# BASES DE DADOS I LEI/2

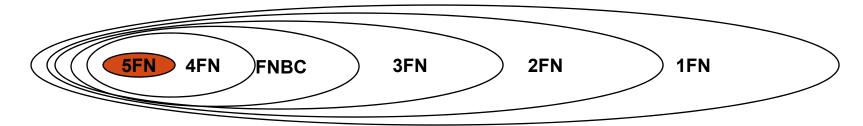
Universidade da Beira Interior, Departamento de Informática Hugo Pedro Proença, 2009/2010

Na sequência do aparecimento do modelo relacional, e uma vez que é necessário organizar os dados por forma a que possam ser tratados relaccionalmente, surge o processo designado de "Normalização".

#### Objectivo:

- Encontrar um esquema de Base de Dados <u>optimizado</u>, capaz de <u>modelar adequadamente</u> os dados relevantes a um determinado universo.
  - Adaptado às restrições do Universo.

- Inicialmente determinava apenas a desagregação dos dados em domínios atómicos.
- Posteriormente desenvolvido por Boyce-Codd dando origem a uma hierarquia de normalização: 1ª Forma Normal, 2ª Forma Normal, 3ª Forma Normal, Forma Normal de Boyce-Codd, 4ª Forma Normal e 5ª Forma Normal.
- □ Trata-se de um processo sistemático (conjunto de regras) que definem os relacionamentos entre os dados (dependências), por forma a minimizar a sua redundância.



- Do processo de normalização emergem três tipos de <u>dependências</u> entre os dados: funcionais, multivalor e de junção.
- □ Diz-se, por definição, que existe uma dependência funcional X →
  Y se uma instância de valores dos atributos de X determina ou identifica univocamente uma instância de valores dos atributos de Y.
  Pode-se dizer que "X determina Y" ou "Y depende de X".
- Exemplo:
  - BI → Nome Pessoa, Naturalidade (N° Factura, Cod\_Produto) → Data, Quantidade, Preço
- Reportando ao modelo relacional, observa-se que a chave primária de uma relação determina sempre os restantes atributos, isto é, todos eles são dependentes funcionalmente da chave.

#### Dependências Funcionais – Regras de Identificação (Separação)

#### $\square$ Se A $\rightarrow$ BC então A $\rightarrow$ B e A $\rightarrow$ C

Exemplo:

Se

 $BI \rightarrow (Nome, Endereço)$ 

Então

BI→Nome e BI →Endereço

Se com o Bl eu conseguir determinar o nome e endereço de uma pessoa, então é lógico que possa determinar cada um deles individualmente.

#### Dependências Funcionais – Regras de Identificação (Acumulação)

#### □Se A $\rightarrow$ B então AC $\rightarrow$ B

- Exemplo:
  - Se
  - BI→Nome
  - Então
  - $\blacksquare \qquad (BI, Idade) \rightarrow Nome$
  - Se com o Bl eu conseguir determinar o nome de uma pessoa, será obviamente possível fazer o mesmo a partir do conjunto de atributos Bl + Idade.

Dependências Funcionais – Regras de Identificação (Transitividade)

#### $\square$ Se A $\rightarrow$ B e B $\rightarrow$ C então A $\rightarrow$ C

**■** Exemplo:

Se

BI→CodigoPostal e CodigoPostal →Cidade

Então

BI → Cidade

Ao saber o BI posso saber o código postal da sua morada. A partir do código postal posso saber a cidade onde habita. Então, aplicando a transitividade, posso concluir que ao saber o BI automaticamente posso saber a cidade onde determinada pessoa habita. Dependências Funcionais – Regras de Identificação (Pseudo-Transitividade)

#### $\square$ Se A $\rightarrow$ B e BC $\rightarrow$ D então AC $\rightarrow$ D

■ Exemplo:

Se

Bl→CodFuncionario e (CodFuncionario, Mês)→ Salario

Então

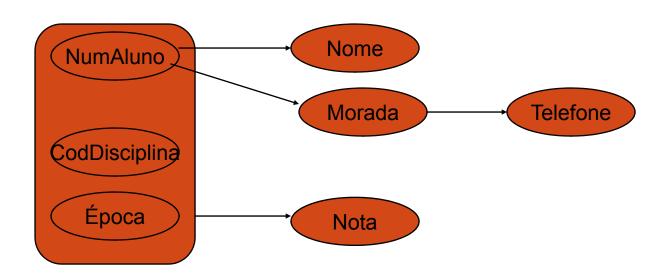
(BI, Mês) → salario

■ Se com o Bl consigo determinar o código de um trabalhador e com esse código mais um determinado mês eu consigo saber o seu vencimento, então com o Bl mais um determinado mês posso igualmente determinar o seu vencimento.

#### Diagramas de Dependências Funcionais

- São importantes para identificar claramente as dependências existentes entre os atributos de uma relação e constituir uma ajuda no processo de normalização de uma base de dados.
  - Exemplo:

Aluno\_Nota (NumAluno, Nome, Morada, Telefone, CodDisciplina, Época, Nota)



- Uma dependência funcional é um caso especial de outra mais abrangente: a dependência <u>multivalor</u>. Relativamente a esta, a sua presença apenas se prova em relações com pelo menos 3 atributos.
  - Definição: A dependência multivalor  $X \rightarrow Y$  existe numa relação R(X,Y,Z) se e só se sempre que (x,y,z) e (x,y',z') são tuplos da relação então (x,y',z) e (x,y,z') também o são.

□ <u>Corolário</u>: Se na relação R(X,Y,Z) existe uma dependência multivalor  $X \rightarrow Y$  então também existirá outra  $X \rightarrow Z$ .

#### Dependência Multivalor

Tome-se o exemplo da relação à direita: O atributo "Disciplina" indica as disciplinas em que um aluno está inscrito, e o atributo "Hobby" indica uma actividade complementar que pratica.

<u>Aluno</u>	<u>Disciplina</u>	<u>Hobby</u>
João	Matemática	Natação
João	Português	Futebol
João	Matemática	Futebol
João	Português	Natação
Ana	Inglês	Teatro
Ana	Inglês	Dança
Ana	Francês	Teatro
Ana	Francês	Dança

O atributo "Disciplina" gera grupos de valores repetidos, o mesmo acontecendo (em sentido inverso) com o atributo "Hobby". Neste caso existe uma dependência multivalor "Aluno" ->> "Disciplina" e "Aluno" ->> "Hobby", originando informação redundante e potenciando a perca de informação após a eliminação inadvertida de tuplos da relação.

#### Dependência Multivalor

- Por observação das instancias de uma relação pode-se concluir da não-existência de dependências multivalor. Contudo essa observação não é suficiente para concluir da sua existência. Para isso, terão que se conhecer as regras que governam a manipulação da relação e os domínios de cada atributo.
- Em todo o caso, parece evidente que só em situações muito específicas surgem dependências multivalor.
- (\*) Se, por exemplo, for retirada a primeira instância da relação anterior, pode-se concluir da não existência de uma dependência multivalor.

Relativamente às dependências de junção, elas são ainda mais raras e, principalmente, mais difíceis de identificar. Diz-se que numa relação existe uma dependência de <u>iunção</u> se, dado um conjunto de projecções sobre a relação, se consegue reconstruir a relação inicial através da junções de algumas delas.

А	В	С
a <sub>1</sub>	b <sub>1</sub>	C <sub>1</sub>
$a_2$	b <sub>1</sub>	C <sub>2</sub>
a <sub>1</sub>	b <sub>2</sub>	C <sub>2</sub>
a <sub>2</sub>	b <sub>3</sub>	C <sub>2</sub>
a <sub>2</sub>	b <sub>4</sub>	C <sub>2</sub>
$a_2$	b <sub>4</sub>	C <sub>4</sub>
$a_2$	<b>b</b> <sub>5</sub>	C <sub>4</sub>
<b>a</b> <sub>3</sub>	b <sub>2</sub>	C <sub>5</sub>

Projectando em (A,B) e (B,C) e (A,C) não é possível reconstruir a relação inicial a partir de qualquer par destas projecções. Apenas é possível fazê-lo através das três projecções.

- Dependências de Junção: Exemplo
- □ Tendo uma relação R com três atributos (A,B,C):
  - □ Podem-se criar 3 relações (R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub>) com atributos
    - (X,Y) tal que X, Y **&** (A,B,C):

Α	В	С
a <sub>1</sub>	b <sub>1</sub>	C <sub>2</sub>
a <sub>1</sub>	b <sub>2</sub>	C <sub>1</sub>
$a_2$	b <sub>1</sub>	C <sub>1</sub>
a <sub>1</sub>	b <sub>1</sub>	C <sub>1</sub>

R1

Α	В
a <sub>1</sub>	b <sub>1</sub>
a <sub>1</sub>	b <sub>2</sub>
a <sub>2</sub>	b <sub>1</sub>

R2

_	-
Α	С
<b>a</b> <sub>1</sub>	C <sub>2</sub>
<b>a</b> <sub>1</sub>	C <sub>1</sub>
a <sub>2</sub>	C <sub>1</sub>

R3

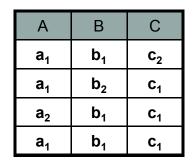
В	С
b <sub>1</sub>	C <sub>2</sub>
b <sub>2</sub>	C <sub>1</sub>
b <sub>1</sub>	C <sub>1</sub>

#### Dependências de Junção: Exemplo



Α	В	С
<b>a</b> <sub>1</sub>	b <sub>1</sub>	C <sub>2</sub>
a <sub>1</sub>	b <sub>1</sub>	C <sub>1</sub>
a <sub>1</sub>	b <sub>2</sub>	C <sub>1</sub>
$a_2$	b <sub>1</sub>	C <sub>2</sub>
<b>a</b> <sub>2</sub>	b <sub>1</sub>	C <sub>1</sub>



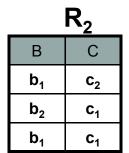


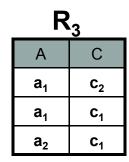
 $a_1$ 

 $a_1$ 

 $a_2$ 

 $b_1$ 

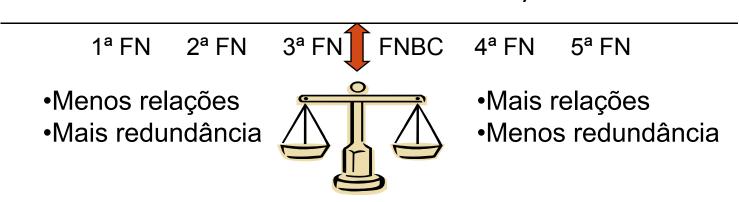




Relação original

Pode-se concluir que existe uma dependência de junção de "R" em relação às projecções  $R_1$ ,  $R_2$  e  $R_3$ .

- Com base nas dependências funcionais, multivalor e de junção define-se o processo de normalização de dados aplicado ao modelo relacional.
- □ A hierarquia é composta por cinco formas normais (1°, 2°, 3°, 4° e 5°
  Forma Normal) e uma intermédia (Forma Normal de Boyce-Codd, entre a 3° e a 4°).
- Na prática, não deve ser levada às ultimas consequências, pois a proliferação de relações pode conduzir à deterioração do desempenho da Base de Dados.
- Na maioria dos casos opta-se por uma solução de compromisso algures entre a 3ª Forma Normal e a Forma Normal de Boyce Codd.



#### □ 1FN − 1<sup>a</sup> Forma Normal

- Este é o passo inicial do processo de normalização de uma base de dados.
- Visa eliminar valores ou grupos de valores repetidos que eventualmente possam existir numa tabela.
- Dada uma relação, a <u>1º forma normal</u> obriga que...
  - Todos os atributos registem apenas valores atómicos.
  - Sejam criadas tantas relações quantos os grupos de valores que se repitam na relação inicial.

#### □ 1FN - 1<sup>a</sup> Forma Normal

**□** Exemplo1:

**Inscrição** (CodAluno, Nome, Morada, Disciplinas)

- Neste caso existe um atributo (Disciplinas) que vai registar os códigos das disciplinas a que o aluno se pode inscrever.
- É um exemplo de um atributo que não contém valores atómicos (Potencialmente irá conter múltiplos valores, tantos quantas as disciplinas a que determinado aluno está inscrito).

#### □ 1FN − 1<sup>a</sup> Forma Normal

**□** Exemplo1:

<u>CodAluno</u>	Nome	Morada	Discip	linas
12300	José António	Rua X	1200, 13	54, 1456
6710	Carla Gomes	Avenida Y	1200,144	15, 1567
15461	Vitor Franco	Bairro Z	<b>1376</b> , 187	77, 1516
			(//	

O atributo "Disciplinas" pode ser decomposto (não contém valores atómicos), pelo que a relação <u>não está</u> na 1° forma normal.

#### □ 1FN − 1<sup>a</sup> Forma Normal

**■** Exemplo2:

<u>CodAluno</u>	Nome	Morada	Disciplina1	Disciplina2	Disciplina3
12300	José António	Rua X	1200	1354	1456
6710	Carla Gomes	Avenida Y	1200	1445	1567
15461	Vitor Franco	Bairro Z	1376	1877	1516
			<	11	

1

□ Neste caso todos os atributos da relação contêm apenas valores atómicos. No entanto, existe um grupo de valores (Disciplina;) que se repete na relação, pelo que esta também não respeita a 1ª forma normal.

#### □ 1FN − 1<sup>a</sup> Forma Normal

Para que este exemplo satisfaça a 1ª forma normal será necessário criar uma nova relação em que cada instância registe a inscrição de um determinado aluno numa única disciplina.

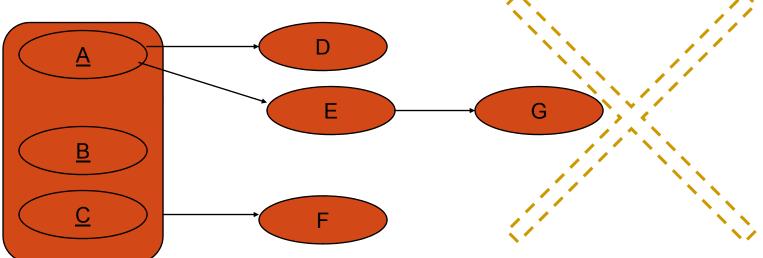
<u>CodAluno</u>	Nome	Morada
12300	José António	Rua X
6710	Carla Gomes	Avenida Y
15461	Vítor Franco	Bairro Z

<u>CodAluno</u>	<u>Disciplina</u>	
12300	1200	
6710	1200	
12300	1376	

À inscrição de um aluno a "N" disciplinas corresponderão "N" instâncias com o código de aluno e o respectivo código de cada disciplina.

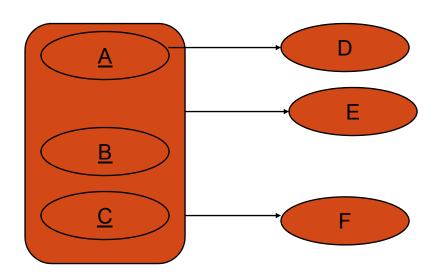
- □ 2FN − 2<sup>a</sup> Forma Normal
- □ Diz-se que uma relação está na 2ª forma normal quando:
  - Está na 1ª forma normal.
  - □Todos os atributos que não pertencem a qualquer chave candidata dependem inteiramente da chave primária, e não apenas de parte dela.

#### □ 2FN - 2<sup>a</sup> Forma Normal



- □ Esta relação não está na 2ª forma normal:
  - Existem atributos que dependem apenas de parte da chave (D, E).
  - Existem atributos que dependem de outros que não são chave (G).

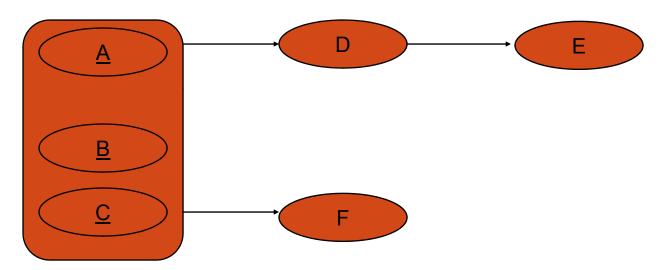
#### □ 2FN − 2<sup>a</sup> Forma Normal





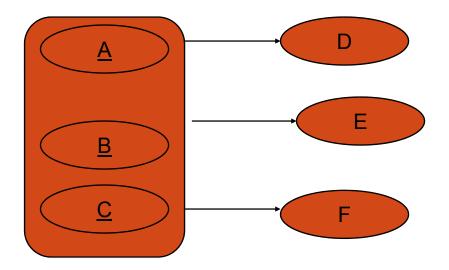
- □ Esta relação também não está na 2ª forma normal:
  - Existem atributos que dependem apenas de parte da chave (D).

#### □ 2FN − 2<sup>a</sup> Forma Normal



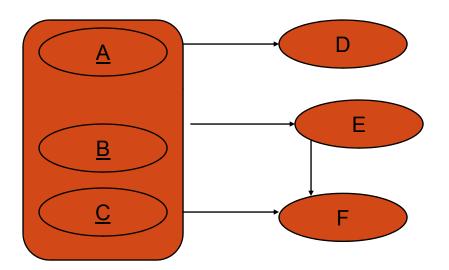
- □ Esta relação já está na 2ª forma normal:
  - Todos os atributos dependem directa ou transitivamente (E) da totalidade da chave primária da relação.

#### □ 2FN − 2<sup>a</sup> Forma Normal



□ Esta relação está na 2ª forma normal, uma vez que não existem atributos que dependam de parte da chave ou de atributos não-chave.

#### □ 2FN - 2<sup>a</sup> Forma Normal



Neste caso a relação também satisfaz as premissas da 2ª forma normal.
 Como se verá de seguida, falhará a restrição associada à próxima forma normal.

- □ 2FN − 2<sup>a</sup> Forma Normal
- Exemplo:

Nota (CodAluno, NomeAluno, CodDisciplina, Nota)

- Esta relação serve para registar as notas de cada aluno nas disciplinas em que encontra inscrito. Uma vez que cada um se pode inscrever em várias disciplinas, a chave primária desta relação é uma chave composta (CodAluno, CodDisciplina).
- Existe um atributo (NomeAluno) que depende <u>apenas</u> de "CodAluno" (O nome de um aluno não depende da disciplina em que se encontra inscrito), pelo que se pode concluir que esta relação <u>não está</u> na 2ª forma normal.

#### □ <u>2FN − 2<sup>a</sup> Forma Normal</u>

□ Para passar a relação anterior à 2ª forma normal seria necessário criar uma nova entidade (Aluno) para guardar informação acerca dos alunos, e retirar o atributo "Nome" da relação "Nota".

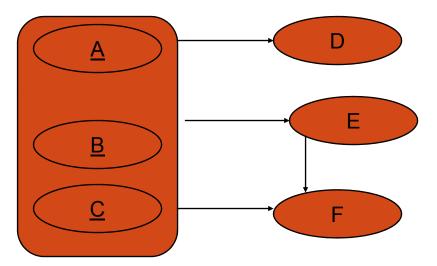
Nota (CodAluno, CodDisciplina, Nota)

**Aluno** (CodAluno, Nome, Naturalidade, ...)

Antes desta transformação, existia uma dependência funcional entre o atributo "Nome" e "CodAluno". Ora numa relação na 2ª forma normal não podem existir dependências funcionais que tenham como premissa apenas parte da chave.

- □ 3FN 3<sup>a</sup> Forma Normal
- □ Uma relação está na 3ª forma normal quando:
  - Está na 2ª forma normal.
  - Não existem dependências funcionais entre os atributos não chave.
- Na prática, uma relação está na 3ª forma normal quando todos os atributos não-chave dependem <u>inteira e exclusivamente</u> da totalidade da chave.

#### □ 3FN − 3<sup>a</sup> Forma Normal



O Atributo "F" depende inteiramente da chave primária da relação. No entanto também existe uma dependência funcional com um atributo não-chave. Então esta relação não satisfaz a 3ª forma normal.

#### □ 3FN − 3<sup>a</sup> Forma Normal

- Exemplo:
  - Imagine-se uma tabela para registar os orientadores dos alunos de projecto do ultimo ano da licenciatura.
  - Uma vez que cada aluno apenas pode ter um único orientador, o atributo "CodAluno" pode ser utilizado como chave primária.

Orientacao (CodAluno, CodDocente, Gabinete, Hora)

#### □ 3FN − 3<sup>a</sup> Forma Normal

**■** Exemplo:

<u>CodAluno</u>	CodDocente	Gabinete	Hora
12300	155	4.1	12:00
6710	255	4.2	12:00
15461	155	4.1	13:00

- □ Todas as restrições impostas pelas 1ª e 2ª formas normais são satisfeitas.
- □ No entanto o atributo "Gabinete" (que depende inteiramente da chave primária), também é dependente de "CodDocente"
  - CodDocente não é chave
  - A relação não está na 3ª forma normal

#### □ 3FN − 3<sup>a</sup> Forma Normal

 Para o exemplo anteriormente ilustrado respeitar a 3<sup>a</sup> forma normal, seria necessário criar uma nova relação para registar o gabinete de cada docente.

Orientacao (CodAluno, CodDocente, Hora)

**Gabinete**(CodDocente, CodGabinete)

- O atributo "Hora" deve manter-se na relação inicial, uma vez que a hora de atendimento depende (é relativa) simultaneamente do aluno e do docente, isto é, varia consoante o aluno e o docente.
- Na maioria das situações a 3<sup>a</sup> forma normal é suficiente para os requisitos pretendidos.

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

- A 3ª forma normal é aquela que, na maioria dos casos, termina o processo de normalização. Em casos bastante específicos ainda transporta algumas anomalias, resolvidas pela <u>Forma Normal de Boyce-Codd</u>.
- □ Por definição, uma relação está na <u>FNBC</u> se:
  - □ Está na 3ª forma normal.
  - □ Todos os determinantes são chaves candidatas.
- Exemplo:

Lecciona (CodAluno, CodDisciplina, CodDocente)

#### FNBC – Forma Normal de Boyce-Codd

A relação anterior serve para registar os alunos das turmas práticas das disciplinas da UBI. Sabe-se que cada disciplina pode ser leccionada por múltiplos docentes. No entanto cada docente só pode leccionar uma disciplina.

<u>codAluno</u>	<u>codDisciplina</u>	codDocente
123	1560	780
671	1560	780
123	1467	766
768	1560	789

A relação satisfaz as premissas relativas às três primeiras formas normais...

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

- Os atributos contêm valores atómicos.
- Não existem grupos de valores repetidos.
- Todos os atributos não-chave dependem inteira e exclusivamente da chave.
- O atributo "CodDocente" não é chave candidata, no entanto é um determinante.
  - □ Uma vez que cada docente só pode leccionar uma disciplina, temos que CodDocente → CodDisciplina.

<u>codAluno</u>	<u>codDisciplina</u>	codDocente
123	1560	780
671	1560	780
123	1467	766
768	1560	789

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

- A passagem da relação anterior para a FNBC exigiria a sua decomposição em duas relações.
  - Uma delas regista os docentes de cada aluno.
  - A outra regista a disciplina leccionada por cada docente.

codAluno	<u>codDocente</u>	
123	780	
671	780	
123	766	
768	789	

codDisciplina	<u>codDocente</u>
1560	780
1467	766
1560	789

- □ 4FN (4<sup>a</sup> Forma Normal)
- Normalmente uma relação na FNBC também já se encontra na 4FN e 5FN, surgindo estas para resolver casos muito raros.
- □ Diz-se que uma relação está na 4ª forma normal se :
  - ■Está na FNBC.
  - □Não existirem dependências multivalor.

- □ 4FN (4<sup>a</sup> Forma Normal)
- □ Exemplo:
- Imagine-se que esta relação regista os produtos que os agentes de uma empresa vendem nas diferentes zonas.
  - Por hipótese, existe uma restrição ao funcionamento da empresa que diz que... "todos os agentes vendem todos os produtos nas zonas em que actuam".

#### 4FN (4<sup>a</sup> Forma Normal)

- Suponha que um determinado "agente" passa a representar um novo "produto". Será necessário inserir várias linhas na tabela, uma para cada zona em que ele actue. Analisando as dependências existentes, temos que:
  - Agente ->> Produto (O conjunto dos produtos é independente das zonas em que este actua, isto é, para cada produto vão existir todos os valores de "zona")
  - Agente ->>Zona (O conjunto das zonas de cada agente é independente dos produtos que venda, ou seja, para cada zona, irão existir todos os "produto" possíveis).
  - Uma vez que existem dependências multivalor, para colocar a relação na 4FN é necessário efectuar a sua decomposição em:

Venda\_Prod (CodAgente, CodProduto) Venda\_Zona (CodAgente, CodZona)

#### □ 4FN (4<sup>a</sup> Forma Normal)

- É importante notar que a decomposição anterior apenas se pode executar tendo por base que "cada agente vende todos os seus produtos em todas as zonas onde actua".
- Esta é também a premissa necessária para verificar a existência de dependências multivalor e requisitar a sua passagem para a 4FN.
- Em situações normais, a FNBC poderia fornecer uma solução que, embora ligeiramente menos eficiente, se poderia revelar bastante mais flexível.

- □ 5FN (5<sup>a</sup> Forma Normal)
- Corresponde ao grau de normalização mais elevado.
- □ Na prática, uma relação na 5° forma normal não pode ser decomposta de nenhuma forma sem perca de informação.
- □ Por definição, uma relação está na 5° forma normal se:
  - Esta na 4<sup>a</sup> forma normal.
  - Não existem dependências de junção relativas a subconjuntos próprios do conjunto de todas as possíveis projecções.

- 5FN (5° Forma Normal)
- Exemplo:

<u>Artista</u>	<u>Filme</u>	<u>Encenador</u>
<b>A</b> 1	F1	E1
<b>A</b> 1	F1	E2
A2	F1	E1
А3	F3	E3
A1	F4	E2
A2	F4	E3

□ Esta relação está na 5° forma normal ?

- 5FN (5<sup>a</sup> Forma Normal)
- Aparentemente, não existem dependências multivalor.
  - Estão satisfeitas as condições da FNBC.

<u>Artista</u>	<u>Filme</u>	<u>Encenador</u>
<b>A</b> 1	F1	E1
<b>A</b> 1	F1	E2
A2	F1	E1
А3	F3	E3
<b>A</b> 1	F4	E2
A2	F4	E3

- □ Falta apenas verificar se existe alguma dependência de junção.
  - Podem-se executar 3 diferentes projecções: {Artista, Filme}, {Artista, Encenador} e {Filme, Encenador}
  - Caso seja possível reconstruir a relação inicial a partir da junção de quaisquer duas projecções, pode-se concluir da existência de dependências de junção relativamente a essas projecções.

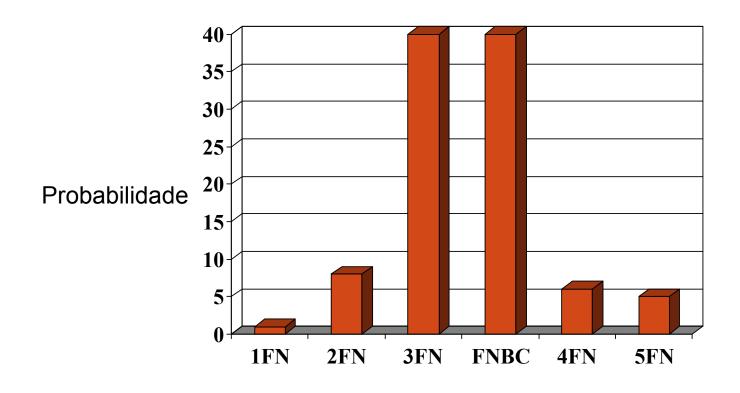
### □ 5FN (5<sup>a</sup> Forma Normal)

- Seja R(A,B,C) uma relação em que foi identificada uma dependência de junção relativamente a {A,B} e {B,C}.
  - □ Significa que é possível reconstruir a relação inicial a partir das duas projecções anteriores.
  - □ Nesse caso, pode-se decompor a relação em duas novas relações com atributos {A,B} e {B,C} sem perca de informação.
- A 5<sup>a</sup> forma normal obriga a efectuar essa decomposição.

#### □ 4FN (4° Forma Normal) e 5FN (5° Forma Normal)

- Na prática, a normalização não deve ser levada às ultimas consequências, pois a proliferação de relações terá consequências ao nível do desempenho da base de dados.
  - A necessidade de manipular múltiplas tabelas para consultar/registar/alterar informação é o principal condicionamento à velocidade de resposta de um SGBD.
- Constitui tarefa do analista encontrar o ponto de equilíbrio que minimize a redundância e maximize a eficiência e integridade da base de dados.

Estado de normalização típico de bases de dados.



Normalização