

Dependências Funcionais e Normalização



Prof. Dr. Renato Bueno

Material elaborado pela Profa. Dra. Marilde Santos

Departamento de Computação – UFSCar



SGBD + Banco de Dados



- Independência de dados
- Consistência de dados



Consistência de Dados



SGBD → Regras de Integridade

- **Validade**
- **Completeza**
- **Consistência**



Consistência de Dados



O controle de consistência pode ser exercido:

- *Pelo gerenciador;*
- *Pelos aplicativos;*
- *Pela própria construção do sistema.*



Consistência de Dados



Pela própria construção do sistema.

- *Controlar a construção do sistema através da criação de tabelas segundo regras que garantam a manutenção de certas propriedades.*



Normalização



- Obtido o esquema relacional do banco de dados fica a questão:
 - *"Como avaliar o esquema obtido?"*
 - *As tabelas que atendem a um determinado conjunto de regras, diz-se estarem em uma determinada **forma normal**.*
- O processo da normalização pode ser utilizado para definir mais precisamente a "excelência" e a "imperfeição" de cada esquema de relação.
 - Baseia-se no conceito de **Forma Normal**



Normalização Primeira Forma Normal (1FN)

- Diz-se que uma tabela está na 1FN quando não contém tabelas aninhadas.
 - Sem atributos multivalorados
 - notação utilizada {atr}
 - Sem atributos compostos
 - notação utilizada (atr:comp1,comp2)
- Exemplo de relação não-normalizada (ññ):
 Aluno(RA, nome, {(e-mail: login, domínio)})

Normalização Primeira Forma Normal (1FN)

ññ: Aluno(RA, nome, {(e-mail: login, domínio)})

DER:

Na 1FN:
 Aluno(RA, nome)
 E_mail(RA, login, domínio)

Formas Normais Primeira Forma Normal

Nome	Idade	DataNasc	DataMatrícula

DataNasc e DataMatrícula serão atributos atômicos se não forem utilizadas "partes" das datas em outras relações do Banco de Dados.

Formas Normais Primeira Forma Normal

Atributos multivalorados

1) Quando a quantidade de valores é pequena e conhecida a priori;

Substitui-se o atributo multivalorado por um conjunto de atributos de mesmo domínio, cada um representando a ocorrência de um valor.

Formas Normais Primeira Forma Normal

Atributos multivalorados.

2) Quando a quantidade de valores é muito grande, variável ou desconhecida.

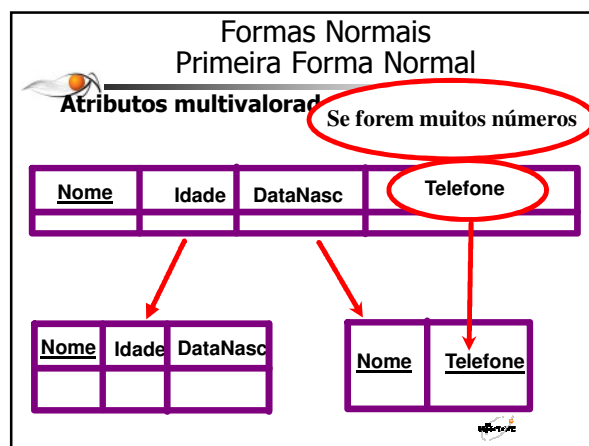
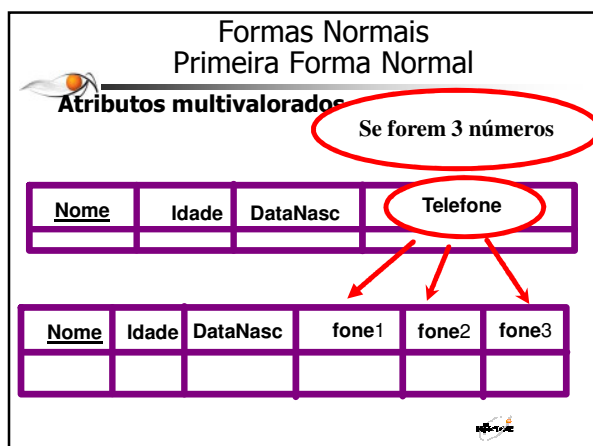
Retira-se da relação o atributo multivalorado, e cria-se uma nova relação que tem o mesmo conjunto de atributos chave, mais o atributo multivalorado como chave, porém tomado como monovalorado.

Formas Normais Primeira Forma Normal

Atributos multivalorados

<u>Nome</u>	Idade	DataNasc	Telefone

Quantos números de telefone?



Normalização
Dependências Funcionais

Se o valor de um conjunto de atributos *A* permite descobrir o valor de um outro conjunto *B*, dizemos que *A* determina funcionalmente *B*, ou que *B* depende de *A*, e denotamos: $A \rightarrow B$

- Exemplo: CEP \rightarrow cidade
 - Sabendo o valor do CEP determino a cidade

Dependências Funcionais

- Propriedade do esquema da relação *R*, não de um estado particular
- Deve ser definida explicitamente por alguém que conheça a semântica dos atributos de *R*
 - projetista especifica as dependências funcionais semanticamente evidente
 - Regras de inferência são usadas para deduzir outras DFs

Dependências Funcionais

Dada uma instância de uma base qualquer, é possível verificar se dois conjuntos de atributos quaisquer atendem à uma Dependência Funcional.

Alunos = {Nome, Curso, Idade}

<Mario, Compt., 21>
<Paulo, Eletr., 22>
<Almir, Fisio., 22>
<Marta, Compt., 21>
<Vânia, Eletr., 22>

OK!!!

Nome \rightarrow curso
Nome \rightarrow idade

Sempre que o Nome se repetir, Curso ou Idade se repete?

Dependências Funcionais

Alunos = {Nome, Curso, Idade}

<Mario, Compt., 21>
<Paulo, Eletr., 22>
<Almir, Fisio., 22>
<Marta, Compt., 21>
<Vânia, Eletr., 22>

OK!!!

curso \rightarrow idade

compt.	21
eletr.	22
fisio.	22

Dependências Funcionais

Alunos = {Nome, Curso, Idade}

<Mario,	Compt.,	21>,
<Paulo,	Eletr.,	22>,
<Almir,	Fisio.,	22>,
<Marta,	Compt.,	21>,
<Vânia,	Eletr.,	22>}

Não!

idade \longrightarrow curso

21 compt.
22 eletr., fisio.

Normalização Dependência Funcional Parcial

Se A for chave da relação e o valor de um subconjunto de atributos de A permite descobrir o valor de um outro conjunto B , dizemos que B possui dependência funcional parcial em relação a A .

Exemplo:

Matricula(RA, SiglaDisc, nome_aluno, descr_disciplina, média)

RA, SiglaDisc \rightarrow média ok

RA \rightarrow nome_aluno Dependência Funcional Parcial

SiglaDisc \rightarrow descr_disciplina Dependência Funcional Parcial

Normalização Atributo **Primo**

- Todo atributo que pertence a uma chave candidata é denominado primo.
- Mas, o que é mesmo chave candidata?
 - É uma superchave mínima!

Normalização Segunda Forma Normal (2FN)

Uma relação está na **2FN** quando:

- está na 1ª. F.N. e;
- todos os seus atributos que não são primos, não dependem parcialmente de qualquer chave candidata da relação.

Normalização Segunda Forma Normal (2FN)

- Exemplo:
Alocação(Sigla, Turma, sala, créditos)

Sigla	Turma	sala	créditos
SBD	A	101	8
SBD	B	110	8
SBD	C	120	4

Sigla, Turma \rightarrow sala

Sigla \rightarrow créditos

Normalização Segunda Forma Normal (2FN)

- Fere a 2FN

Alocação

Sigla	Turma	sala	créditos
SBD	A	101	8
SBD	B	110	8
SBD	C	120	4

- Na 2FN:

Alocação

Sigla	Turma	sala
SBD	A	101
SBD	B	110
SBD	C	120

Disciplina

Sigla	créditos
SBD	8

Considerações sobre a 2FN

- 2FN evita:
 - Inconsistências devido a duplicidade de informações
 - Perda de dados em operações de remoções / alteração na relação
- Se chaves candidatas contém somente 1 atributo, a necessidade do teste não se aplica
- Normalização:
 - Criação de relações na 2FN nas quais os atributos não-primos só estarão associados à parte da chave com a qual possuírem DF total

2FN e 3FN

- 2FN e 3FN abordam problemas diferentes
 - Mas historicamente é habitual seguir essa sequência
- Presumiremos então que uma relação na 3FN *já satisfaz* a 2FN

Normalização Terceira Forma Normal (3FN)

Uma relação está na **3FN** quando:

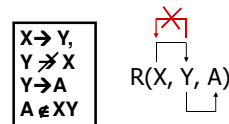
- Está na 2ª F.N. E;
- Todos os seus atributos não primos são dependentes não transitivos de uma chave candidata.

Mas o que é
Dependência Funcional Transitiva?

Normalização Terceira Forma Normal (3FN)

Dependência Funcional Transitiva

- Seja a relação $R(X, Y, A)$, A é transitivamente dependente de X , se existe Y tal que:



Normalização Terceira Forma Normal (3FN)

- As relações não devem ter atributos que não pertençam a uma chave funcionalmente determinados por outro atributo que também não pertença a uma chave

Normalização Terceira Forma Normal (3FN)

- Exemplo:
Alocação(Sigla, Turma, Sala, Prédio)

Sigla	Turma	sala	prédio
SBD	A	101	AT1
SBD	B	101	AT4
SBD	C	101	AT1

Sigla, Turma → Sala ok
Sala → Prédio atributo não primo dependente transitivo

- Fere a 3FN

Normalização Terceira Forma Normal (3FN)

- Fere a 3FN

Sigla	Turma	sala	prédio
SBD	A	101	AT1
SBD	B	101	AT1
SBD	C	101	AT1

- Na 3FN:

Sigla	Turma	sala
SBD	A	101
SBD	B	101
SBD	C	101

Sala	prédio
101	AT1

Normalização Terceira Forma Normal (3FN)

- Intuitivamente, podemos notar é problemática qualquer DF em que:
 - O lado esquerdo for parte de uma chave
 - O lado esquerdo for um atributo não chave

Normalização Terceira Forma Normal (3FN)

Uma relação está na 3^a Forma normal quando:

- Está na 2^a F.N. e;
- Todos os seus atributos não primos são dependentes transitivos de uma chave candidata.

Em outras palavras, uma relação R está na 3FN se:

- para toda dependência funcional $X \rightarrow A$ de R,
 - X for superchave ou
 - A for atributo primo

Normalização Terceira Forma Normal (3FN)

Relembrando: Dependência Funcional Transitiva

- Seja a relação R(X, Y, A), A é transitivamente dependente de X, se existe Y tal que:

$X \rightarrow Y$
 $Y \rightarrow X$
 $Y \rightarrow A$
 $A \notin XY$

R(X, Y, A)

Normalização Forma Normal de Boyce-Codd (FNBC)

- Uma relação R está na FNBC, se:

Para toda dependência funcional $X \rightarrow A$ de R, X for superchave

Não adianta A ser primo!!!

Normalização Forma Normal de Boyce-Codd (FNBC)


Propriedade

Id_Propr	Nome_Região	Lote	Área
prop1	Central	L231	500
prop2	Central	L30	500
prop3	Sul	L140	1000
prop4	Sul	L231	1000
prop5	Leste	L616	1500
prop6	Central	L247	1500

Id_Propr → Nome_Região, Lote, Área
 Nome_Região, Lote → Id_Propr, Área
 Área → Nome_Região

Id_Propr é superchave
 (Nome_Região, Lote) é superchave
 Área não é superchave
Fere a FNBC

Normalização Forma Normal de Boyce-Codd (FNBC)

 Fere a FNBC

Id_Propr	Nome_Região	Lote	Área
prop1	Central	L231	500
prop2	Central	L30	500
prop3	Sul	L140	1000
prop4	Sul	L231	1000
prop5	Leste	L616	1500
prop6	Central	L247	1500

Propriedade


Id_Propr	Lote	Área
prop1	L231	500
prop2	L30	500
prop3	L140	1000
prop4	L231	1000
prop5	L616	1500
prop6	L247	1500

Divisão

Área	Nome_Região
500	Central
1000	Sul
1500	Leste


Considerando que a informação de área estava correta!

Normalização Dependências Multivaloradas (DMV)

 Se um conjunto de atributos A restringe os valores possíveis para os atributos de um outro conjunto B , diz-se que A multidetermina B , ou que B é multidependente de A , e denota-se:

$$A \twoheadrightarrow B$$

Normalização Dependências Multivaloradas (DMV)

 Se existe $X \twoheadrightarrow Y$ em R , onde X e Y são atributos de R , então existe a seguinte restrição sobre quaisquer duas tuplas $T1$ e $T2$ de R :


Se $T1[X] = T2[X]$ então duas tuplas $T3$ e $T4$ deverão também existir com as seguintes propriedades, onde $Z = (R - (X \cup Y))$:

$$T3[X] = T4[X] = T1[X] = T2[X]$$

$$T3[Y] = T1[Y] \text{ e } T4[Y] = T2[Y]$$

$$T3[Z] = T2[Z] \text{ e } T4[Z] = T1[Z]$$

Normalização Dependências Multivaloradas (DMV)


 Exemplo:

Empregado(NomeEmpr, NomeProj, NomeDep)

NomeEmpr \twoheadrightarrow NomeProj
NomeEmpr \twoheadrightarrow NomeDep

NomeEmpr	NomeProj	NomeDep
DePaula	ProjX	João
DePaula	ProjY	Ana
DePaula	ProjX	Ana
DePaula	ProjY	João


Normalização Quarta Forma Normal (4FN)

 $A \twoheadrightarrow B$ é uma Dependência Multivalorada Trivial se:

$$B \subseteq A \text{ ou } A \cup B = R$$

- Uma relação está na 4FN quando:
 - dado um conjunto completo de dependências multivaloradas não triviais para essa relação:
 - Para todas as $A \twoheadrightarrow B$, A é uma superchave da relação.

Normalização Dependências Multivaloradas (DMV)

 Exemplo:


Empregado(NomeEmpr, NomeProj, NomeDep)

NomeEmpr \twoheadrightarrow NomeProj NomeEmpr não é superchave
NomeEmpr \twoheadrightarrow NomeDep e NomeProj é independente de NomeDep

NomeEmpr	NomeProj	NomeDep
DePaula	ProjX	João
DePaula	ProjY	Ana
DePaula	ProjX	Ana
DePaula	ProjY	João

Normalização

Dependências Multivaloradas (DMV)



- Empregado(NomeEmpr, NomeProj, NomeDep)

NomeEmpr → NomeProj NomeEmpr não é superchave
 NomeEmpr → NomeDep e NomeProj é independente de NomeDep

Fere a 4FN

Empregado

NomeEmpr	NomeProj	NomeDep
DePaula	ProjX	João
DePaula	ProjY	Ana
DePaula	ProjX	Ana
DePaula	ProjY	João


Na 4FN:

Empr_Proj

NomeEmpr	NomeProj
DePaula	ProjX
DePaula	ProjY


Empr_Dependente

NomeEmpr	NomeDep
DePaula	Ana
DePaula	João



Normalização

Considerações Finais



- Normalizar evita introduzir inconsistências quando se alteram relações
 - porém obriga a execução de custosas operações de junção para a consulta de informações.
- A decisão deve ser tomada considerando-se o compromisso entre a garantia de eliminação de inconsistências na base e a eficiência de acesso.

