



# Engenharia Reversa de Arquivos e Normalização

Instituto Federal de Educação, Ciência e Tecnologia do Triângulo Mineiro  
Prof. Edwar Saliba Júnior



## Sistemas Legados

- Sistemas de informação antigos;
- geralmente construídos sem utilização de BD relacionais;
- dados armazenados em arquivos ou bancos de dados “pré-relacionais”, como **IMS** ou ADABAS;
- sistemas sem documentação;
- sistemas com alto nível de complexidade e de manutenção cara;
- exemplos de linguagens de programação da época:
  - COBOL,
  - NATURAL,
  - BASIC e
  - etc.



## Engenharia Reversa

- Sistemas antigos usavam:
  - arquivos binários e/ou
  - arquivo texto.
- Obter descrição dos arquivos do jeito que se encontram;
- gerar um modelo ER com as tabelas (não normalizadas) representando os arquivos;
- passar o diagrama gerado pelo processo de normalização, eliminando redundâncias;
- gerar modelo lógico e físico.



## Normalização

- Este processo baseia-se nos conceitos de formas normais.
- Uma forma normal é uma regra que deve ser obedecida por uma tabela para que esta seja considerada “bem projetada”;
- aprenderemos 4 formas normais, denominadas: 1FN, 2FN, 3FN e 4FN.



## 1FN

- Primeira Forma Normal.
- Diz-se que uma tabela está na 1FN quando ela não possui tabelas aninhadas.

## Dados do Arquivo

### Relatório de Alocação a Projeto

Código do Projeto: LSC001

Tipo: Novo Desenv.

Descrição: Sistema de Estoque

Código do Empregado	Nome	Categoria Funcional	Salário	Data Início no Projeto	Tempo Alcoado ao Projeto
2146	João	A1	4	01/11/91	24
3145	Sílvio	A2	4	02/10/91	24
6126	José	B1	9	03/10/92	18
1214	Carlos	A2	4	04/10/92	18
8191	Mário	A1	4	01/11/92	12

Código do Projeto: PAG02

Tipo: Manutenção

Descrição: Sistema de RH

Código do Empregado	Nome	Categoria Funcional	Salário	Data Início no Projeto	Tempo Alcoado ao Projeto
8191	Mário	A1	4	01/05/93	12
4112	João	A2	4	04/01/91	24
6126	José	B1	9	01/11/92	12

Fonte: HEUSER, Carlos Alberto. **Projeto de Banco de Dados**. 6. ed., Porto Alegre: Bookman, 2009, p. 187.



## Dados Tabulados

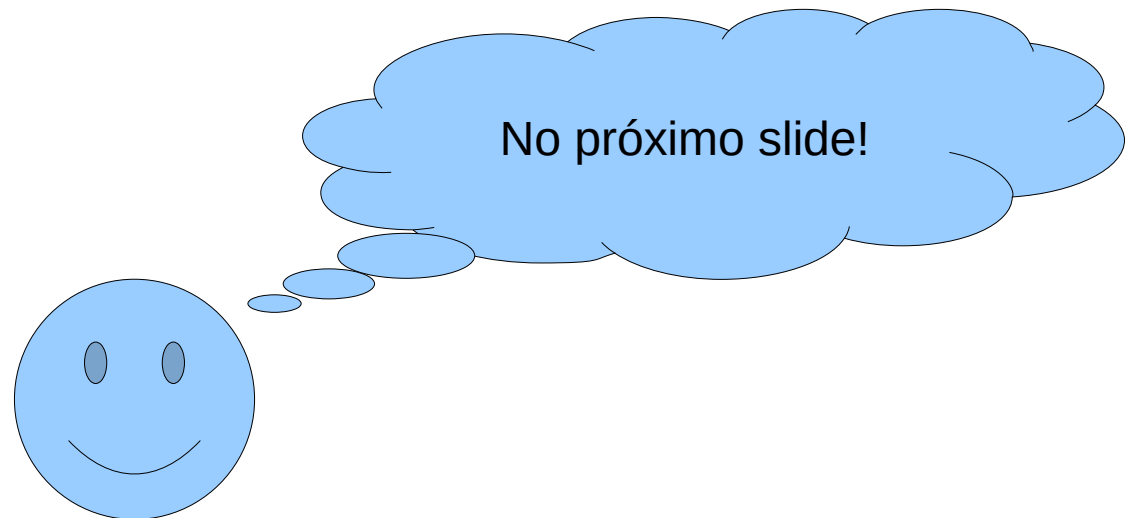
CódProj	Tipo	Descr	Emp					
			CodEmp	Nome	Cat	Sal	DataIni	TempAl
LSC001	Novo Desenv.	Sistema de Estoque	2146	João	A1	4	01/11/91	24
			3145	Sílvio	A2	4	02/10/91	24
			6126	José	B1	9	03/10/92	18
			1214	Carlos	A2	4	04/10/92	18
			8191	Mário	A1	4	01/11/92	12
PG02	Manutenção	Sistema de RH	8191	Mário	A1	4	01/05/93	12
			4112	João	A2	4	04/01/91	24
			6126	José	B1	9	01/11/92	12

Fonte: HEUSER, Carlos Alberto. **Projeto de Banco de Dados**. 6. ed., Porto Alegre: Bookman, 2009, p. 188.



## Exemplo

- Vamos tomar a tabela mostrada no *slide* anterior e colocá-la na 1FN;
- para isto, vamos dividi-la em duas tabelas: **Proj** e **ProjEmp**.





## Tabela na 1FN

**Proj:**

CódProj	Tipo	Descr
LSC001	Novo Desenv.	Sistema
PAG02	Manutenção	Sistema de RH

**ProjEmp:**

CódProj	CodEmp	Nome	Cat	Sal	DataIni	TempAl
LSC001	2146	João	A1	4	01/11/91	24
LSC001	3145	Sílvio	A2	4	02/10/91	24
LSC001	6126	José	B1	9	03/10/92	18
LSC001	1214	Carlos	A2	4	04/10/92	18
LSC001	8191	Mário	A1	4	01/11/92	12
PAG02	8191	Mário	A1	4	01/05/93	12
PAG02	4112	João	A2	4	04/01/91	24
PAG02	6126	José	B1	9	01/11/92	12

Fonte: HEUSER, Carlos Alberto. **Projeto de Banco de Dados**. 6. ed., Porto Alegre: Bookman, 2009, p. 192.



## Dependência Funcional

- Em uma tabela relacional, diz-se que uma coluna  $C_2$  depende funcionalmente de uma coluna  $C_1$  (ou que a coluna  $C_1$  determina a coluna  $C_2$ ) quando, em todas as linhas da tabela, para cada valor de  $C_1$  que aparece na tabela, aparecer o mesmo valor de  $C_2$ .



O resumo disto, no próximo *slide*.

## Parte da Tabela Com Dependências Funcionais

Repare que há uma correlação entre os valores destas duas colunas da tabela

...	Código	...	Salário	...
	E1	→	10	
	E3	→	10	
	E1	→	10	
	E2	→	5	
	E3	→	10	
	E2	→	5	
	E1	→	10	

Fonte: HEUSER, Carlos Alberto. **Projeto de Banco de Dados**. 6. ed., Porto Alegre: Bookman, 2009, p. 196.



## 2FN

- Uma tabela encontra-se na segunda forma normal quando, além de estar na 1FN, não contém ***dependências parciais***;
- uma ***dependência (funcional) parcial*** ocorre quando uma coluna depende de parte de uma chave primária composta.
- Ou seja, para estar na 2FN cada atributo não-chave deve ser dependente da chave primária inteira, isto é, cada atributo não-chave não poderá ser dependente de apenas parte da chave.

## Dependência Parcial

Proj:

CódProj	Tipo	Descr
LSC001	Novo Desenv.	Sistema
PAG02	Manutenção	Sistema de RH

ProjEmp:

CódProj	CodEmp	Nome	Cat	Sal	DataIni	TempAl
LSC001	2146	João	A1	4	01/11/91	24
LSC001	3145	Sílvio	A2	4	02/10/91	24
LSC001	6126	José	B1	9	03/10/92	18
LSC001	1214	Carlos	A2	4	04/10/92	18
LSC001	8191	Mário	A1	4	01/11/92	12
PAG02	8191	Mário	A1	4	01/05/93	12
PAG02	4112	João	A2	4	04/01/91	24
PAG02	6126	José	B1	9	01/11/92	12

Nestas 3 colunas os dados estão redundantes para os empregados que trabalham em mais de um projeto.

Fonte: HEUSER, Carlos Alberto. **Projeto de Banco de Dados**. 6. ed., Porto Alegre: Bookman, 2009, p. 192.

## Eliminando as Dependências Parciais

- Observe a tabela **ProjEmp** nos próximos *slides*.



FREEPIK. **Homem com lupa**. Disponível em: <[https://br.freepik.com/fotos-premium/homem-com-lupa\\_5443460.htm](https://br.freepik.com/fotos-premium/homem-com-lupa_5443460.htm)>. Acesso em: 01 fev. 2024.

Formato	Tabela (CPM)	Determinação	Expansível	Diária	Dimensões**
01 - Super Leader Board	R\$ 30,00	+ 20%	+ 30%	R\$ 459,00	970x90
02 - Retângulo Superior	R\$ 16,00	+ 20%	+ 30%	R\$ 459,00	300x250
03 - Retângulo Inferior	R\$ 13,00	+ 20%	+ 30%	R\$ 369,00	300x250
04 - Retângulo Inferior	R\$ 13,00	+ 20%	+ 30%	R\$ 369,00	300x250
05 - Shopping	R\$ 100,00*	+ 20%			
06 - Flutuante	R\$ 26,00	+ 20%		R\$ 828,00	300x250
07 - Quadrado Notícia	R\$ 30,00	+ 20%		R\$ 459,00	200x200

\* Shopping - Valor mensal por produto    \*\* Dimensões em pixel

## Dividindo a Tabela “ProjEmp”

Proj:

CódProj	Tipo	Descr
LSC001	Novo Desenv.	Sistema
PAG02	Manutenção	Sistema de RH

Podemos dividir a tabela **ProjEmp** em duas. Sendo que com estas 4 colunas nós criaremos uma nova tabela que chamaremos de **Emp** (Empregado).

ProjEmp:

CódProj	CodEmp	Nome	Cat	Sal	DataIni	TempAl
LSC001	2146	João	A1	4	01/11/91	24
LSC001	3145	Sílvio	A2	4	02/10/91	24
LSC001	6126	José	B1	9	03/10/92	18
LSC001	1214	Carlos	A2	4	04/10/92	18
LSC001	8191	Mário	A1	4	01/11/92	12
PAG02	8191	Mário	A1	4	01/05/93	12
PAG02	4112	João	A2	4	04/01/91	24
PAG02	6126	José	B1	9	01/11/92	12

Fonte: HEUSER, Carlos Alberto. **Projeto de Banco de Dados**. 6. ed., Porto Alegre: Bookman, 2009, p. 192.

## Dividindo a Tabela ProjEmp

Proj:

CódProj	Tipo	Descr
LSC001	Novo Desenv.	Sistema
PAG02	Manutenção	Sistema de RH

E somente estas 4 colunas comporão a nova tabela **ProjEmp** (Projeto x Empregado).

ProjEmp:

CódProj	CodEmp	Nome	Cat	Sal	DataIni	TempAl
LSC001	2146	João	A1	4	01/11/91	24
LSC001	3145	Sílvio	A2	4	02/10/91	24
LSC001	6126	José	B1	9	03/10/92	18
LSC001	1214	Carlos	A2	4	04/10/92	18
LSC001	8191	Mário	A1	4	01/11/92	12
PAG02	8191	Mário	A1	4	01/05/93	12
PAG02	4112	João	A2	4	04/01/91	24
PAG02	6126	José	B1	9	01/11/92	12

Fonte: HEUSER, Carlos Alberto. **Projeto de Banco de Dados**. 6. ed., Porto Alegre: Bookman, 2009, p. 192.



## Após a Divisão da Tabela ProjEmp Temos em 2FN:

Fonte: HEUSER, Carlos Alberto. **Projeto de Banco de Dados**. 6. ed., Porto Alegre: Bookman, 2009, p. 199.

### Proj:

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

### ProjEmp:

CódProj	CodEmp	DataIni	TempAl
LSC001	2146	01/11/91	24
LSC001	3145	02/10/91	24
LSC001	6126	03/10/92	18
LSC001	1214	04/10/92	18
LSC001	8191	01/11/92	12
PAG02	8191	01/05/93	12
PAG02	4112	04/01/91	24
PAG02	6126	01/11/92	12

### Emp:

CodEmp	Nome	Cat	Sal
2146	João	A1	4
3145	Sílvio	A2	4
6126	José	B1	9
1214	Carlos	A2	4
8191	Mário	A1	4
4112	João	A2	4



## 3FN

- Uma tabela encontra-se na terceira forma normal quando, além de estar na 2FN, não contém dependências transitivas;
- uma dependência transitiva ocorre quando uma coluna, além de depender da chave primária da tabela, depende de outra coluna ou conjunto de colunas da tabela.

## Dependência Transitiva

- Exemplo:

Itens do pedido				
Pedido	Item	Preço	Quantidade	Total
15	102	9,25	2	18,5
15	132	1,3	5	6,5

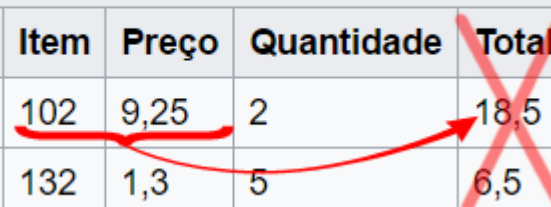


Tabela **com** Dependência Transitiva

Itens do pedido			
Pedido	Item	Preço	Quantidade
15	102	9,25	2
15	132	1,3	5

Tabela **sem** Dependência Transitiva

Fonte: NORMALIZAÇÃO DE DADOS. In: WIKIPEDIA: a enciclopédia livre. Disponível em: <[https://pt.wikipedia.org/wiki/Normaliza%C3%A7%C3%A3o\\_de\\_dados](https://pt.wikipedia.org/wiki/Normaliza%C3%A7%C3%A3o_de_dados)>. Acesso em 23 nov. 2018.

## Dependência Transitiva

Observe que o campo "Sal" (Salário) tem uma relação direta com o campo "Cat" (Categoria).

**Emp:**

CodEmp	Nome	Cat	Sal
2146	João	A1	4
3145	Sílvio	A2	4
6126	José	B1	9
1214	Carlos	A2	4
8191	Mário	A1	4
4112	João	A2	4

## Eliminando as Dependências Transitivas

- Observe a tabela **Emp** nos próximos *slides*.



FREEPIK. **Homem com lupa**. Disponível em: [https://br.freepik.com/fotos-premium/homem-com-lupa\\_5443460.htm](https://br.freepik.com/fotos-premium/homem-com-lupa_5443460.htm). Acesso em: 01 fev. 2024.

Formato	Tabela (CPM)	Determinação	Expansível	Diária	Dimensões**
01 - Super Leader Board	R\$ 30,00	+ 20%	+ 30%	R\$ 459,00	970x90
02 - Retângulo Superior	R\$ 16,00	+ 20%	+ 30%	R\$ 459,00	300x250
03 - Retângulo Inferior	R\$ 13,00	+ 20%	+ 30%	R\$ 369,00	300x250
04 - Retângulo Inferior	R\$ 13,00	+ 20%	+ 30%	R\$ 369,00	300x250
05 - Shopping	R\$ 100,00*	+ 20%			
06 - Flutuante	R\$ 26,00	+ 20%		R\$ 828,00	300x250
07 - Quadrado Notícia	R\$ 30,00	+ 20%		R\$ 459,00	200x200

\* Shopping - Valor mensal por produto    \*\* Dimensões em pixel



## Dividindo a Tabela “Emp”

Podemos dividir a tabela **Emp** em duas. Sendo que com estas 2 colunas nós criaremos uma nova tabela que chamaremos de **Cat** (Categoria).

**Emp:**

CodEmp	Nome	Cat	Sal
2146	João	A1	4
3145	Sílvio	A2	4
6126	José	B1	9
1214	Carlos	A2	4
8191	Mário	A1	4
4112	João	A2	4



## Dividindo a Tabela “Emp”

**Emp:**

E somente estas 3 colunas comporão a nova tabela **Emp** (Empregado).

CodEmp	Nome	Cat	Sal
2146	João	A1	4
3145	Sílvio	A2	4
6126	José	B1	9
1214	Carlos	A2	4
8191	Mário	A1	4
4112	João	A2	4

Proj:

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

ProjEmp:

CódProj	CodEmp	DataIni	TempAl
LSC001	2146	01/11/91	24
LSC001	3145	02/10/91	24
LSC001	6126	03/10/92	18
LSC001	1214	04/10/92	18
LSC001	8191	01/11/92	12
PAG02	8191	01/05/93	12
PAG02	4112	04/01/91	24
PAG02	6126	01/11/92	12

Emp:

CodEmp	Nome	Cat
2146	João	A1
3145	Sílvio	A2
6126	José	B1
1214	Carlos	A2
8191	Mário	A1
4112	João	A2

Cat:

Cat	Sal
A1	4
A2	4
B1	9

**Após a  
Divisão da  
Tabela  
“Emp”  
Temos em  
3FN:**





## Forma Normal Boyce-Codd

- Alguns autores citam a Forma Normal de Boyce-Codd (FNBC) em seus trabalhos;
- a FNBC postula o seguinte:
  - por definição uma relação está na BCNF se todos os atributos são funcionalmente dependentes da chave, de toda a chave e nada mais do que a chave.
- A FNBC foi proposta como uma forma mais simples da 3FN, [...], cada relação em BCNF também está na 3FN.(ELMASRI, NAVATHE, 2011, P. 355)

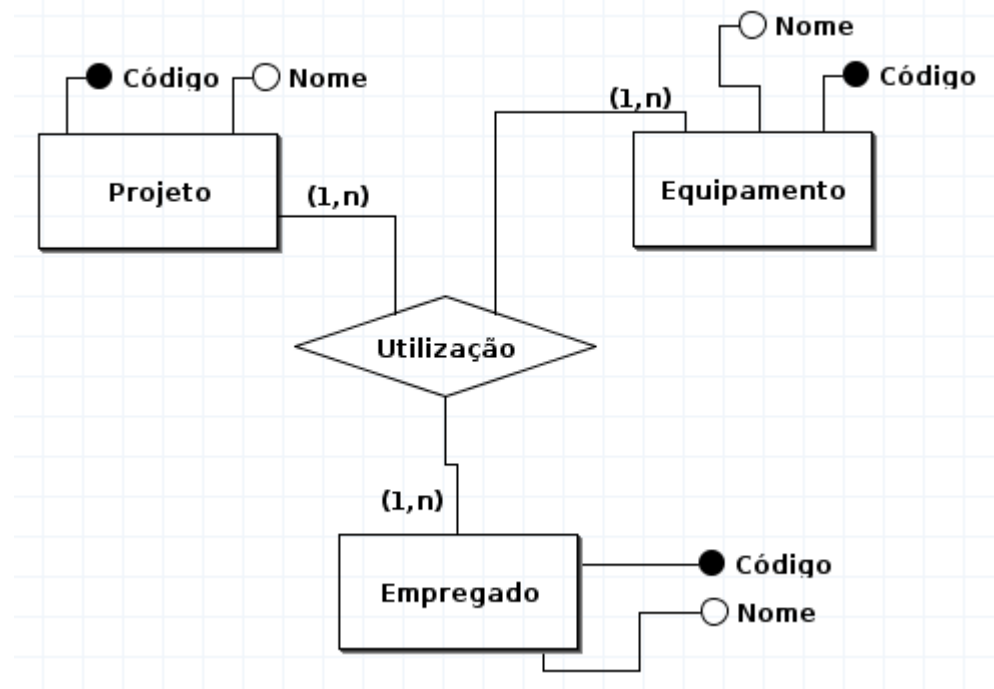


## 4FN

- Uma tabela está na quarta forma normal quando, além de estar na 3FN, não contém dependências multivaloradas.

## Sistema

- Analise o diagrama e dê uma descrição simplificada de sua funcionalidade.
- Ou seja, o que exatamente o relacionamento “Utilização” nos diz?





## Regra de Negócio

- No *slide* anterior o diagrama quer dizer o seguinte:
  - o relacionamento “Utilização” indica que deseja-se manter a informação de qual “empregado” usa qual “equipamento” em qual “projeto”.

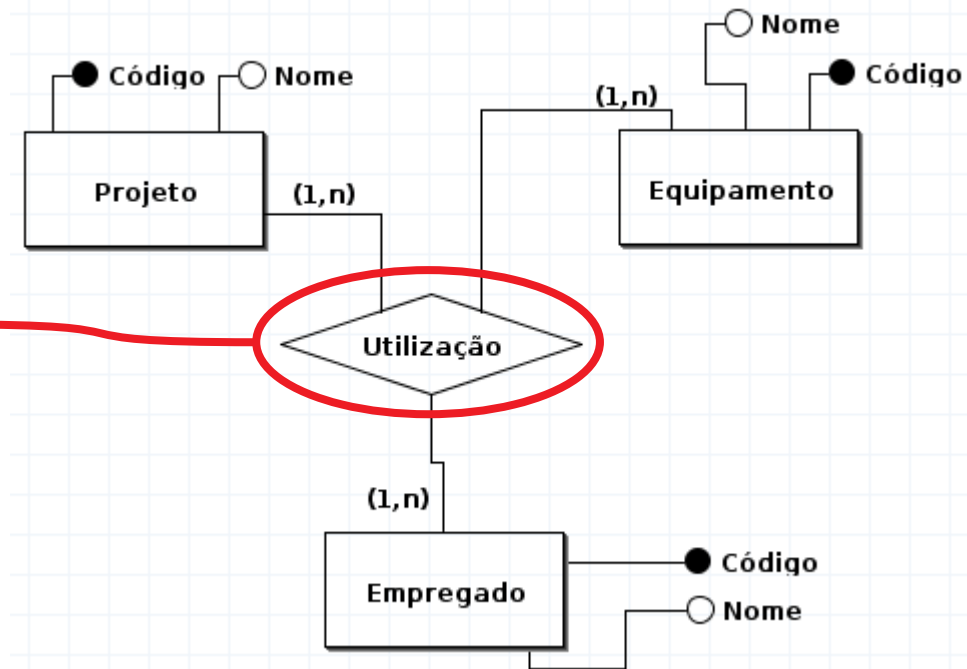


# Relacionamento “Utilização”

- Por ser um relacionamento N para N, então, este relacionamento se transformará em uma tabela que chamaremos de “Utili”;
- a tabela “Utili” terá os seguintes campos:
  - CodProj (Código do Projeto),
  - CodEmp (Código do Empregado) e
  - CodEquip (Código do Equipamento).

## Exemplo de Tuplas na Tabela “Utili”

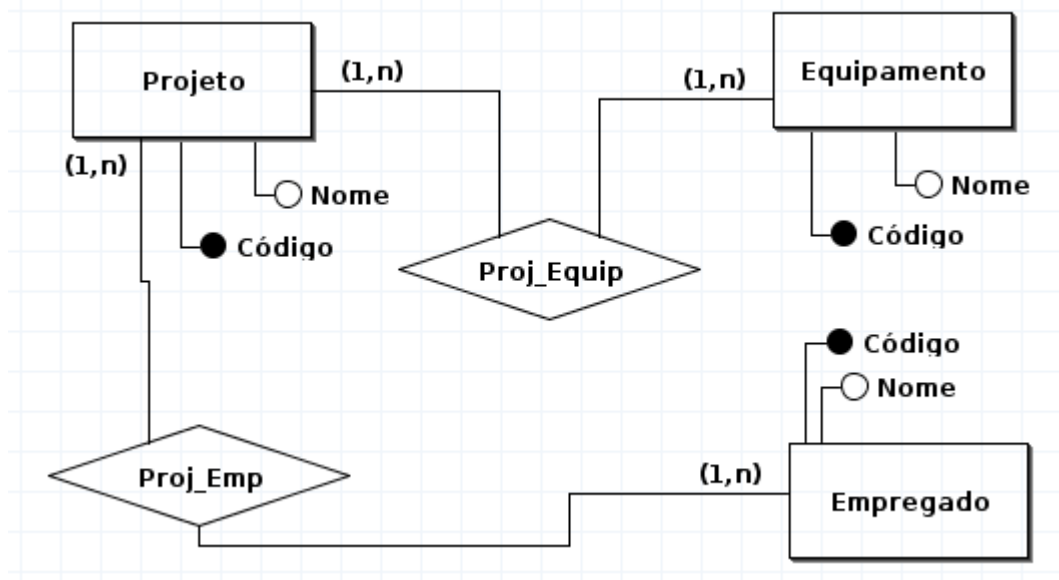
CodProj	CodEmp	CodEquip
1	1	1
1	2	1
1	3	1
1	1	2
1	2	2
1	3	2
2	2	2
2	2	4
3	3	1
3	4	1
3	3	3
3	4	3
3	3	5
3	4	5
4	2	5



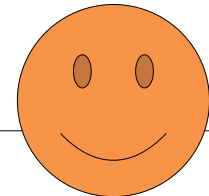
Fonte: HEUSER, Carlos Alberto. **Projeto de Banco de Dados**. 6. ed., Porto Alegre: Bookman, 2009, p. 205.

## Mudança da Regra de Negócio

- A regra de negócio do sistema mudou. O sistema agora contempla o diagrama abaixo.



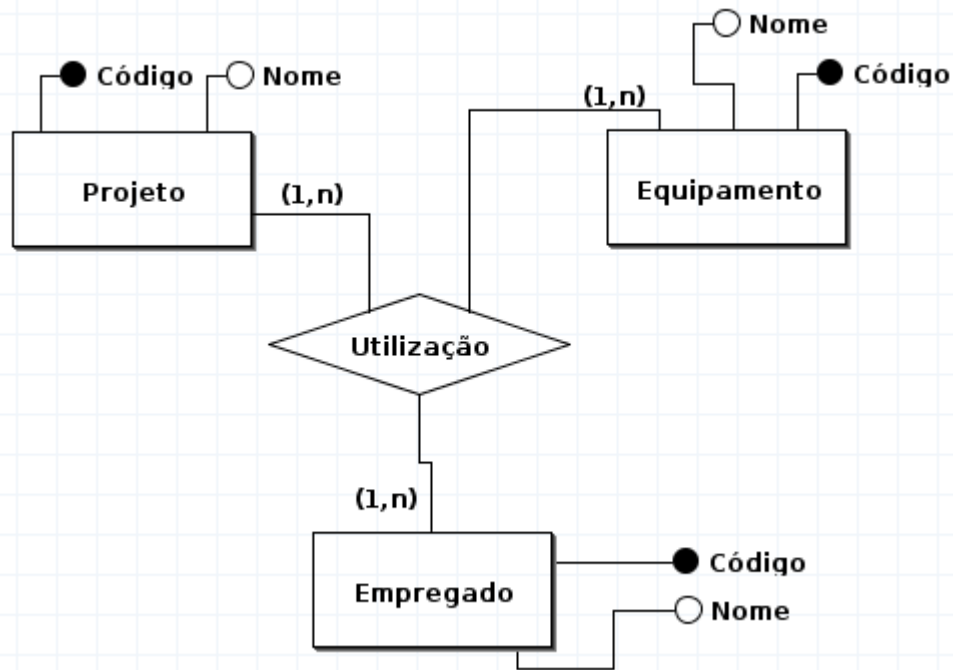
Vamos comparar o diagrama anterior com o diagrama atual. No próximo *slide*!



## O que mudou na Regra de Negócio ?

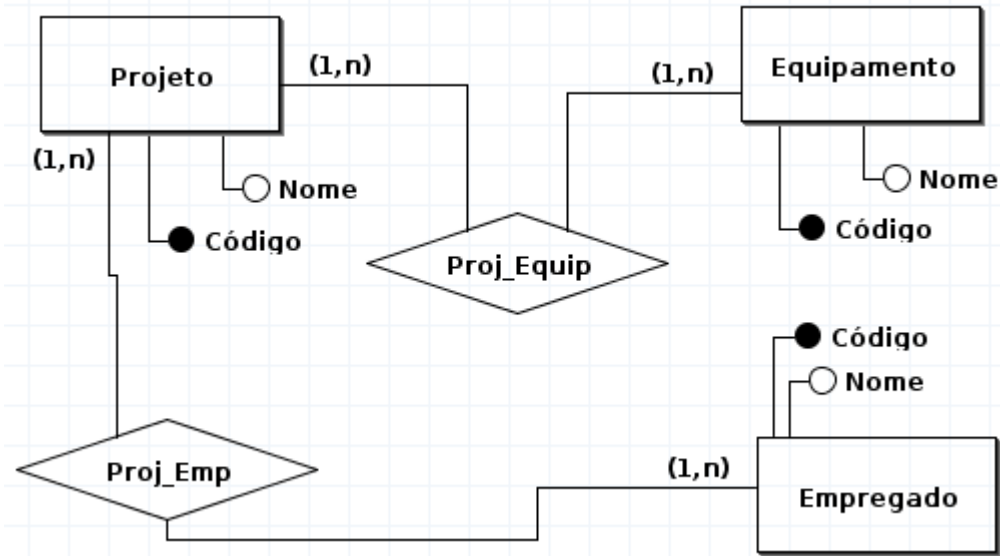
- Diagrama conceitual do sistema:

\* antes \*



e

\* depois \*







## Suposição

- A regra de negócio mudou, como visto nos diagramas conceituais (*antes e depois*) no *slide* anterior;
- porém, suponha que o sistema ainda não tenha sido atualizado e a tabela “Utili” ainda continua da mesma forma.

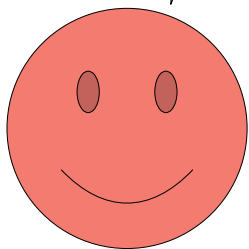


## 4FN

- A regra nos diz que:
  - Uma tabela encontra-se na quarta forma normal quando, além de estar na 3FN, não contém dependências multivaloradas;
- perceba no próximo *slide*, que na tabela “Utili” existem repetições de dependências multivaloradas. Isto ocorre porque a regra de negócio do sistema mudou e o banco de dados não foi refeito.

## Repetições de Dependências Multivaloradas

Para corrigir esta situação, esta tabela terá que ser dividida.



CodProj	CodEmp	CodEquip
1	1	1
1	2	1
1	3	1
1	1	2
1	2	2
1	3	2
2	2	2
2	2	4
3	3	1
3	4	1
3	3	3
3	4	3
3	3	5
3	4	5
4	2	5

Fonte: HEUSER, Carlos Alberto. **Projeto de Banco de Dados**. 6. ed., Porto Alegre: Bookman, 2009, p. 205.

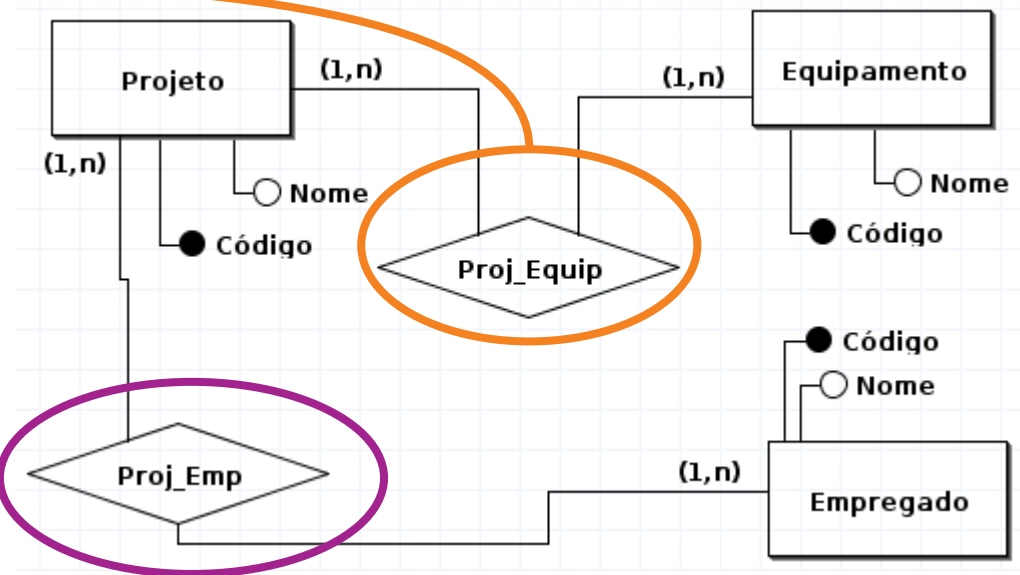
## Eliminando a Repetição

ProjEquip:

CodProj	CodEquip
1	1
1	2
2	2
2	4
3	1
3	3
3	5
4	5

ProjEmp:

CodProj	CodEmp
1	1
1	2
1	3
2	2
3	3
3	4
4	2





## 5FN

- Uma relação estará na 5FN se não puder ser mais decomposta sem perda de informação.
- Se puder ser reconstruída sem perda de informação a partir de algumas das suas projeções, então existe uma Dependência de Junção (DJ) e, portanto, a relação não se encontra na 5FN, devendo ser decomposta segundo essa dependência.

(CARRIÇO, José António; CARRIÇO, António João, 2004, Pág. 50).



## 5FN

- DEVMEDIA - SQL Magazine 7
- Exemplo



## Desnormalização

- A normalização de um Banco de Dados significa não ter redundância de dados entre suas tabelas;
- redundância de dados é indesejável em bancos de dados relacionais por N motivos:
  - ocupação de espaço físico desnecessário;
  - atualização, quando realizada, deve ser feita em mais de um lugar;
  - risco de se ter inconsistência no processamento ou apresentação dos dados e
  - etc.

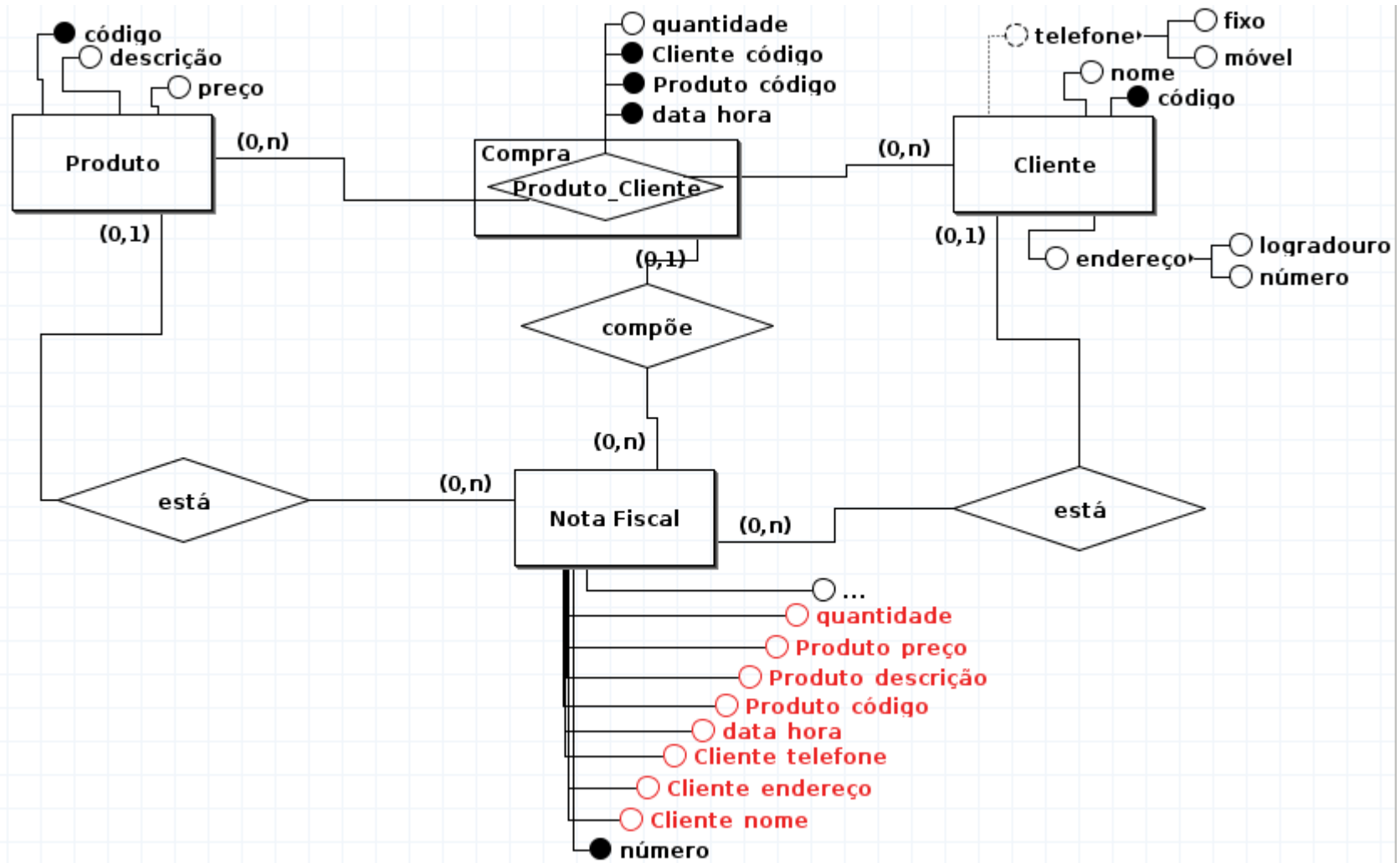


## Porque Desnormalizar?

- Existem situações onde a desnormalização se faz necessária. Um ótimo exemplo disto é a emissão de uma Nota Fiscal (NF) eletrônica.
- Os dados que são emitidos numa NF não podem ser alterados com o tempo. Ou seja:
  - o nome, endereço e dados pessoais do cliente devem ser sempre os mesmos da data de emissão da NF,
  - o código, a descrição, a quantidade comprada e o preço dos produtos/serviços adquiridos pelo cliente, também devem ser sempre os mesmos da data de emissão da NF e
  - também os impostos calculados, dados da empresa e tudo mais que há numa NF, devem permanecer “eternamente” os mesmos da data de emissão da NF.
- A resposta para se conseguir tudo isto é: **desnormalizar!**
- Vejamos o diagrama conceitual no próximo slide.



## Nota Fiscal





## Exercícios

- HEUSER, Carlos Alberto. **Projeto de Banco de Dados**. 6. ed., Porto Alegre: Bookman, 2009, p. 215.



## Bibliografia

- DBMS | INTRODUCTION OF 4th AND 5th NORMAL FORM. In: GeeksforGeeks: a computer science portal for geeks. Disponível em: <<https://www.geeksforgeeks.org/dbms-introduction-of-4th-and-5th-normal-form/>>. Acesso em: 23 Nov. 2018.
- ELMASRI, R. NAVATHE, S. B. **Sistema de Banco de Dados**. 6. ed. São Paulo: Pearson, 2011.
- SQL MAGAZINE 7: formas normais superiores. In: Devmedia. Disponível em: <<https://www.devmedia.com.br/artigo-sql-magazine-7-formas-normais-superiores/7474>>. Acesso em: 25 Nov. 2018.
- HEUSER, Carlos Alberto. **Projeto de Banco de Dados**. 6. ed., Porto Alegre: Bookman, 2009.
- INFORMATION MANAGEMENT SYSTEM. In: WIKIPEDIA: a enciclopédia livre. Disponível em: <[https://pt.wikipedia.org/wiki/Information\\_Management\\_System](https://pt.wikipedia.org/wiki/Information_Management_System)>. Acesso em: 15 Out. 2018.
- IMS DB MANIPULAÇÃO DE DADOS. In: Tutorialspoint: simply easy learning. Disponível em: <[https://www.tutorialspoint.com/pg/ims\\_db/ims\\_db\\_data\\_manipulation.htm](https://www.tutorialspoint.com/pg/ims_db/ims_db_data_manipulation.htm)>. Acesso em: 15 Out. 2018.
- MACÊDO, D. **Normalização de um Banco de Dados**. Disponível em: <<https://www.diegomacedo.com.br/normalizacao-de-um-banco-de-dados/>>. Acesso em: 14 Nov. 2018.
- NORMALIZAÇÃO DE DADOS. In: WIKIPEDIA: a enciclopédia livre. Disponível em: <[https://pt.wikipedia.org/wiki/Normaliza%C3%A7%C3%A3o\\_de\\_dados](https://pt.wikipedia.org/wiki/Normaliza%C3%A7%C3%A3o_de_dados)>. Acesso em 23 Nov. 2018.



## IMS DB

- O **Information Management System (IMS)** da IBM é um sistema de **banco de dados hierárquico** e de gerenciamento de informações com extensos recursos de processamento de transações. (Fonte: <[https://pt.wikipedia.org/wiki/Information\\_Management\\_System](https://pt.wikipedia.org/wiki/Information_Management_System)>. Acesso em: 15 Out. 2018).
- Tutorial sobre IMS: [download](#).

