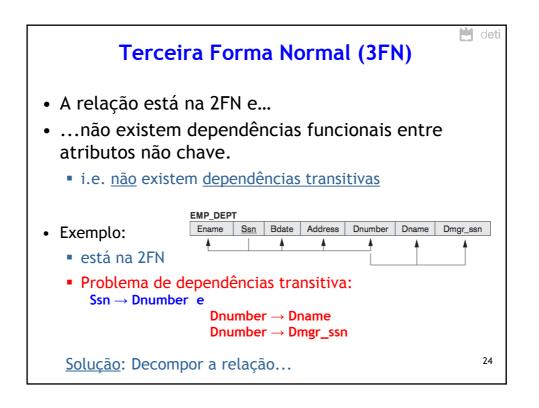


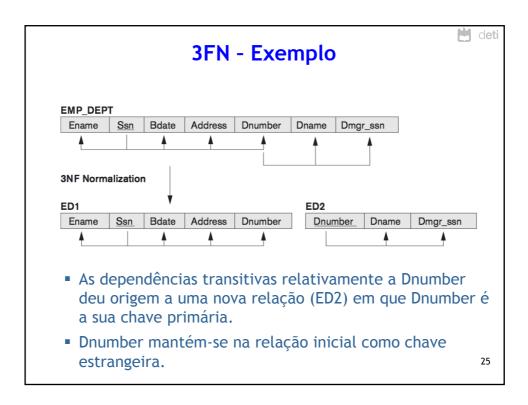


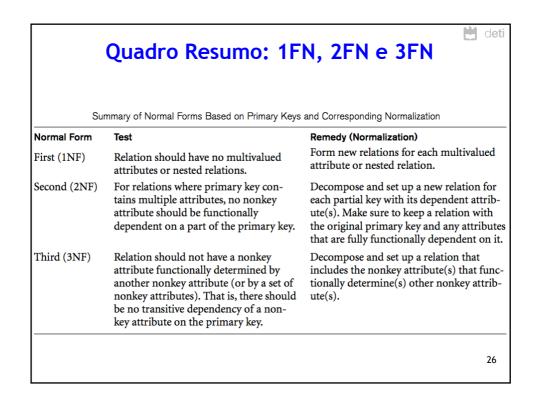












Boyce-Codd Normal Form (BCNF)

- Usualmente, a 3FN é aquela que termina o processo de normalização.
 - No entanto, em algumas situações a 3FN ainda apresenta algumas anomalias.
- BCNF é mais restritiva que a 3FN
 - BCNF => 3FN
- Definição:

Todos os atributos são funcionalmente dependentes da chave da relação, de toda a chave e de nada mais.

- Exemplo:
 - está na 3FN
 - FD2 viola a BCNF



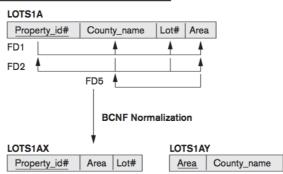
27

deti

deti

BCNF - Exemplo

Base de dados de uma imobiliária:



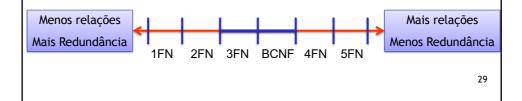
Chaves candidatas: Property_id# e {County_name, Lot#}

- Relações LOTS1A, LOTS1B e LOTS2 estão na 3FN
- FD5 viola BCNF

Solução: Decomposição de LOTS1A em LOTS1AX e LOTS1AY Reverso: Perdemos a FD2

Normalização - Ponto de Equilíbrio

- Como verificámos no exemplo de BCNF, perdeu-se uma dependência funcional importante (deduzida da semântica dos atributos).
 - Que deverá ser tratada ao nível aplicacional.
- Assim, existe um ponto de equilíbrio no processo de Normalização que tipicamente fica entre a 3FN e a BCNF.



4FN e 5FN



deti

- <u>Usualmente uma relação na BCNF também se</u> encontra na 4FN e 5FN.
 - 4FN são raros e 5FN ainda mais raros
- Definição 4FN:
 - Está na BCNF
 - Não existem dependências multivalor
- Definição 5FN:
 - Está na 4FN
 - A relação não pode ser mais decomposta sem haver perda de informação
 - Não existem dependências de junção

Dependências Multivalor

- deti
- Dependência multivalor X -» Y em R(X,Y,Z)
- Garantir a seguinte restrição em qualquer instância r(R):
 - Se dois tuplos t1 e t2 existem em r(R) tal que t1[X]=t2[X]
 - Então também devem existir dois tuplos t3 e t4 em r(R) com as seguintes características:
 - t4[X] = t3[X] = t1[X] = t2[X]
 - t3[Y] = t1[Y] e t4[Y] = t2[Y]
 - t3[Z] = t2[Z] e t4[Z] = t1[Z]

mesn	mesma r(R)		
Х	Υ	Z	
x1	y1	z1	
x1	y1	z2	
x1	y2	z1	
x1	y2	z2	

- Exemplo:
 - X -» Y
 - X -» Z
- Outras palavras...

X multidetermina Y se, para cada par de tuplos de R contendo os mesmo valores de X, 31 existe em R um par de tuplos correspondentes à troca dos valores de Y no par original.

4FN: Dependências Multivalor - Exemplo

ЕМР

Ename	<u>Pname</u>	<u>Dname</u>
Smith	X	John
Smith	Y	Anna
Smith	Х	Anna
Smith	Y	John

Dependências Multivalor:

Ename -» Pname

Ename -» Dname

• Solução: decomposição da relação EMP

EMP_PROJECTS

Ename	Pname	
Smith	X	
Smith	Y	

EMP_DEPENDENTS

Ename Dname		
Smith	John	
Smith	Anna	

Dependências de Junção

- Existe uma dependência de junção em R se, dadas algumas projeções de R, apenas se reconstrói R através de algumas junções bem definidas, mas não de todas.
- Muito rara na prática
 - difícil de detectar
- Exemplo:
 - Projetando R em (X,Y), (X,Z) e (Y,Z)
 - Verificamos que não é possível reconstruir R por junção de qualquer umas das projeções.
 - Só com a junção das 3 projeções é que conseguimos reconstruir R.

r(R)		
Х	Y	Z
x1	y1	z1
x1	y1	z2
x1	y2	z2
x2	у3	z2
x2	y4	z2
x2	y4	z4
x2	у5	z4
х3	y2	z5

33

deti

deti

5FN: Dependência Junção - Exemplo

SUPPLY

Sname	Part_name	Proj_name
Smith	Bolt	ProjX
Smith	Nut	ProjY
Adamsky	Bolt	ProjY
Walton	Nut	ProjZ
Adamsky	Nail	ProjX
Adamsky	Bolt	ProjX
Smith	Bolt	ProjY

Vamos Criar 3 Projecções de Supply:
R1(Sname, Part_name)
R2(Sname, Proj_name)
R3(Part_name, Proj_name)

 Sname
 Part_name

 Smith
 Bolt

 Smith
 Nut

 Adamsky
 Bolt

 Walton
 Nut

 R2

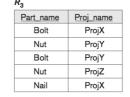
 Sname
 Proj_name

 Smith
 ProjX

 Smith
 ProjY

 Adamsky
 ProjZ

 Adamsky
 ProjX



- A relação SUPPLY, com dependência de junção, pode ser decomposta em 3 relações R1, R2 e R3 cada uma na 5FN.
 - Só reconstruímos Supply com a junção das 3 relações R1, R2 e R3.



Normalização - Caso de Estudo

Gestão de Encomendas

35

Esquema de Base de Dados - Início

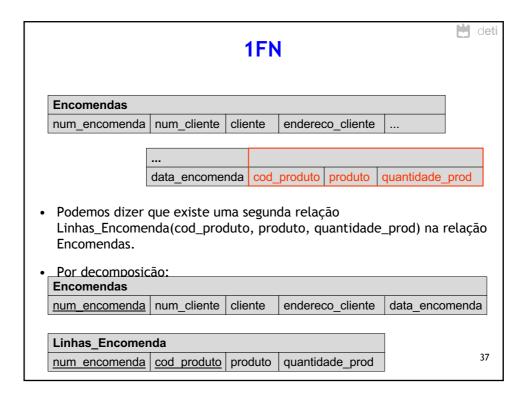


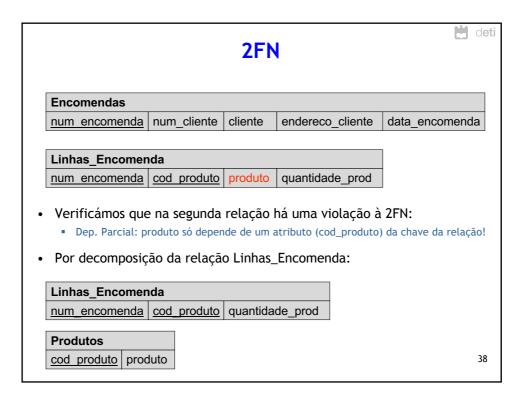


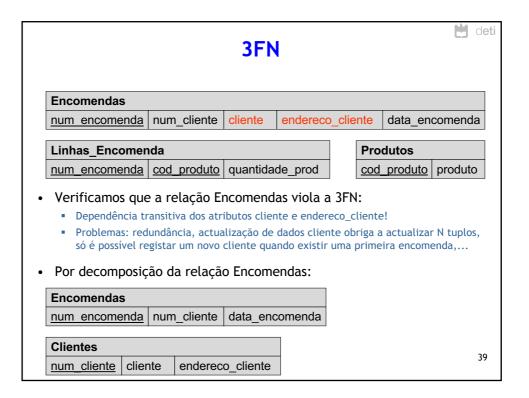


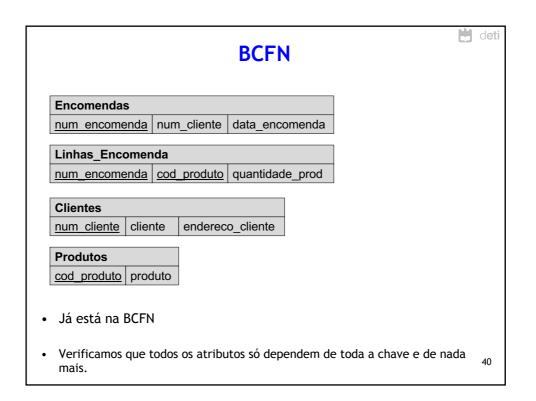
- É notório que o designer não tem conhecimentos de desenho de base de dados...
- Problemas:
 - Mistura de grupos de atributos de entidades (claramente) distintas.
 - Redundância de informação nos tuplos
 - Temos de repetir num_encomenda, num_cliente, cliente, endereco_cliente e data_encomenda para registar várias linhas de uma encomenda!

O









Resumo

- deti
- Qualidade do Desenho de Base de Dados Relacionais
- Critérios Informais
- Dependências Funcionais
- Normalização (Formas Normais)