PCS3413

Engenharia de Software e Banco de Dados

Aula 21

Escola Politécnica da Universidade de São Paulo

NORMALIZAÇÃO

NORMALIZAÇÃO

- Conceito próprio de BDR que objetiva eliminar redundância de dados de arquivos, permitindo a construção de relações que não apresentam anomalias quando são manipuladas.

- Permite o reagrupamento de informações para a obtenção de um modelo E-R.

- Formas Normais:
  - regras que originam tabelas bem projetadas
  - 1FN, 2FN, 3FN, BCNF, 4NF e 5NF

Dependência Funcional

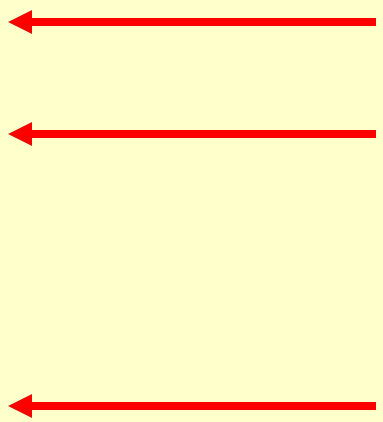
- Dada uma relação (tabela) R , o atributo Y de R é funcionalmente dependente do Atributo X de R (**R.X** $\rightarrow$ **R.Y**) , se e somente se para cada valor de X tem a ele associado apenas um valor de Y.

Neste caso, diz-se que Y é dependente funcional de X.

ou que X determina o valor de Y.

Exemplo:

•••	Matricula	Nome	•••	Salário	•••
	E1	**João**		**10**	
	E3	Ana		10	
	E1	**João**		**10**	
	E2	Maria		5	
	E3	Ana		10	
	E2	Maria		5	
	E1	**João**		**10**	



- Diz-se que a coluna **Salário e Nome *depende funcionalmente* da coluna Matrícula** (ou que a coluna **Matrícula *determina* as colunas Salário e Nome**) porque **cada valor de matríccula está associado sempre ao mesmo valor de salário e aos mesmo valor de nome.**

  Matricula → Salário (indica a dependência funcional)

  Matricula → Nome

Dependência Funcional da Chave Completa

AlocaçãoProjeto (**matrícula, codigo_projeto**, horas_projeto)

(matrícula, codigo_projeto) → horas_projeto

Exemplo de documento a ser normalizado: Arquivo com dados de NotasFiscais em PASCAL

```
type pedido  = record
      num: integer;
      data: char;
      RG_cli: char_60;
      end_cli: char;
      total: real;
      discriminação_itens: array [1..10] of record
            cod_item: integer;
            descr_item: integer;
            preço_unit: real;
            qtde_item: integer;
            subtotal: real;
            end;
end;
arq_pedido = file of  pedido;
```

Pedido: estrutura não normalizada

Campos compostos por outros campos

Tabela aninhada

Pedido

Num	Data	RG_Cli	EndCli	Discriminação					Total
				Cod_item	Descr_item	preço unit	Qtde_item	SubTotal	
123	07/01/01	357769475	R. X,101	a1	calça	80	2	160	470
				b2	meia	10	1	10	
				c3	gravata	30	10	300	
124	01/01/01	405467485	R. AA, 1	b2	meia	10	5	50	50

multivaloramento

Pedido (num, data, RG_Cli, end_cli, total,
(cod_item, descr_item, preço_unti, qtde_item, subtotal))

Primeira Forma Normal (1FN)

Diz-se que uma relação está na 1FN se e somente se todos os atributos contém somente valores atômicos.

Banco de Dados Relacionais: as relações devem estar na 1FN

Como tornar a estrutura com tabelas aninhadas (não atômico) em uma estrutura atômica?

eliminando campos compostos por outros campos e campos com valores multivalorados

para o exemplo do arquivo pedido:

num	data	RG_cli	end_cli	**cod_ item**	descr_ item	preço unit	qtde_ item	subTotal	total
123	07/01/01	357769475	R. X,101	a1	calça	80	2	160	470
123	07/01/01	357769475	R. X,101	b2	meia	10	1	10	470
123	07/01/01	357769475	R. X,101	c3	gravata	30	10	300	470
124	01/01/01	405467485	R. AA, 1	b2	meia	10	5	50	50

redundância

2FN

PEDIDO (1FN):

Pedido (**num, cod_item**, data,RG_cli, end_cli, qtd_item, descr_item, preço_unit, subtotal, total)

Anomalias na 1FN

- Insersão: não é possível incluir um novo item (cod e descr) sem existir um pedido.
- Eliminação: pode ocorrer a perda de informação sobre um item ao eliminar-se um pedido.
- Modificação: na alteração do preço unitário de um item, deve-se alterar todos os pedidos em que este aparece.

Outro exemplo

RELATÓRIO DE ALOCAÇÃO A PROJETO

CÓDIGO DO PROJETO: LSC001 **TIPO**: Novo Desenv.

DESCRIÇÃO: Sistema de Estoque

CÓDIGO DO EMPREGADO	NOME	CATEGORIA FUNCIONAL	SALÁRIO	DATA DE INÍCIO NO PROJETO	TEMPO ALOCADO AO PROJETO
2146	João	A1	4	1/11/91	24
3145	Sílvio	A2	4	2/10/91	24
6126	José	B1	9	3/10/92	18
1214	Carlos	A2	4	4/10/92	18
8191	Mário	A1	4	1/11/92	12

CÓDIGO DO PROJETO: PAG02 TIPO: Manutenção

DESCRIÇÃO: Sistema de RH

CÓDIGO DO EMPREGADO	NOME	CATEGORIA FUNCIONAL	SALÁRIO	DATA DE INÍCIO NO PROJETO	TEMPO ALOCADO AO PROJETO
8191	Mário	A1	4	1/05/93	12
4112	João	A2	4	4/01/91	24
6126	José	B1	9	1/11/92	12

1º Passo: transformar o documento numa Tabela Não Normalizada

tabela aninhada ou grupo repetitivo ou
coluna multivalorada ou coluna não atômica

Proj

CodProj	Tipo	Descr	Emp					
			CodEmp	**Nome**	**Cat**	**Sal**	**DataIni**	**TempAl**
LSC001	Novo Desenv.	Sistema de Estoque	2146	João	A1	4	1/11/91	24
			3145	Sílvio	A2	4	2/10/91	24
			6126	José	B1	9	3/10/92	18
			1214	Carlos	A2	4	4/10/92	18
			8191	Mário	A1	4	1/11/92	12
PAG02	Manut.	Sistema de RH	8191	Mário	A1	4	1/05/93	12
			4112	João	A2	4	4/01/91	24
			6126	José	B1	9	1/11/92	12

Proj (CodProj, Tipo, Descr,
(CodEmp, Nome, Cat, Sal, DataIni, TempAl))

Transformar numa estrutura normalizada ou na 1FN:

Proj (**CodProj**, **CodEmp**, Tipo, Descr, Nome, Cat, Sal, DataIni, TempAl)

CodProj	Tipo	Descr	**CodEmp**	Nome	Cat	Sal	DataIni	TempAl
LSC001	Novo Desenv.	Sistema de Estoque	2146	João	A1	4	1/11/91	24
LSC001	Novo Desenv.	Sistema de Estoque	3145	Sílvio	A2	4	2/10/91	24
LSC001	Novo Desenv.	Sistema de Estoque	6126	José	B1	9	3/10/92	18
LSC001	Novo Desenv.	Sistema de Estoque	1214	Carlos	A2	4	4/10/92	18
LSC001	Novo Desenv.	Sistema de Estoque	8191	Mário	A1	4	1/11/92	12
PAG02	Manutenção	Sistema de RH	8191	Mário	A1	4	1/05/93	12
PAG02	Manutenção	Sistema de RH	4112	João	A2	4	4/01/91	24
PAG02	Manutenção	Sistema de RH	6126	José	B1	9	1/11/92	12

2ª opção para geração na 1FN:

- Construir uma tabela para cada tabela aninhada

1. cria-se uma tabela na 1FN referente a tabela externa (sem tabelas aninhadas);
2. para cada tabela aninhada cria-se uma tabela na 1FN com os campos: chave-primária de cada uma das tabelas da qual ela é aninhada + colunas da própria tabela
3. defini-se as chaves primárias de cada uma das tabelas criadas

Proj (CodProj, Tipo, Descr,
(CodEmp, Nome, Cat, Sal, DataIni, TempAl))

Proj (**CodProj**, Tipo, Descr)

CodProj	Tipo	Descr
LSC001	Novo Desenv.	Sistema de Estoque
PAG02	Manutenção	Sistema de RH

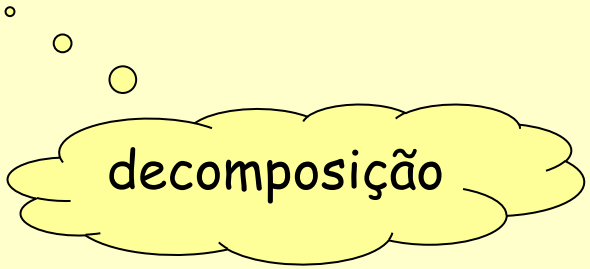
ProjEmp (**CodProj**, **CodEmp**, Nome, Cat, Sal, DataIni, TempAl)

CodProj	**CodEmp**	Nome	Cat	Sal	DataIni	TempAl
LSC001	2146	João	A1	4	1/11/91	24
LSC001	3145	Sílvio	A2	4	2/10/91	24
LSC001	6126	José	B1	9	3/10/92	18
LSC001	1214	Carlos	A2	4	4/10/92	18
LSC001	8191	Mário	A1	4	1/11/92	12
PAG02	8191	Mário	A1	4	1/05/93	12
PAG02	4112	João	A2	4	4/01/91	24
PAG02	6126	José	B1	9	1/11/92	12

Segunda Forma Normal (2FN)

Segunda Forma Normal (2FN)

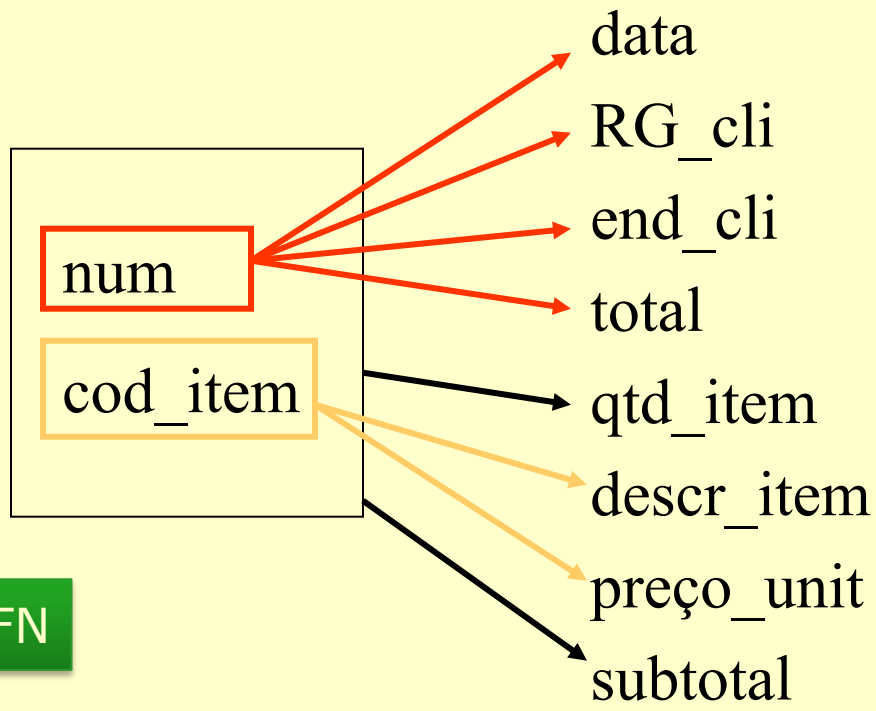
Diz-se que uma tabela está na 2FN se e somente se ela estiver na 1FN e cada coluna não chave for dependente funcional completo da chave principal.



Dependência Funcional para o exemplo

PEDIDO (1FN):

Pedido (**num, cod_item**, data, RG_cli, end_cli, qtd_item, descr_item, preço_unit, subtotal, total)



2FN

Nota (**num**, data, RG_cli, end_cli,total)

Item (**cod_item**, descr_item, preço_unit)

Pedido (**num, cod_item**, qtd_item, subtotal)

Tabelas na 2FN

NOTA

num	data	RG_cli	end_cli	total
123	07/01/01	357769475	R. X,101	470
124	01/01/01	405467485	R. AA, 1	50

ITEM

cod_ item	descr_ item	preço unit
a1	calça	80
b2	meia	10
c3	gravata	30

PEDIDO

num	**cod_ item**	qtde_ item	subTotal
123	a1	2	160
123	b2	1	10
123	c3	10	300
124	b2	5	50

NOTA (**num**, data, RG_cli, end_cli,total)

ITEM (**cod_item**, descr_item, preço_unit)

PEDIDO (**num, cod_item**, qtd_item, subtotal)

3FN

Algumas anomalias na 2FN

- Inserção: não é possível registrar o endereço de um cliente, a não ser que ele tenha feito um pedido.
- Eliminação: se for eliminado um pedido, é possível a perda de informação prospectiva sobre um determinado cliente.
- Modificação: se o cliente mudar de endereço, é preciso alterar todos os pedidos em que este cliente aparece.

Terceira Forma Normal (3FN)

Terceira Forma Normal (3FN)

Diz-se que uma tabela está na 3FN se e somente se ela estiver na 2FN e todo atributo não chave não depender de outro atributo não chave.

- eliminação de dependências entre atributos não chave
- eliminação de atributos derivados

Para o exemplo, estão na 3FN:

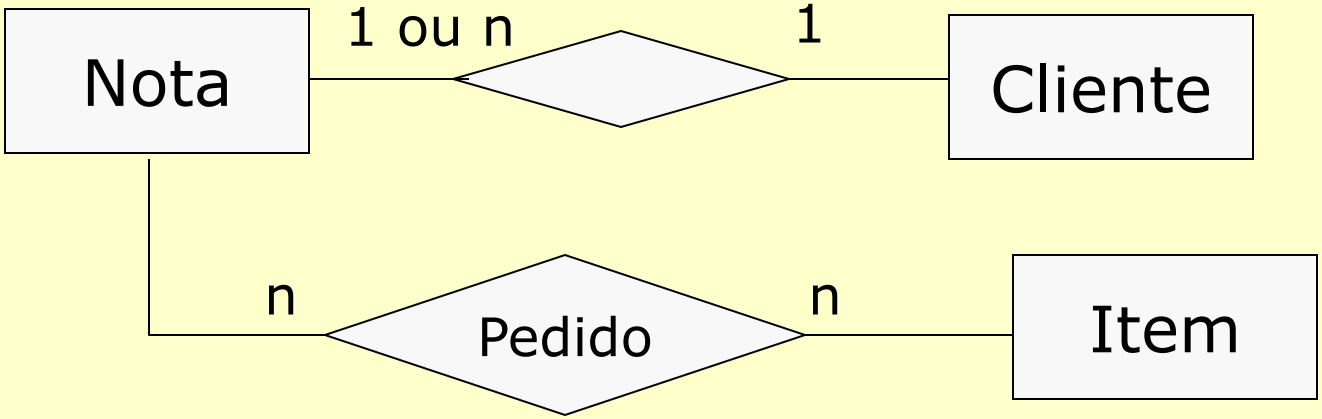
NOTA (**num**, data, RG_cli)

CLIENTE (**RG_cli**, end_cli)

ITEM (**cod**, descr_item, preço_unit)

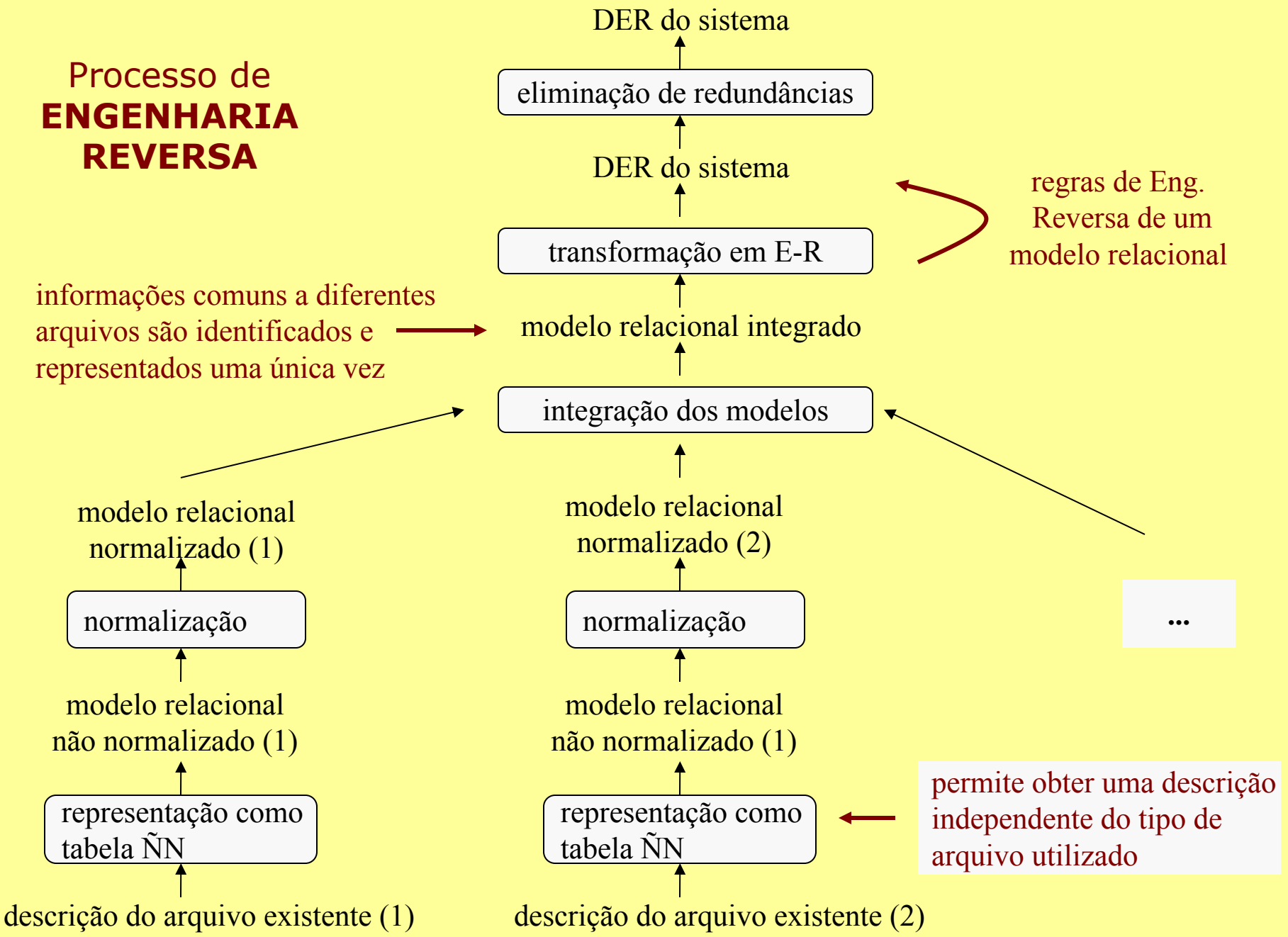
PEDIDO (**num, cod**, qtd_item)

Modelo E-R para o exemplo:



Nota (**num**, data, nome_cli)
CLIENTE (**nome_cli**, end_cli)
ITEM (**cod**, descr_item, preço_unit)
Pedido (**num, cod**, qtd_item)

Processo de **ENGENHARIA REVERSA**



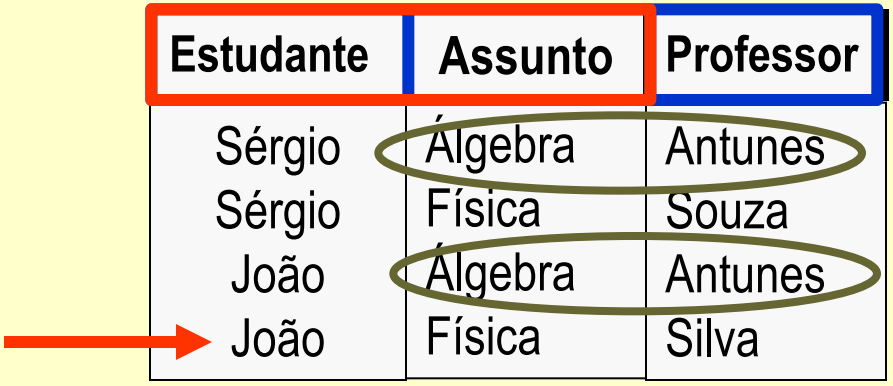
Forma Normal de Boyce/Cood (BCNF)

Diz-se que uma tabela está BCNF se e somente se os únicos determinantes são chaves candidatas.

- tabelas que tenham duas ou mais chaves candidatas e
- cada chave candidata (uma chave composta) apresente superposição em relação a algum atributo

Exemplo

- considerar a tabela Estudante com as seguintes restrições:
  - para cada assunto, cada estudante aprende esse assunto lecionado por um único professor;
  - cada professor leciona apenas um assunto (mas cada assunto pode ser lecionado por diversos professores)



Estudante	Assunto	Professor
Sérgio	Álgebra	Antunes
Sérgio	Física	Souza
João	Álgebra	Antunes
João	Física	Silva

está na 3FN

anomalias na 3NF:

eliminar que João estuda Física: elimina prof. Silva

redundância da informação

duas chaves candidatas superpostas:

- {Estudante, Assunto}
- {Estudante, Professor}
- Porém,

DF {Professor} → {Assunto}

DF {Assunto} $\not\rightarrow$ {Professor}

cada professor leciona apenas um assunto (mas cada assunto pode ser lecionado por diversos professor).

Estudante	Assunto	Professor
Sérgio	Álgebra	Antunes
Sérgio	Física	Souza
João	Álgebra	Antunes
João	Física	Silva

Projeções na BCNF:

E_P (estudante, professor)

P_A (professor, assunto)

Estudante	Professor
Sérgio	Antunes
Sérgio	Souza
João	Antunes
João	Silva

Assunto	Professor
Álgebra	Antunes
Física	Souza
Física	Silva

Problemas: as tabelas não são independentes! É preciso avaliar as duas tabelas para certas operações!

ex. inserir (Sergio, Silva)

deve ser rejeitada, pois Silva leciona Física e Sergio já estuda Física com Souza

projeto do BD: inclusão de procedimento para avaliação da inclusão em E_P

Quarta Forma Normal (4FN)

Diz-se que uma tabela está 4FN se e somente se estiver na 3FN e não apresentar dependências multivaloradas.

- dependência multivalorada: ocorre quando um dado valor da chave principal identifica repetidas vezes um conjunto de valores de atributos dependentes.

Exemplo

Curso	**Professor**	**Texto**
Física	Souza	Mecânica Básica
Física	Souza	Óptica
Física	Silva	Mecânica Básica
Física	Silva	Óptica
Matemática	Antunes	Matemática Básica
Matemática	Antunes	Análise Vetorial
Matemática	Antunes	Trigonometria

o curso pode ser ministrado por qualquer dos professores que podem usar qualquer dos livros definidos como referências

- Anomalias:

  para inserir que o curso de Física pode ser ministrado por um novo professor é preciso inserir duas linhas, uma para cada livro texto

professores e textos são independentes:

$$\text{Curso} \rightarrow\!\!\!\!\rightarrow \text{Professor}$$

$$\text{Curso} \rightarrow\!\!\!\!\rightarrow \text{Texto}$$

Estão na 4FN

Curso	Professor
Física	Souza
Física	Silva
Matemática	Antunes

Curso	Texto
Física	Mecânica Básica
Física	Óptica
Matemática	Matemática Básica
Matemática	Análise Vetorial
Matemática	Trigonometria