

Bases de Dados

Normalização de Esquemas Relacionais

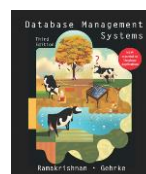
FCUL, Departamento de Informática

Ano Letivo 2015/2016

Ana Paula Afonso

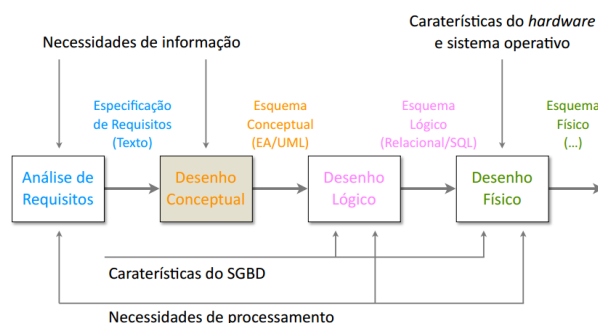
Sumário e Referências

- Sumário
 - Motivação para a normalização de esquemas relacionais
 - Problemas da redundância de dados
 - Exemplo de relação com redundância
 - Dependências funcionais
 - Etapas da normalização
 - Principais formas normais
 - Exemplo de normalização
- Referências
 - R. Ramakrishnan (**capítulo 19, secção 19.1 e 19.4**)



Motivação da Normalização

- Após a construção do modelo concetual dos dados (Modelo E/A) é feita a transformação para um modelo lógico (Esquema Relacional)
- O esquema relacional obtido representa a estrutura da informação de um modo natural e completo. Mas terá o mínimo de redundância possível?



© 2015 - Docentes SI - DI/FCUL

Fonte: António Ferreira, Guião SIBD, 2015

3

Normalização

- Dado um esquema relacional por vezes
 - Os mesmos dados existem armazenados em múltiplos locais (**redundância**)
 - Existência de **anomalias** aquando da escrita de dados (incoerências dos dados)
- **Normalização** permite melhorar a qualidade do esquema através da
 - Eliminação da redundância dos dados
 - Prevenção de anomalias
- **Normalização** é uma abordagem que envolve a
 - Decomposição sucessiva de relações até se obter um conjunto de relações
 - ... sem redundâncias e que permitam inserções, atualizações e remoções sem incoerências

© 2015 - Docentes SI - DI/FCUL

4

Exemplo de Relação não Normalizada

Nota

<u>nº estudante</u>	nome estudante	curso	nº disciplina	nome disciplina	cod professor	nome professor	grau professor	nota
21934	Antunes	Informática	04	Álgebra	21	Gil Algébrico	PA	15
			14	Análise Sist.	87	Ana Listada	PC	12
			23	P.Linear	43	Plínio	AS	16
42346	Bernardo	Matemática	08	Topologia	32	Topo Lógico	AE	10
			04	Álgebra	21	Gil Algébrico	PA	12
			12	Geometria	21	Gil Algébrico	PA	18
			16	Lógica	32	Topo Lógico	AE	13
54323	Correia	Estatística	04	Álgebra	21	Gil Algébrico	PA	11
			08	Topologia	32	Topo Lógico	AE	10
...

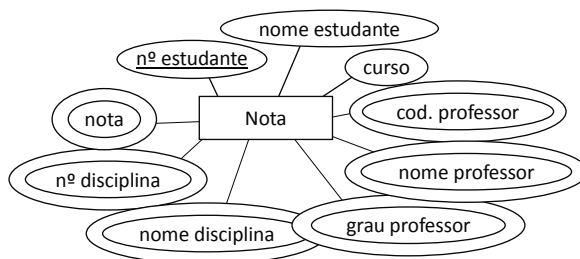
Anomalias de Relação não Normalizada

<u>nº estudante</u>	nome estudante	curso	nº disciplina	nome disciplina	cod professor	nome professor	grau professor	nota
21934	Antunes	Informática	04	Álgebra	21	Gil Algébrico	PA	15
			14	Análise Sist.	87	Ana Listada	PC	12
			23	P.Linear	43	Plínio	AS	16
42346	Bernardo	Matemática	08	Topologia	32	Topo Lógico	AE	10
			04	Álgebra	21	Gil Algébrico	PA	12
			12	Geometria	21	Gil Algébrico	PA	18
			16	Lógica	32	Topo Lógico	AE	13
54323	Correia	Estatística	04	Álgebra	21	Gil Algébrico	PA	11
			08	Topologia	32	Topo Lógico	AE	10

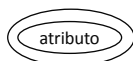
- **Inserção** de um novo professor requer indicação de outros dados
- **Atualização** do grau do professor tem de afetar várias linhas
- **Remoção** de um professor elimina os dados da disciplina a que está afeto

Exemplo de Relação não Normalizada

- Possível EA para Notas



Notação: atributo com vários valores



© 2015 - Docentes SI - DI/FCUL

7

Problemas da Redundância dos Dados

- A redundância introduz problemas (**anomalias**) de coerência e manutenção
 - **Anomalia de inserção** - informação que é independente não pode ser inserida de forma separada na Base de Dados
 - **Anomalia de remoção** - a remoção de informação acarreta a perda de outra informação independente contida na Base de Dados
 - **Anomalia de atualização** - a modificação de informação num conjunto de ocorrências implica a criação de inconsistências ou a necessidade de alterar informação noutras instâncias da Base de Dados que são independentes das primeiras

© 2015 - Docentes SI - DI/FCUL

8

Conceitos associados à Normalização

- Dependência funcional (DF) de $X \rightarrow Y$ de uma relação R
 - X e Y são conjuntos de colunas (atributos) de R
 - Lê-se: qualquer valor de X determina univocamente o valor de Y
Para quaisquer dois tuplos t1 e t2 de R se:
 $t1.X = t2.X \Rightarrow t1.Y = t2.Y$
 - Lado esquerdo da DF designa-se por **determinante**
 - Exemplos: $n_aluno \rightarrow nome_aluno$; $n_aluno \rightarrow sigla$
 $n_aluno \rightarrow nome_aluno, sigla$
- Formas normais (FN)
 - 1FN, 2FN, 3FN, FNBC, outras (não estudadas em BD)
 - Quanto maior a FN menor a redundância nos dados

© 2015 - Docentes SI - DI/FCUL

9

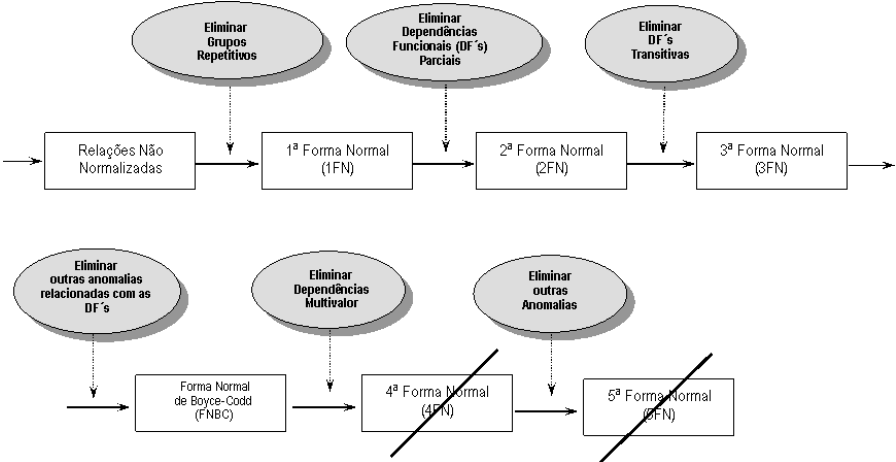
Principais Formas Normais

- **Primeira** forma normal: 1FN
 - Colunas da relação guardam apenas **um** valor por linha
 - Tipicamente verificado em BD relacionais
- **Segunda** forma normal: 2FN = 1FN +
 - Colunas não pertencentes às chaves candidatas da relação...
 - ...dependem da **totalidade** das colunas de cada chave
 - Trivial, se chaves da relação tiverem apenas uma coluna
- **Terceira** forma normal: 3FN = 2FN +
 - Colunas não pertencentes às chaves candidatas...
 - ...dependem **apenas** das chaves candidatas
- Forma normal de **Boyce-Codd**: FNBC = 3FN +
 - Todos os determinantes são chaves candidatas da relação

© 2015 - Docentes SI - DI/FCUL

10

Etapas da Normalização



1ª Forma Normal (1FN)

<u>nº</u> estudante	nome estudante	curso	nº disciplina	nome disciplina	cod professor	nome professor	grau professor	nota
21934	Antunes	Informática	04	Álgebra	21	Gil Algébrico	PA	15
42346	Bernardo	Matemática	14	Análise Sist.	87	Ana Listada	PC	12
54323	Correia	Estatística	23	P.Linear	43	Plínio	AS	16
...	08	Topologia	32	Topo Lógico	AE	10
...	04	Álgebra	21	Gil Algébrico	PA	12
...	12	Geometria	21	Gil Algébrico	PA	18
...	16	Lógica	32	Topo Lógico	AE	13
...	04	Álgebra	21	Gil Algébrico	PA	11
...	08	Topologia	32	Topo Lógico	AE	10
...

Anomalias com a 1ª Forma Normal

- **Inserção**
 - Não se pode inserir informação sobre uma nova disciplina ...
 - enquanto não existirem alunos inscritos para a nova disciplina (o atributo N.º Estudante faz parte da chave da relação)
- **Remoção**
 - Quando se apaga a informação sobre todos os alunos que têm uma determinada disciplina
 - perde-se toda a informação dessa disciplina e do respetivo professor
- **Atualização**
 - Quando se modifica o nome de uma disciplina ...
 - ... é necessário percorrer toda a relação e fazer essa modificação para todos os alunos que tivessem essa disciplina
 - No caso de falhar a aplicação de modificação em alguma ocorrência, então ter-se-iam dados inconsistentes

© 2015 - Docentes SI - DI/FCUL

13

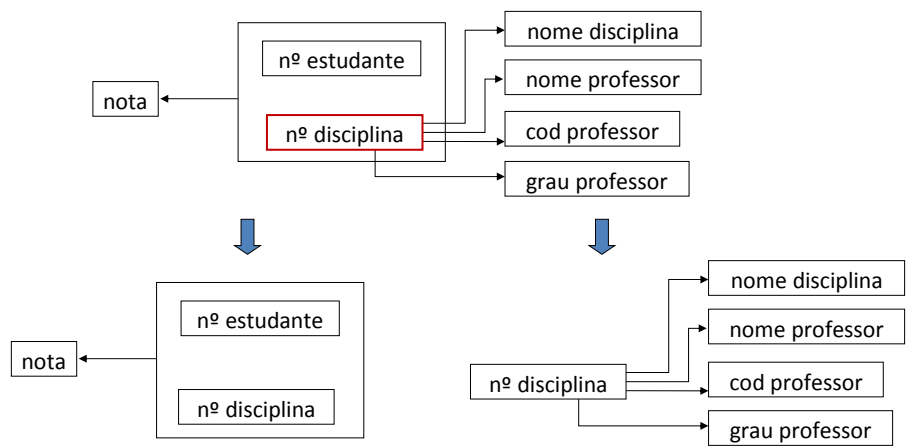
Principais Formas Normais

- **Primeira forma normal: 1FN**
 - Colunas da relação guardam apenas **um** valor por linha
 - Tipicamente verificado em bases de dados relacionais
- **Segunda forma normal: 2FN = 1FN +**
 - Se **não existirem dependências** funcionais entre **subconjuntos próprios da chave e atributos não chave**
 - Colunas não pertencentes às chaves candidatas da relação...
 - ...dependem da **totalidade** das colunas de cada chave
 - As relações com chaves simples (1 único atributo) que estejam na 1FN estão sempre na 2FN

© 2015 - Docentes SI - DI/FCUL

14

Passagem (Decomposição) de Relações para 2FN



© 2015 - Docentes SI - DI/FCUL

15

Decomposição na 2FN

<u>nº estudante</u>	<u>nº disciplina</u>	nome disciplina	cod professor	nome professor	grau professor	nota
---------------------	----------------------	-----------------	---------------	----------------	----------------	------

Nota

<u>nº estudante</u>	<u>nº disciplina</u>	nota
21934	04	15
21934	14	12
21934	23	16
42346	08	10
42346	04	12
42346	12	18
42346	16	13
54323	04	11
54323	08	10
...

Disciplina

<u>nº disciplina</u>	nome disciplina	cod professor	nome professor	grau professor
04	Álgebra	21	Gil Algébrico	PA
14	Análise Sist.	87	Ana Listada	PC
23	P.Linear	43	Plínio	AS
08	Topologia	32	Topo Lógico	AE
12	Geometria	21	Gil Algébrico	PA
16	Lógica	32	Topo Lógico	AE
...

© 2015 - Docentes SI - DI/FCUL

16

Decomposição na 2FN

- O esquema atual tem atualmente 3 tabelas
 - Estudante, Disciplina e Nota. Inicialmente era apenas 1 tabela
 - As 3 tabelas estão na 2FN ...
 - ... colunas não chave dependem da totalidade da chave
- Contudo, ainda existem anomalias na tabela **Disciplina**
 - **Inserção** de um novo professor exige que lhe seja distribuída pelo menos uma disciplina (a chave é N^o Disciplina)
 - **Remoção** de uma disciplina elimina a informação do professor se esse professor não dá aulas a outras disciplinas, então perde-se a sua informação
 - **Actualização** do grau de um professor exige efectuar a alteração em todas as disciplinas que esse professor dá

© 2015 - Docentes SI - DI/FCUL

17

Principais Formas Normais

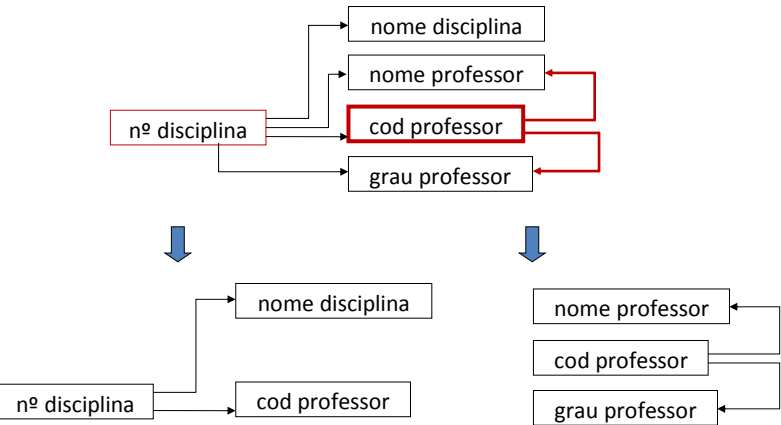
- **Primeira** forma normal: 1FN
 - Colunas da relação guardam apenas **um** valor por linha
 - Tipicamente verificado em bases de dados relacionais
- **Segunda** forma normal: 2FN = 1FN +
 - Colunas não pertencentes às chaves candidatas da relação...
 - ...dependem da **totalidade** das colunas de cada chave
 - Trivial, se chaves da relação tiverem apenas uma coluna
- **Terceira** forma normal: 3FN = 2FN +
 - Colunas não pertencentes às chaves candidatas...
 - ...dependem **apenas** das chaves candidatas
 - As relações com um único atributo não chave que estejam na 2FN estão sempre na 3FN

© 2015 - Docentes SI - DI/FCUL

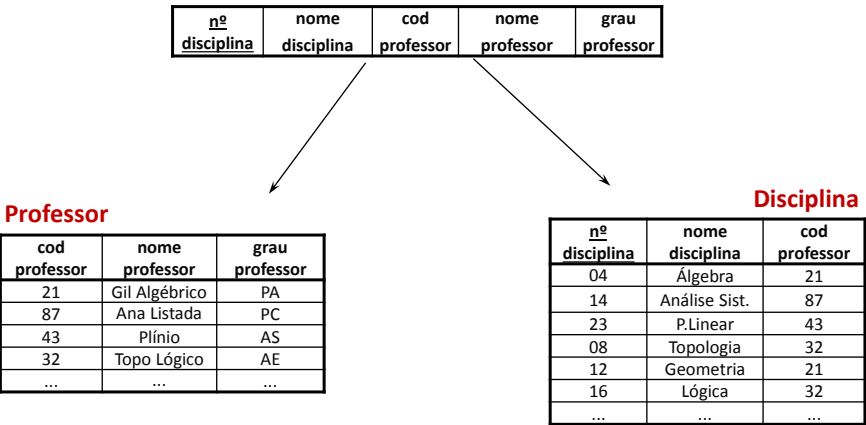
18

Decomposição na 3FN

Nºdisciplina: chave candidata da tabela Disciplina



Decomposição na 3FN



Decomposição na 3FN

- Esquema atual tem 4 tabelas
 - Estudante (número aluno, nome aluno, sigla curso)
 - Nota (número aluno, número disciplina, nota)
 - Disciplina (número disciplina, nome disciplina, número professor)
 - Professor (número professor, nome professor, grau professor)
- Esquema relacional está na 3FN
 - Para cada relação, não existem dependências entre colunas não chave
 - Todas as dependências são das chaves candidatas para as outras colunas
- Também está na forma normal de Boyce-Codd
 - Para cada relação, todos os determinantes são chaves candidatas
- Esquema normalizado é que seria traduzido para comandos SQL

© 2015 - Docentes SI - DI/FCUL

21

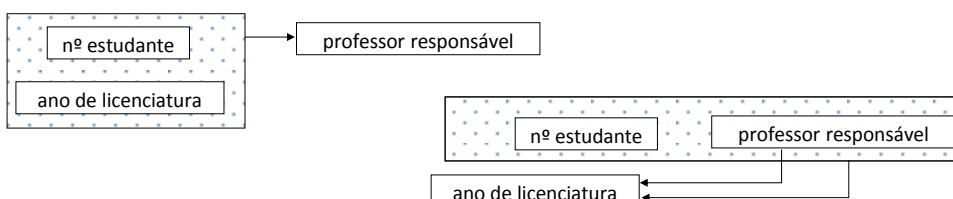
Anomalias com a 3ª Forma Normal

Novo problema

- Um estudante pode frequentar vários anos simultaneamente
- A cada estudante está afeto um tutor em cada ano que frequenta
- A cada ano estão afetos vários professores como tutor
- Cada professor é tutor apenas de um ano
- Cada professor pode ser tutor de vários estudantes num ano

Chaves candidatas: {nº estudante, ano_de_licenciatura} {nºestudante, professor_responsável}

Tutores (nºestudante, ano de licenciatura, professor_responsável)

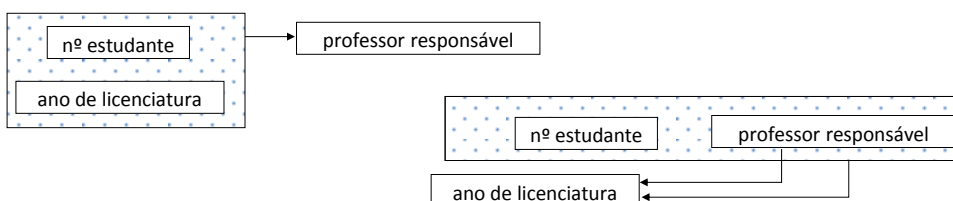


© 2015 - Docentes SI - DI/FCUL

22

Anomalias com a 3ª Forma Normal

- **A Inserção** da informação de que um professor é tutor de um ano não pode ser efetuada enquanto não for afeto a pelo menos um aluno
- **A Remoção** (anular a inscrição) da inscrição de um aluno num ano, implica a perda da informação que esse professor é tutor nesse ano, caso o seu tutor não tiver na altura mais alunos desse ano
- **A Atualização** de ano de licenciatura de um aluno para outro ano, pode implicar a perda da informação de que esse professor é tutor do primeiro ano (caso análogo à remoção).



© 2015 - Docentes SI - DI/FCUL

23

Principais Formas Normais

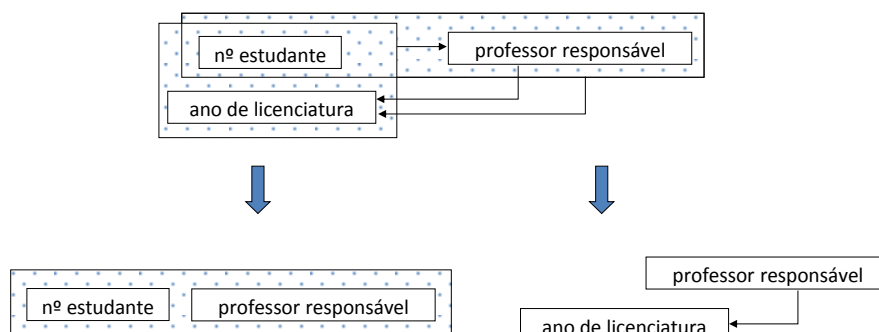
- **Primeira** forma normal: 1FN
 - Colunas da relação guardam apenas **um** valor por linha
 - Tipicamente verificado em BD relacionais
- **Segunda** forma normal: 2FN = 1FN +
 - Colunas não pertencentes às chaves candidatas da relação...
 - ...dependem da **totalidade** das colunas de cada chave
 - Trivial, se chaves da relação tiverem apenas uma coluna
- **Terceira** forma normal: 3FN = 2FN +
 - Colunas não pertencentes às chaves candidatas...
 - ...dependem **apenas** das chaves candidatas
- Forma normal de **Boyce-Codd**: FNBC = 3FN +
 - Todos os determinantes são chaves candidatas da relação

© 2015 - Docentes SI - DI/FCUL

24

Decomposição na BCNF

Na dependência funcional $\text{professor_responsavel} \rightarrow \text{ano de licenciatura}$ da relação Tutores, o determinante $\text{professor_responsavel}$ não é chave candidata



© 2015 - Docentes SI - DI/FCUL

25

FN de Boyce Codd (BCNF)

- A FNBC só é distinta da 3FN quando
 - tem mais do que uma chave candidata, tal que
 - duas chaves candidatas têm mais do que um atributo, e
 - existem atributos comuns nessas chaves

© 2015 - Docentes SI - DI/FCUL

26

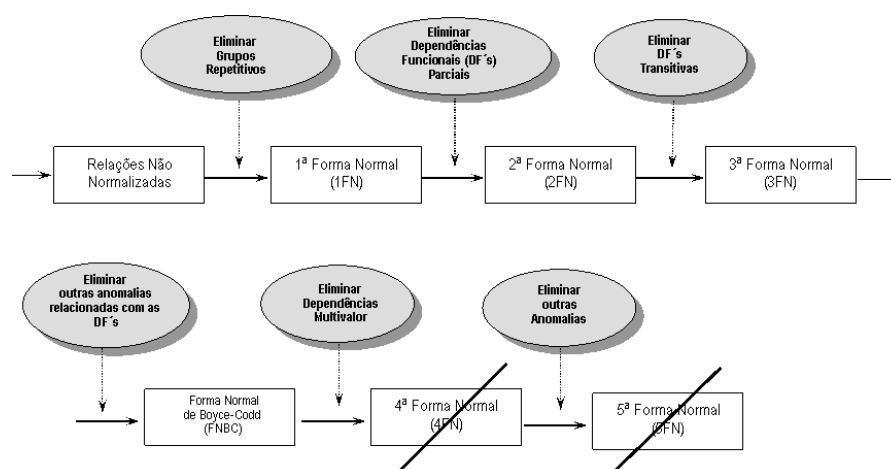
Principais Formas Normais

- **Primeira** forma normal: 1FN
 - Colunas da relação guardam apenas **um** valor por linha
 - Tipicamente verificado em BD relacionais
- **Segunda** forma normal: 2FN = 1FN +
 - Colunas não pertencentes às chaves candidatas da relação...
 - ...dependem da **totalidade** das colunas de cada chave
 - Trivial, se chaves da relação tiverem apenas uma coluna
- **Terceira** forma normal: 3FN = 2FN +
 - Colunas não pertencentes às chaves candidatas...
 - ...dependem **apenas** das chaves candidatas
- Forma normal de **Boyce-Codd**: FNBC = 3FN +
 - Todos os determinantes são chaves candidatas da relação

© 2015 - Docentes SI - DI/FCUL

27

Etapas da Normalização



© 2015 - Docentes SI - DI/FCUL

28