

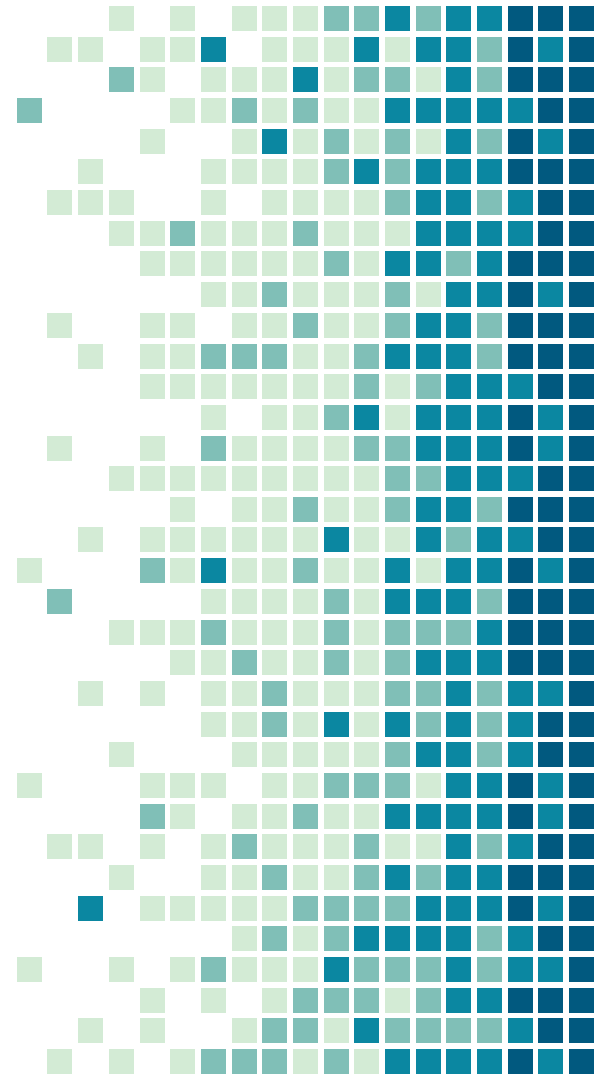
Banco de Dados I

2018.2

Rebeca Barros



Normalização de Dados



Diretrizes Informais

- Alguns fatores que podem ser usados para determinar a qualidade um projeto relacional:
 - Garantir uma semântica clara aos atributos;
 - Reduzir a informação **redundante** nas tuplas (linhas)
 - Reduzir os valores NULL;
 - Reprovar a possibilidade de gerar tuplas falsas.

Tipos de Anomalias

<u>Cpf</u>	Fnome	Datanasc	Endereco	Dnumero	Dnome	cpf gerente
111	João Silva	09-01-1965	Rua das Flores, 751 – São Paulo	5	Pesquisa	222
222	Fernando Wong	08-12-1955	Rua da Lapa – São Paulo	5	Pesquisa	222
333	Alice Souza	19-01-1968	Av. Arthur de Lima, Curitiba	5	Pesquisa	222
444	Ronaldo Lima	15-09-1962	Rua Rebouças – Vitória	4	Administração	333
555	Joice Almeida	31-07-1972	Rua do Horto, São Paulo	4	Administração	333
666	Jorge Brito	10-11-1937	Rua Souza Lima, Salvador	1	Matriz	666

Func_Dep (cpf, Fnome, Datanasc, Endereco, Dnumero, Dnome, cpf_gerente),

Anomalia de Inserção

<u>Cpf</u>	Fnome	Datanasc	Endereco	Dnumero	Dnome	cpf gerente
111	João Silva	09-01-1965	Rua das Flores, 751 – São Paulo	5	Pesquisa	222
222	Fernando Wong	08-12-1955	Rua da Lapa – São Paulo	5	Pesquisa	222
333	Alice Souza	19-01-1968	Av. Arthur de Lima, Curitiba	5	Pesquisa	222
444	Ronaldo Lima	15-09-1962	Rua Rebouças – Vitória	4	Administração	333
555	Joice Almeida	31-07-1972	Rua do Horto, São Paulo	4	Administração	333
666	Jorge Brito	10-11-1937	Rua Souza Lima, Salvador	1	Matriz	666

Não ser capaz de armazenar alguma informação a não ser que outra, não relacionada, seja armazenada também.

Anomalia de Exclusão

<u>Cpf</u>	Fnome	Datanasc	Endereco	Dnumero	Dnome	cpf gerente
111	João Silva	09-01-1965	Rua das Flores, 751 – São Paulo	5	Pesquisa	222
222	Fernando Wong	08-12-1955	Rua da Lapa – São Paulo	5	Pesquisa	222
333	Alice Souza	19-01-1968	Av. Arthur de Lima, Curitiba	5	Pesquisa	222
444	Ronaldo Lima	15-09-1962	Rua Rebouças – Vitória	4	Administração	333
555	Joice Almeida	31-07-1972	Rua do Horto, São Paulo	4	Administração	333
666	Jorge Brito	10-11-1937	Rua Souza Lima, Salvador	4	Matriz	666

Não ser capaz de deletar alguma informação sem perder outra, não relacionada, também.

Anomalia de Atualização

<u>Cpf</u>	Fnome	Datanasc	Endereco	Dnumero	Dnome	cpf gerente
111	João Silva	09-01-1965	Rua das Flores, 751 – São Paulo	5	Pesquisa	222
222	Fernando Wong	08-12-1955	Rua da Lapa – São Paulo	5	Pesquisa	222
333	Alice Souza	19-01-1968	Av. Arthur de Lima, Curitiba	5	Pesquisa	222
444	Ronaldo Lima	15-09-1962	Rua Rebouças – Vitória	4	Administração	333
555	Joice Almeida	31-07-1972	Rua do Horto, São Paulo	4	Administração	333
666	Jorge Brito	10-11-1937	Rua Souza Lima, Salvador	1	Matriz	666

Se uma cópia de algum desses dados repetidos é atualizada, uma inconsistência é criada a não ser que todas as cópias sejam atualizadas também.

Dependência Funcional

- O conceito isolado mais importante na teoria de projeto de esquema relacional é o de **dependência funcional (DF)**.
- Uma dependência funcional indicada por $X \rightarrow Y$, entre dois conjuntos de atributos X e Y que são subconjuntos da relação (tabela) R , especifica uma restrição sobre possíveis tuplas (linhas) que podem formar um estado de relação r de R . A restrição é que, para quaisquer tuplas t_1 e t_2 em r que tenham $t_1[X] = t_2[X]$, elas também devem ter $t_1[Y] = t_2[Y]$.

Dependência Funcional

- Essencialmente isso diz que se duas linhas concordam nos valores dos atributos X , eles também devem concordar nos valores nos atributos Y .

A	B	C	D
a1	b1	c1	d1
a1	b1	c1	d2
a1	b2	c2	d1
a2	b1	c3	d1

$(A,B) \rightarrow C$

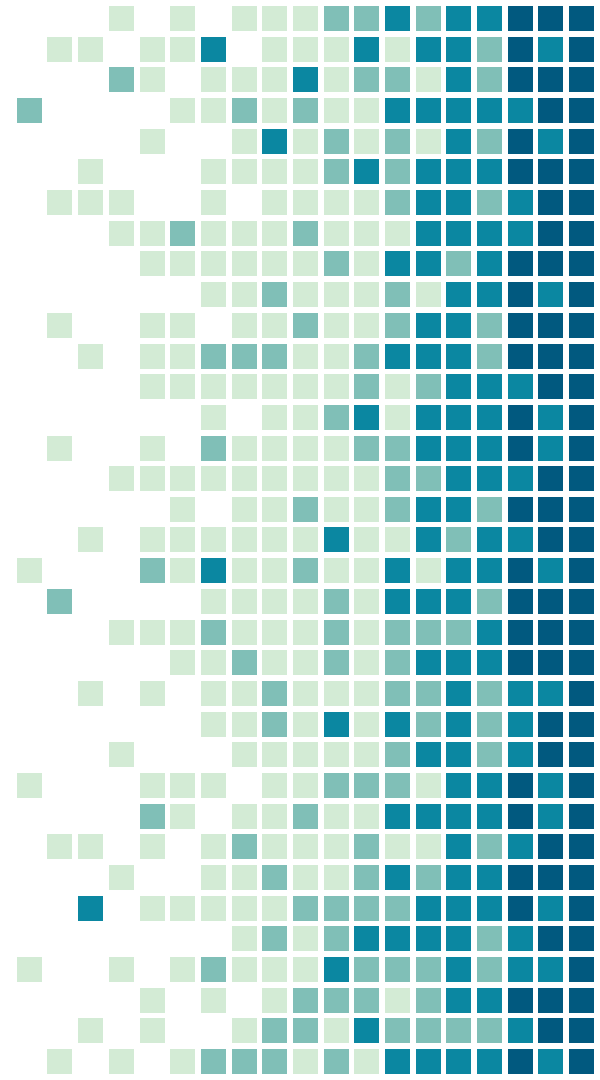
Dependência Funcional

<u>Cpf</u>	<u>Projnumero</u>	Horas	Fnome	Projnome	ProjLocal
111	1	32,5	João Silva	ProdutoX	Santo André
111	2	7,5	João Silva	ProdutoY	Itu
222	3	40	Ronaldo Lima	ProdutoZ	São Paulo
333	1	20	Joice Leite	ProdutoX	Santo André
333	2	10	Joice Leite	ProdutoY	Itu
444	3	10	Fernando Wong	ProdutoZ	São Paulo
444	10	10	Fernando Wong	Informatização	Mauá

FUNC_PROJ (cpf, projnumero, horas, Fnome, Projnome, Projlocal)

Normalização de Dados

Formas Normais baseadas em chave
primária (1FN à 3FN)

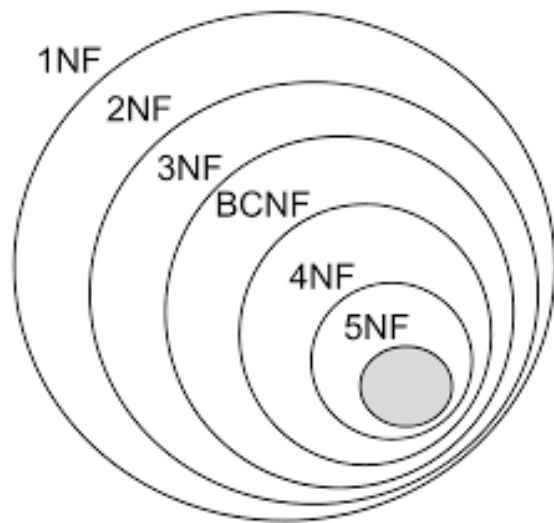


Normalização

- Considerando que um conjunto de DF's é dado para cada tabela de um banco de dados e que cada tabela tem uma chave primária, essa informação combinada com testes (condições) para formas normais é o que chamamos de *processo de normalização*.
- A **normalização de dados** pode ser considerada um processo de analisar os esquemas de relação com base em suas DFs e chaves primárias para conseguir as propriedades desejadas de (1) minimizar a redundância, (2) minimizar as anomalias de inserções, exclusão e atualização.

Formas Normais

- A **forma normal** de uma tabela refere-se a condição de forma normal mais alta que ela atende, indicando portando o grau ao qual ela foi normalizada.



1ª Forma Normal (1FN)

- Uma tabela encontra-se na **1FN** quando possui apenas atributos atômicos (simples, indivisíveis), ou seja, não pode ter atributos multivalorados, compostos ou atributos compostos multivalorados.

1ª Forma Normal (1FN)

- Considere o seguinte esquema supondo que nesse minimundo cada departamento pode ter *certo número de locais*.

<u>Dnumero</u>	Dnome	cpf gerente	Dlocal
5	Pesquisa	222	Santo Andre, Itu, São Paulo
4	Administração	333	Mauá
1	Matriz	666	São Paulo

1ª Forma Normal (1FN)

- Passos:
 - Remover os atributos compostos/multivalorados da tabela inicial e criar uma nova tabela referente aos mesmos;
 - A chave primária dessa nova tabela passa a ser composta pela chave primária da tabela inicial + um atributo super-chave que identifique o atributo composto/multivalorado.

1ª Forma Normal (1FN)

- **Antes**
 - Dept (Dnumero, Dnome, cpf-gerente, Dlocal)
- **Depois – na 1FN**
 - Dept(Dnmero, Dnome, cpf-gerente)
 - DeptLocal (Dnumero, Dlocal)

<u>Dnumero</u>	Dnome	cpf gerente
5	Pesquisa	222
4	Administração	333
1	Matriz	666

<u>Dnumero</u>	<u>Dlocal</u>
5	Santo Andre
5	Itu
5	São Paulo
4	Mauá
1	São Paulo

1ª Forma Normal (1FN)

- FUNC-PROJ (cpf, Fnome, {Projs (Projnumero, Horas)})

<u>cpf</u>	Fnome	Projnumero	Horas
111	João Silva	1 2	32,5 7,5
222	Ronaldo Lima	3	40
333	Joice Leite	1 2	20 20

1ª Forma Normal (1FN)

- **Antes**
 - Func-Proj (cpf, Fnome, {Projs (Projnumero, Horas)})
- **Depois – na 1FN**
 - Func-Proj1 (cpf, Fnome)
 - Func-Proj2 (cpf, Projnumero, Horas)

2ª Forma Normal (2FN)

- Uma tabela encontra-se na **2FN** quando está na 1FN e não possui **dependência funcional parcial**. Ou seja, todos os atributos da tabela devem ser dependente totais da chave primária.
- Tabelas com chave primária simples já encontram-se automaticamente na 2FN.

2ª Forma Normal (2FN)



<u>Cpf</u>	<u>Projnumero</u>	Horas	Fnome	Projnome	ProjLocal
111	1	32,5	João Silva	ProdutoX	Santo André
111	2	7,5	João Silva	ProdutoY	Itu
222	3	40	Ronaldo Lima	ProdutoZ	São Paulo
333	1	20	Joice Leite	ProdutoX	Santo André
333	2	10	Joice Leite	ProdutoY	Itu
444	3	10	Fernando Wong	ProdutoZ	São Paulo
444	10	10	Fernando Wong	Informatização	Mauá

2ª Forma Normal (2FN)

- Passos:
 - Mantenha na tabela inicial apenas os atributos que dependem de toda a chave primária;
 - Criar uma nova tabela para cada dependência funcional parcial. Nessa nova tabela a chave primária será aquela parcial da qual alguns atributos são dependentes.

2ª Forma Normal (2FN)

- **Antes**

- Func-Proj (cpf, projnumero, horas, Fnome, Projnome, Projlocal)

- **Depois – na 2FN**

- Func-Proj (cpf, Projnumero, Horas)
- Func(cpf, Fnome)
- Proj (Projnumero, Projnome, ProjLocal)

3ª Forma Normal (3FN)

- Uma tabela encontra-se na **3FN** quando está na 2FN e não possui **dependência transitiva**. Ou seja, nenhum atributo não-chave depende de outro não-chave.

3ª Forma Normal (3FN)

<u>Cpf</u>	Fnome	Datanasc	Endereco	Dnumero	Dnome	cpf gerente
111	João Silva	09-01-1965	Rua das Flores, 751 – São Paulo	5	Pesquisa	222
222	Fernando Wong	08-12-1955	Rua da Lapa – São Paulo	5	Pesquisa	222
333	Alice Souza	19-01-1968	Av. Arthur de Lima, Curitiba	5	Pesquisa	222
444	Ronaldo Lima	15-09-1962	Rua Rebouças – Vitória	4	Administração	333
555	Joice Almeida	31-07-1972	Rua do Horto, São Paulo	4	Administração	333
666	Jorge Brito	10-11-1937	Rua Souza Lima, Salvador	1	Matriz	666

3ª Forma Normal (3FN)

- Passos:
 - Criar uma nova tabela para cada dependência transitiva. Nessa nova tabela a chave primária será o atributo não-chave do qual alguns atributos são dependentes.

3ª Forma Normal (3FN)

- **Antes**

- Func-Dep (cpf, Fnome, Datanasc, Endereco, Dnumero, Dnome, cpf_gerente)

- **Depois – na 3FN**

- Func-Dep (cpf, Fnome, Datanasc, Endereco, Dnumero)
- Dept (Dnumero, Dnome, cpf-gerente)

Exercícios



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

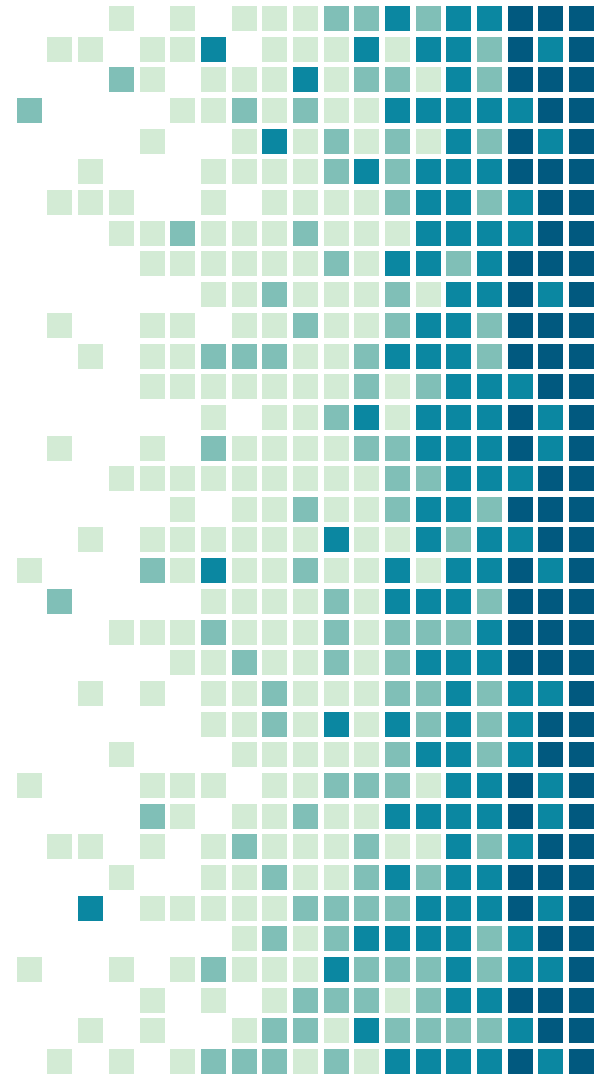
DESCRIÇÃO: Sistema de RH

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

Figura 1: Exemplo de documento a ser normalizado

Normalização de Dados

Formas Normais superiores



Forma Normal Boyce-Codd (FNBC)

- Uma tabela encontra-se na **FNBC** quando está na 3FN e se toda vez que uma dependência funcional $X \rightarrow A$ se mantiver na tabela, então X é uma superchave. Ou, simplificando, quando não existem dependências funcionais onde uma chave candidata depende de uma não-chave.
- Toda tabela que está na FNBC também está na 3FN porém o contrário não é verdade.

Forma Normal Boyce-Codd (FNBC)

<u>Aluno</u>	<u>Disciplina</u>	Professor
Lima	Banco de Dados	Marcos
Silva	Banco de Dados	Navathe
Silva	Sistemas operacionais	Omar
Silva	Teoria	Charles
Souza	Banco de Dados	Marcos
Souza	Sistemas Operacionais	Antonio
Wong	Banco de Dados	Gomes
Zelaya	Banco de Dados	Navathe
Lima	Sistemas Operacionais	Omar

(Aluno, Disciplina) -> Professor

Professor -> Disciplina*

*Nesse cenário existe a restrição de que um professor só ensina uma disciplina

Forma Normal Boyce-Codd (FNBC)

- Passos:
 - O atributo determinado por um campo que não é chave candidata vai para outra tabela.

Forma Normal Boyce-Codd (FNBC)

- **Antes**

- Ensina (Aluno, Disciplina, Professor)

- **Depois – na FNBC**

- ProfessorDisciplina (Professor, Disciplina)
- ProfessorAluno (Professor, Aluno)

4ª Forma Normal (4FN)

- Uma tabela encontra-se na **4FN** quando está na 3FN e não possui duas ou mais **dependência multivalorada (DMV)** independentes.
- Chama-se de **dependência multivalorada** quando um atributo **não chave** recebe valores múltiplos para um mesmo valor de chave. Ou seja, quando o valor de um atributo determina um conjunto de valores de outro atributo.
- $X \twoheadrightarrow Y$

4ª Forma Normal (4FN)

<u>Fnome</u>	<u>Projnome</u>	<u>Dnome</u>
Silva	X	João
Silva	Y	Ana
Silva	X	Ana
Silva	Y	João

Fnome ->> Projnome
Fnome ->> Dnome

4ª Forma Normal (4FN)

- Passos:
 - Decompor em tabelas distintas onde cada dependência multivalorada independente vire uma dependência multivalorada trivial.

4ª Forma Normal (4FN)

- **Antes**
 - Func (Fnome, Projnome, Dnome)
- **Depois – na 4FN**
 - FuncProj (Fnome, Projnome)
 - FuncDep (Fnome, Dnome)

<u>Fnome</u>	<u>Projnome</u>
Silva	X
Silva	Y

<u>Fnome</u>	<u>Dnome</u>
Silva	João
Silva	Ana