

DCC011

Introdução a Banco de Dados

Normalização

Mirella M. Moro
Departamento de Ciência da Computação
Universidade Federal de Minas Gerais
mirella@dcc.ufmg.br

Normalização

1. Normalização
 - A. Semântica dos atributos
 - B. Anomalias
 - C. Valores nulos
 - D. Valores espúrios
2. Dependências Funcionais
3. Formas Normais

1. Normalização

Noção central: qualidade do projeto

- Medir formalmente porque um agrupamento de atributos em um esquema de relação é melhor que outro.
- Dois níveis:
 - Lógico: interpretação dos esquema e atributos
 - Armazenamento: como as tuplas são armazenadas e atualizadas

DCC011 - profa Mirella M. Moro

3

Normalização : na Prática

POR EXEMPLO (existem muitos outros casos)

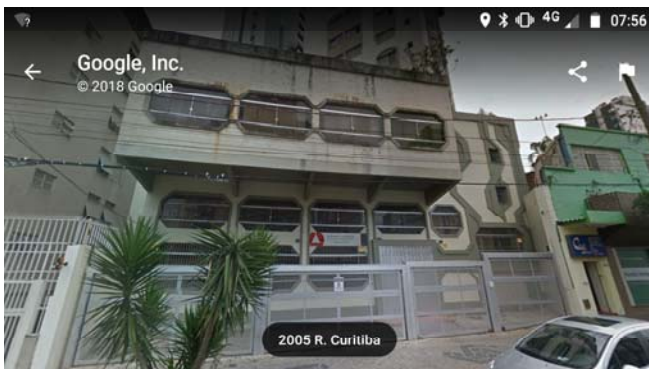
- Informatização de processos
- Download de datasets (csv, txt, etc)
- Aprimoramento de planilhas eletrônicas
- Organização de papeis

➔ Próximos slides

DCC011 - profa Mirella M. Moro

4

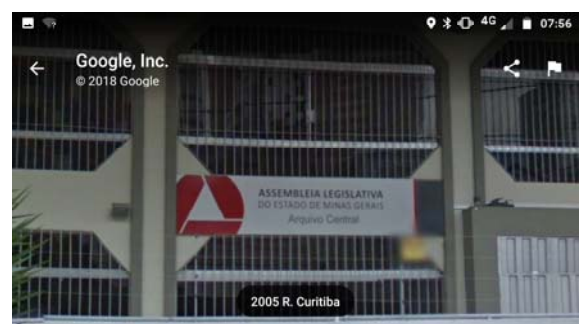
Qual o problema desta imagem?



DCC011 - profa Mirella M. Moro

5

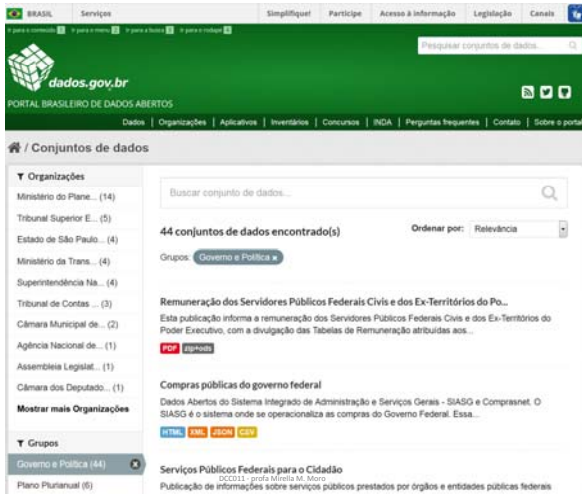
Zoom



DCC011 - profa Mirella M. Moro

6

Qual o problema desta imagem?



Normalização

- Processo através do qual esquemas de relação são sucessivamente decompostos até que satisfaçam determinadas propriedades ou **formas normais**.
- Medidas informais de qualidade do projeto de um esquema relacional:
 - Atributos das relações semanticamente compatíveis
 - Ausência de valores redundantes nas relações (inexistência de anomalias de atualização)
 - Número reduzido de valores nulos nas relações
 - Sem possibilidade de gerar tuplas espúrias/falsas

A. Semântica dos atributos

Semântica dos atributos da relação

- Atributos são agrupados em uma relação
- Por quê?

Atributos em uma relação têm um significado no mundo real e uma interpretação associada

O agrupamento “faz sentido”

Semântica dos atributos

- Especifica como interpretar os valores dos atributos armazenados nas relações

Regra 1: projete o esquema de uma relação que seja fácil de explicar.

- Um esquema de relação que representa um tipo de entidade ou relacionamento no mundo real tende a ser claro
- Não combinar atributos de diferentes entidades e relacionamentos em uma tabela

Semântica dos atributos da relação

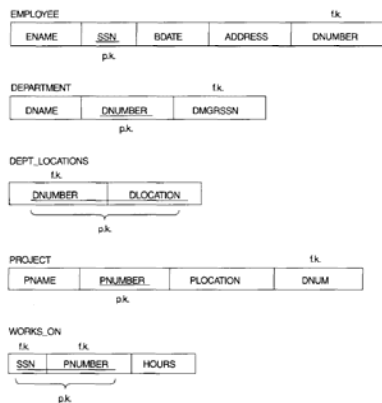


Figura 10.1 Um esquema simplificado do banco de dados relacional EMPRESA.

Semântica dos atributos da relação

Emp_Dept

SSN	EName	Bdate	Address	DNumber	DName	DMgrSSN
-----	-------	-------	---------	---------	-------	---------

Emp_Proj

SSN	PNumber	Hours	EName	PName	PLocation
-----	---------	-------	-------	-------	-----------

B. Espaço e Anomalias de Atualização

- Minimizar espaço de armazenamento
- Agrupar atributos no esquema de uma relação tem um efeito significativo no espaço de armazenamento
- Veja o exemplo a seguir

Minimizar espaço

EMP_DEPT						
ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
Smith, John B.	123456789	1965-01-09	731 FORDEN, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1952-12-08	6384 Keesi, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Nagura, Ramesh K.	666884444	1962-09-15	975 Five, Oak, Humble, TX	5	Research	333445555
Englin, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jacob, Ahmad V.	98787987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1957-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

EMP_PROJ					
SSN	PNUMBER	HOURS	ENAME	PNAME	PLLOCATION
123456789	1	32.5	Smith, John B.	ProductX	Bellaire
123456789	2	7.5	Smith, John B.	ProductY	Sugarland
988664444	3	40.0	Nagura, Ramesh K.	ProductZ	Houston
453453453	1	20.0	Englin, Joyce A.	ProductX	Bellaire
453453453	2	20.0	Englin, Joyce A.	ProductY	Sugarland
333445555	2	10.0	Wong, Franklin T.	ProductY	Sugarland
333445555	3	10.0	Wong, Franklin T.	ProductZ	Houston
333445555	10	10.0	Wong, Franklin T.	Computerization	Stafford
333445555	20	10.0	Wong, Franklin T.	Reorganization	Houston
999877777	30	30.0	Zelaya, Alicia J.	NewGenetics	Stafford
999877777	10	10.0	Zelaya, Alicia J.	Computerization	Stafford
98787987	10	35.0	Jacob, Ahmad V.	Computerization	Stafford
98787987	30	5.0	Jacob, Ahmad V.	NewGenetics	Stafford
987654321	30	20.0	Wallace, Jennifer S.	NewGenetics	Stafford
987654321	20	15.0	Wallace, Jennifer S.	Reorganization	Houston
888665555	20	null	Borg, James E.	Reorganization	Houston

Figura 10.4 Exemplo de estado para EMP_DEPT e EMP_PROJ resultantes da aplicação do NATURAL JOIN nas relações da Figura 10.2. Elas podem ser armazenadas por razões de desempenho.

Anomalias de Atualização

- Outro problema: anomalias de atualização
- Anomalia de Inserção
 - Consistência nos dados do departamento
 - Inserção de departamento sem empregados
 - Inserção de empregados sem departamento

Opção 1

EMP_DEPT						
ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN

EMPLOYEE						fk.
ENAME	SSN	BDATE	ADDRESS	DNUMBER		

Opção 2

DEPARTMENT			fk.
DNAME	DNUMBER	DMGRSSN	

Anomalias de Atualização

- Anomalia de Remoção
 - Remover o último empregado de um departamento, remove os dados do depto
- Anomalia de Modificação
 - Modificar um atributo do depto obriga a atualização de todos os empregados

Opção 1

EMP_DEPT						
ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN

Informação redundante em tuplas e Anomalias de atualização

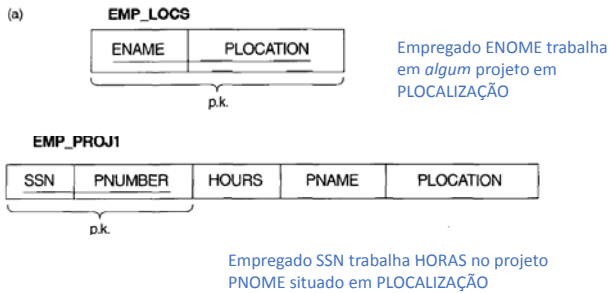
• **Regra 2:** Projetar os esquema de relação tal que **nenhuma** anomalia de inserção, remoção, ou modificação esteja presente nas relações.

- Se quaisquer anomalias existem, anote-as claramente e esteja certo que os programas que atualizam o banco de dados operem corretamente

C. Valores nulos nas tuplas

- Problemas de:
 - Espaço de armazenamento
 - Interpretação de valores
 - Junções
 - Operações de Agregação
- **Regra 3:** Evite ao máximo inserir atributos em relações que podem assumir valores nulos frequentemente
- Se esses valores são inevitáveis, tenha certeza qe eles são a exceção e não a regra.
 - Office_Number

D. Geração de Tuplas espúrias



DCC011 - profa Mirella M. Moro

19

Geração de Tuplas espúrias

(b)

EMP_LOCS	
ENAME	PLOCATION
Smith, John B.	Bellaire
Smith, John B.	Sugarland
Narayan, Ramesh K.	Houston
English, Joyce A.	Bellaire
English, Joyce A.	Sugarland
Wong, Franklin T.	Sugarland
Wong, Franklin T.	Houston
Wong, Franklin T.	Stafford
Zelaya, Aldo J.	Stafford
Jabbar, Ahmad Y.	Stafford
Wallace, Jennifer S.	Stafford
Wallace, Jennifer S.	Houston
Borg, James E.	Houston

Figura 10.5 Projeto particularmente pobre para a relação EMP_PROJ da Figura 10.3b. (a) Os dois esquemas de relações EMP_LOCS e EMP_PROJ1 [SLIDE ANTERIOR]

(b) O resultado da projeção de EMP_PROJ, da Figura 10.4, para as relações EMP_LOCS e EMP_PROJ1.

EMP_PROJ1				
SSN	PNUMBER	HOURS	PNAME	PLOCATION
123456789	1	32.5	Product X	Bellaire
123456789	2	7.5	Product Y	Sugarland
666894444	3	40.0	Product Z	Houston
453453453	1	20.0	Product X	Bellaire
453453453	2	20.0	Product Y	Sugarland
333445555	2	10.0	Product Y	Sugarland
333445555	3	10.0	Product Z	Houston
333445555	10	10.0	Computerization	Stafford
333445555	20	10.0	Reorganization	Houston
999887777	30	30.0	Newbenefits	Stafford
999887777	10	10.0	Computerization	Stafford
987987987	10	35.0	Computerization	Stafford
987987987	30	5.0	Newbenefits	Stafford
987987987	30	20.0	Newbenefits	Stafford
987654321	20	15.0	Reorganization	Houston
888665555	20	null	Reorganization	Houston

DCC011 - profa Mirella M. Moro

20

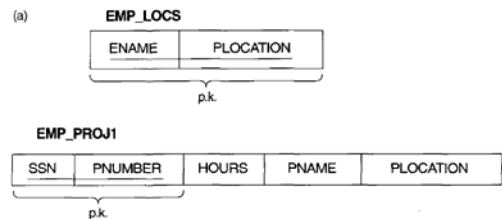
Geração de Tuplas espúrias

SSN	PNUMBER	HOURS	PNAME	PLOCATION	ENAME
123456789	1	32.5	ProductX	Bellaire	Smith,John B.
123456789	1	32.5	ProductX	Bellaire	English,Joyce A.
123456789	2	7.5	ProductY	Sugarland	Smith,John B.
123456789	2	7.5	ProductY	Sugarland	English,Joyce A.
123456789	2	7.5	ProductY	Sugarland	Wong,Franklin T.
866884444	3	40.0	ProductZ	Houston	Narayan,Ramesh K.
866884444	3	40.0	ProductZ	Houston	Wong,Franklin T.
453453453	1	20.0	ProductX	Bellaire	Smith,John B.
453453453	1	20.0	ProductX	Bellaire	English,Joyce A.
453453453	2	20.0	ProductY	Sugarland	Smith,John B.
453453453	2	20.0	ProductY	Sugarland	English,Joyce A.
453453453	2	20.0	ProductY	Sugarland	Wong,Franklin T.
333445555	2	10.0	ProductY	Sugarland	Smith,John B.
333445555	2	10.0	ProductY	Sugarland	English,Joyce A.
333445555	2	10.0	ProductY	Sugarland	Wong,Franklin T.
333445555	3	10.0	ProductZ	Houston	Narayan,Ramesh K.
333445555	3	10.0	ProductZ	Houston	Wong,Franklin T.
333445555	10	10.0	Computerization	Stafford	Wong,Franklin T.
333445555	20	10.0	Reorganization	Houston	Narayan,Ramesh K.
333445555	20	10.0	Reorganization	Houston	Wong,Franklin T.

Figura 10.6 Resultado da aplicação de NATURAL JOIN nas tuplas acima da linha pontilhada de EMP_PROJ1 e EMP_LOCS da Figura 10.5. As tuplas ilegítimas geradas estão marcadas com asteriscos.

21

Geração de Tuplas espúrias



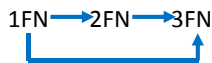
• **Regra 4:** Projete o esquema de uma relação tal que junções operem em condições de igualdade apenas em atributos que são chaves primárias ou chaves estrangeiras.

DCC011 - profa Mirella M. Moro

22

2. Dependências Funcionais

- Processo de normalização:



Formas normais definidas a partir do conceito de dependência funcional.

DCC011 - profa Mirella M. Moro

23

Dependências Funcionais

- Definição

Uma dependência funcional entre dois atributos (ou conjuntos de atributos) X e Y de um esquema de relação R, denotada por $X \rightarrow Y$, é uma restrição de integridade que estabelece que, para quaisquer tuplas t1 e t2 de uma instância de R tal que $t1[X] = t2[X]$, $t1[Y] = t2[Y]$.

- Em outras palavras:

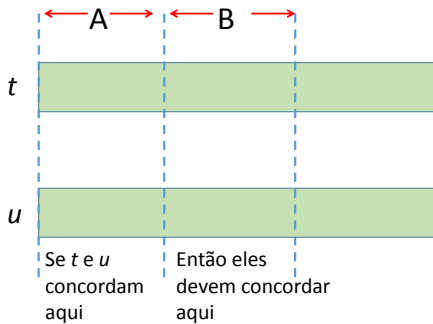
Dados os atributos "A" e "B" de uma entidade, diz-se que "B" é funcionalmente dependente de "A" se e somente se, a cada valor de "A" está associado um único valor de "B". se conhecermos o valor de "A" então podemos encontrar o valor de "B" associado a ele.

DCC011 - profa Mirella M. Moro

24

Dependências Funcionais

Se duas tuplas em R concordam nos atributos A_1, A_2, \dots, A_N (i.e. as tuplas têm os mesmos valores), então eles devem concordar em um outro atributo, B



Dependências Funcionais

Definição:

Uma dependência funcional entre dois atributos (ou conjuntos de atributos) X e Y de um esquema de relação R , denotada por $X \rightarrow Y$, é uma restrição de integridade que estabelece que, para quaisquer tuplas t_1 e t_2 de uma instância de R tal que $t_1[X] = t_2[X]$, $t_1[Y] = t_2[Y]$.

Exemplos:

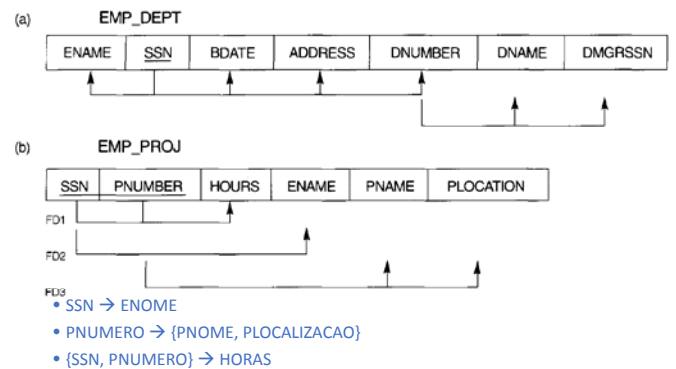
EMP-DEPT(ENAME,SSN,BDATE,ADDRESS,DNUMBER,DNAME,DMGSSN)
 $SSN \rightarrow \{ENAME, BDATE, ADDRESS, DNUMBER\}$
 $DNUMBER \rightarrow \{DNAME, DMGSSN\}$

Atenção: Não confundir a notação $X \rightarrow Y$ com a notação $R_1[X] \rightarrow R_2[Y]$ que denota a RIR entre os esquemas de relação R_1 e R_2 .

Dependências Funcionais

- Se X é um chave candidata
- Isso implica que $X \rightarrow Y$ para qualquer subconjunto de atributos Y de R
- A dependência vira uma *constraint*
- Propriedade da semântica ou significado dos atributos
 - {Orgão Emissor, Identidade} \rightarrow CPF
 - CEP \rightarrow Rua
 - Professor \rightarrow Curso
- Deve ser definida explicitamente

Dependências Funcionais



Regras de Inferência para Dependências Funcionais

- $F = \{SSN \rightarrow \{ENAME, BDATE, ADDRESS, DNUMBER\}, DNUMBER \rightarrow \{DNAME, DMGRSSN\}\}$
- Podemos inferir que:
 - $SSN \rightarrow \{DNAME, DMGRSSN\}$
 - $SSN \rightarrow SSN$
 - $DNUMBER \rightarrow DNAME$

Regras de Inferência para Dependências Funcionais

- RI1 (reflexiva): If $X \supseteq Y$, então $X \rightarrow Y$
- RI2 (aumento): $\{X \rightarrow Y\} \models XZ \rightarrow YZ$
- RI3 (transitiva): $\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$
- RI4 (decomposição): $\{X \rightarrow YZ\} \models X \rightarrow Y$
- RI5 (união): $\{X \rightarrow Y, X \rightarrow Z\} \models X \rightarrow YZ$
- RI6 (pseudotransitiva): $\{X \rightarrow Y, WY \rightarrow Z\} \models WX \rightarrow Z$

Exercícios

Considere a figura ao lado.

X	Y	Z
x ₁	y ₁	z ₁
x ₁	y ₁	z ₂
x ₂	y ₁	z ₁
x ₂	y ₁	z ₃

- 1) Liste todas as dependências funcionais que a instância dessa relação satisfaz.
- 2) Assuma que o valor do atributo Z na última tupla tenha sido atualizado de z₃ para z₂. Liste novamente as dependências funcionais que a instância dessa relação satisfaz.

Exercícios

Considere a figura ao lado.

X	Y	Z
x ₁	y ₁	z ₁
x ₁	y ₁	z ₂
x ₂	y ₁	z ₁
x ₂	y ₁	z ₃

- 1) Liste todas as dependências funcionais que a instância dessa relação satisfaz.
 $R: Z \rightarrow Y; X \rightarrow Y; XZ \rightarrow Y$
- 2) Assuma que o valor do atributo Z na última tupla tenha sido atualizado de z₃ para z₂. Liste novamente as dependências funcionais que a instância dessa relação satisfaz.

Não há mudança

3. Normalização

- Processo através do qual esquemas de relação são sucessivamente decompostos até que satisfaçam determinadas propriedades ou **formas normais**.
- Medidas informais de qualidade do projeto de um esquema relacional:
 - Atributos das relações semanticamente compatíveis
 - Ausência de valores redundantes nas relações (inexistência de anomalias de atualização)
 - Número reduzido de valores nulos nas relações
 - Sem possibilidade de gerar tuplas espúrias/falsas

Primeira Forma Normal (1NF)

Um esquema de relação R está na 1FN se todos os seus atributos forem **atômicos**.

Department

Dnumber	DName	DMgrSSN	DLocation
5	Research	333445555	{Bellaire, Sugarland, Houston}
4	Administration	987654321	{Stafford}
1	Headquarters	888666555	{Houston}

1NF

Uma entidade está na 1NF se ela não contém grupos de atributos repetitivos (multivalorados).

Entidade não normalizada:

PEDIDO (NUMERO_PEDIDO, DATA_PEDIDO, NUMERO_CLIENTE, NOME_CLIENTE, ENDERECO_CLIENTE, (NUMERO_PRODUTO, NOME_PRODUTO, QTDE_PEDIDA, PRECO_PRODUTO, TOTAL_PRODUTO), TOTAL_PEDIDO) →

Compras registradas em planilha excel ou papel

PEDIDO DE COMPRA				
NRO: 2016102598381		DATA: 25/10/2016		
CLIENTE: #892853		NOME: Mirella M. Moro		
ENDEREÇO: DCC, ICEx sala 6326, UFMG				
#	PRODUTO	QTDE	PREÇO	TOTAL
192HP40	Impressora HP40	2	R\$3.500,00	R\$7.000,00
08340P	T HP40 Preto	10	R\$540,00	R\$5.400,00
08340M	T HP40 Amarelo	4	R\$540,00	R\$2.160,00
08340Z	T HP40 Azul	4	R\$540,00	R\$2.160,00
08340V	T HP40 Vermelho	4	R\$540,00	R\$2.160,00
827500	Papel 500	10	R\$95,00	R\$950,00
			TOTAL FINAL: R\$19.830,00	

Compras registradas em planilha excel ou papel

PEDIDO DE COMPRA				
NRO: _____	DATA: __/__/__			
CLIENTE: # _____	NOME: _____			
ENDEREÇO: _____				
#	PRODUTO	QTDE	PREÇO	TOTAL
		(N)		
TOTAL FINAL: _____				

PEDIDO (NUMERO_PEDIDO, DATA_PEDIDO, NUMERO_CLIENTE, NOME_CLIENTE, ENDEREÇO_CLIENTE, (NUMERO_PRODUTO, NOME_PRODUTO, QTDE_PEDIDA, PRECO_PRODUTO, TOTAL_PRODUTO), TOTAL_PEDIDO)

1NF

PEDIDO (NUMERO_PEDIDO, DATA_PEDIDO, NUMERO_CLIENTE, NOME_CLIENTE, ENDEREÇO_CLIENTE, (NUMERO_PRODUTO, NOME_PRODUTO, QTDE_PEDIDA, PRECO_PRODUTO, TOTAL_PRODUTO), TOTAL_PEDIDO)

- PEDIDO (NUMERO_PEDIDO, DATA_PEDIDO, NUMERO_CLIENTE, NOME_CLIENTE, ENDEREÇO_CLIENTE, TOTAL_PEDIDO)
- PRODUTO_PEDIDO (NUMERO_PEDIDO, NUMERO_PRODUTO, NOME_PRODUTO, QTDE_PEDIDA, PRECO_PRODUTO, TOTAL_PRODUTO)

Segunda Forma Normal (2FN)

• Um esquema de relação R está na 2FN se todo atributo de R não pertencente a uma de suas chaves for totalmente dependente da chave primária.

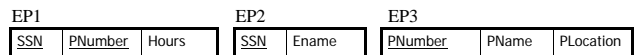
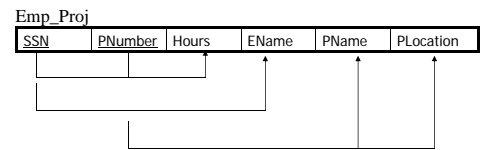
• O esquema de relação

EMP-PROJ(SSN,PNUMBER,HOURS,ENAME,PNAME,PLOCATION)

não está na 2FN porque

SSN → ENAME e PNUMBER → {PNAME,PLOCATION}, ou seja, há atributos em R que não são totalmente dependentes da chave primária {SSN,PNUMBER}.

2NF



2NF, outro exemplo

PEDIDO (NUMERO_PEDIDO, DATA_PEDIDO, NUMERO_CLIENTE, NOME_CLIENTE, ENDEREÇO_CLIENTE, TOTAL_PEDIDO)

PRODUTO_PEDIDO (NUMERO_PEDIDO, NUMERO_PRODUTO, NOME_PRODUTO, QTDE_PEDIDA, PRECO_PRODUTO, TOTAL_PRODUTO)

PEDIDO

NUMERO_PEDIDO
DATA_PEDIDO
NUMERO_CLIENTE
NOME_CLIENTE
ENDEREÇO_CLIENTE
TOTAL_PEDIDO

PRODUTO_PEDIDO

NUMERO_PEDIDO
NUMERO_PRODUTO
NOME_PRODUTO
QTDE_PEDIDA
PRECO_PRODUTO
TOTAL_PRODUTO

2NF, outro exemplo

- PEDIDO (NUMERO_PEDIDO, DATA_PEDIDO, NUMERO_CLIENTE, NOME_CLIENTE, ENDEREÇO_CLIENTE, TOTAL_PEDIDO)
- PRODUTO_PEDIDO (NUMERO_PEDIDO, NUMERO_PRODUTO, NOME_PRODUTO, QTDE_PEDIDA, PRECO_PRODUTO, TOTAL_PRODUTO)

PEDIDO (NUMERO_PEDIDO, DATA_PEDIDO, NUMERO_CLIENTE, NOME_CLIENTE, ENDEREÇO_CLIENTE, TOTAL_PEDIDO)

PRODUTO_PEDIDO (NUMERO_PEDIDO, NUMERO_PRODUTO, QTDE_PEDIDA, TOTAL_PRODUTO)

PRODUTO (NUMERO_PRODUTO, NOME_PRODUTO, PRECO_PRODUTO)

Terceira Forma Normal (3FN)

- Um esquema de relação R está na 3FN se estiver na 2FN e nenhum atributo de R não pertencente a uma de suas chaves for transitivamente dependente da chave primária.
- Uma dependência $X \rightarrow Y$ é transitiva se existe um conjunto de atributos Z que não é chave candidata ou um subconjunto de de qualquer chave de R para os qual ambas as dependências existem: $X \rightarrow Z$ e $Z \rightarrow Y$

3NF

- O esquema de relação

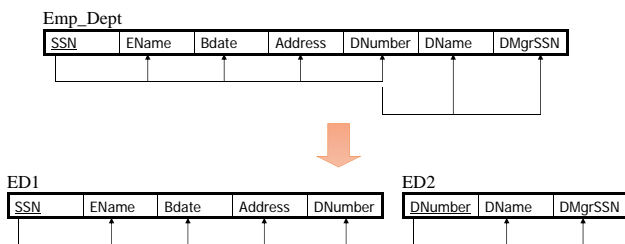
EMP-DEPT(ENAME, SSN, BDATE, ADDRESS, DNUMBER, DNAME, DMGRSSN)

não está na 3FN porque

$SSN \rightarrow DNUMBER$ e $DNUMBER \rightarrow \{DNAME, DMGRSSN\}$

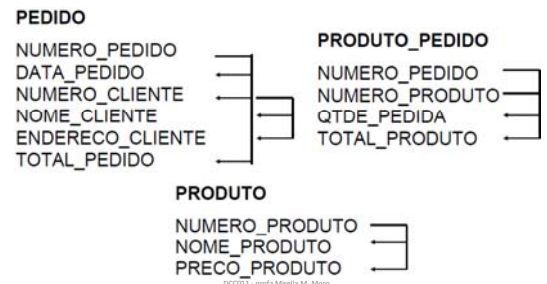
ou seja, há atributos em R que são transitivamente dependentes da chave primária SSN.

3NF



3NF, outro exemplo

- PEDIDO (NUMERO_PEDIDO, DATA_PEDIDO, **NUMERO_CLIENTE**, **NOME_CLIENTE**, **ENDERECO_CLIENTE**, TOTAL_PEDIDO)
- PRODUTO_PEDIDO (NUMERO_PEDIDO, NUMERO_PRODUTO, QTDE_PEDIDA, TOTAL_PRODUTO)
- PRODUTO (NUMERO_PRODUTO, NOME_PRODUTO, PRECO_PRODUTO)

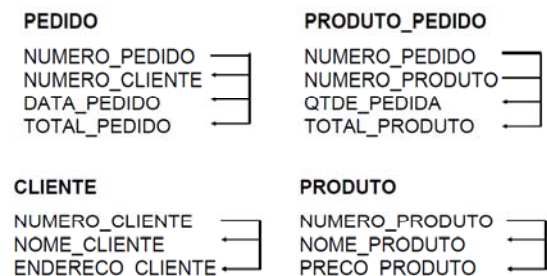


3NF, outro exemplo

- PEDIDO (NUMERO_PEDIDO, DATA_PEDIDO, **NUMERO_CLIENTE**, **NOME_CLIENTE**, **ENDERECO_CLIENTE**, TOTAL_PEDIDO)
 - PRODUTO_PEDIDO (NUMERO_PEDIDO, NUMERO_PRODUTO, QTDE_PEDIDA, TOTAL_PRODUTO)
 - PRODUTO (NUMERO_PRODUTO, NOME_PRODUTO, PRECO_PRODUTO)
-
- PEDIDO (NUMERO_PEDIDO, NUMERO_CLIENTE, DATA_PEDIDO, TOTAL_PEDIDO)
 - CLIENTE (**NUMERO_CLIENTE**, **NOME_CLIENTE**, **ENDERECO_CLIENTE**)
 - PRODUTO_PEDIDO (NUMERO_PEDIDO, NUMERO_PRODUTO, QTDE_PEDIDA, TOTAL_PRODUTO)
 - PRODUTO (NUMERO_PRODUTO, NOME_PRODUTO, PRECO_PRODUTO)

3NF

- PEDIDO (NUMERO_PEDIDO, NUMERO_CLIENTE, DATA_PEDIDO, TOTAL_PEDIDO)
- CLIENTE (NUMERO_CLIENTE, NOME_CLIENTE, ENDERECO_CLIENTE)
- PRODUTO_PEDIDO (NUMERO_PEDIDO, NUMERO_PRODUTO, QTDE_PEDIDA, TOTAL_PRODUTO)
- PRODUTO (NUMERO_PRODUTO, NOME_PRODUTO, PRECO_PRODUTO)



Revisão

Forma Normal	Teste	Remédio (normalização)
1NF	Relação não tem atributos multivalorados ou relações aninhadas	Definir novas relações para cada atributo multivalorado ou relação aninhada
2NF	Para relações onde chave primária contém múltiplos atributos, nenhum atributo não-chave deve ser dependente funcionalmente de uma parte da chave primária	Decompor e definir uma nova relação para cada chave parcial com seus atributos dependentes. Mantenha a relação original com a sua chave e seus atributos funcionalmente dependentes
3NF	Relação não pode ter um atributo não-chave determinado funcionalmente por outro atributo não-chave. Não tem dependência transitiva de um atributo não chave.	Decompor e definir uma relação que inclua os atributos não-chave que determinam funcionalmente os outros atributos não-chave.

DCC011 - profa Mirella M. Moro

50

Definição Geral

- Informalmente, podemos dizer que um esquema de relação R está normalizado (ou seja, na 3FN) se todos os seus atributos são dependentes exclusivamente de suas chaves.

DCC011 - profa Mirella M. Moro

51

Forma Normais

- A **forma normal de uma relação** refere à maior forma normal à qual a relação satisfaz, e então indica a qual grau a relação foi normalizada
- Não necessariamente 3NF

DCC011 - profa Mirella M. Moro

52

Formas Normais

- Formas normais sozinhas não garantem um bom projeto de BD
- Não é suficiente conferir cada relação individualmente verificando se satisfaz a 3NF
- Outras duas propriedades necessárias
 - *Lossless join*: que garante que tuplas espúrias não são geradas quando se faz uma junção
 - *Preservação de dependência*: que garante que cada dependência funcional é representada em alguma relação

DCC011 - profa Mirella M. Moro

53

Fechamento (Closure)

- Algoritmo 14.1 Determinar X^+ , o fechamento de X sobre um conjunto de dependências F
- ```

 $X^+ := X;$
repeat
 $oldX^+ := X^+;$
 for each functional dependency $Y \rightarrow Z$ in F do
 if $X^+ \supseteq Y$ then $X^+ := X^+ \cup Z;$
until $(X^+ = oldX^+)$

```

DCC011 - profa Mirella M. Moro

54

## Fechamento (Closure)

```

F = {SSN → ENAME,
 PNUMBER → {PNAME, PLOCATION}
 {SSN, PNUMBER} → HOURS}

```

```

SSN+ = {SSN, ENAME}
PNUMBER+ = {PNUMBER, PNAME, PLOCATION}
{SSN, PNUMBER}+ = {SSN, PNUMBER, ENAME, PNAME, PLOCATION, HOURS}

```

DCC011 - profa Mirella M. Moro

55