

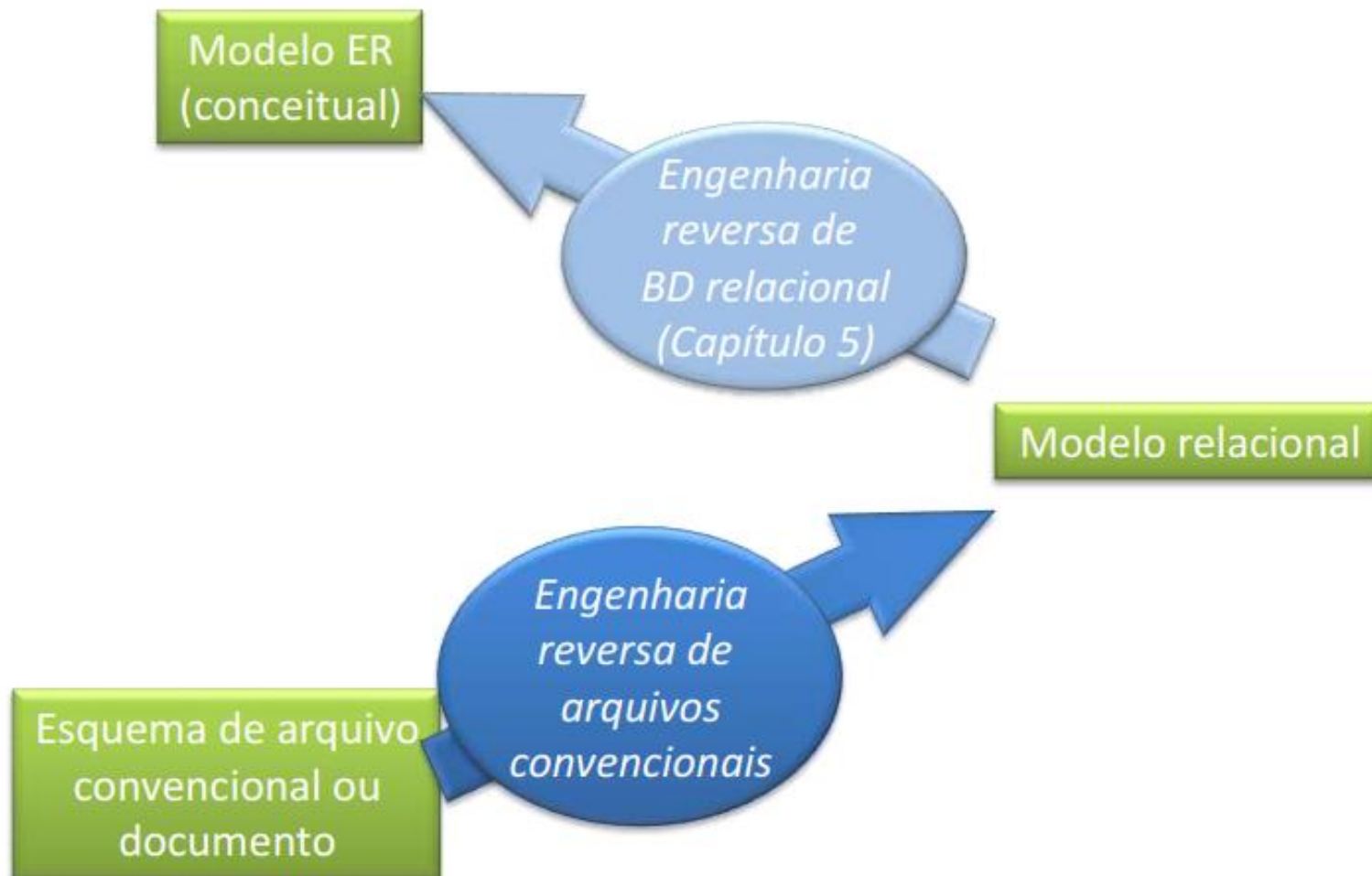
Banco de Dados I

Prof. Diego Buchinger
diego.buchinger@outlook.com
diego.buchinger@udesc.br

Profa. Rebeca Schroeder Freitas
Prof. Fabiano Baldo

Normalização

Engenharia Reversa



- Entrada do processo:
 - ❖ Documentos que disponham de dados ou descrição destes:
 - Documentos
 - Arquivos manuais
 - Banco de dados não gerenciados por SGBDs
 - ...
 - ❖ Sistemas legados:
 - Necessidade de modelo ER para manutenção, migração ou integração (dificuldade: raramente documentados)

Engenharia Reversa

- Normalização:

“Processo que transforma um esquema de dados, com base em regras, em um modelo relacional/tabular bem projeto.”

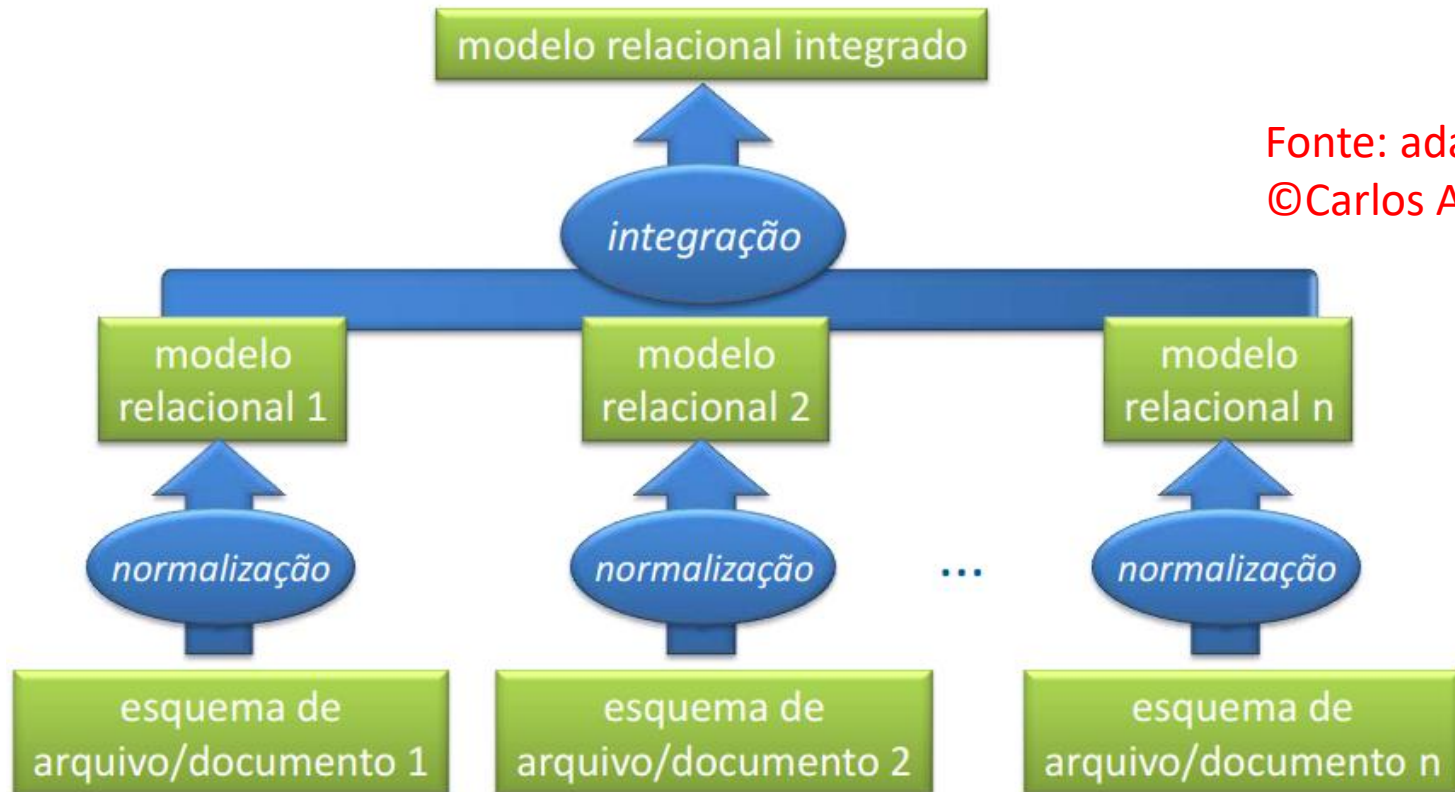
- Objetivos:

- Reagrupar informações de maneira conveniente
- Eliminar/reduzir **redundância** de dados



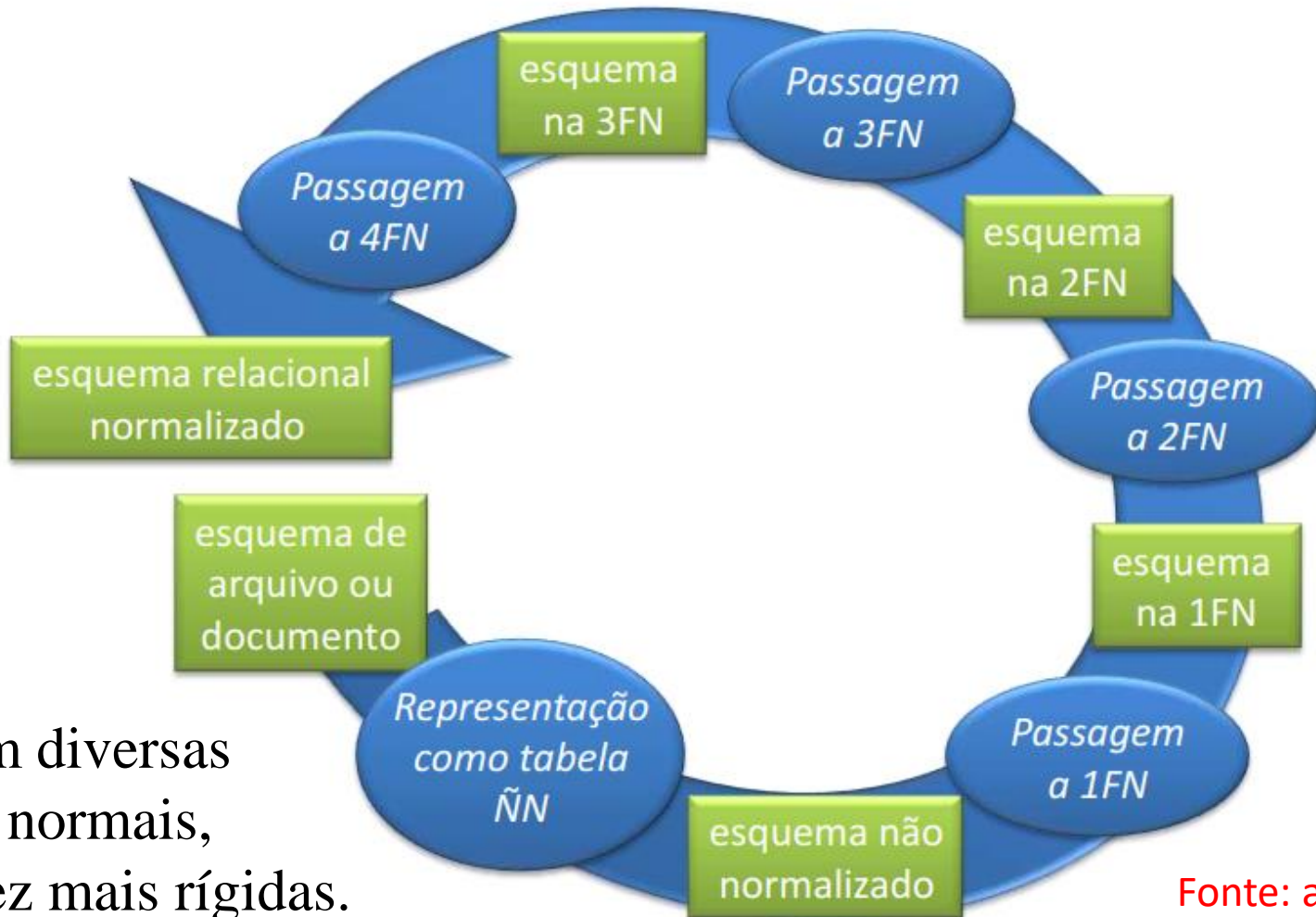
Engenharia Reversa

- Deve-se normalizar todos os esquemas de documentos disponíveis e ao final integrá-los.



Fonte: adaptado de
©Carlos A. Heuser

Passos da Normalização



Existem diversas formas normais, cada vez mais rígidas.

Exemplo - Documento

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

Exemplo - FNN

- **Tabela não-normalizada:** possui uma ou mais tabelas aninhadas – geralmente associadas a colunas ou campo multivalorados.
 - Ações:
 - Manter tabelas aninhadas na representação
 - Criar tabelas aninhadas para valores multivalorados
 - Identificar e representar as chaves primárias (geralmente códigos ou siglas)

Exemplo - FNN

- Os projetos fazem parte da mesma tabela (possuem os mesmos campos!)

Fonte: adaptado de
©Carlos A. Heuser

CódProj	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	Manutenção	Sistema de Pagamento	1091	Mário	A1	4	1/05/93	12
				João	A2	4	4/01/91	24
				José	B1	9	1/11/92	12

tabela
aninhada

Exemplo - FNN

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

CódProj	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	Manutenção	Sistema de Pagamento	1091	Mário	A1	4	1/05/93	12
				João	A2	4	4/01/91	24
				José	B1	9	1/11/92	12

tabela
aninhada

Exemplo - 1FN

- **Primeira Forma Normal:** dizemos que um esquema está na primeira forma normal quando não existem tabelas aninhadas.
- Ações:
 - Opção 1: construir uma única tabela com redundância
 - Opção 2: construir uma tabela para cada tabela aninhada, adicionar chaves primárias da tabela não aninhada e repensar a chave primária nas novas tabelas

Exemplo - 1FN

- **Primeira Forma Normal – Opção 1**
 - Garante que nenhuma relação intertabelas se perca
 - Dados do projeto serão repetidos para cada empregado do projeto (redundância)

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



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

Exemplo - 1FN

- **Primeira Forma Normal – Opção 2**
 - Não gera redundância
 - Pode-se perder algum vínculo entre tabelas

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



Proj (#CodProj, Tipo, Descr)

ProjEmp (#&CodProj, #CodEmp, Nome, Cat, Sal, DataIni, TempAl)

Exemplo - 1FN

- **Primeira Forma Normal – Opção 2**

```
Proj ( #CodProj, Tipo, Descr )  
ProjEmp ( #&CodProj, #CodEmp, Nome, Cat, Sal, DataIni, TempAl )
```

Um valor CodEmp pode aparecer apenas uma vez na tabela ou pode aparecer várias vezes (associado a n projetos)?

- Se aparece apenas uma vez, apenas CodEmp precisa ser chave primária
- Se pode aparecer mais de uma vez, então tanto CodEmp quanto CodProj devem ser chaves primárias

(neste caso o relatório apresenta a resposta)

Exemplo - 1FN

Proj:

<u>CódProj</u>	Tipo	Descr
LSC001	Novo Desenv.	Sistema de Estoque
PAG02	Manutenção	Sistema de RH

ProjEmp:

<u>CódProj</u>	<u>CodEmp</u>	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

Exercício

Matrícula: 21043

Nome: Mathias Alonso Neto

Endereço: R. Moçambique, n. 415, Bananal

Telefone: 99963-1100

Curso: (01) – Ciência da Computação

Ingresso: 2017/1

Disciplina	Semestre	Frequência	Nota	Exame
(0203) LPG – Linguagem de Programação	2017/1	98%	8,6	-
(0204) ALG – Álgebra Linear	2017/1	72%	4,2	3,1
(0216) – ANN – Análise Numérica	2017/1	88%	7,2	-
(0204) ALG – Álgebra Linear	2017/2	91%	6,2	6,3
(0211) POO – Programação Orientada a Objetos	2017/2	93%	8,4	-

Curso: (05) – Engenharia Elétrica

Ingresso: 2016/2

Disciplina	Semestre	Frequência	Nota	Exame
(0052) ALP – Algoritmos e Lógica de Prog.	2016/2	99%	9,9	-
(0162) CDI – Cálculo Diferencial e Integral	2016/2	79%	6,8	4,3

Continua...

Exercício

Matrícula: 21199 **Nome:** Mayara Kendra Blokowsky
Endereço: R. Armação, n. 91, Judiá **Telefone:** 98352-4122

Curso: (06) – Engenharia Civil

Ingresso: 2017/1

Disciplina	Semestre	Frequência	Nota	Exame
(0052) ALP – Algoritmos e Lógica de Prog.	2017/1	82%	7,9	-
(0252) CDI – Cálculo Diferencial e Integral	2017/1	88%	8,1	-
(0253) DBA – Desenho Básico	2017/1	89%	8,6	-
(0254) IEC – Introdução à Engenharia Civil	2017/1	92%	7,5	-
(0255) QEC – Química para Eng. Civil	2017/1	81%	8,3	-
(0262) ALI – Álgebra Linear	2017/2	86%	8,7	-
(0263) DAT – Des. Arquitetônico e Topográfico	2017/2	90%	8,9	-
(0264) FEX – Física Experimental I	2017/2	92%	9,5	-
(0265) FGE – Física Geral I	2017/2	92%	9,0	--

Exemplo - 2FN

- **Segunda Forma Normal** : dizemos que um esquema está na segunda forma normal quanto, além de estar na 1FN, não há dependências funcionais parciais.
 - O que é uma dependência funcional?
 - Dizemos que uma coluna C_2 depende funcionalmente de uma coluna C_1 quando o valor de C_1 determina o valor de C_2 – ou seja, para cada valor de C_1 , aparece um mesmo valor em C_2 .

Exemplo - 2FN

Fonte: adaptado de
©Carlos A. Heuser

- Exemplos dependência funcional

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

Código → Salário

~~A → B~~

A → D

(A,B) → C

A	B	C	D
B	5	2	20
C	4	2	15
B	6	7	20
B	5	2	20
C	2	2	15
C	4	2	15
A	10	5	18
A	12	3	18
A	10	5	18
B	5	2	20
C	4	2	15
A	10	5	18
C	4	2	15

As vezes ter apenas os valores da tabela não é suficiente para descobrir relação de dependência

Exemplo - 2FN

- O que é uma dependência funcional parcial?
 - É uma dependência funcional, na qual um valor depende apenas de parte da chave primária
 - Só existe quando há uma chave primária composta!
- O que é uma dependência funcional total?
 - É uma dependência funcional, na qual um valor depende de toda a chave primária composta
- Ações:
 - Criar uma nova tabela para os dados redundantes que apresentam dependência funcional parcial com a(s) chave(s) primária(s) que os determinam

Exemplo - 2FN

Proj:

<u>CódProj</u>	Tipo	Descr
LSC001	Novo Desenv.	Sistema de Estoque
PAG02	Manutenção	Sistema de RH

Fonte: adaptado de ©Carlos A. Heuser

ProjEmp:

<u>CódProj</u>	<u>CodEmp</u>	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

Existe chave primária composta na primeira ou segunda tabela?

Existe alguma coluna que tem dependência funcional parcial?

Exemplo - 2FN

```
Proj ( #CodProj, Tipo, Descr )  
ProjEmp ( #&CodProj, #CodEmp, Nome, Cat, Sal, DataIni, TempAl )
```

Pergunta a ser feita:
“O valor da coluna
depende de toda a chave
ou só de parte dela?”

Dependência Funcional Parcial

Dependência Funcional Total

```
ProjEmp ( #&CodProj, #CodEmp, Nome, Cat, Sal, DataIni, TempAl )  
Proj ( #CodProj, Tipo, Descr )
```

Exemplo - 2FN

- **Segunda Forma Normal**
 - Elimina a redundância de dados com dependência funcional parcial

```
Proj ( #CodProj, Tipo, Descr )  
ProjEmp ( #&CodProj, #<u>CodEmp</u>, Nome, Cat, Sal, DataIni, TempAl )
```



```
Proj ( #CodProj, Tipo, Descr )  
ProjEmp ( #&CodProj, #<u>CodEmp</u>, DataIni, TempAl )  
Empregado ( #<u>CodEmp</u>, Nome, Cat, Sal )
```


Exemplo - 2FN

Proj:

<u>CódProj</u>	Tipo	Descr
LSC001	Novo Desenv.	Sistema de Estoque
PAG02	Manutenção	Sistema de RH

Fonte: adaptado de
©Carlos A. Heuser

ProjEmp:

<u>CódProj</u>	<u>CodEmp</u>	DataIni	TempAl
LSC001	2146	1/11/91	24
LSC001	3145	2/10/91	24
LSC001	6126	3/10/92	18
LSC001	1214	4/10/92	18
LSC001	8191	1/11/92	12
PAG02	8191	1/05/93	12
PAG02	4112	4/01/91	24
PAG02	6126	1/11/92	12

Emp:

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

Exemplo - 3FN

- **Terceira Forma Normal:** dizemos que um esquema está na terceira forma normal quando não existem dependências funcionais transitivas
 - O que é dependência funcional transitiva:
 - É quando o valor de uma coluna depende de uma coluna que não compõem a chave primária, dependendo indiretamente da chave primária do registro.
 - Ações:
 - Criar uma nova tabela para os dados redundantes que apresentam dependência funcional transitiva

Exemplo - 3FN

Proj:

<u>CódProj</u>	Tipo	Descr
LSC001	Novo Desenv.	Sistema de Estoque
PAG02	Manutenção	Sistema de RH

Existe alguma dependência funcional transitiva em alguma destas tabelas?

ProjEmp:

<u>CódProj</u>	<u>CodEmp</u>	DataIni	TempAl
LSC001	2146	1/11/91	24
LSC001	3145	2/10/91	24
LSC001	6126	3/10/92	18
LSC001	1214	4/10/92	18
LSC001	8191	1/11/92	12
PAG02	8191	1/05/93	12
PAG02	4112	4/01/91	24
PAG02	6126	1/11/92	12

Emp:

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

Exemplo - 3FN

O valor de salário depende apenas da categoria

Emp:

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



Exemplo - 3FN

- **Terceira Forma Normal**
- Elimina a redundância de dados com dependência funcional transitiva

```
Proj ( #CodProj, Tipo, Descr )  
ProjEmp ( #&CodProj, #&CodEmp, DataIni, TempAl )  
Empregado ( #CodEmp, Nome, Cat, Sal )
```



```
Proj ( #CodProj, Tipo, Descr )  
ProjEmp ( #&CodProj, #&CodEmp, DataIni, TempAl )  
Empregado ( #CodEmp, Nome, &Cat )  
Categoria ( #Cat, Sal )
```

Exemplo - 3FN

Fonte: adaptado de
©Carlos A. Heuser

Proj:

<u>CódProj</u>	Tipo	Descr
LSC001	Novo Desenv.	Sistema de Estoque
PAG02	Manutenção	Sistema de RH

Cat:

<u>Cat</u>	Sal
A1	4
A2	4
B1	9

ProjEmp:

<u>CódProj</u>	<u>CodEmp</u>	DataIni	TempAl
LSC001	2146	1/11/91	24
LSC001	3145	2/10/91	24
LSC001	6126	3/10/92	18
LSC001	1214	4/10/92	18
LSC001	8191	1/11/92	12
PAG02	8191	1/05/93	12
PAG02	4112	4/01/91	24
PAG02	6126	1/11/92	12

Emp:

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

Exercício

- Escreva o esquema para o projeto do exercício anterior na Segunda Forma Normal (2FN) e na Terceira Forma Normal (3FN).

Normalização Posterior

- Para a maioria dos casos a decomposição até a 3FN é suficiente mas existem outras formas normais na literatura (forma normal de Boyce/Codd, 4FN, 5FN)
 - Na 4FN, por exemplo, exige-se que não exista dependência multi-valorada.
 - Boyce/Codd restringe um pouco mais a 3FN, evitando certas ambiguidades em dependência entre chaves.

Problemas na Normalização

- Chaves primárias omitidas ou incorretas
 - Nem sempre os registros em arquivos apresentam uma chave primária => neste caso deve-se analisar o contexto e inseri-las apropriadamente já na FNN
- Atributos relevantes implicitamente representados
 - Arquivos podem conter atributos de forma implícita como a ordem dos registros => deve-se adicionar explicitamente tais atributos já na FNN (ex: lista de vestibular)
- Atributos irrelevantes, redundantes ou derivados:
 - Devem ser eliminados já na FNN

Integração de modelos: cada documento é normalizado e posteriormente um último modelo é gerado.

- Passos:
 - Integração de tabelas com mesma chave
 - Integração de tabelas com chave contida
 - Verificação da 2FN e 3FN
- Dificuldades:
 - Atributos e tabelas podem aparecer com nomes diferentes nos documentos e nos modelos (sinônimos/homônimos)
 - Tabelas entre diferentes modelos podem ter redundância entre si => deve-se integrá-las e eliminar as redundâncias

Considerações

- Integração de tabelas com mesma chave

Fonte: adaptado de
©Carlos A. Heuser

Documento 1:

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

Documento2:

```
Proj (CodProj, DataInicio, Descr, CodDepto)
Depto (CodDepto, NomeDepto)
ProjEquipamento (CodProj, CodEquipam, DataIni)
ProjEmp (CodProj, CodEmp, FunçãoEmpProj)
Equipamento (CodEquipam, Descrição)
```

Considerações

- Integração de tabelas com mesma chave

Fonte: adaptado de
©Carlos A. Heuser

Modelo integrado:

```
Proj (CodProj,Tipo,Descr,DataInicio,CodDepto)
ProjEmp (CodProj,CodEmp,DataIni,TempAl,FunçãoEmpProj)
Emp (CodEmp,Nome,Cat)
Cat (Cat,Sal)
Depto (CodDepto,NomeDepto)
ProjEquipamento (CodProj,CodEquipam,DataIni)
Equipamento (CodEquipam,Descrição)
```

Considerações

- Integração de tabelas com chave contida

Fonte: adaptado de
©Carlos A. Heuser

Modelo #1:

AlunoDisc (Cod-Al, Cod-Disc)

Modelo #2:

AlunoDiscSem (Cod-Al, Cod-Disc, Sem-Disc-Cursada,
Nota-Disc)

AlunoDisc: informa que um aluno cursou uma disciplina

AlunoDiscSem: informa a nota obtida pelo aluno em uma determinada disciplina em um semestre

Considerações

- Não integrar quando tabela contém dados além da chave primária

Fonte: adaptado de
©Carlos A. Heuser

Modelo #1:

AlunoDisc (Cod-Al, Cod-Disc, BolsaSimNao)

Modelo #2:

AlunoDiscSem (Cod-Al, Cod-Disc, Sem-Disc-Cursada,
Nota-Disc)

AlunoDisc: informa que um aluno cursou uma disciplina

AlunoDiscSem: informa a nota obtida pelo aluno em uma determinada disciplina em um semestre

Considerações

- Verificação da 2FN e 3FN

Fonte: adaptado de
©Carlos A. Heuser

Modelo #1:

Departamento (CodDepto, NomeDepto, CodGerenteDepto)

Modelo # 2:

Departamento (CodDepto, LocalDepto, NomeGerenteDepto)



Modelo #1:

Departamento (CodDepto, NomeDepto, CodGerenteDepto,
LocalDepto, NomeGerenteDepto)