

# Normalização

- O que é?
  - Processo para obtenção de um esquema relacional **livre de redundâncias**
  - método de projeto ou mesmo de validação do projeto de um BD relacional
- Quando a Normalização se faz necessária?
  - Migração de sistemas legados
  - Documentação do sistema - Obtenção de esquemas
  - Adaptação de outros modelos para o relacional
  - Reengenharia ou engenharia reversa

# Processo de Normalização

- Entrada
  - fonte de dados estruturados da organização
    - Relatório
    - Fichário
    - documento estruturado
    - tabelas relacionais (reengenharia)
- Saída
  - esquema relacional para a fonte de dados

# Processo de Normalização

- Baseado no conceito de formas normais
  - processo sistemático de geração de tabelas
- Uma tabela  $T$  está em uma forma normal se atende uma certa regra
  - se  $T$  está na Forma Normal (FN)  $i$ , então  $T$  está automaticamente nas FNs  $i - 1, i - 2, \dots$
- Três FNs são mais utilizadas na prática
- A teoria nem sempre é ideal do ponto de vista prático
  - proliferação de tabelas!

# Exemplo de Fonte de Dados: Relatório

## 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

# Passo 1 – Conversão para Tabela Não-Normalizada (ÑN)

- Objetivo
  - obtenção de uma representação padrão para as fontes de dados
    - facilita o processo de normalização
- Forma de representação: Tabela ÑN
  - pode ter uma ou mais tabelas aninhadas
  - tabela aninhada:
    - possui atributos multivalorados
      - atributo que ao invés de conter valores atômicos, pode conter:
        - » múltiplos valores
        - » OU uma tabela que pode ter outras possíveis tabelas dentro dela

# Exemplo de Tabela ÑÑ

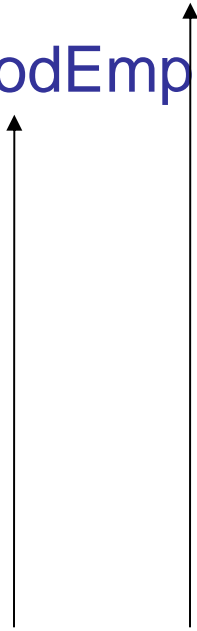
CódProj	Tipo	Descr	Emp					
			CodEmp	Nome	Cat	Sal	DataIni	TempAl
LSC001	Novo Desenv.	Sistema de Estoque	2146	João	A1	4	1/11/91	24
			3145	Sílvio	A2	4	2/10/91	24
			6126	José	B1	9	3/10/92	18
			1214	Carlos	A2	4	4/10/92	18
			8191	Mário	A1	4	1/11/92	12
PAG02	Manutenção	Sistema de RH	8191	Mário	A1	4	1/05/93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

Tabela aninhada em uma linha de projeto

# Representação na Forma de Tabela ÑÑ

Projetos(#codProj, tipo, descr,

(#codEmp, nome, categ, sal, dataIni, tempoAloc))



Indicam as chaves primárias de cada tabela

# 1ª Forma Normal (1FN)

- *“Uma tabela está na 1FN se ela não possui tabelas aninhadas”*
- Procedimento usual
  - gerar uma tabela para cada aninhamento



# 1FN - Aplicação

ÑN: Projetos (#codProj, tipo, descr,  
          (#codEmp, nome, categ, sal, dataIni, tempoAloc))



1FN: Projetos (#codProj, tipo, descr)  
      Alocações (&codProj, #codEmp, nome, categ, sal,  
                  dataIni, tempoAloc)

- CP da tabela externa migra para a tabela aninhada
- Qual a CP da tabela aninhada?

# 1FN - Aplicação

ÑN: Projetos (#codProj, tipo, descr,  
(#codEmp, nome, categ, sal, dataIni, tempoAloc))



1FN: Projetos (#codProj, tipo, descr)  
Alocações (&codProj, #codEmp, nome, categ, sal,  
dataIni, tempoAloc)

- CP da tabela aninhada
  - codEmp identifica unicamente uma tupla em Alocações?  
(analisar os dados da tabela aninhada na fonte de dados como um todo)

# 1FN - Aplicação

ÑN: Projetos (#codProj, tipo, descr,  
(#codEmp, nome, categ, sal, dataIni, tempoAloc))



1FN: Projetos (#codProj, tipo, descr)  
Alocações (&codProj, &codEmp, nome, categ, sal,  
dataIni, tempoAloc)

- CP da tabela aninhada
  - codEmp identifica unicamente uma tupla em Alocações?  
**resposta: NÃO**
  - logo, codProj deve fazer parte da CP da tabela aninhada

# 1FN – Outro Exemplo

ÑN: Departamentos (#codD, nome,  
(#CPF, nome, salário))



1FN: Departamentos (#codD, nome)  
Empregados(&codD, #CPF, nome, salário)



- CP da tabela aninhada (**Empregados**)
  - CPF identifica unicamente uma tupla em Empregados?  
**resposta: SIM**
  - logo, CPF é suficiente como CP da tabela aninhada

# Dependência Funcional

- Conceito necessário para o entendimento da segunda e terceira formas normais
- Definição
  - um atributo  $A_2$  *depende funcionalmente* de um atributo  $A_1$  (ou um atributo  $A_1$  *determina* um atributo  $A_2$ ) quando, em todas linhas da tabela, para cada valor de  $A_1$  que aparece na tabela, aparece o mesmo valor de  $A_2$

# Dependência Funcional - Exemplo

...	Código	....	Salário
	E1		500
	E3		450
	E2		500
	E1		500
	E3		450
	E2		500
	...		

# DF Total e DF Parcial

- DF Total
  - se um atributo  $A_x$  depende funcionalmente de todos os atributos que compõem a CP de uma tabela T, diz-se que  $A_x$  possui DF **total** da CP de T
- DF Parcial
  - se um atributo  $A_x$  depende funcionalmente apenas de alguns atributos (não todos!) que compõem a CP de uma tabela T, diz-se que  $A_x$  possui DF **parcial** da CP de T

## 2ª Forma Normal (2FN)

- *“Uma tabela está na 2FN se ela estiver na 1FN e não possuir DFs parciais”*
  - tabelas com DFs parciais devem ser desmembradas em tabelas que possuam DFs totais
- Tabelas cuja CP possui apenas um atributo estão automaticamente na 2FN



# 2FN - Aplicação

**1FN: Projetos (#CodProj, Tipo, Descr)**

**Alocações (#&CodProj, #CodEmp, Nome, Cat, Sal,  
DataIni, TempoAloc)**

**DFs:** CodEmp  $\rightarrow$  Nome, Cat, Sal (DF Parcial!)  
(CodProj, CodEmp)  $\rightarrow$  DataIni, TempoAloc



**2FN: Projetos (#CodProj, Tipo, Descr)**

**Alocações (#&CodProj, #&CodEmp, DataIni, TempoAloc)**

**Empregados (#CodEmp, Nome, Cat, Sal)**

# DF Transitiva ou Indireta

- Se um atributo não-chave  $A_x$  possui DF total da CP de uma tabela T e também possui DF total de um ou mais atributos não-chave de T, então diz-se que  $A_x$  possui DF transitiva ou indireta da CP de T

# 3ª Forma Normal (3FN)

- *“Uma tabela está na 3FN se ela estiver na 2FN e não possuir DFs indiretas”*
  - tabelas com DFs indiretas devem ser desmembradas em tabelas que não possuam tais DFs
- Tabelas que possuem zero ou apenas um atributo que não faz parte da CP estão automaticamente na 3FN

# 3FN - Aplicação

**2FN: Projetos (#CodProj, Tipo, Descr)**

**Alocações (#&CodProj, #&CodEmp, DataIni, TempoAloc)**

**Empregados (#CodEmp, Nome, Cat, Sal)**

**DFs:** CodEmp  $\rightarrow$  Sal

CodEmp  $\rightarrow$  Cat  $\rightarrow$  Sal (DF indireta!)

**3FN: Projetos (#CodProj, Tipo, Descr)**

**Alocações (#&CodProj, #&CodEmp, DataIni, TempoAloc)**

**Empregados (#CodEmp, Nome, &Cat)**

**CategoriasFuncionais(#Cat, Sal)**

# Normalização - Questões

- Análise de chaves primárias (CPs)
  - tabelas podem ou não ter atributos que garantam identificação única de suas tuplas ou ter uma CP muito extensa
    - sugestão: definir uma CP

**ÑN: Projetos (#CodProj, Tipo, Descr,  
(Nome, Cat, Sal, DataIni, TempoAloc))**



**ÑN: Projetos (#CodProj, Tipo, Descr,  
(#CodEmp, Nome, Cat, Sal, DataIni, TempoAloc))**

# Normalização - Questões

- Dados irrelevantes
  - tabelas podem ter atributos que não precisam ser mantidos necessariamente no BD
    - sugestão: eliminar estes atributos

ÑN: Projetos (#CodProj, Tipo, Descr, **NroEmps, DataRel,**  
(#CodEmp, Nome, Cat, Sal, DataIni, TempoAloc))



ÑN: Projetos (#CodProj, Tipo, Descr,  
(#CodEmp, Nome, Cat, Sal, DataIni, TempoAloc))

# Normalização - Questões

- Dados relevantes, porém implícitos
  - sugestão: definir tais dados

ÑN: Aprovação (#CodCurso, Nome, *a ordem determina a classificação do candidato*  
(#CodCand, Nome, Endereço))



ÑN: Aprovação (#CodCurso, Nome,  
(#CodCand, Nome, Endereço, **OrdemClass**))

# Normalização - Questões

- Relacionamentos incorretos
  - sugestão: validar as tabelas ao final do processo!

## Relatório de Pedidos

ÑN: Pedidos (#nroPed, dataPed, codCli, nomeCli,  
(#nroTel), (#codPeça, descrPeça, qtdePedida))



1FN: Pedidos (#nroPed, dataPed, codCli, nomeCli)

Telefones (#&nroPed, #nroTel)

Peças (#&nroPed, #codPeça, nomePeça, qtde))



... (validação)

Telefones (#&codCli, #nroTel) - definir tabela cliente



# DF Multivalorada

- Se um atributo  $A_{x_1}$  de  $T$  determina um conjunto finito de valores para os outros atributos  $A_{x_2}, \dots, A_{x_n}$  de  $T$ , então diz-se que  $A_{x_2}, \dots, A_{x_n}$  possuem DF multivalorada de  $A_{x_1}$  em  $T$

# Exemplo de Normalização até 3FN

## Ficha de Estante e seus Livros

ÑN: Estantes (#número, capacidade, (#ISBN, título, ano  
(#codAutor, nome, nacionalidade)))

1FN: Estantes (#número, capacidade)

DistribuiçãoLivros (#&número, #ISBN, título, ano)

DistribuiçãoLivrosAutor (#&número, #&ISBN, #codAutor,  
nome, nacionalidade)

2FN = 3FN: Estantes (#número, capacidade)

Livros (#ISBN, título, ano)

DistribuiçãoLivros (#&número, #&ISBN)

Autores (#codAutor, nome, nacionalidade)

DistribuiçãoLivrosAutor(#&número, #&ISBN, #&codAutor)

relacionamento  
contido em outra  
tabela! (pode ser  
removido)



# Exemplo de DF Multivalorada

## DistribuiçãoLivrosAutor

número	ISBN	codAutor
E1	L1	A1
E1	L1	A2
E1	L2	A7
E1	L2	A8
E1	L2	A9
E2	L1	A1
E2	L1	A2
E3	L2	A7
E3	L2	A8
E3	L2	A9

$ISBN = L1 \Rightarrow \text{codAutor} = \{A1, A2\}$

$ISBN = L2 \Rightarrow \text{codAutor} = \{A7, A8, A9\}$

$\Rightarrow$  Redundância de dados  
para representar as  
associações  
multivaloradas!

DF multivalorada:

$ISBN \rightarrow\rightarrow\text{codAutor}$

# 4ª Forma Normal (4FN)

- *“Uma tabela está na 4FN se ela estiver na 3FN e não possuir DFs multivaloradas”*
  - tabelas com DFs multivaloradas devem ser desmembradas em tabelas que não possuam tais DFs
- Tabelas que possuem CP composta por um ou dois atributos, ou que possuem atributos não-chave estão automaticamente na 4FN

# 4FN - Aplicação

3FN: Estantes (#número, capacidade)

Livros (#ISBN, título, ano)

Autores (#codAutor, nome, nacionalidade)

DistribuiçãoLivrosAutor(#&número, #&ISBN, #&codAutor)

**DF Multivalorada: ISBN →→codAutor**



4FN: Estantes (#número, capacidade)

Livros (#ISBN, título, ano)

Autores (#codAutor, nome, nacionalidade)

DistribuiçãoLivros(#&número, #&ISBN)

Autoria(#&ISBN, #&codAutor)