

DCC011: Introdução a Banco de Dados

Rodrygo Santos

rodrygo@dcc.ufmg.br

Departamento de Ciência da Computação
Universidade Federal de Minas Gerais

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

Emp_Dept

<u>SSN</u>	EName	Bdate	Address	DNumber	DName	DMgrSSN
------------	-------	-------	---------	---------	-------	---------

Emp_Proj

<u>SSN</u>	<u>PNumber</u>	Hours	EName	PName	PLocation
------------	----------------	-------	-------	-------	-----------

Semântica dos atributos da relação

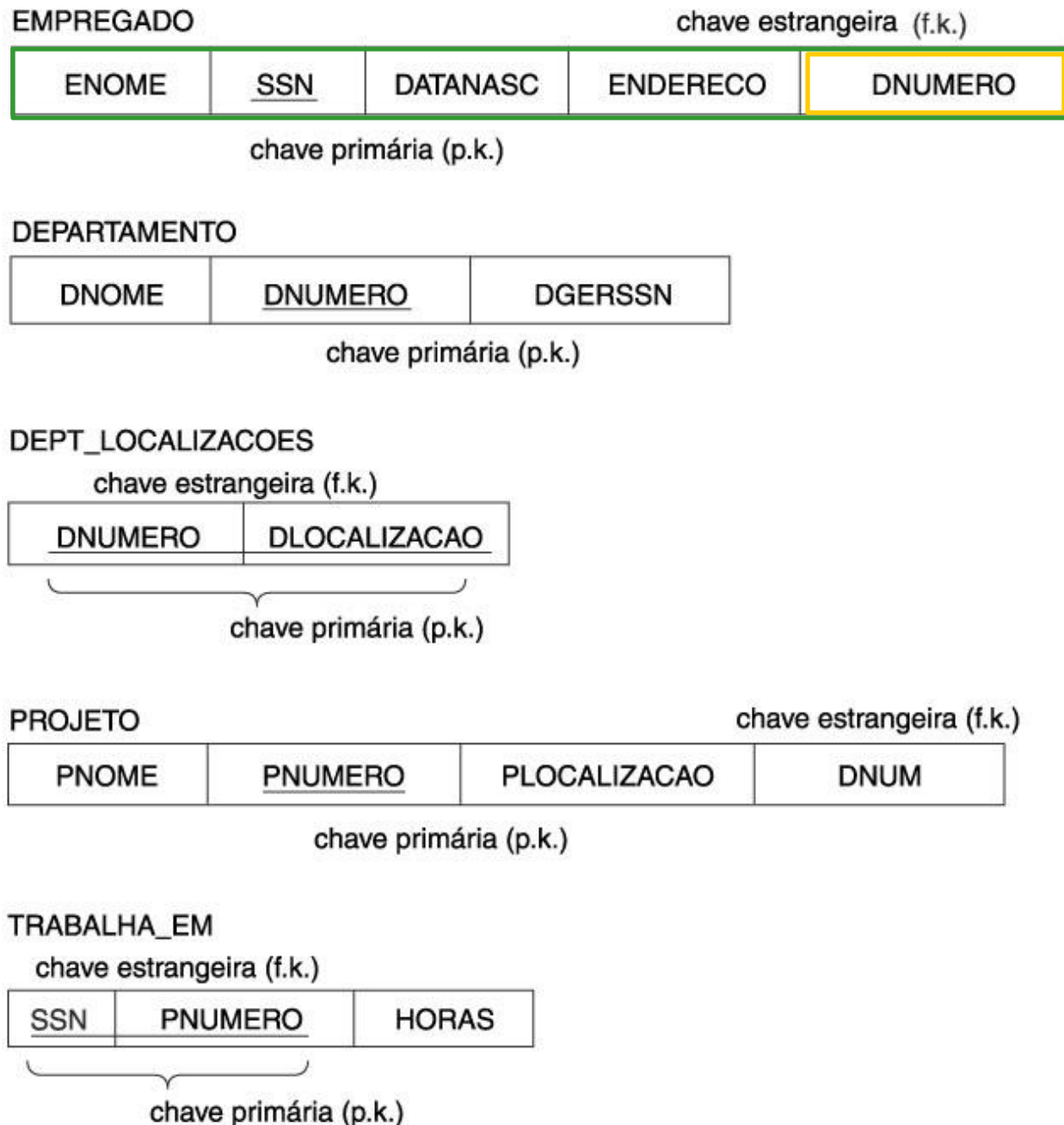


Figura 10.1 Um esquema simplificado do banco de dados relacional EMPRESA.

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

redundância

EMP_DEPT

ENOME	SSN	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

EMP_PROJ

redundância

redundância

SSN	PNUMERO	HORAS	ENOME	PNOME	PLOCALIZACAO
123456789	1	32.5	Smith,John B.	ProdutoX	Bellaire
123456789	2	7.5	Smith,John B.	ProdutoY	Sugarland
666884444	3	40.0	Narayan,Ramesh K.	ProdutoZ	Houston
453453453	1	20.0	English,Joyce A.	ProdutoX	Bellaire
453453453	2	20.0	English,Joyce A.	ProdutoY	Sugarland
333445555	2	10.0	Wong,Franklin T.	ProdutoY	Sugarland
333445555	3	10.0	Wong,Franklin T.	ProdutoZ	Houston
333445555	10	10.0	Wong,Franklin T.	Automação	Stafford
333445555	20	10.0	Wong,Franklin T.	Reorganização	Houston
999887777	30	30.0	Zelaya,Alicia J.	NovosBenefícios	Stafford
999887777	10	10.0	Zelaya,Alicia J.	Automação	Stafford
987987987	10	35.0	Jabbar,Ahmad V.	Automação	Stafford
987987987	30	5.0	Jabbar,Ahmad V.	Novos benefícios	Stafford
987654321	30	20.0	Wallace,Jennifer S.	Novos benefícios	Stafford
987654321	20	15.0	Wallace,Jennifer S.	Reorganização	Houston
888665555	20	null	Borg,James E.	Reorganização	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 como relações básicas 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

EMPREGADO

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

chave primária (p.k.)

Opção 2

EMPREGADO

chave estrangeira (f.k.)

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

chave primária (p.k.)

DEPARTAMENTO

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

chave primária (p.k.)

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

EMPREGADO

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

chave primária (p.k.)

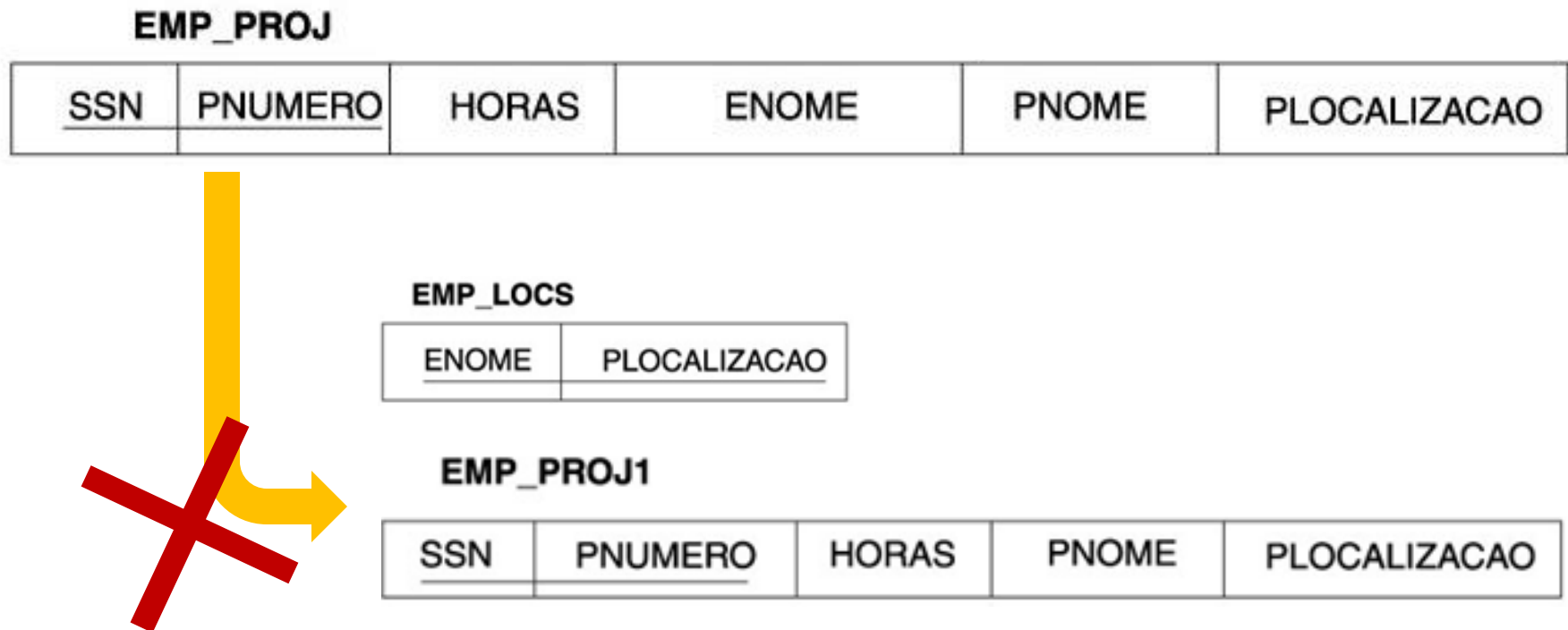
Informação redundante em tuplas e Anomalias de atualização

- **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:
 - Interpretação de valores
 - Junções (INNER e OUTER)
 - 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 que eles são a exceção e não a regra.
 - Office_number for 15% of employees? New relation!

D. Geração de Tuplas espúrias



Junção natural pode acidentalmente casar atributo não-único (PLOCALIZACAO)

Geração de Tuplas espúrias

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_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
Zelaya, Alicia J.	Stafford
Jabbar, Ahmad V.	Stafford
Wallace, Jennifer S.	Stafford
Wallace, Jennifer S.	Houston
Borg, James E.	Houston

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
999887777	30	30.0	Novos benefícios	Stafford
999887777	10	10.0	Automação	Stafford
987987987	10	35.0	Automação	Stafford
987987987	30	5.0	Novos Benefícios	Stafford
987654321	30	20.0	Novos Benefícios	Stafford
987654321	20	15.0	Reorganização	Houston
888665555	20	null	Reorganização	Houston

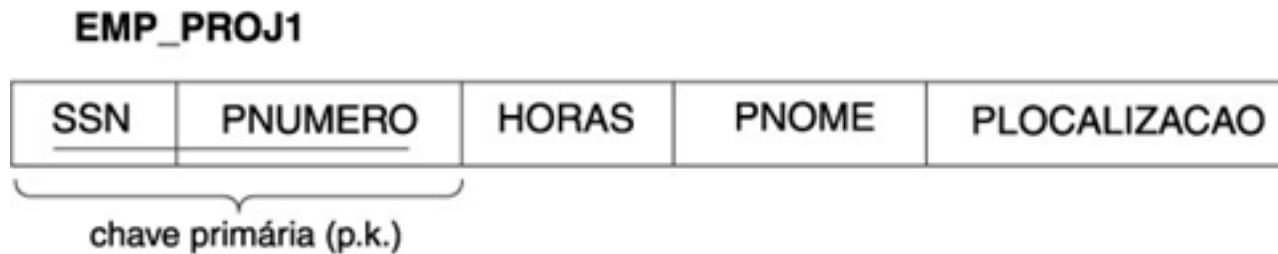
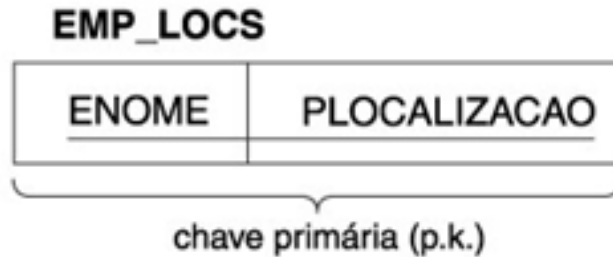
Geração de Tuplas espúrias

SSN	PNUMERO	HORAS	PNOME	PLOCALIZACAO	ENOME
123456789	1	32.5	Produto X	Bellaire	Smith,John B.
* 123456789	1	32.5	Produto Y	Bellaire	English,Joyce A.
123456789	2	7.5	Produto Y	Sugarland	Smith,John B.
* 123456789	2	7.5	Produto Y	Sugarland	English,Joyce A.
* 123456789	2	7.5	Produto Y	Sugarland	Wong,Franklin T.
666884444	3	40.0	Produto Z	Houston	Narayan,Ramesh K.
* 666884444	3	40.0	Produto Z	Houston	Wong,Franklin T.
* 453453453	1	20.0	Produto X	Bellaire	Smith,John B.
453453453	1	20.0	Produto X	Bellaire	English,Joyce A.
* 453453453	2	20.0	Produto Y	Sugarland	Smith,John B.
453453453	2	20.0	Produto Y	Sugarland	English,Joyce A.
* 453453453	2	20.0	Produto Y	Sugarland	Wong,Franklin T.
* 333445555	2	10.0	Produto Y	Sugarland	Smith,John B.
* 333445555	2	10.0	Produto Y	Sugarland	English,Joyce A.
333445555	2	10.0	Produto Y	Sugarland	Wong,Franklin T.
* 333445555	3	10.0	Produto Z	Houston	Narayan,Ramesh K.
333445555	3	10.0	Produto Z	Houston	Wong,Franklin T.
333445555	10	10.0	Automação	Stafford	Wong,Franklin T.
* 333445555	20	10.0	Reorganização	Houston	Narayan,Ramesh K.
333445555	20	10.0	Reorganização	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.

Geração de Tuplas espúrias

(a)



- **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.

Formas Normais

Primeira Forma Normal (1FN)

- Primeira Forma Normal (1FN):

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

= simples e monovalorados

- Contra-exemplo: atributo composto

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

Primeira Forma Normal (1FN)

- Primeira Forma Normal (1FN):

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

= simples e monovalorados

- Contra-exemplo: atributo multivalorado

Department

Dnumber	Dname	DMgrSSN	Dlocation
5	Research	33344555 5	{Bellaire, Sugarland, Houston}
4	Administration	98765432 1	{Stafford}
1	Headquarters	88866655 5	{Houston}

1FN

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



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

Dependências Funcionais

- Processo de normalização:

1FN \longrightarrow 2FN \longrightarrow 3FN

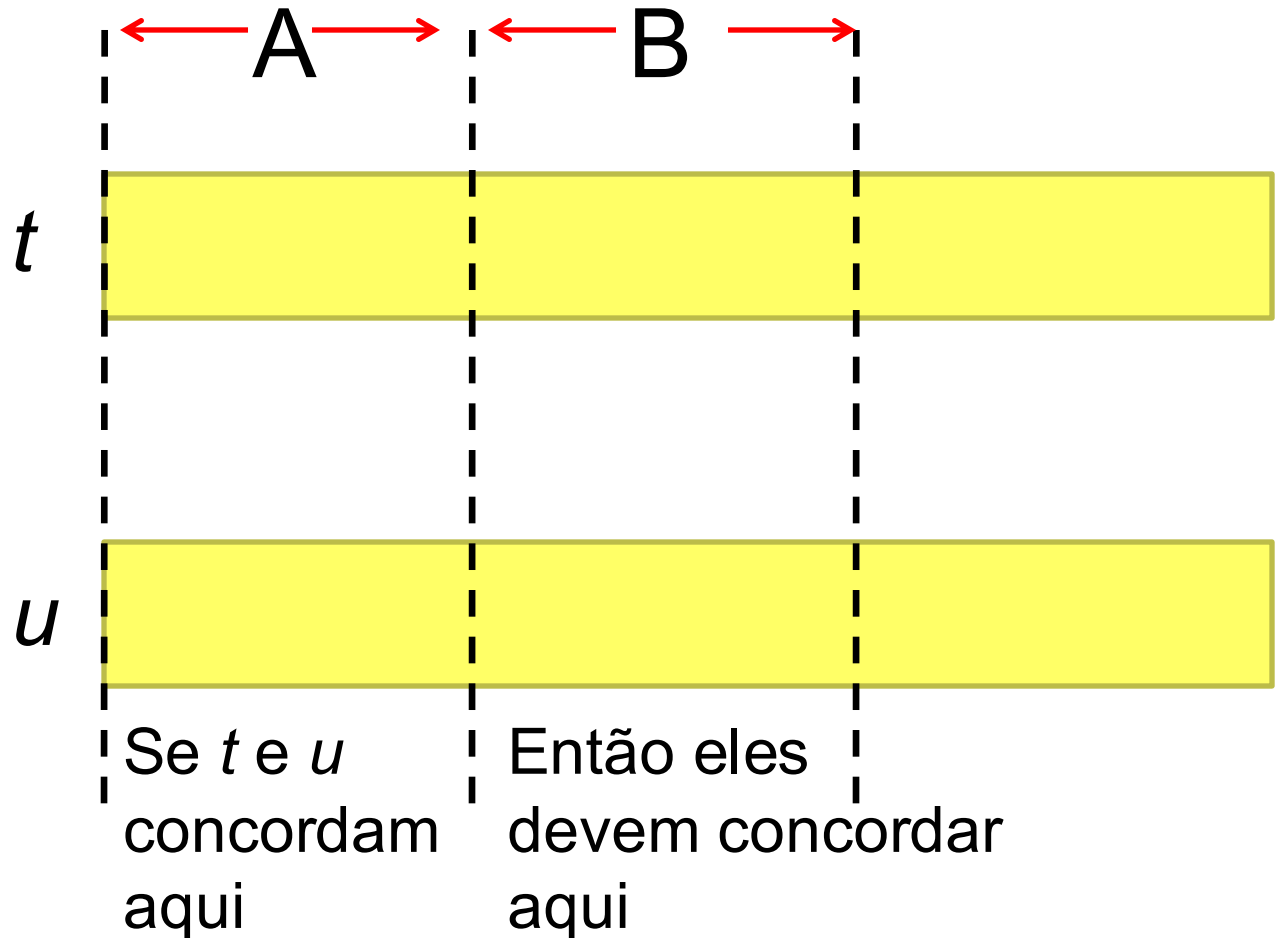
Formas normais 2FN e 3FN definidas a partir do conceito de **dependência funcional**.

Dependências Funcionais

- Definição informal
 - 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"
- Em outras palavras
 - Se conhecermos o valor de "A" então podemos encontrar o valor de "B" associado a ele

Dependências Funcionais

Se duas tuplas em R concordam nos atributos $A1, A2, \dots, AN$ (i.e. as tuplas têm os mesmos valores), então eles devem concordar em um outro atributo, B



Dependências Funcionais

- Definição formal:

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 $t1[X] = t2[X]$, $t1[Y] = t2[Y]$.

- Exemplos:

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

$SSN \rightarrow \{ENAME, BDATE, ADDRESS, DNUMBER\}$

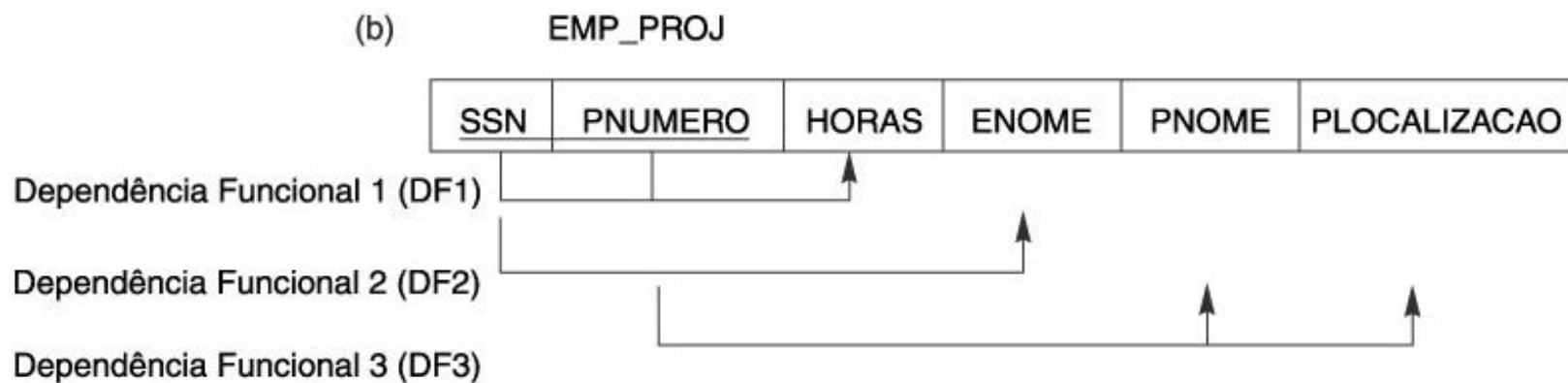
$DNUMBER \rightarrow \{DNAME, DMGSSN\}$

Atenção: Não confundir a notação $X \rightarrow Y$ com a notação $R1[X] \rightarrow R2[Y]$ que denota a RIR entre os esquemas de relação $R1$ e $R2$.

Dependências Funcionais

- Se X é uma chave candidata
 - Isso implica que $X \rightarrow Y$ para qualquer subconjunto de atributos Y de R
- Deve ser definida explicitamente no processo de projeto e normalização
- Devem refletir a semântica dos atributos
 - $\{\text{Orgão Emissor, Identidade}\} \rightarrow \text{CPF}$
 - $\text{CEP} \rightarrow \text{Rua}$
 - $\text{Professor} \not\rightarrow \text{Curso}$

Dependências Funcionais



- DF1: {SSN, PNUMERO} → HORAS
- DF2: SSN → ENOME
- DF3: PNUMERO → {PNOME, PLOCALIZACAO}

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)

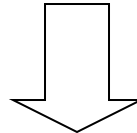
não está na 2FN porque

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

2FN

Emp_Proj

<u>SSN</u>	<u>PNumber</u>	Hours	EName	PName	PLocation



EP1

<u>SSN</u>	<u>PNumber</u>	Hours
------------	----------------	-------

EP2

<u>SSN</u>	Ename
------------	-------

EP3

<u>PNumber</u>	PName	PLocation
----------------	-------	-----------

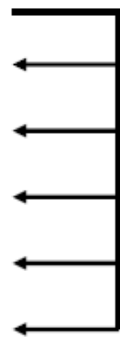
2FN, outro exemplo

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

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

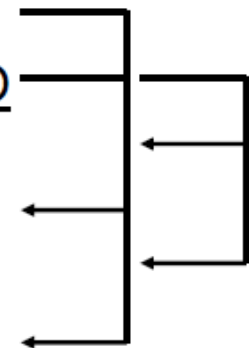
PEDIDO

NUMERO_PEDIDO
DATA_PEDIDO
NUMERO_CLIENTE
NOME_CLIENTE
ENDERECO_CLIENTE
TOTAL_PEDIDO




PRODUTO_PEDIDO

NUMERO_PEDIDO
NUMERO_PRODUTO
NOME_PRODUTO
QTDE_PEDIDA
PRECO_PRODUTO
TOTAL_PRODUTO



2FN, outro exemplo

- PEDIDO (NUMERO_PEDIDO, DATA_PEDIDO, NUMERO_CLIENTE, NOME_CLIENTE, ENDERECO_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, ENDERECO_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 não-chave de R depender de outro atributo não-chave

Terceira Forma Normal (3FN)

- Um esquema de relação R está na 3FN se estiver na 2FN e nenhum atributo não-chave de R for **transitivamente dependente** da chave primária
- Uma dependência $X \rightarrow Y$ é transitiva se existe um conjunto de atributos não-chave Z para o qual ambas as dependências existem: $X \rightarrow Z$ e $Z \rightarrow Y$

3FN

- O esquema de relação

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

não está na 3FN porque?

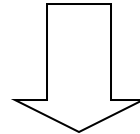
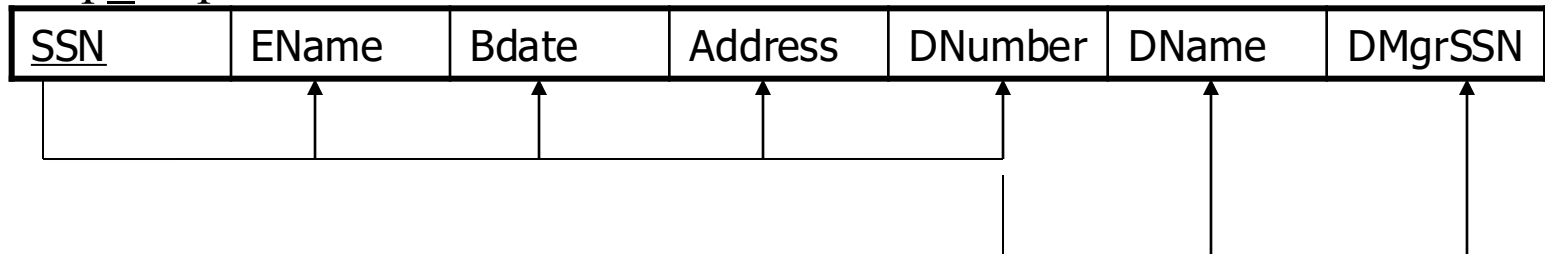
$SSN \rightarrow DNUMBER$ e $DNUMBER \rightarrow \{DNAME, DMGRSSN\}$

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

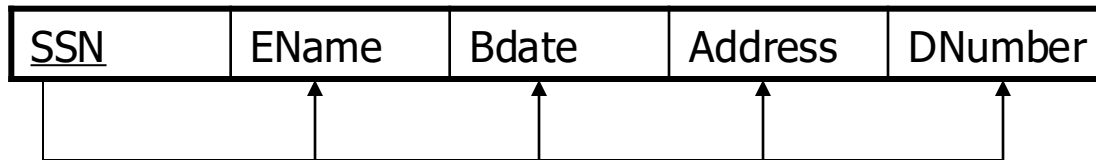
ou, há atributos em R que são dependentes de atributos não-chave

3FN

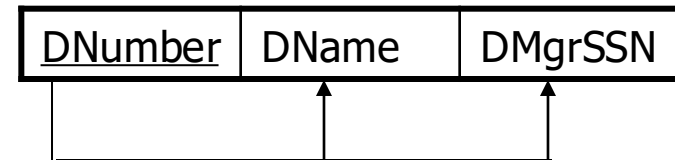
Emp_Dept



ED1



ED2



3FN, 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

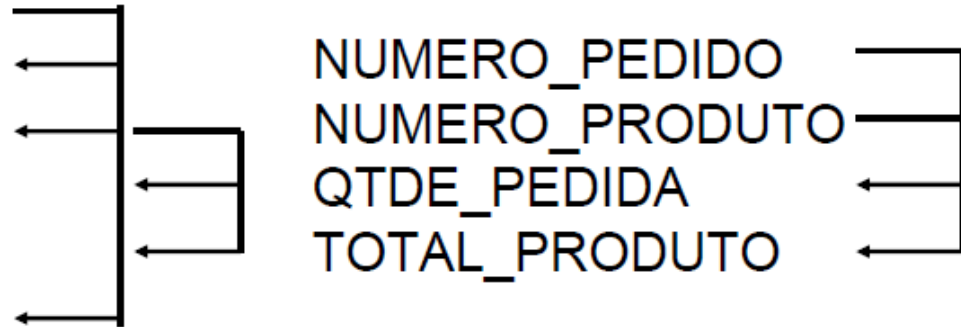
NUMERO_PEDIDO
DATA_PEDIDO
NUMERO_CLIENTE
NOME_CLIENTE
ENDERECO_CLIENTE
TOTAL_PEDIDO

PRODUTO_PEDIDO


NUMERO_PEDIDO
NUMERO_PRODUTO
QTDE_PEDIDA
TOTAL_PRODUTO

PRODUTO

NUMERO_PRODUTO
NOME_PRODUTO
PRECO_PRODUTO



3FN, 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)

3FN

PEDIDO (NUMERO_PEDIDO, NUMERO_CLIENTE,
DATA_PEDIDO, TOTAL_PEDIDO)

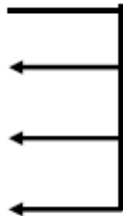
CLIENTE (NUMERO_CLIENTE, NOME_CLIENTE,
ENDERECO_CLIENTE)

PRODUTO_PEDIDO (NUMERO_PEDIDO,
NUMERO_PRODUTO, QTDE_PEDIDA,
TOTAL_PRODUTO)

PRODUTO (NUMERO_PRODUTO, NOME_PRODUTO,
PRECO_PRODUTO)

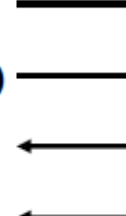
PEDIDO

NUMERO_PEDIDO
NUMERO_CLIENTE
DATA_PEDIDO
TOTAL_PEDIDO




PRODUTO_PEDIDO

NUMERO_PEDIDO
NUMERO_PRODUTO
QTDE_PEDIDA
TOTAL_PRODUTO



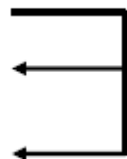
CLIENTE

NUMERO_CLIENTE
NOME_CLIENTE
ENDERECO_CLIENTE



PRODUTO

NUMERO_PRODUTO
NOME_PRODUTO
PRECO_PRODUTO



Revisão

Forma Normal	Diagnóstico	Remédio (normalização)
1FN	Relação não tem atributos compostos nem multivalorados	Definir novas relações para cada atributo composto ou multivalorado
2FN	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
3FN	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.