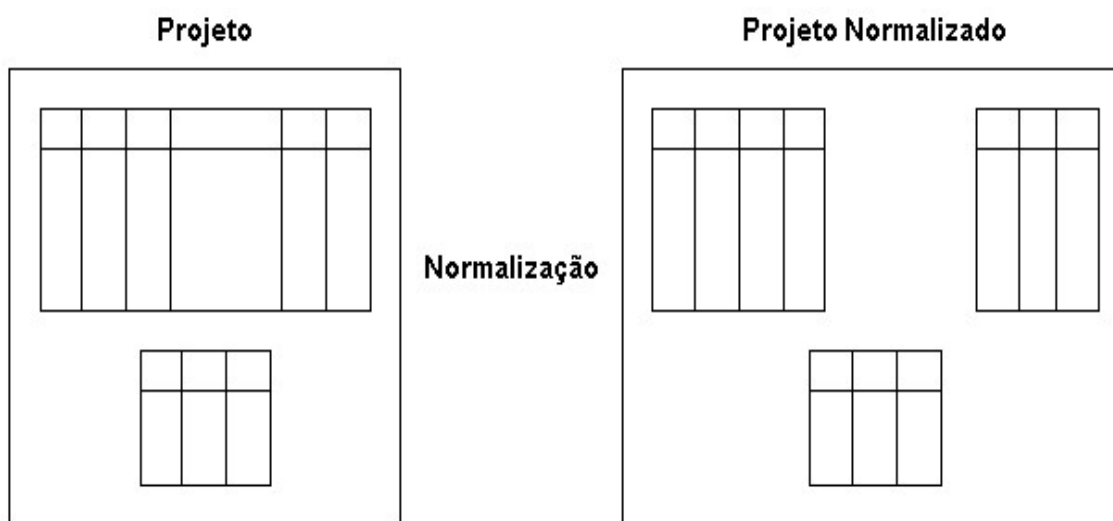


## BANCO DE DADOS - NORMALIZAÇÃO

Profs. Raimundo Cláudio Vasconcelos

### Normalização:

Técnica para simplificar um projeto de **Bando de Dados Relacional**, usada para eliminar redundâncias, evitar problemas nas **atualizações** e implementar estruturas mais estáveis.



Se o projeto do Banco de Dados resultado do **MER** é bom, ou seja, não misturou informações, não possui entidades nem **relacionamentos inconvenientes**, as relações (tabelas) representantes das entidades e relacionamentos estão naturalmente **normalizadas**.

Para verificar se um projeto de BD relacional está normalizado ou se existe um conjunto de tabelas mais estável para representá-lo, basta aplicar um conjunto de regras de Forma Normal.

Para mostrar a aplicação das Regras de Forma Normal será utilizado o seguinte exemplo:

### NOTA FISCAL

Nº: 123		Data : 15/03/2013		
Destinatário: Comércio de Parafusos S.A.				
Endereço: Rua Blah, 100 cidade: Avanhandava – UF: SP				
CNPJ Dest.: 123456789-01				
Código	Descrição	Preço	Qtde	Subtotal
1	martelo	3.00	2	6.00
121	alicate	4.00	3	12.00
131	pregos	0.10	20	2.00
Total				20.00

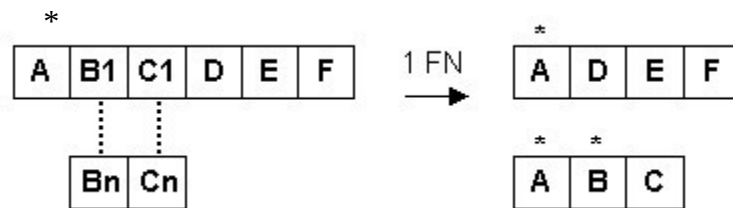
A forma mais simples de representar os dados de uma Nota Fiscal (NF) na forma de tabelas é criar uma única tabela contendo todos estes dados, embora está provavelmente seja uma das piores representações:

**NotaFiscal** (Número, data, destinatário, endereço\_dest, cidade\_dest, UF\_dest, CNPJ\_dest, (código, descrição, preço, qtde, subtotal) total)

OBS: Os atributos (código, descrição, preço, qtde, subtotal) são mantidos entre parênteses para representar que os mesmos são multivalorados.

## 1ª Forma Normal (1FN)

Um projeto de Banco de Dados relacional está na 1FN se para cada tabela todas as linhas são diferentes e não contém grupos repetitivos (ou, analisando de outra maneira, não há atributos multivalorados).



\* Chave Primária

Exemplo:

**NotaFiscal** (Número, data, destinatário, endereço\_dest, cidade\_dest, UF\_dest, CNPJ\_dest, (código, descrição, preço, qtde, subtotal) total)

### 1FN

**NotaFiscal** (Número, data, destinatário, endereço\_dest, cidade\_dest, UF\_dest, CNPJ\_dest, total)

**Itens** (Número, código, descrição, preço, qtde, subtotal)

OBS: Se houver atributos multivalorados, uma nova tabela é gerada para eles levando junto a chave primária de tabela original.

## Dependência:

Um atributo B depende de outro atributo A ( $A \rightarrow B$ ) se, e somente se, para cada valor de A corresponde um único valor de B.

Exemplos:

número\_nota  $\rightarrow$  data  
data  $\nrightarrow$  número\_nota (  $\nrightarrow$  não há dependência)  
código\_produto  $\rightarrow$  descrição  
CNPJ  $\rightarrow$  destinatário

Tipos de Dependência:  $\left\{ \begin{array}{l} \text{Parcial } (A,B) \sqsubset C, \text{ e } A \sqsubset C \text{ ou } B \sqsubset C \\ \text{Total } (A,B) \sqsubset C, \text{ mas } A \not\sqsubset C \text{ e } B \not\sqsubset C \end{array} \right.$

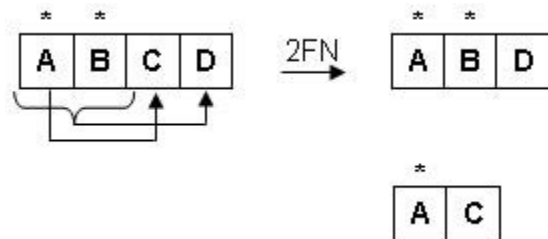
Exemplo:

TOTAL: (número\_nota, código) -> qtde , mas  
número\_nota /-> qtde e código /-> qtde

PARCIAL: (número\_nota, código) -> descrição, mas  
código -> descrição

## 2ª Forma Normal (2FN)

Um projeto de Banco de Dados relacional esta na 2FN se, e somente se, ele estiver na **1FN**, e em cada tabela não existir atributos que dependam apenas de parte da chave primária (dependência parcial).



OBS: para que haja dependência parcial com relação à chave primária, é necessário que a chave primária da tabela seja composta por dois ou mais atributos. Portanto, pode-se concluir que tabelas cuja chave primária é formada por um único atributo já estão na 2FN.

Exemplo:

**2FN**

**NotaFiscal** já está na 2FN

**Itens** (Número, código, qtde, sub-total)

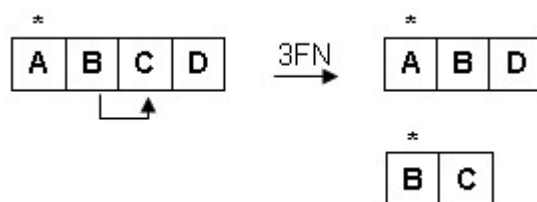
**Produtos** (Código, descrição, preço)

OBS: Para colocar uma tabela na 2FN:

- 1) A tabela deve estar na 1FN
- 2) Remover os dados da estrutura dependentes de parte da chave primária
- 3) Criar uma nova estrutura para eles contendo só a parte da chave primária e os atributos que dela dependam.

## 3ª Forma Normal (3FN)

Um projeto de Banco de Dados Relacional está na 3FN se, e somente se, estiver na **2FN** e em cada tabela não existem atributos que **dependem de** outros atributos que não sejam a chave primária (ou uma superchave).



Outra forma de avaliar se uma

relação (tabela) esta na 3FN é verificar se não há uma dependência transitiva indesejável, por exemplo:

**Relação R (A, B, C, D) não está na 3FN se:**

$A \rightarrow B, B \rightarrow A, B \rightarrow C$

Um relação R esta na 3FN se para toda dependência funcional não trivial  $X \rightarrow Y$ , válida para R, então X é superchave ou Y é atributo primo (atributo contido em alguma chave de R).

Exemplo:

### 3FN

**NotaFiscal** (Número, data, CNPJ\_dest, total)

**Destinatário** (CNPJ, nome, endereço, cidade, uf)

**Itens** (Número, código, qtde, subtotal)

**Produtos** (Código, descrição, preço)

OBS: Para colocar uma tabela na 3FN:

- 1) A tabela deve estar na 2FN
- 2) Remover os dados da estrutura dependentes dos atributos que não são parte da chave e criar uma nova estrutura para armazená-los. O atributo do qual os outros dependem deve permanecer na tabela original também.

### Forma Normal de Boyce-Codd (FNBC)

Um projeto de um BD Relacional esta na FNBC se, e somente se, estiver na 3FN e, em cada relação (tabela) R, para toda dependência funcional não trivial  $X \rightarrow Y$  válida de R, X é uma superchave de R.

Esta restrição expressa uma propriedade muito simples e intuitiva: numa relação que está na FNBC, as únicas dependências funcionais não triviais são derivadas de chaves da relação.

Excepcionalmente, não se podem ter todas as relações de uma BD na FNBC. Porém estas exceções são muito raras. Exemplo clássico:

Considere a relação Endereços (Cidade, Rua, CEP), com as seguintes dependências funcionais:

$(Cidade, Rua) \rightarrow CEP$

$CEP \rightarrow Cidade$

Estas dependências são normalmente esperadas em um sistema postal, isto é:

- o par Cidade, Rua determina o CEP, mas nenhum deles isoladamente determina o CEP
- dado um CEP determina-se a Cidade, mas não vice-versa, nem o CEP determina a RUA.

(Cidade,rua) é a chave da relação Endereços, mas esta relação não está na FNBC, pois CEP não é chave ou superchave e  $CEP \rightarrow Cidade$ . No entanto, a relação Endereços está na 3FN, pois ela não tem outras chaves e cidade é um atributo primo, assim a dependência funcional  $CEP \rightarrow Cidade$  não viola a 3FN.

### 4ª Forma Normal (4FN)

Relações que violam a 4FN possuem redundâncias indesejáveis, normalmente advindas do mapeamento na mesma relação (tabela) de dois ou mais atributos multivalorados e independentes entre si. Um exemplo simples é:

Livros(no\_tombo, título, edição, editora, ano\_publicação)  
 AutoresAssuntosLivros (no\_tombo, autor, assunto)  
 Um dado livro pode ter vários autores e tratar de vários assuntos, por exemplo:  
 AutoresAssuntosLivros

No_tombo	Autor	Assunto
.....	.....	.....
321321	Silberschatz, A.	Banco de Dados
321321	Korth, H.F.	Banco de Dados
321321	Sudarshan, S	Banco de Dados
321321	Silberschatz, A.	SQL
321321	Korth, H.F.	SQL
321321	Sudarshan, S	SQL
.....	.....	.....

A relação AutoresAssuntosLivros está na FNBC, mas existe grande redundância na representação acima um livro com k autores e m assuntos teria k\*m linhas nesta tabela e inconsistências poderiam ocorrer, por exemplo, se o assunto SQL for substituído por "Linguagem de Consulta" em apenas uma das linhas onde aparece o autor Korth.

A solução para evitar este tipo de redundância seria decompor a tabela de AutoresAssuntosLivros em duas:

AutoresLivros(no\_tombo, autor)  
 AssuntosLivros (no\_tombo, assunto)  
 AutoresLivros

No_tombo	Autor
.....	.....
321321	Silberschatz, A.
321321	Korth, H.F.
321321	Sudarshan, S
.....	.....

AssuntosLivros

No_tombo	Assunto
.....	.....
321321	Banco de Dados
321321	SQL
.....	.....

Um projeto de um BD Relacional esta na 4FN se, e somente se, estiver na FNBC e, em cada relação (tabela) R, para toda dependência multivalorada não trivial  $X \twoheadrightarrow Y$  válida de R, X é uma superchave de R.

Dependências multivaloradas explicitam a independência entre os atributos e possuem regras de inferência não óbvias. No exemplo da relação AutoresAssuntosLivros teríamos as seguintes dependências multivaloradas:

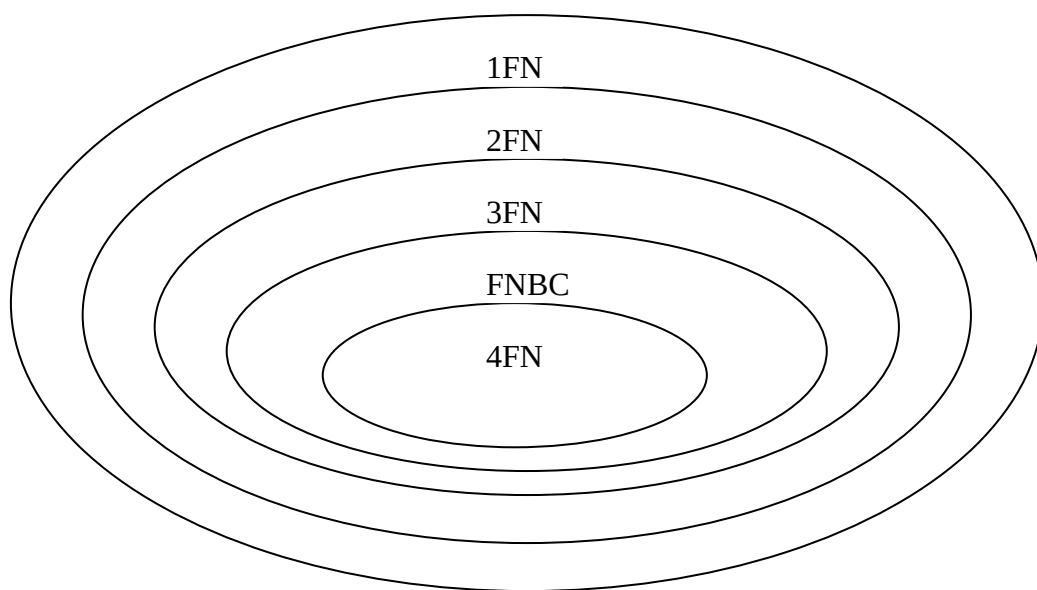
no\_tombo ↔ autor e no\_tombo ↔ assunto

Uma dependência multivalorada  $X \leftrightarrow Y$  é dita trivial quando todos os atributos de R estão na união de X com Y, e não trivial caso contrário.

Na relação AutoresAssuntosLivros temos as dependências multivaloradas não triviais no\_tombo ↔ autor e no\_tombo ↔ assunto e o no\_tombo não é superchave de R, portanto esta relação não está na 4FN.

Nas relações AutoresLivros e AssuntosLivros temos as mesmas dependências multivaloradas válidas, porém elas são triviais (envolvem todos os atributos), portanto estas relações estão na 4FN.

Famílias das Formas Normais vistas como conjuntos encaixados:



## Resumindo

Um projeto de um BD Relacional está Normalizado se, e somente se, não existirem em nenhuma tabela atributos multivalorados, e todos os atributos em cada tabela dependem da chave primária completa e somente dela.

## Engenharia Reversa:

Processo no qual a partir da implementação ou do projeto de um sistemas monta-se a modelagem do mesmo. Neste caso, pretende-se a partir do projeto de um BD relacional definir o MER que melhor o represente.

Exemplo:

**NotaFiscal** (Número, data, CNPJ\_dest, total)

**Destinatário** (CNPJ, nome, endereço, cidade, uf)

**Itens** (Número, código, qtde, subtotal)

**Produtos** (Código, descrição, preço)

