



IBD012
BANCOS & ARMAZENS DE DADOS

NORMALIZAÇÃO

✕ Programação

+ Introdução

✕ Por que?

+ Conceitos Iniciais

✕ Multivaloração

✕ Dependências funcionais

✕ Multivaloração

✕ Dep. Funcionais

✕ Parcial

✕ Transitiva

+ As Formas Normais (Definição e procedimento)

✕ 1ª Forma Normal

✕ 2ª Forma Normal

✕ 3ª Forma Normal

+ Observações

+ Exercício

IBD012-NORMALIZAÇÃO

✖ Introdução

✖ Por que desenvolveram a normalização?

✖ Dê um nome para esta tabela:

CdCli	NmCli	cdVend	NmVend	NuPed	vtPed	dtPed	Qtd	vUnit	vTot	cdmat	nmMat	cdGru	nmGru
A	Ana	1	Luiz	1	44	02/set	2	10	20	A	Martelo	2	Marcenaria
A	Ana	1	Luiz	1	44	02/set	2	12	24	B	Formão	2	Marcenaria
B	Luiz	1	Luiz	2	78	02/set	3	10	30	A	Martelo	2	Marcenaria
B	Luiz	1	Luiz	2	78	02/set	4	12	48	C	Eletrodo	1	Serralheria
A	Ana	2	Claudio	3	91	02/set	3	11	33	D	Rebites	1	Serralheria
A	Ana	2	Claudio	3	91	02/set	2	9	18	C	Eletrodo	1	Serralheria
A	Ana	2	Claudio	3	91	02/set	4	10	40	B	Formão	2	Marcenaria
C	Luiz	2	Claudio	4	62	03/set	2	11	22	A	Martelo	2	Marcenaria
C	Luiz	2	Claudio	4	62	03/set	4	10	40	D	Rebites	1	Serralheria
B	Luiz	1	Luiz	5	81	03/set	1	12	12	C	Eletrodo	1	Serralheria
B	Luiz	1	Luiz	5	81	03/set	3	9	27	D	Rebites	1	Serralheria
B	Luiz	1	Luiz	5	81	03/set	2	10	20	A	Martelo	2	Marcenaria
D	José	3	Rui	6	90	04/set	4	10	40	B	Formão	2	Marcenaria
D	José	3	Rui	6	90	04/set	2	10	20	D	Rebites	1	Serralheria
D	José	3	Rui	6	90	04/set	3	10	30	A	Martelo	2	Marcenaria
A	Ana	3	Rui	7	92	05/set	2	11	22	A	Martelo	2	Marcenaria
A	Ana	3	Rui	7	92	05/set	7	10	70	D	Rebites	1	Serralheria

✖ Para o MR esta tabela está 'certa'! Não tem linhas repetidas, não tem nomes de campos repetidos...

✖ Esta tabela está meio 'bagunçada' porque tem muitos campos de uma mesma área funcional.

✗ Introdução

- ✗ Quando o MR surgiu foi uma euforia. Todos viam uma ferramenta de construção de bancos de dados eficiente.
- ✗ Mas não atentaram aos cuidados com a modelagem.
- ✗ Consequência: tabelas 'bagunçadas', 'esquisitas', etc.
- ✗ No fundo estas tabelas não eram representações de um só conjunto, mas sim de um grupo de conjuntos (importantes para uma área funcional que estava sendo modelada).
- ✗ Mas sobre estas tabelas já haviam se desenvolvidos muitos programas... Como resolver estes problemas de modelagem?
 - + Estudando um método de modelagem →

}	Criar novas ideias de modelagem:
	Mod. Entidade-Relacionamento
 - + Corrigindo as tabelas já existentes.
 - Criar procedimentos que ajustem as Tabelas com erro!

- ✖ Introdução
- ✖ As tabelas apresentam 'Anomalias' (não chega a ser um "erro")...
- ✖ Anomalias são 'situações de definições de dados que dificultam a manutenção dos dados da tabela e dos programas aplicativos'.
- ✖ Ah!... Ok... Mas então quais são as anomalias e como eliminá-las?
- ✖ Entendendo um pouco mais de conceitos e estabelecendo classificação (das tabelas) e procedimentos de ajustes (das tabelas).

- ✗ Estudando aspectos dos dados das tabelas.
- ✗ Sendo T uma Tabela qualquer e sendo C1 e C2 dois de seus campos, diz se que C2 é DEPENDENTE FUNCIONAL de C1 SE, e somente se, sempre sabemos os valores de C2 quando informamos os valores de C1.
- ✗ Dizemos que C1 é **DETERMINANTE** de C2 ou que C2 é **DEPENDENTE** de C1.

C1	C2
10	100
10	100
20	200
20	200
30	100
20	200
30	100

Dado o valor '10' de C1 SEMPRE (para todos os valores 10 de C1) vemos em C2 o mesmo Valor 100, dado o valor '20' de C1 vemos o mesmo valor em todas as tuplas no campo C2.

Tendo C2 o valor 100, temos o valor de C1 Constante?

Não! Portanto, C2 não determina C1.

C2 depende de C1 mas C1 não depende de C2

Notação: $C1 \rightarrow C2$ (Leia-se: C1 Determina C2)

Neste caso: $C2 \not\rightarrow C1$

- ✗ Estudando aspectos dos dados das tabelas.
- ✗ Uma regra observada durante todo o procedimento de Normalização: *É desejável que todos os campos sejam dependentes funcionais de toda a Chave Primária.*
- ✗ Então....
- ✗ Pode ocorrer Dependência Funcional entre campos fora da Chave Primária com uma parte da Chave Primária (no caso da primária ser composta)? SIM. Essa 'anomalia' se chama Dependência Funcional Parcial.
- ✗ Pode ocorrer Dependência Funcional entre campos fora da chave? SIM. Essa 'anomalia' se chama Dependência Funcional Transitiva
- ✗ Hum... Então... Saber quem é a CP de uma tabela já é importante para saber se existe DFTransitiva ou DFParcial? SIM.
- ✗ Estas duas 'anomalias' não são boas para a tabela.
- ✗ Além destas... Pode existir a Multivaloração.

- ✘ Estudando aspectos dos dados das tabelas.
- ✘ Não bom ter multivaloração. Multivaloração?
- ✘ Na definição do modelo relacional cada campo de uma tupla deve ter valor único!
Mas esta máxima pode às vezes não ser seguida... (!)
- ✘ Em uma tabela pessoas o analista definiu um campo NuTelefone para o número de telefone das pessoas.
- ✘ Mas HOJE uma pessoa pode ter muitos números de telefone.

✗ Estudando aspectos dos dados das tabelas.

✗ Multivaloração:

✗ O que era assim:

cdCli	NuTelefone
A	325612
B	215421
C	241214
D	415241

Passa a ser assim:

cdCli	NuTelefone
A	325612 / 214521 / ...
B	215421 / 785458
C	241214 / 856475 / ...
D	415241 / 954585

✗ Esta é uma anomalia não desejável.

- ✗ Estudando aspectos dos dados das tabelas.
- ✗ Explicando a DFParcial
- ✗ Em tabelas com Chave Primária Composta, pode acontecer de um campo fora da CP ser determinado por uma parte da chave.

C1	C2	C3	<u>C4</u>	<u>C5</u>
10	100		A	B
15	100		A	C
20	400		C	B
20	200		D	E
30	200		D	C
20	300		B	A
30	100		A	D
40	300		B	E


CP

C4 → C2 (C4 é parte da CP),

Mas

C5 ~~→~~ C2 (C5 é a outra parte da CP)

Nesta situação dizemos que a Tabela apresenta Dependência funcional Parcial.

- ✗ Estudando aspectos dos dados das tabelas.
- ✗ Explicando a DFTransitiva
- ✗ Uma tabela que apresentem Dependência Funcional entre campos fora da Chave Primária, diz-se que a Tabela tem DFTransitiva.

C1	C2	C3	<u>C4</u>	<u>C5</u>
10	100	AB	A	B
15	100	AB	A	C
20	400	ER	C	B
20	200	ER	D	E
30	200	ER	D	C
20	300	WE	B	A
30	100	AB	A	D
40	300	WE	B	E

↖
↗
CP

C2 → C3 (campo fora da Chave)

Note que C3 ~~→~~ C2

Nesta situação a Tabela apresenta Dependência Funcional Transitiva.

- ✗ Estudando aspectos dos dados das tabelas.
- ✗ Explicando a DFTransitiva... Note o exemplo:

C1	C2	C3	<u>C4</u>	<u>C5</u>
10	10	100	A	B
15	10	150	A	C
20	40	800	C	B
20	20	400	D	E
30	20	600	D	C
20	30	600	B	A
30	10	300	A	D
40	30	1200	B	E


CP

C1 E C2 $\rightarrow$ C3 (campo fora da Chave)
Note que C3 tem valores determinados por:
C1 * C2 (processo matemático)

Nesta situação a Tabela tem
Dependência Transitiva (de cálculo).

Note que:

C3 E C2 $\rightarrow$ C1 (C1=C3 $\div$ C2)

e que

C3 E C1 $\rightarrow$ C2. (C2=C3 $\div$ C1)

- ✗ Definindo as Formas Normais.
- ✗ É uma Classificação do Estado da tabela:
- ✗ Primeira Forma Normal (1FN):
 - + Uma Tabela está na 1FN quando:
 - ✗ NÃO apresenta Multivaloração em seus campos.
- ✗ Segunda Forma Normal (2FN):
 - + Uma Tabela está na 2FN quando:
 - ✗ Está na Primeira Forma Normal E
 - ✗ NÃO existe Dependência Parcial.
- ✗ Terceira Forma Normal (3FN):
 - + Uma Tabela está na 3FN quando:
 - ✗ Está na Segunda Forma Normal E
 - ✗ NÃO existe Dependência Transitiva. (de cálculo, inclusive!)

IBD012-NORMALIZAÇÃO

- ✗ Então... Já sabemos quais são as anomalias
- ✗ Por que podem ocorrer estes problemas?
 - + Estas anomalias aparecem por falhas na consideração de um campo para uma Tabela.
- ✗ Como eliminá-las?
 - + Aplicando uma série de procedimentos sobre as tabelas (extraíndo dados – e campos – simplificando o esquema e facilitando o 'entendimento' do conjunto representado na tabela. → Melhorando o Modelo de Dados.
- ✗ Isso é a NORMALIZAÇÃO (análise das anomalias e aplicação de procedimentos corretivos nas tabelas).

IBD012-NORMALIZAÇÃO

- ✖ Procedimento: FNN → 1FN
- ✖ Antes de qualquer procedimento, deve-se identificar qual é a CP da tabela. SE a CP for 'imposta' (indicada previamente), deve-se ver se existe a multivaloração. SE Não houver, ENTÃO 1FN=Ok!
- ✖ Agora pode acontecer isso:

T1

QtdItem	<u>NumPed</u>	DtPed	CdMat
10 / 15 / 17	10	11-Ago-2002	A / B / C
14 / 12	11	11-Ago-2002	B / D
20 / 10 / 20	12	12-Ago-2002	A / C / D
15 / 20	13	12-Ago-2002	B / C

- ✖ CP existente e campos multivalorados ocorrem:
1FN=Ñ-Ok!

- ✗ Procedimento: FNN → 1FN
- ✗ Eliminando a Multivaloração

T1

QtyItem	<u>NumPed</u>	DtPed	CdMat
10 / 15 / 17	10	11-Ago-2002	A / B / C
14 / 12	11	11-Ago-2002	B / D
20 / 10 / 20	12	12-Ago-2002	A / C / D
15 / 20	13	12-Ago-2002	B / C

- Se T1 apresentar Multivaloração:
 - Crie uma outra Tabela T2 e leve para T2 a CP da Tabela T1 e os campos multivalorados (T1 fica com a CP e os campos não multivalorados).
 - Na Tabela T2 devemos ter um valor em cada tupla para os campos multivalorados (deve-se 'desmembrar' as combinações, mas somente com uma 'reanalise' dos valores correlatos).

IBD012-NORMALIZAÇÃO

- ✗ Procedimento: FNN → 1FN
- ✗ Eliminando a Multivaloração

T1

<u>NumPed</u>	DtPed
10	11-Ago-2002
11	11-Ago-2002
12	12-Ago-2002
13	12-Ago-2002

T2

<u>NumPed</u>	QtdItem	CdMat
10	10 / 15 / 17	A / B / C
11	14 / 12	B / D
12	20 / 10 / 20	A / C / D
13	15 / 20	B / C

- Na Tabela T2, se existirem dois ou mais campos multivalorados devemos fazer uma análise que indique a correlação de valores dos campos multivalorados.
- Na Tabela T2 que é criada não temos definida a Chave Primária.

IBD012-NORMALIZAÇÃO

- ✗ Procedimento: FNN → 1FN
- ✗ Analisando a correlação de valores em T2

<u>NumPed</u>	CdMat	QtdItem
10	A / B / C	10 / 15 / 17
11	B / D	14 / 12
12	A / C / D	20 / 10 / 20
13	B / C	15 / 20

<u>NumPed</u>	CdMat	QtdItem
10	A	10
10	B	17
10	C	15
11	B	14
11	D	12
12	A	10
12	D	20
12	C	20
13	B	20
13	C	15

- ✗ Note:
- ✗ Não é posicional.
- ✗ O analista deve 'voltar' ao contexto 'real' e identificar as Correlações existentes entre os dados.

IBD012-NORMALIZAÇÃO

- ✖ Procedimento: 1FN → 2FN
- ✖ Se a tabela tem uma CP Simples (um só campo), então NÃO tem como ter DFParcial.
- ✖ Desse modo, SE ESTÁ na 1FN (com CP simples) está na 2FN.
- ✖ Certo... Parece um jogo de palavras, mas temos conceitos envolvidos. Por isso vimos o conceito de Dependência Funcional e o conceito de Chave Primária (simples e composta).
- ✖ Ok! Vamos no caso da tabela com CP Composta E já está na 1FN. Temos que provar que Não Tem DFParcial para estar na 2FN.

IBD012-NORMALIZAÇÃO

- ✗ Procedimento: 1FN → 2FN
- ✗ Eliminando a Dep Func. Parcial.

T1

Seção	NomeSec	<u>NumReq</u>	<u>DtReq</u>	<u>CdMat</u>	DescMatPrima	QtdReq	VlrUnit	Vlr
020	Manutencao	101	19930601	100	Enxofre	8	12500	100000
020	Manutencao	101	19930601	101	Areia Monazitica	5	10000	50000
020	Manutencao	106	19930602	100	Enxofre	15	15000	225000
020	Manutencao	106	19930602	101	Areia Monazitica	4	11000	44000
021	Producao-1	102	19930601	101	Areia Monazitica	6	11500	69000
021	Producao-1	102	19930601	102	Argila Moida	25	3400	85000
021	Producao-1	109	19930604	101	Areia Monazitica	5	12000	60000

- ✗ T1 tem CP composta e apresenta Dependência Funcional Parcial.
 - + 1) Criar uma nova Tabela para cada parte determinante da CP de T1 transportando o **DETERMINANTE** todos os *DEPENDENTES*.
 - + 2) Retirar de T1 SOMENTE os *dependentes*. A CP em T permanece inteira.

IBD012-NORMALIZAÇÃO

- ✗ Procedimento: 1FN → 2FN
- ✗ Eliminando a Dep Func. Parcial.

T1

Seção	NomeSec	<u>NumReq</u>	<u>DtReq</u>	<u>CdMat</u>	DescMatPrima	QtdReq	VlrUnit	Vlr
020	Manutencao	101	19930601	100	Enxofre	8	12500	100000
020	Manutencao	101	19930601	101	Areia Monazitica	5	10000	50000
020	Manutencao	106	19930602	100	Enxofre	15	15000	225000
020	Manutencao	106	19930602	101	Areia Monazitica	4	11000	44000
021	Producao-1	102	19930601	101	Areia Monazitica	6	11500	69000
021	Producao-1	102	19930601	102	Argila Moida	25	3400	85000
021	Producao-1	109	19930604	101	Areia Monazitica	5	12000	60000

- ✗ Numreq → seção, nomesec, dtreq (tabela T2)
- ✗ Cdmat → descmatprima (tabela T3)

IBD012-NORMALIZAÇÃO

✗ Procedimento: 1FN → 2FN

T1

Seção	NomeSec	<u>NumReq</u>	<u>DtReq</u>	<u>CdMat</u>	DescMatPrima	QtdReq	VlrUnit	Vlr
020	Manutencao	101	19930601	100	Enxofre	8	12500	100000
020	Manutencao	101	19930601	101	Areia Monazitica	5	10000	50000
020	Manutencao	106	19930602	100	Enxofre	15	15000	225000
020	Manutencao	106	19930602	101	Areia Monazitica	4	11000	44000
021	Producao-1	102	19930601	101	Areia Monazitica	6	11500	69000
021	Producao-1	102	19930601	102	Argila Moida	25	3400	85000
021	Producao-1	109	19930604	101	Areia Monazitica	5	12000	60000

Numreq →
seção, nomesec, dtreq (tabela T2)

T2

Seção	NomeSec	<u>NumReq</u>	<u>DtReq</u>
020	Manutencao	101	19930601
021	Producao-1	102	19930601
020	Manutencao	106	19930602
021	Producao-1	109	19930604

Cdmat →
descmatprima (tabela T3)

T3

<u>CdMat</u>	DescMatPrima
100	Enxofre
102	Argila Moida
101	Areia Monazitica

T1

<u>NumReq</u>	<u>CdMat</u>	QtdReq	VlrUnit	Vlr
101	100	8	12500	100000
101	101	5	10000	50000
102	101	6	11500	69000
102	102	25	3400	85000
106	100	15	15000	225000
106	101	4	11000	44000
109	101	5	12000	60000

Repare:

Cada Determinante é uma nova CP
nas Tabelas 'desmembradas'. E são Chaves Estrangeiras em T1.

IBD012-NORMALIZAÇÃO

- ✗ Procedimento: 2FN → 3FN
- ✗ Eliminando a Dependência Transitiva (DF fora da CP)
- ✗ Desconsidere como determinante de dependência transitiva qualquer campo, fora da chave primária, que possa ser candidato à ser chave primária (em T1, o campo DescMatPrima pode ser chave).

T1

<u>CdMat</u>	<u>DescMatPrima</u>	<u>CdTipoMat</u>	<u>DescTipo</u>	<u>QtdEst</u>
100	Enxofre	10	Químico	1500
101	Areia Monazitica	20	Natural	1600
102	Argila Moida	20	Natural	1450
103	Sódio	10	Químico	1600
104	Cloro	10	Químico	1300
105	Flúor	10	Químico	1250

CdTipoMat ~~→~~ QtdEst, DescMatPrima

CdTipoMat → DescTipo

IBD012-NORMALIZAÇÃO

- ✖ Procedimento: 2FN → 3FN
- ✖ Eliminando a Dependência Transitiva (DF fora da CP)
- ✖ Analisar todas as possíveis combinações de dependência.
- ✖ Determine o campo que tem mais dependentes.
- ✖ Criar uma nova Tabela (T1, T2, etc) para cada campo determinante. Para cada nova Tabela leva-se o campo Determinante e os Dependentes.
- ✖ Retirar da Tabela de origem somente os campos dependentes. O campo DETERMINANTE fica na Tabela

IBD012-NORMALIZAÇÃO

✗ Procedimento: 2FN → 3FN

✗ Eliminando a Dependência Transitiva (DF fora da CP)

T1

<u>CdMat</u>	DescMatPrima	CdTipoMat	DescTipo	QtdEst
100	Enxofre	10	Químico	1500
101	Areia Monazitica	20	Natural	1600
102	Argila Moida	20	Natural	1450
103	Sódio	10	Químico	1600
104	Cloro	10	Químico	1300
105	Flúor	10	Químico	1250

A tabela desmembrada fica assim:

T1 ficar assim:

<u>CdMat</u>	DescMatPrima	CdTipoMat	QtdEst
100	Enxofre	10	1500
101	Areia Monazitica	20	1600
102	Argila Moida	20	1450
103	Sódio	10	1600
104	Cloro	10	1300
105	Flúor	10	1250

<u>CdTipoMat</u>	DescTipo
10	Químico
20	Natural

- ✗ Procedimento: 2FN → 3FN
- ✗ Eliminando Dependência Transitiva de cálculo

T2

<u>NumReq</u>	<u>CdMat</u>	QtdReq	VlrUnit	Vlr
101	100	8	12500	100000
101	101	5	10000	50000
106	100	15	15000	225000
106	101	4	11000	44000
102	101	6	11500	69000
102	102	25	3400	85000
109	101	5	12000	60000

Neste caso podemos retirar no máximo 1 campo
(1 entre qtdReq, vlrUnit ou vlr)

Eliminamos nesta etapa também os campos resultantes de processamento matemático. Temos que eliminar o maior número possível de campos de modo que, com o que sobrar, ainda seja possível reconstituir os demais.

T2

<u>NumReq</u>	<u>CdMat</u>	QtdReq	VlrUnit
101	100	8	12500
101	101	5	10000
106	100	15	15000
106	101	4	11000
102	101	6	11500
102	102	25	3400
109	101	5	12000

✗ Observações Finais

- ✗ Note que as etapas de normalização devem ser cumpridas em ordem:
 - + Estude e elimine a multivaloração (tabela FNN → 1FN).
 - + Estude e elimine a Dep. Func Parcial (tabela 1FN → 2FN)
 - + Estude e elimine a Dep. Func Transitiva (tabela 2FN → 3FN)
Também eliminando os campos de cálculo.
- ✗ As tabelas desmembradas sempre carregam um campo de ligação para a Tabela origem.
 - + Este campo é uma CP na Tabela desmembrada
 - ✗ Se desmembramento acontece na 2ª ou 3ª forma,
 - + Este campo é um campo de ligação
 - ✗ Se desmembramento acontece na 1ª Forma.
- ✗ As CPs e CEs possibilitam a ligação
 - + A presença da CP indica a entidade (tabelas) que recebe a cardinalidade de 1 e a CE, qual recebe a de N.

IBD012-NORMALIZAÇÃO

✖ Exercício

- ✖ Normalize a tabela abaixo até que todas as desmembradas estejam na 3FN e desenhe o Modelo Lógico usando o MR como representação

Dica:

Copie estes dados em uma planilha Excell.

Nomeie esta tabela e normalize-a até a 3FN sem se preocupar com as desmembradas.

Depois normalize cada desmembrada da FNN→3FN.

Terminada a Normalização identifique as CPs e as CEs.

CdCli	NmCli	cdVend	NmVend	NuPed	vtPed	dtPed	Qtd	vUnit	vTot	cdmat	nmMat	cdGru	nmGru
A	Ana	1	Luiz	1	44	02/set	2	10	20	A	Martelo	2	Marcenaria
A	Ana	1	Luiz	1	44	02/set	2	12	24	B	Formão	2	Marcenaria
B	Luiz	1	Luiz	2	78	02/set	3	10	30	A	Martelo	2	Marcenaria
B	Luiz	1	Luiz	2	78	02/set	4	12	48	C	Eletrodo	1	Serralheria
A	Ana	2	Claudio	3	91	02/set	3	11	33	D	Rebites	1	Serralheria
A	Ana	2	Claudio	3	91	02/set	2	9	18	C	Eletrodo	1	Serralheria
A	Ana	2	Claudio	3	91	02/set	4	10	40	B	Formão	2	Marcenaria
C	Luiz	2	Claudio	4	62	03/set	2	11	22	A	Martelo	2	Marcenaria
C	Luiz	2	Claudio	4	62	03/set	4	10	40	D	Rebites	1	Serralheria
B	Luiz	1	Luiz	5	81	03/set	1	12	12	C	Eletrodo	1	Serralheria
B	Luiz	1	Luiz	5	81	03/set	3	9	27	D	Rebites	1	Serralheria
B	Luiz	1	Luiz	5	81	03/set	2	10	20	A	Martelo	2	Marcenaria
D	José	3	Rui	6	90	04/set	4	10	40	B	Formão	2	Marcenaria
D	José	3	Rui	6	90	04/set	2	10	20	D	Rebites	1	Serralheria
D	José	3	Rui	6	90	04/set	3	10	30	A	Martelo	2	Marcenaria
A	Ana	3	Rui	7	92	05/set	2	11	22	A	Martelo	2	Marcenaria
A	Ana	3	Rui	7	92	05/set	7	10	70	D	Rebites	1	Serralheria

Dê Nomes às Tabelas e faça o desenho do Modelo.



IBD012
BANCOS & ARMAZENS DE DADOS