



# Bancos de Dados

---

Profa. Patrícia R. Oliveira  
EACH - USP

Normalização e Dependências Funcionais em  
BD's Relacionais – Parte 1

slides parcialmente baseados em material de aula  
dos Profs. José E. Ferreira e José F. Rodrigues Junior



# Qualidade do projeto lógico

---

- Primeiramente, pode-se considerar um guia informal para projetos de BD's Relacionais.
  - semântica dos atributos;
  - redução de informação redundante em tuplas e anomalias de atualização;
  - redução de valores nulos em tuplas;
  - impedimento para geração de tuplas ilegítimas.



# Guia informal

---

## ■ Importante:

■ Projetar um BD Relacional com qualidade significa:

Agrupar atributos para formar  
“bons” esquemas de relações.



# Guia informal

---

- O que é um “bom” esquema?
  - No nível lógico:
    - é fácil de entender;
    - ajuda a formular consultas corretas.
  - No nível físico:
    - tuplas são armazenadas eficientemente;
    - tuplas são acessadas com eficiência.

# Guia informal – semântica dos atributos



---

- Diretriz 1: Cada tupla de uma relação deve representar uma entidade ou instância de relacionamento.
  - atributos de entidades distintas não devem estar na mesma relação.
  - apenas chaves estrangeiras devem ser usadas para referenciar outras entidades.
  - atributos de diferentes entidades e relacionamentos devem ser mantidos tão separadamente quanto possível.



# Guia informal – semântica dos atributos

---

## ■ Dicas:

- projete um esquema que possa ser facilmente explicado.
- a semântica dos atributos deve ser de fácil interpretação.

# Exemplo

- O que é mais claro? O esquema abaixo,

## EMPREGADO

ENOME	<u>SSN</u>	DATANASC	ENDERECO	DNUMERO
-------	------------	----------	----------	---------

## PROJETO

PNAME	<u>PNUMERO</u>	PLOCALIZACAO	DNUM
-------	----------------	--------------	------

## TRABALHA\_EM

<u>SSN</u>	<u>PNUMERO</u>	HORAS
------------	----------------	-------

- ou o esquema a seguir?

## EMP\_PROJ

<u>SSN</u>	<u>PNUMERO</u>	HORAS	ENOME	PNAME	PLOCALIZACAO
------------	----------------	-------	-------	-------	--------------

# Exemplo

- Qual esquema precisa de maior espaço de armazenamento? O esquema abaixo,

EMPREGADO

ENOME	<u>SSN</u>	DATANASC	ENDERECO	DNUMERO
-------	------------	----------	----------	---------

DEPARTAMENTO

DNOME	<u>DNUMERO</u>	DGERSSN
-------	----------------	---------

- ou o esquema a seguir?

EMP\_DEPT

ENOME	<u>SSN</u>	DATANASC	ENDERECO	DNUMERO	DNOME	DGERSSN
-------	------------	----------	----------	---------	-------	---------



**EMPREGADO**

ENOME	<u>SNN</u>	DATANASC	ENDERECO	DNUMERO
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5
Zelaya,Alicia J.	999887777	1968-07-19	3321 Castle,Spring,TX	4
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4
Narayan,Remesh K.	666884444	1962-09-15	975 Fire Oak,Humble,TX	5
English,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1

**DEPARTAMENTO**

DNOME	<u>DNUMERO</u>	DGERSSN
Pesquisa	5	333445555
Administração	4	987654321
Diretoria	1	888665555

redundância

**EMP\_DEPT**

ENOME	<u>SSN</u>	DATANASC	ENDERECO	DNUMERO	DNOME	DGERSSN
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5	Pesquisa	333445555
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5	Pesquisa	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle,Spring,TX	4	Administracao	987654321
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4	Administracao	987654321
Narayan,Ramesh K.	666884444	1962-09-15	975 FireOak,Humble,TX	5	Pesquisa	333445555
English,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5	Pesquisa	333445555
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4	Administracao	987654321
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1	Sede Administrativa	888665555



# Guia informal – informação redundante e anomalias de atualização

---

- Informações redundantes desperdiçam espaço de armazenamento.
- A mistura de atributos de várias entidades pode gerar problemas conhecidos como anomalias de atualização
  - anomalias de inserção
  - anomalias de eliminação
  - anomalias de modificação.

EMP\_DEPT

ENOME	<u>SSN</u>	DATANASC	ENDERECO	DNUMERO	DNOME	DGERSSN
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5	Pesquisa	333445555
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5	Pesquisa	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring,TX	4	Administracao	987654321
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4	Administracao	987654321
Narayan,Ramesh K.	666884444	1962-09-15	975 FireOak,Humble,TX	5	Pesquisa	333445555
English,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5	Pesquisa	333445555
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4	Administracao	987654321
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1	Sede Administrativa	888665555

## ■ Anomalia de inserção (1):

- para inserir dados de um empregado, temos que inserir dados para o departamento em que ele trabalha ou *nulls*.
- se o empregado trabalhar no departamento 5, devemos informar os valores dos atributos desse departamento de forma correta e consistente com as outras tuplas.

EMP\_DEPT

ENOME	<u>SSN</u>	DATANASC	ENDEREÇO	DNUMERO	DNOME	DGERSSN
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5	Pesquisa	333445555
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5	Pesquisa	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring,TX	4	Administracao	987654321
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4	Administracao	987654321
Narayan,Ramesh K.	666884444	1962-09-15	975 FireOak,Humble,TX	5	Pesquisa	333445555
English,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5	Pesquisa	333445555
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4	Administracao	987654321
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1	Sede Administrativa	888665555

## ■ Anomalia de inserção (2):

- não é possível inserir um novo departamento que ainda não tenha empregado.
- Por que?

EMP\_DEPT

ENOME	<u>SSN</u>	DATANASC	ENDereco	DNUMERO	DNOME	DGERSSN
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5	Pesquisa	333445555
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5	Pesquisa	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring,TX	4	Administracao	987654321
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4	Administracao	987654321
Narayan,Ramesh K.	666884444	1962-09-15	975 FireOak,Humble,TX	5	Pesquisa	333445555
English,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5	Pesquisa	333445555
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4	Administracao	987654321
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1	Sede Administrativa	888665555

- Suponha que quiséssemos mudar o gerente de um determinado departamento.
- Qual seria a principal dificuldade encontrada?

EMP\_DEPT

ENOME	<u>SSN</u>	DATANASC	ENDERECO	DNUMERO	DNOME	DGERSSN
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5	Pesquisa	333445555
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5	Pesquisa	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring,TX	4	Administracao	987654321
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4	Administracao	987654321
Narayan,Ramesh K.	666884444	1962-09-15	975 FireOak,Humble,TX	5	Pesquisa	333445555
English,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5	Pesquisa	333445555
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4	Administracao	987654321
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1	Sede Administrativa	888665555

- Anomalia de modificação: a mudança de um gerente de departamento implica na atualização das tuplas de todos os empregados que trabalham nesse departamento.

EMP\_DEPT

ENOME	<u>SSN</u>	DATANASC	ENDERECO	DNUMERO	DNOME	DGERSSN
Smith,John B.	123456789	1965-01-09	731 Fondren,Houston,TX	5	Pesquisa	333445555
Wong,Franklin T.	333445555	1955-12-08	638 Voss,Houston,TX	5	Pesquisa	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring,TX	4	Administracao	987654321
Wallace,Jennifer S.	987654321	1941-06-20	291 Berry,Bellaire,TX	4	Administracao	987654321
Narayan,Ramesh K.	666884444	1962-09-15	975 FireOak,Humble,TX	5	Pesquisa	333445555
English,Joyce A.	453453453	1972-07-31	5631 Rice,Houston,TX	5	Pesquisa	333445555
Jabbar,Ahmad V.	987987987	1969-03-29	980 Dallas,Houston,TX	4	Administracao	987654321
Borg,James E.	888665555	1937-11-10	450 Stone,Houston,TX	1	Sede Administrativa	888665555

## ■ Anomalia de remoção:

- quando uma tupla de empregado é removida, pode ser que ela represente o último empregado que trabalha nesse departamento.
  - O que acontece nesse caso?
  - E se um departamento tiver que ser extinto?



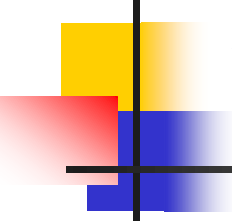
# Guia informal – informação redundante e anomalias de atualização

---

- Diretriz 2: projete um esquema que não sofra de anomalias de inserção, remoção e de modificação.
- Se existir alguma, então assegure-se de que as aplicações levem tais anomalias em consideração.



# Guia informal – valores nulos em tuplas



---

- Diretriz 3: relações devem ser projetadas de forma que suas tuplas tenham a menor quantidade possível de valores nulos.
- Normalmente, os atributos que possuem valores nulos podem ser colocados em uma relação separada (com uma chave primária).

# Exemplo

EMPREGADO

ENOME	<u>SSN</u>	DATANASC	ENDERECO	DNUMERO
-------	------------	----------	----------	---------

- Se só 10% dos empregados tiverem escritórios particulares:
  - há pouca justificativa para incluir um atributo NUM\_ESCRITORIO na relação EMPREGADO.
  - pode-se criar uma relação EMP\_ESCRITORIO(ESSN, NUM\_ESCRITORIO), contendo apenas tuplas de empregado que têm escritório.



# Guia informal – tuplas ilegítimas

---

- Projetos incorretos podem gerar resultados inválidos em operações de junção.
- Diretriz 4: evite relações que contenham relacionamentos entre atributos que não sejam combinações *chave estrangeira – chave primária*.
  - junções sobre esses atributos podem produzir tuplas ilegítimas.

## EMP\_PROJ

<u>SSN</u>	<u>PNUMERO</u>	HORAS	ENOME	PNOME	PLOCALIZACAO
------------	----------------	-------	-------	-------	--------------

- O que acontecesse se o esquema abaixo for usado ao invés desse acima?

## EMP\_LOCS

ENOME	PLOCALIZACAO
-------	--------------

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

## EMP\_PROJ1

SSN	PNUMERO	HORAS	PNOME	PLOCALIZACAO
123456789	1	32.5	Produto X	Bellaire
123456789	2	7.5	Produto Y	Sugarland
666884444	3	40.0	Produto Z	Houston
453453453	1	20.0	Produto X	Bellaire
453453453	2	20.0	Produto Y	Sugarland
333445555	2	10.0	Produto Y	Sugarland
333445555	3	10.0	Produto Z	Houston
333445555	10	10.0	Automação	Stafford
333445555	20	10.0	Reorganização	Houston

■ Podem ocorrer tuplas ilegítimas (\*) em uma operação de junção natural!

	<u>NSS</u>	<u>PNÚMERO</u>	<u>HORAS</u>	<u>PNOME</u>	<u>PLOCALIZAÇÃO</u>	<u>ENAME</u>
	123456789	1	32.5	ProdutoX	Bellaire	John Smith
*	123456789	1	32.5	ProdutoX	Bellaire	Joyce English
	123456789	2	7.5	ProdutoY	Sugarland	John Smith
*	123456789	2	7.5	ProdutoY	Sugarland	Joyce English
*	123456789	2	7.5	ProdutoY	Sugarland	Franklin Wong
	666884444	3	40.0	ProdutoZ	Houston	Ramesh Narayan
*	666884444	3	40.0	ProdutoZ	Houston	Franklin Wong
*	453453453	1	20.0	ProdutoX	Bellaire	John Smith
	453453453	1	20.0	ProdutoX	Bellaire	Joyce English
*	453453453	2	20.0	ProdutoY	Sugarland	John Smith
	453453453	2	20.0	ProdutoY	Sugarland	Joyce English
*	453453453	2	20.0	ProdutoY	Sugarland	Franklin Wong
*	333445555	2	10.0	ProdutoY	Sugarland	John Smith
*	333445555	2	10.0	ProdutoY	Sugarland	Joyce English
	333445555	2	10.0	ProdutoY	Sugarland	Franklin Wong
*	333445555	3	10.0	ProdutoZ	Houston	Ramesh Narayan
	333445555	3	10.0	ProdutoZ	Houston	Franklin Wong
	333445555	10	10.0	Automação	Stafford	Franklin Wong
*	333445555	20	10.0	Reorganização	Houston	Ramesh Narayan
	333445555	20	10.0	Reorganização	Houston	Franklin Wong



# Dependências funcionais

---

- Dependências Funcionais (DF's) são usadas para medir formalmente a qualidade do projeto lógico.
- As DF's e chaves são usadas para definir formas normais de relações.
- As DF's são restrições que são derivadas do significado e do inter-relacionamento dos dados de atributos.



# Dependências funcionais

---

- Um conjunto de atributos X determina funcionalmente um conjunto de atributos Y se o valor de X determinar um único valor Y.
- Exemplos:
  - o número do Seguro Social (SSN) determina o nome do empregado.
  - o número do projeto determina o nome do projeto e a sua localização.



# Dependências funcionais

---

- X determina funcionalmente Y é representado por:  
 $X \rightarrow Y$ .
- $X \rightarrow Y$  diz que se duas tuplas tiverem o mesmo valor para X, elas devem ter o mesmo valor para Y, ou seja:

Se  $X \rightarrow Y$ , então, para quaisquer tuplas  $t_1$  e  $t_2$  de  $r(R)$ :  
se  $t_1[X] = t_2[X]$ , então,  $t_1[Y] = t_2[Y]$ .





# Exemplos

---

- 1) O número do Seguro Social determina o nome do empregado:
  - $SSN \rightarrow ENOME$
- 2) O número do projeto determina o seu nome e a sua localização:
  - $PNUMERO \rightarrow \{PNOME, PLOCALIZACAO\}$
- 3) O SSN de um empregado e o número do projeto determinam as horas em que o empregado trabalha no projeto:
  - $\{SSN, PNUMERO\} \rightarrow HORAS$



# Dependências funcionais

---

## ■ Importante:

- $X \rightarrow Y$  especifica uma restrição sobre todas as instâncias de  $r(R)$ .
- As DF's são derivadas das restrições do mundo real
  - podem ser verificadas nas instâncias do BD, mas nunca definidas a partir delas.



# Exemplo

---

ENSINA		
PROFESSOR	CURSO	TEXTOS
Smith	Estruturas de dados	Bartram
Smith	Administração de dados	Al-Nour
Hall	Compiladores	Hoffman
Brown	Estruturas de dados	Augenthaler

- Pode-se pensar que TEXTO → CURSO
  - isso tem sentido semântico?
- Com certeza PROFESSOR não determina funcionalmente CURSO!



# Dependências funcionais

---

## ■ Controle de consistência:

- É necessário conhecer todas as dependências funcionais.

- informação semântica fornecida pelo projetista.

- DF's podem ser inferidas a partir de outras.

- regras de inferência.



# Regras de inferência para DF's

---

## ■ Regras de inferência de Armstrong:

■ RI1. (Reflexiva) Se  $X \supseteq Y$ , então,  $X \rightarrow Y$

■ Um conjunto de atributos sempre determina seus subconjuntos (ou determina a si próprio se  $X = Y$ ).

■ RI2. (Aumentativa) Se  $X \rightarrow Y$ , então,  $XZ \rightarrow YZ$ .

■ Adicionar um conjunto de atributos a ambos os lados de uma dependência resulta em uma outra dependência válida.

■ Notação:  $XZ$  significa  $\{X, Z\}$



# Regras de inferência para DF's

---

- RI3. (Transitiva) Se  $X \rightarrow Y$  e  $Y \rightarrow Z$ , então,  $X \rightarrow Z$ .
- Outras regras de inferência úteis:
  - (Decomposição) Se  $X \rightarrow YZ$ , então,  $X \rightarrow Y$  e  $X \rightarrow Z$ .
  - (Aditiva) Se  $X \rightarrow Y$  e  $X \rightarrow Z$ , então,  $X \rightarrow YZ$ .
- Obs: As duas regras acima podem ser deduzidas a partir de RI1, RI2 e RI3.



# Regras de inferência para DF's

---

- O projetista do esquema define as DF's que são mais óbvias.
- Entretanto, existem outras DF's que podem ser deduzidas.
- Exemplo:

$$F = \{SSN \rightarrow \{ENOME, DATANASC, ENDERECO, DNUMERO\}, \\ DNUMERO \rightarrow \{DNOME, DGERSSN\}\}$$

É possível inferir, por exemplo, que:

$$SSN \rightarrow SSN$$
$$SSN \rightarrow \{DNOME, DGERSSN\}$$
$$DNUMERO \rightarrow DNOME$$



# Normalização de relações

---

- Normalização: processo de decompor relações “ruins”, dividindo os seus atributos em relações menores.
- Forma normal: indica o nível de qualidade de uma relação.



# Uso prático das formas normais



---

- Na prática, a normalização é realizada para obter projetos de alta qualidade e atender às propriedades desejáveis.
- Os projetistas de BD's não precisam normalizar todas as relações na maior forma normal possível.
- Desnormalização: processo de armazenar junções de relações de forma normal superior como uma nova relação que terá uma forma normal inferior (Ex: slide 8).



# Revisão

---

- Uma superchave de uma relação esquema  $R = \{A_1, A_2, \dots, A_n\}$  é um conjunto de atributos  $S$ , subconjunto de  $R$ , com a propriedade de que:

$t_1[S] \neq t_2[S]$  para qualquer extensão  $r(R)$ .

- Uma superchave  $K$  é uma chave se  $K$  é uma superchave mínima.



# Revisão

---

- Se uma relação esquema tiver mais de uma chave, cada chave será chamada de chave-candidata.
  - Uma das chaves-candidatas é arbitrariamente escolhida para ser a chave-primária e as outras são chamadas de chaves-secundárias.

# Atributos primários e não primários



---

- Um atributo primário (ou primo) é membro da chave primária da relação.
- Um atributo não primário é um atributo que não é primário – isto é, não é membro da chave primária da relação.



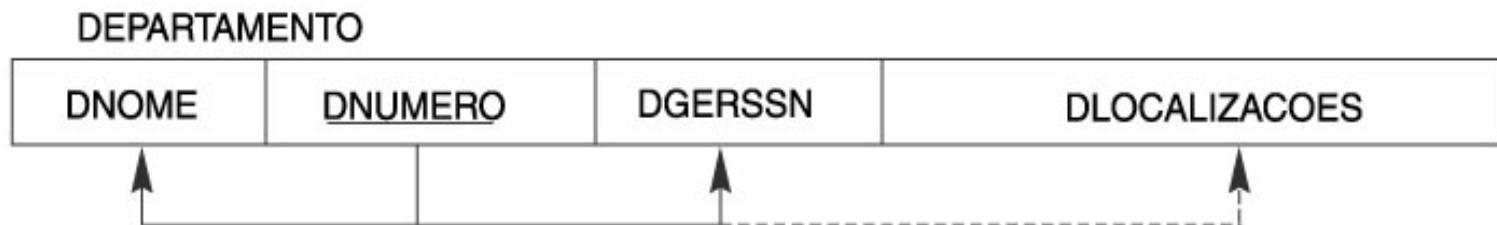
# Primeira forma normal

---

- Proíbe atributos compostos, atributos multivalorados e relações aninhadas.
  - ou seja, permite apenas atributos que sejam atômicos.
- Considerada como sendo parte da definição de relação.

# Normalização na 1FN

(a)



(b)

DEPARTAMENTO

DNOME	<u>DNUMERO</u>	DGERSSN	DLOCALIZACOES
Pesquisa	5	333445555	{Bellaire, Sugarland, Houston}
Administração	4	987654321	{Stafford}
Diretoria	1	888665555	{Houston}

(c)

DEPARTAMENTO

DNOME	<u>DNUMERO</u>	DGERSSN	<u>DLOCALIZACAO</u>
Pesquisa	5	333445555	Bellaire
Pesquisa	5	333445555	Sugarland
Pesquisa	5	333445555	Houston
Administração	4	987654321	Stafford
Diretoria	1	888665555	Houston

EMP\_PROJ (SSN, ENOME, {PROJS (PNUMERO, HORAS)})

(a) **EMP\_PROJ**

SSN	ENOME	PROJS	
		PNUMERO	HORAS

(b) **EMP\_PROJ**

SSN	ENOME	PNUMERO	HORAS
123456789	Smith, John B.	1	32.5
		2	7.5
666884444	Narayan, Ramesh K.	3	40.0
453453453	English, Joyce A.	1	20.0
		2	20.0
333445555	Wong, Franklin T.	2	10.0
		3	10.0
		10	10.0
		20	10.0
999887777	Zelaya, Alicia J.	30	30.0
		10	10.0
987987987	Jabbar, Ahmad V.	10	35.0
		30	5.0
987654321	Wallace, Jennifer S.	30	20.0
		20	15.0
888665555	Borg, James E.	20	null

■ No caso de relações aninhadas (atributos multivalorados compostos).

(c) **EMP\_PROJ1**

<u>SSN</u>	ENOME
------------	-------

**EMP\_PROJ2**

<u>SSN</u>	<u>PNUMERO</u>	HORAS
------------	----------------	-------



# Dependência funcional total

- Dependência funcional total – uma DF,  $Y \rightarrow Z$ , tal que a remoção de qualquer atributo de  $Y$  invalida a DF.
- Exemplos:
  - $\{ \text{SSN}, \text{PNUMERO} \} \rightarrow \text{HORAS}$  é uma DF total, uma vez que SSN não determina HORAS e nem PNUMERO determina HORAS.
  - $\{ \text{SSN}, \text{PNUMERO} \} \rightarrow \text{ENOME}$  não é uma DF total (é uma DF parcial) pois  $\text{SSN} \rightarrow \text{ENOME}$ .

EMP\_PROJ

<u>SSN</u>	<u>PNUMERO</u>	HORAS	ENOME	PNOME	PLOCALIZACAO
------------	----------------	-------	-------	-------	--------------





# Segunda forma normal

---

- Um esquema de relação  $R$  está na 2FN se todo atributo não primário  $A$  em  $R$  tem dependência funcional total da chave primária de  $R$ .
- $R$  pode ser decomposto em relações que estejam na 2FN por meio do processo de normalização.



# Segunda forma normal

---

- Exemplo:

- Turma = (CodDisc, Numero, Sala, No.Horas)



Suponha que cada disciplina tem sua quantidade de horas bem definida como ocorre no mundo real.

Assim, o modelo não deve permitir que duas disciplinas sejam armazenadas com número de horas diferentes.



# Segunda forma normal

---

- Exemplo:
  - Turma = (CodDisc, Numero, Sala, No.Horas)

<u>CodDisc</u>	<u>Numero</u>	Sala	No.Horas
Matemática	1	24	60
Computação	1	13	30
Computação	2	25	30
Matemática	2	31	50

# Segunda forma normal

- Exemplo:

- Turma = (CodDisc, Numero, Sala, No.Horas)

<u>CodDisc</u>	<u>Numero</u>	Sala	No.Horas
Matemática	1	24	60
Computação	1	13	30
Computação	2	25	30
Matemática	2	31	50

Dependência  
funcional  
parcial  
à chave





# Segunda forma normal

- Exemplo:

- $Turma = (\underline{CodDisc}, \underline{Numero}, Sala, No.Horas)$

<u>CodDisc</u>	<u>Numero</u>	Sala	No.Horas
Matemática	1	24	60
Computação	1	13	30
Computação	2	25	30
Matemática	2	31	50

Se  $CodDisc = Computação \rightarrow No. Horas = 30$

Se  $CodDisc = Matemática \rightarrow No. Horas = 60 \text{ ou } 50?$

# Segunda forma normal

- Exemplo:

- Turma = (CodDisc, Numero, Sala, No.Horas)

<u>CodDisc</u>	<u>Numero</u>	Sala	No.Horas
Matemática	1	24	60
Computação	1	13	30
Computação	2	25	30
Matemática	2	31	50

Redundância e  
Anomalia  
de inserção

- Como normalizar esta relação?

Turma = (CodDisc, Numero, Sala, No.Horas)



1) Nova relação:

parte da chave que define a dependência  
(chave da nova relação)

+

atributos dependentes desta parte da chave

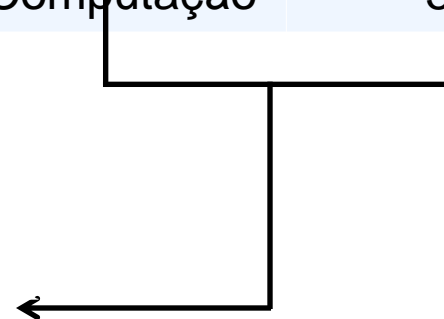
2) Atributos parcialmente dependentes da  
chave saem da relação

# Segunda forma normal

- Relação normalizada – sem DFs parciais à chave primária:

<u>CodDisc</u>	<u>Numero</u>	Sala
Matemática	1	24
Computação	1	13
Computação	2	25
Matemática	2	31

<u>CodDisc</u>	No.Horas
Matemática	60
Computação	30



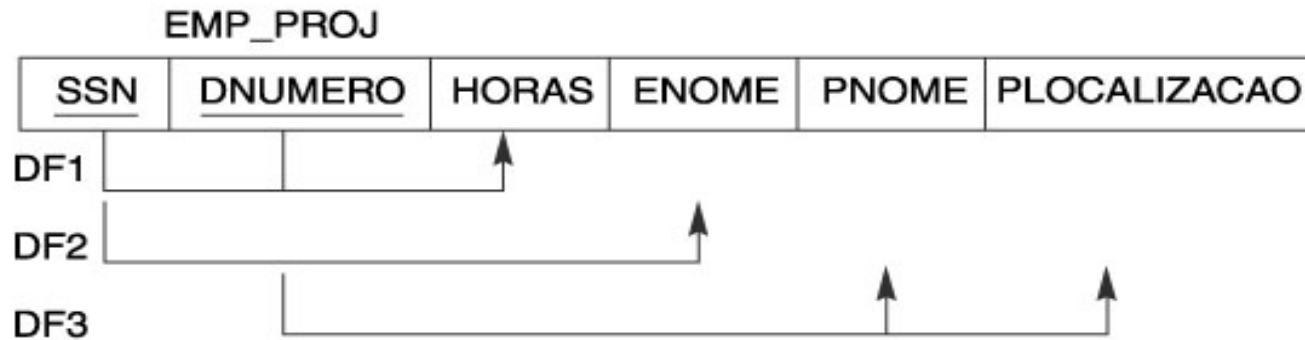
A própria estrutura do esquema:

- garante a não redundância
- garante a consistência (apenas uma versão dos dados)

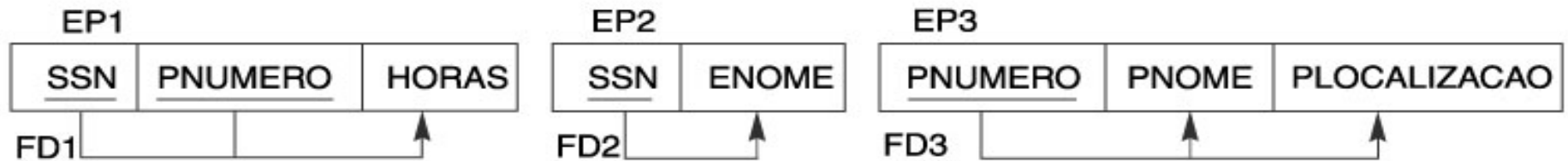


# Normalização na 2FN

(a)



NORMALIZAÇÃO 2FN



# Dependência funcional transitiva

- Dependência funcional transitiva – uma DF,  $X \rightarrow Z$ , que pode ser derivada a partir de duas DFs  $X \rightarrow Y$  e  $Y \rightarrow Z$ .
- Exemplos:
  - $SSN \rightarrow DGERSSN$  é uma DF transitiva pois  $SSN \rightarrow DNUMERO$  e  $DNUMERO \rightarrow DGERSSN$
  - $SSN \rightarrow ENOME$  não é transitiva pois não existe um conjunto de atributos  $X$  em que  $SSN \rightarrow X$  e  $X \rightarrow ENOME$

EMP_DEPT						
ENOME	<u>SSN</u>	DATANASC	ENDERECO	DNUMERO	DNOME	DGERSSN



# Terceira forma normal

---

- Um esquema de relação  $R$  está na 3FN se estiver na 2FN e nenhum atributo não-primário,  $A$ , for transitivamente dependente da chave-primária.
- $R$  pode ser decomposto em relações que estejam na 3FN via o processo de normalização.
- NOTA: Em  $X \rightarrow Y$  e  $Y \rightarrow Z$ , sendo  $X$  a chave-primária, pode ser considerado um problema se, e somente se,  $Y$  não for uma chave-candidata.
  - Quando  $Y$  é uma chave-candidata, não existe problema com a dependência transitiva.

# Terceira forma normal

- Exemplo:

Campeoes(Competicaold, Ano, Vencedor, DataNascVenc)



Como:

$(\text{Competicaold}, \text{Ano}) \rightarrow \text{Vencedor}$  e  $\text{Vencedor} \rightarrow \text{DataNascVenc}$

Então, por transitividade:

$(\text{Competicaold}, \text{Ano}) \rightarrow \text{DataNascVenc}$

# Terceira forma normal

- Exemplo:

Campeoes(CompeticaoId, Ano, Vencedor, DataNascVenc)




<u>CompeticaoId</u>	<u>Ano</u>	Vencedor	DataNascVenc
C21	2001	Miguel	04/04/1975
C34	2002	César	09/12/1980
C21	2003	Miguel	04/04/1975
C57	2004	Fabiano	06/02/1978

# Terceira forma normal

- Exemplo:

Campeoes(CompeticaoId, Ano, Vencedor, DataNascVenc)



<u>CompeticaoId</u>	<u>Ano</u>	Vencedor	DataNascVenc
C21	2001	Miguel	04/04/1975
C34	2002	César	09/12/1980
C21	2003	Miguel	04/04/1975
C57	2004	Fabiano	06/02/1978


Atributo não-primário DataNascVenc depende transitivamente da chave {CompeticaoId, Ano}



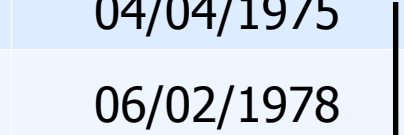
# Terceira forma normal

- Exemplo:

Campeoes(CompeticaoId, Ano, Vencedor, DataNascVenc)



<u>CompeticaoId</u>	<u>Ano</u>	Vencedor	DataNascVenc
C21	2001	Miguel	04/04/1975
C34	2002	César	09/12/1980
C21	2003	Miguel	04/04/1975
C57	2004	Fabiano	06/02/1978




Redundância

Se Vencedor = Miguel → DataNascVenc = 04/04/1975

# Terceira forma normal

Campeoes(CompeticaoId, Ano, Vencedor, DataNascVenc)



<u>CompeticaoId</u>	<u>Ano</u>	Vencedor	DataNascVenc
C21	2001	Miguel	14/04/1975
C34	2002	César	09/12/1980
C21	2003	Miguel	04/04/1975
C57	2004	Fabiano	06/02/1978

- O vencedor da competição C21 de 2001 forneceu data de nascimento errada – precisamos atualizar os dados

UPDATE Campeoes

SET DataNascVenc='14/04/1975'

WHERE CompeticaoId = C21 AND ANO=2001



# Terceira forma normal

Campeoes(CompeticaoId, Ano, Vencedor, DataNascVenc)

<u>CompeticaoId</u>	<u>Ano</u>	Vencedor	DataNascVenc
C21	2001	Miguel	14/04/1975
C34	2002	César	09/12/1980
C21	2003	Miguel	04/04/1975
C57	2004	Fabiano	06/02/1978

Anomalia de  
atualização




# Terceira forma normal

---

.Como normalizar essa relação?

Campeoes(CompeticaoId, Ano, Vencedor, DataNascVenc)



1) Nova relação:

atributos dependentes transitivamente

+

atributos dos quais eles dependem  
diretamente

(chave da nova relação)

2) Atributos dependentes transitivamente saem  
da relação

# Terceira forma normal

.Como normalizar essa relação?

Campeoes(CompeticaoId, Ano, Vencedor, DataNascVenc)

1) Nova relação:

atributos dependentes transitivamente

+

atributos dos quais eles dependem  
diretamente

(chave da nova relação)

2) Atributos dependentes transitivamente saem  
da relação

# Terceira forma normal

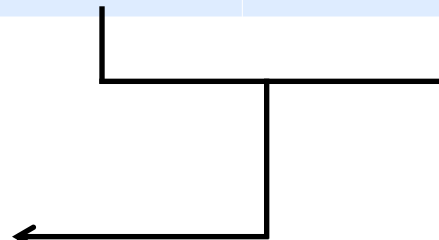
- Relação normalizada sem dependências funcionais transitivas a qualquer chave.

<u>Competicao</u>	<u>Ano</u>	<u>Vencedor</u>
C21	2001	Miguel
C34	2002	César
C21	2003	Miguel
C57	2004	Fabiano

<u>Vencedor</u>	<u>DataNascVenc</u>
Miguel	14/04/1975
César	09/12/1980
Fabiano	06/02/1978

A própria estrutura do esquema:

- garante a não redundância
- garante a consistência de atualização





# Exercício

---

■ Considere a seguinte relação e suas dependências funcionais. Indique se as formas normais 1NF, 2NF e 3F são atendidas. Se necessário, normalize a relação para atender a cada uma destas formas.

Livro = {ISBN, autor, instituicaoAutor, titulo, editora, tipo, preço}

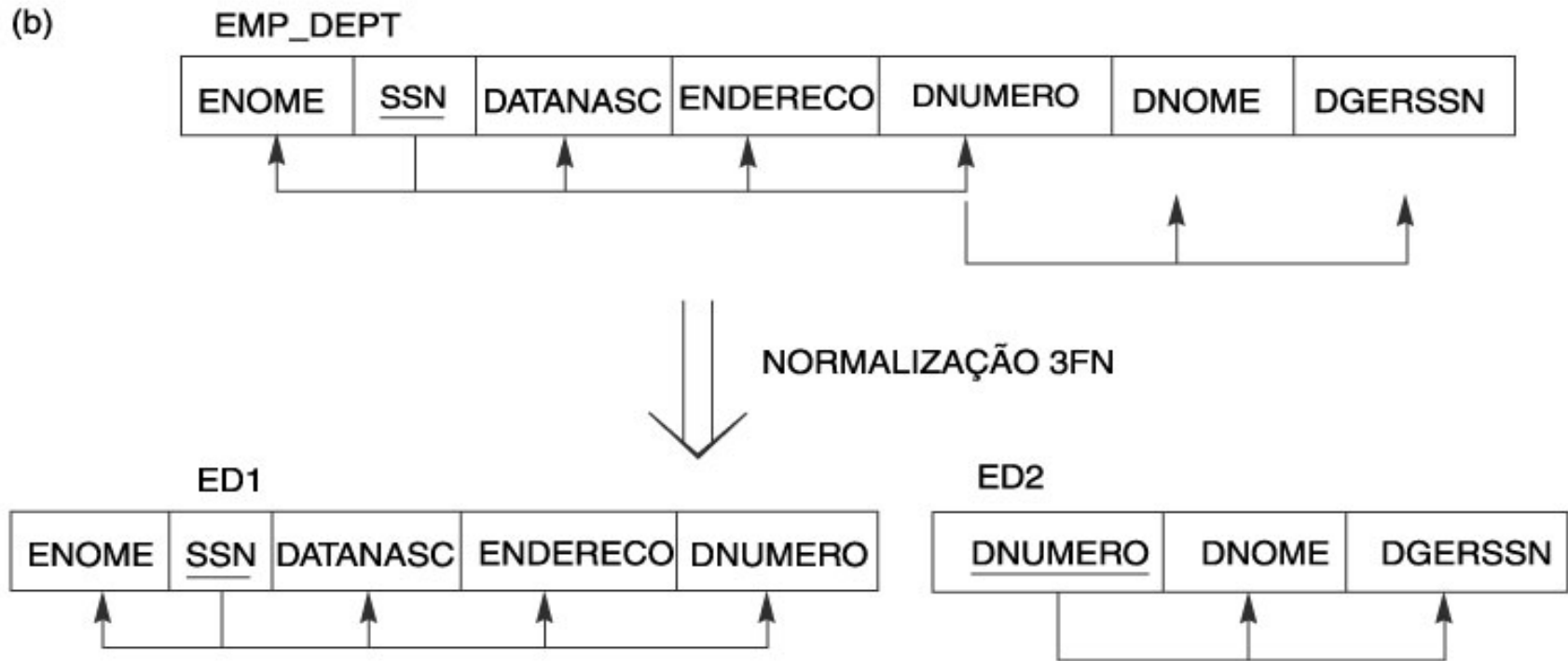
ISBN  $\rightarrow$  autor, titulo, editora, tipo

autor  $\rightarrow$  instituiçãoAutor

tipo  $\rightarrow$  preço

autor, titulo  $\rightarrow$  ISBN, editora

# Normalização na 3FN



# Definição geral de formas normais



---

- As definições anteriores consideravam somente a chave primária.
- As próximas definições levarão em consideração as várias chaves candidatas.
- Será considerado como primário, um atributo que faça parte de qualquer chave candidata.



# Definição geral de formas normais

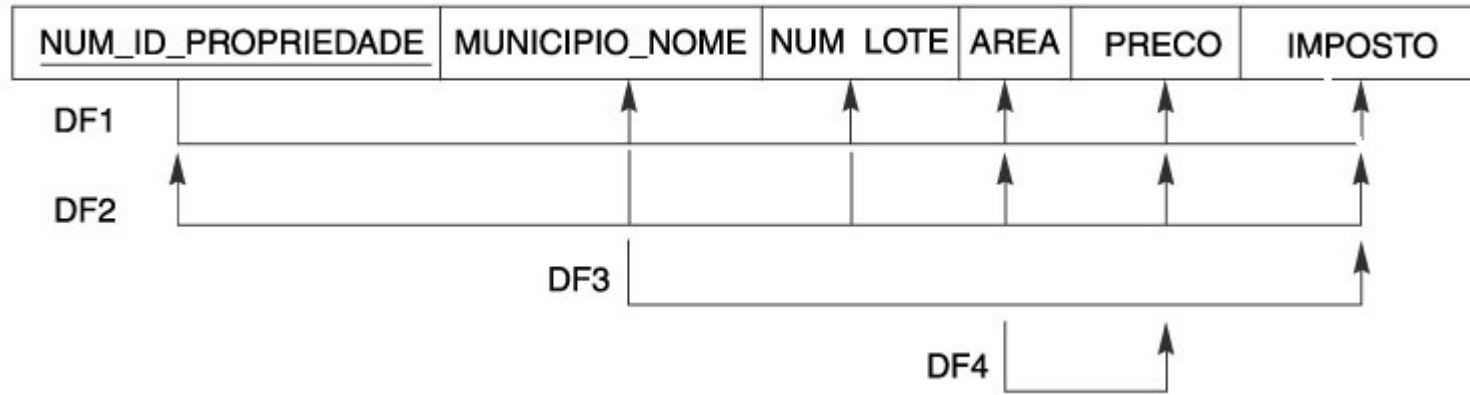
---

## ■ Redefinição da 2FN:

- Uma relação  $R$  está na 2FN se cada atributo não primário  $A$  de  $R$  não for parcialmente dependente de nenhuma chave candidata de  $R$ .

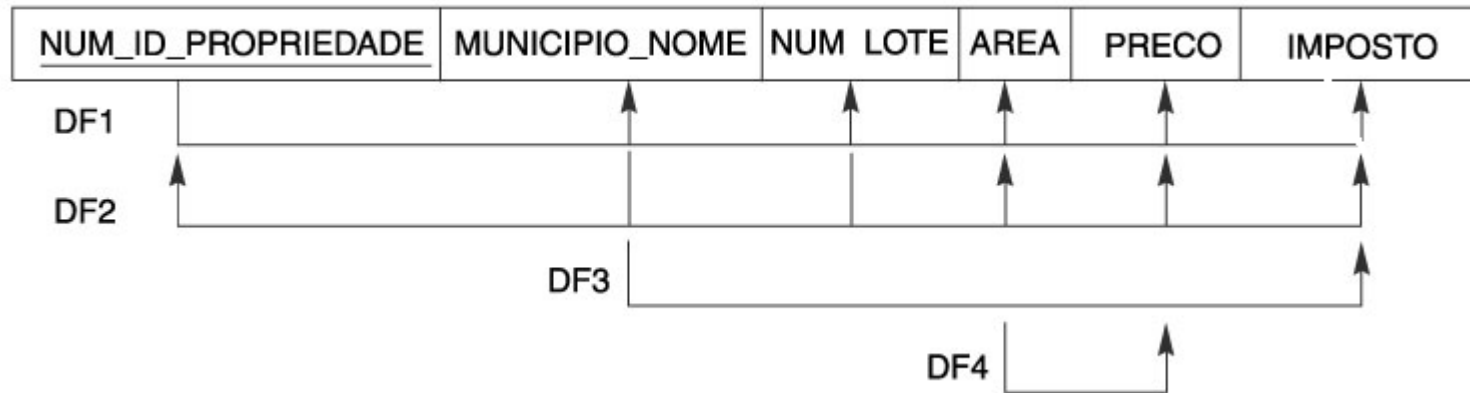


(a) LOTES



- LOTES descreve os lotes de terra à venda em vários municípios de um estado.
- Duas chaves candidatas:
  - NUM\_ID\_PROPRIEDADE
  - {MUNICIPIO\_NOME, NUM\_LOTE}

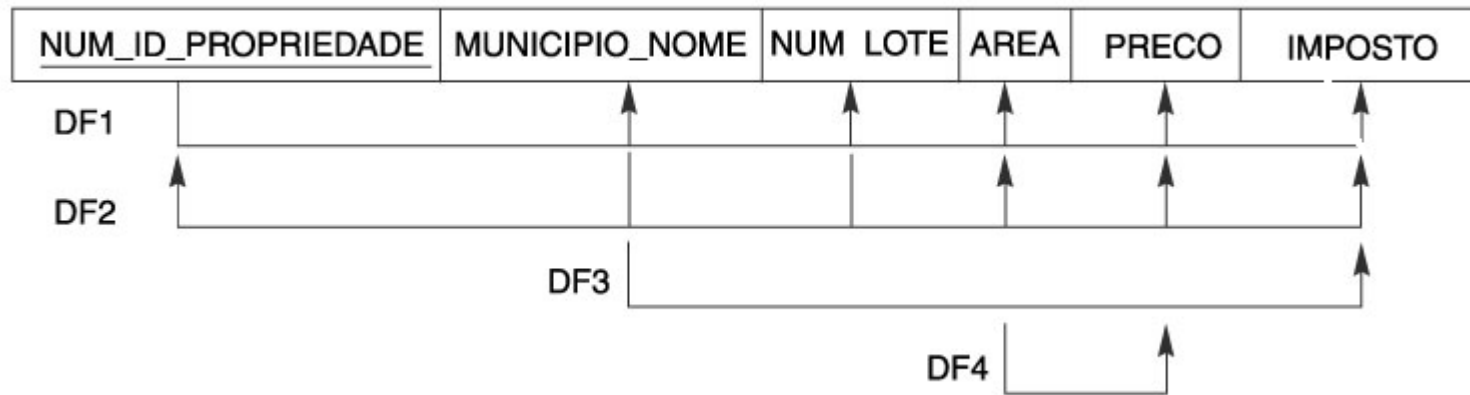
(a) LOTES



- DF1 e DF2: dependências funcionais dos atributos não primários em relação às chaves candidatas.
- DF3: MUNICIPIO\_NOME → IMPOSTO
  - a taxa de imposto é fixa para um dado município.
- DF4: AREA → PRECO
  - o preço de um lote é determinado por sua área, independente do município a quem pertença.

(a)

LOTES



■ DF3: viola a definição geral da 2FN!!

■ IMPOSTO parcialmente dependente da chave candidata {MUNICIPIO\_NOME, NUM\_LOTE}.

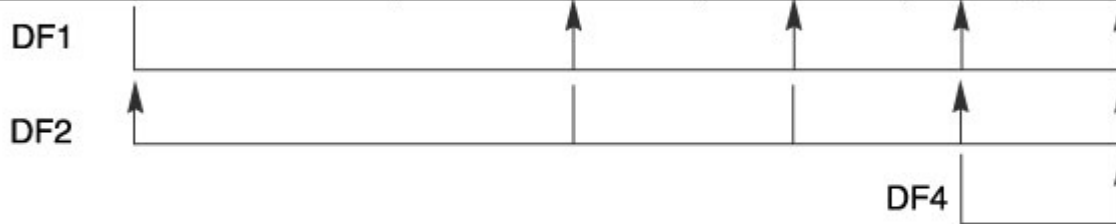
■ Solução: decompor LOTES em duas relações:

■ LOTES1: resulta da remoção de IMPOSTO da relação LOTES.

■ LOTES2: resulta de uma nova relação com os atributos IMPOSTO e MUNICIPIO\_NOME (chave primária de LOTES2).

(b) LOTES1

<u>NUM_ID_PROPRIEDADE</u>	MUNICIPIO_NOME	NUM_LOTE	AREA	PRECO
---------------------------	----------------	----------	------	-------



LOTES2

<u>MUNICIPIO NOME</u>	IMPOSTO
-----------------------	---------



■ Nota: DF4 não viola a definição geral da 2FN!!

# Definição geral de formas normais

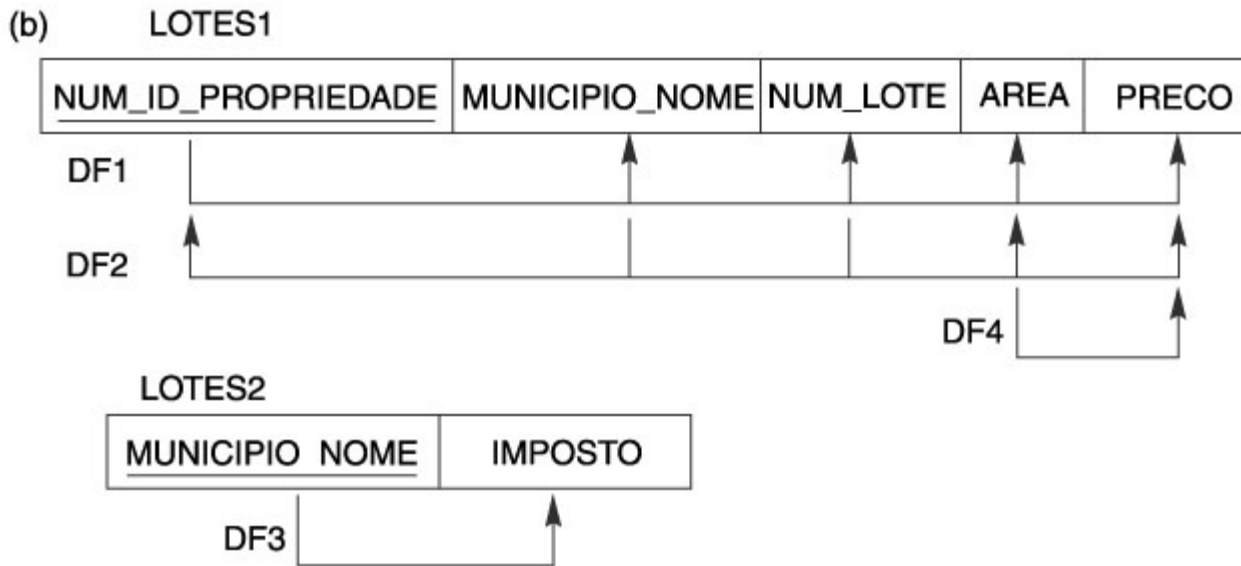
## ■ Redefinição da 3FN:

■ Uma relação esquema  $R$  está na 3FN se, sempre que houver uma DF  $X \rightarrow A$ , então uma das duas condições são válidas:

■  $X$  é uma superchave de  $R$ , ou

■  $A$  é um atributo primário de  $R$ .

■ NOTA: A Forma normal de Boyce-Codd não admite a segunda condição.

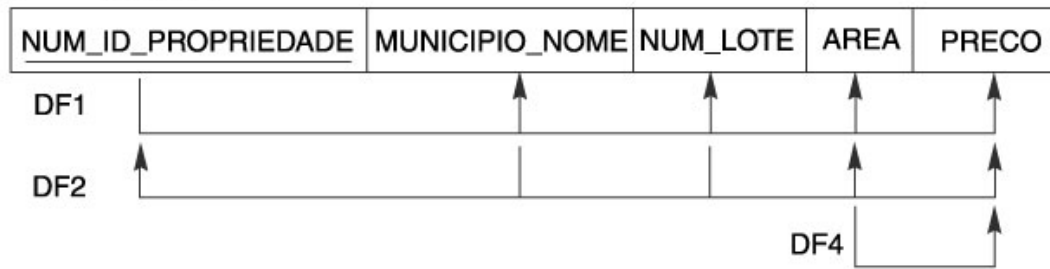


■ LOTES2 está na 3FN.

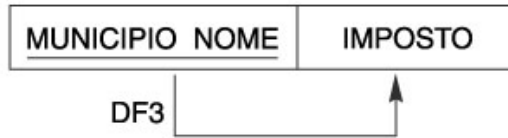
■ DF4 em LOTES1 viola a 3FN:

■ AREA não é uma superchave e PRECO não é atributo primário.

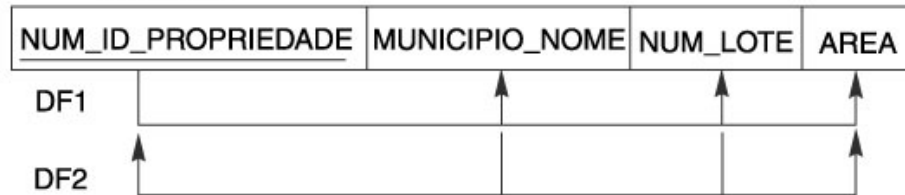
(b) LOTES1



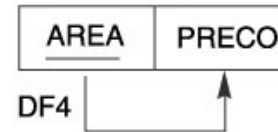
LOTES2



(c) LOTES1A



LOTES1B



■ Solução: decompor LOTES1 em duas relações:

■ LOTES1A: resulta da remoção de PRECO da relação LOTES1.

■ LOTES1B: resulta de uma nova relação com os atributos PRECO e AREA (chave primária de LOTES1A).

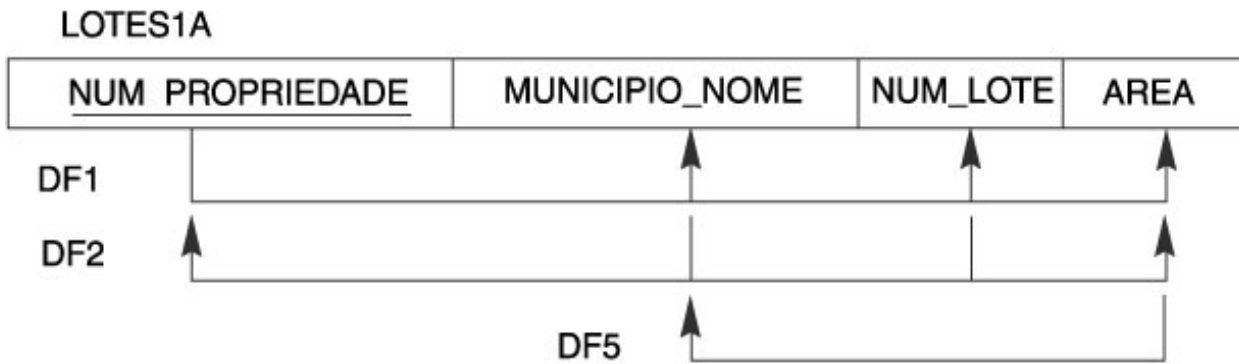
# Forma normal de Boyce-Codd (BCNF)



---

- Um esquema de relação  $R$  está na BCNF se, sempre que houver uma DF  $X \rightarrow A$  em  $R$ , então  $X$  é uma superchave de  $R$ .
- Cada FN engloba a FN anterior:
  - Toda relação em 2FN está na 1FN.
  - Toda relação em 3FN está na 2FN.
  - Toda relação em BCNF está na 3FN.
- Existem relações que estão na 3FN mas não em BCNF.





■ LOTES1A está na 3FN, mas não está na BCNF.

■ por causa da DF5: AREA → MUNICIPIO\_NOME.

■ Solução: decompor LOTES1A em duas relações:

■ LOTES1AX: resulta da remoção de MUNICIPIO\_NOME da relação LOTES1A.

■ LOTES1AY: resulta de uma nova relação com os atributos MUNICIPIO\_NOME e AREA (chave primária de LOTES1AY).



ENSINA

ALUNO	CURSO	INSTRUTOR
Narayan	Banco de dados	Mark
Smith	Banco de dados	Navathe
Smith	Sistemas operacionais	Ammar
Smith	Teoria	Schulman
Wallace	Banco de dados	Mark
Wallace	Sistemas operacionais	Ahamad
Wong	Banco de dados	Omiecinski
Zelaya	Banco de dados	Navathe

- $\{\underline{\text{ALUNO}}, \underline{\text{CURSO}}\}$ : chave candidata
- A dependência  $\text{INSTRUTOR} \rightarrow \text{CURSO}$  viola a BCNF.
- Normalização para BCNF:
  - decompor os dois esquemas de relação em:
  - $\{\underline{\text{INSTRUTOR}}, \underline{\text{CURSO}}\}$  e  $\{\underline{\text{ALUNO}}, \underline{\text{INSTRUTOR}}\}$