

Material para a formación profesional inicial

A02. Normalización de relacións

Familia profesional	IFC	Informática e comunicacións
Ciclo formativo	CSIFC03 CSIFC02	Desenvolvemento de aplicacións web Desenvolvemento de aplicacións multiplataforma
Grao		Superior
Módulo profesional	MP0484	Bases de datos
Unidade didáctica	UD03	Deseño lóxico de base de datos
Actividade	A02	Normalización de relacións
Autores		Marta Fernández García María del Carmen Fernández Lameiro Miguel Fraga Vila María Carmen Pato González Andrés del Río Rodríguez
Nome do arquivo		CSIFC02_ MP0484_ V000302_UD03_A02_Normalizar.docx
<p>© 2015 Xunta de Galicia. Consellería de Cultura, Educación e Ordenación Universitaria.</p> <p>Este traballo foi realizado durante unha licenza de formación retribuída pola Consellería de Cultura, Educación e Ordenación Universitaria e ten licenza CreativeCommons BY-NC-SA (recoñecemento - non comercial - compartir igual). Para ver unha copia desta licenza, visitar a ligazón http://creativecommons.org/licenses/by-nc-sa/3.0/es/.</p>		

1.	A02. Normalización de relacións.....	5
1.1	Introdución	5
1.2	Actividade 02. Normalización de relacións	5
1.2.1	Normalización e formas normais (FN).....	5
1.2.2	Obxectivos da normalización.....	7
1.2.3	Dificultades no Deseño de BDR	7
1.2.4	Dependencias funcionais	11
	Definición formal de dependencia funcional (DF).....	11
1.2.5	Proceso de normalización dunha relación.....	12
1.2.5.1	Primeira forma normal (1FN ou 1NF). Atributos multivaluados ou repetitivos.....	12
	Pasos para converter unha relación en 1FN:.....	13
	Pasos:.....	13
1.2.5.2	Segunda forma normal (2FN ou 2NF). Dependencia de clave	15
	Proceso de descomposición formal:.....	16
	Pasos para converter relacións en 1FN a 2FN	16
1.2.5.3	Terceira forma normal (3FN ou 3NF). Dependencias transitivas	17
	Proceso de descomposición formal:.....	18
	Pasos.....	18
1.2.5.4	Forma normal de Boyce-Codd (FNBC ou BCNF). Dependencia de la clave	21
1.2.5.5	Cuarta forma normal (4FN ou 4NF) e as dependencias multivaluadas	22
1.2.5.6	Quinta forma normal (5FN ou 5NF) dependencias de JOIN ou reunión	23
	Operación JOIN ou de reunión	23
	Operación de proxección.....	23
	Dependencias de JOIN ou reunión	23
1.2.5.7	Forma normal de dominio clave (FNDC ou KDFN)	24
1.2.6	Outros criterios de deseño	25
1.2.7	Desnormalización	26
1.3	Tarefas.....	26
1.3.1	Tarefa 1. Normalizar relacións a 1FN	27
	Solución.....	27

1.3.2	Tarefa 2. Normalizar relacións a 2FN	27
	Solución.....	27
1.3.3	Tarefa 3. Normalizar relacións a 3FN	28
	Solución.....	29
1.3.4	Tarefa 4 Normalizar relacións a FNBC.....	29
	Solución.....	30
2.	Materiais	31
2.1	Documentos de apoio ou referencia	31

1. A02. Normalización de relacións

1.1 Introducción

O obxectivo desta actividade é a xustificación da importancia do deseño de datos empregando relacións normalizadas, para elo:

- Describírase e empregárase as formas normais.
- Describírase e empregárase as técnicas de normalización de base de datos.

1.2 Actividade 02. Normalización de relacións

O obxectivo do desenvolvemento dun esquema de BD relacional é crear unha representación precisa e adecuada dos datos, as súas relación e as súas restricións. Esta representación realízase mediante un conxunto de esquemas de relación que permitirán almacenar a información do “minimundo” cun mínimo de redundancias e facilitando a recuperación da información, evitando problemas de actualización, posibles incoherencias, etc.

1.2.1 Normalización e formas normais (FN)

En 1972 E.F. Codd definiu a Normalización como unha técnica para producir un conxunto de relacións con propiedades desexables, dados os requisitos de datos dunha organización.

Unha vez obtido o esquema de BD Relacional, formado por todos os esquemas de relación da BD, a partir da transformación do esquema conceptual (modelo entidade-interrelación estendido) necesitamos pautas ou teorías que axuden ao deseñador a identificar o agrupamento óptimo dos atributos para cada relación no esquema. Estas pautas poden ser:

- **Informais:**
 - Diseñar esquemas de relación con significado fácil de entender e claras.
 - Eficiencias de consulta e aproveitamento de espazo en disco.
 - Redución de valores redundantes en tuplas para evitar anomalías de actualización, facilitando o mantemento e a evolución do deseño.
 - Boa selección de claves.
 - Redución de nulos nas tuplas, empregándoos en casos excepcionais (estes valores desperdician espazo de almacenamento e dificultan as operacións). Pode considerarse tamén un método formal.
- **Formais:** aplicando técnicas formais de Normalización. Este proceso identifica relacións baseándose na súa clave primaria (ou candidatas na FNBC) e das dependencias funcionais entre os seus atributos. O esquema relacional de BD pode ser normalizado ata unha forma normal específica, para evitar “propiedades non desexables”. As formas normais pódense clasificar en:
 - 1ª, 2ª, 3ª Forma Normal (Codd, 1970/1971) e BNFC (R.F. Boyce e Codd, 1974) que constitúe unha definición máis forte da 3FN baseadas en dependencias funcionais. Eliminan case totalmente a posibilidade de anomalías de actualización. Estas formas normais resúmense requirindo que todos os atributos non clave sexan dependentes da "clave, clave completa, e nada máis excepto a clave".
 - 4ª, e 5ª Forma Normal creadas por Fagin en 1977/1979, baseándose nos conceptos de dependencias multivaluadas e de reunión, respectivamente e 6ª Forma Normal (2002),

que se basea no principio de que existen máis de dúas claves candidatas nunha táboa se terán que crear outras táboas con estes atributos claves. Estas últimas formas normais eliminan anomalías de actualización excepcionais.

Na seguinte táboa resúmense as distintas formas normais, indicando os seus creadores e a data en que se propuxeron:

Forma normal	Definida por
Primera Forma Normal (1FN ou 1NF)	Existen dos versiones, la de E. F. Codd de 1970 y la de C. J. Date de 2003.
Segunda Forma Normal (2FN ou 2NF)	E. F. Codd en 1971.
Tercera Forma Normal (3FN ou 3NF)	E. F. Codd en 1971.
Forma Normal de Boyce-Codd (FNBC ou BCNF)	Raymond F. Boyce y E. F. Codd en 1974.
Cuarta Forma Normal (4FN ou 4NF)	Ronald Fagin en 1977.
Quinta Forma Normal (5FN ou 5NF)	Ronald Fagin en 1979.
Forma Normal de Dominio/Clave (FNDc ou DKNF)	Ronald Fagin en 1981.
Sexta Forma Normal (6FN ou 6NF)	C. J. Date, Hugh Darwen y Nikos Lorentzos en 2002.

Táboa 2.1. Autores das Formas Normais

A normalización poderíase definir dunha maneira sinxela como o proceso de organizar e estruturar os datos de forma que se minimice a redundancia. Este proceso baséase nun conxunto de pautas, as regras de normalización, que diminúen o risco de ter un deseño de base de datos defectuoso. Estas regras aplícanse ao modelo relacional da base de datos, normalmente obtido a partir do modelo entidade-interrelación.

As formas normais son incrementais, é dicir, se se cumpre a forma normal (n-ésima) cúmprese tamén a (n-1)-ésima. Unha relación pode atoparse en primeira forma normal e non en segunda forma normal, pero non ao contrario, é dicir os números altos de formas normais son mais restritivos.

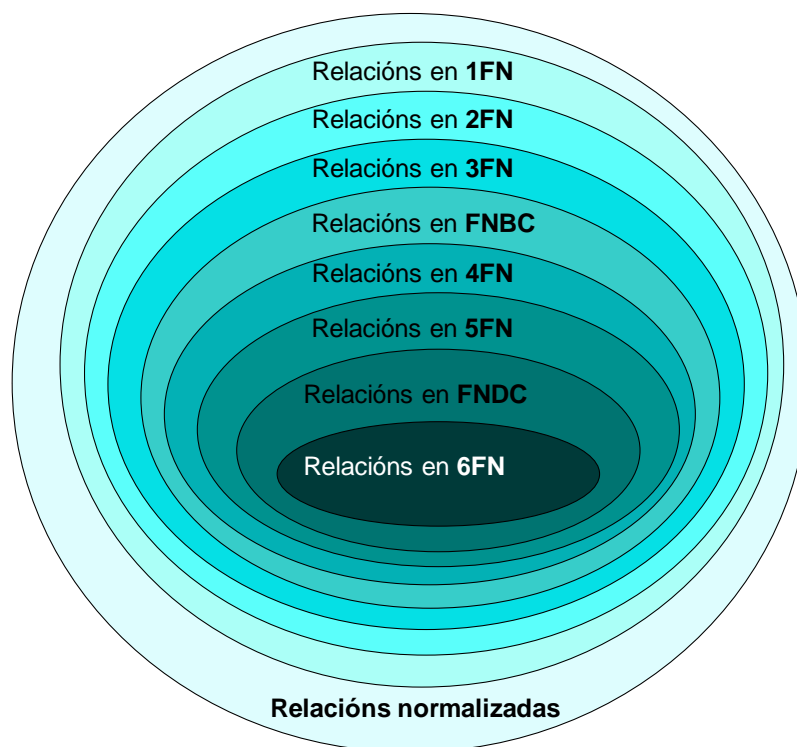


Figura 2.1. Relacións normalizadas en formas normais

O itinerario que se segue á hora de normalizar unha relación é o que se mostra no seguinte diagrama:

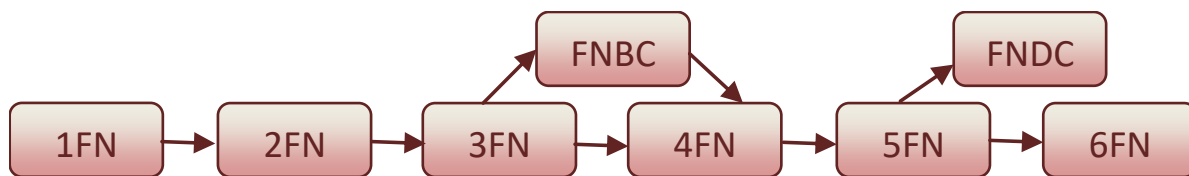


Figura 2.2. As formas normais

Existe un certo consenso en normalizar as táboas da base de datos ata a 3FN ou superior, de feito hai autores que opinan que as bases de datos en cuarta ou quinta forma teñen un redeamento peor que as que se atopan na terceira o una FNBC. En cualquier caso debería ser obligatorio para cualquier diseñador chegar ata a forma normal de Boyce-Codd.

1.2.2 Obxectivos da normalización

Os principais obxectivos que persegue o proceso de normalización son:

- Evitar a redundancia de datos.
- Protexer e dar un mellor soporte á integridade dos datos.
- Eliminar as anomalías nos datos, tanto nas actualizacións como nas insercións e os borrados.
- Reducir na medida do posible o redeseño da base de datos cando esta se amplía.
- Facer máis entendible o modelo de datos a quenes vaian a utilizalo, posto que se modeliza mellor a realidade, o dominio do problema.
- Proporcionar independencia das linguaxes específicas para a consulta de datos.

Mellorar o rendimento da base de datos non está na lista de obxectivos, xa que a normalización céntrase exclusivamente en almacenar a información de maneira correcta.

1.2.3 Dificultades no Deseño de BDR

As bases de datos mal deseñadas teñen problemas de:

- Almacenamento redundante (varias copias da mesma información).
- Perdas non desexadas de información, ao modificar tuplas base, ou borrar tuplas poden levar a un estado non consistente.
- Imposibilidade de almacenar certa información.

Estas dificultades poden dar lugar a anomalías de actualización que aínda que non constitúen un erro en si mesmas (si se fai ben a actualización non habería problemas) son un factor de erro, xerando inconsistencias semánticas e outros problemas que o modelo relacional podería non detectar, estas anomalías non detectadas evitaríanse coa normalización e permitirán un deseño máis robusto.

Supoña o seguinte exemplo dunha táboa que contén a información sobre os pedidos de produtos que fan unha serie de clientes:

PEDIDO (codPedido, codProducto, descripcion, prezoUnidade, unidades, total, data, idCliente, nomeCliente, cidade)

PEDIDO									
CodPedido	codProducto	descripcion	prezoUnidade	cantidade	total	data	idCliente	nomeCliente	cidade
22546	P111345	silla	180	12	2160	11/11/2014	G948499	Pirrel	Ourense
23766	P748765	bolígrafo	0,9	30	27	12/07/2015	G948499	Pirrel	Ourense
17676	P847877	folios	3	20	60	23/08/2014	G948499	Pirrel	Ourense
34355	P111894	Silla	180	10	1800	25/08/2013	D482049	Xen	Lugo
77244	P111684	Silla	180	15	2700	02/01/2015	F985980	Carpazo	Pontevedra
86987	P748786	bolígrafo	0,9	33	29,7	12/12/2012	D482049	Xen	Lugo
55356	P748866	bolígrafo	0,9	22	19,8	07/09/2014	F985980	Carpazo	Pontevedra
56435	P847689	folios	3	21	63	15/04/2015	F985980	Carpazo	Pontevedra
12135	P135455	grapas	0,75	23	17,25	02/01/2015	J445532	Lana	A Coruña

Figura 2.3. Relación PEDIDO non normalizada

Na relación obsérvase que a información dun cliente con máis dun pedido aparece repetida (idCliente, nomeCliente e cidade do mesmo). Isto pode conlevar certas anomalías de actualización:

- **Modificación do valor dun campo:** ¿qué ocorre se queremos modificar o id do cliente Carpazo.?

PEDIDO									
CodPedido	codProducto	descripcion	prezoUnidade	cantidade	total	data	idCliente	nomeCliente	cidade
22546	P111345	silla	180	12	2160	11/11/2014	G948499	Pirrel	Ourense
23766	P748765	bolígrafo	0,9	30	27	12/07/2015	G948499	Pirrel	Ourense
17676	P847877	folios	3	20	60	23/08/2014	G948499	Pirrel	Ourense
34355	P111894	Silla	180	10	1800	25/08/2013	D482049	Xen	Lugo
77244	P111684	Silla	180	15	2700	02/01/2015	A124476	Carpazo	Pontevedra
86987	P748786	bolígrafo	0,9	33	29,7	12/12/2012	D482049	Xen	Lugo
55356	P748866	bolígrafo	0,9	22	19,8	07/09/2014	F985980	Carpazo	Pontevedra
56435	P847689	folios	3	21	63	15/04/2015	F985980	Carpazo	Pontevedra
12135	P135455	grapas	0,75	23	17,25	02/01/2015	J445532	Lana	A Coruña

Tuplas a modificar

Figura 2.4. Relación PEDIDO non normalizada que mostra anomalías de modificación

Ao modificar os valores dos atributos, debido a redundancia poderían xurdir problemas ao modificar e se non actualizamos os campos en todas as filas levaríamos a base de datos a un estado inconsistente.

- **Inserción dunha tupla:**
 - ¿Qué acontece si engadimos outro pedido, con código 90457, feito polo cliente Pirrel (con id de cliente G948499) de Ourense, de 45 unidades do produto P748786 Ficon prezo 0.90 € na data 01 de Xaneiro de 2015?.

PEDIDO									
CodPedido	codProducto	descripcion	prezoUnidade	cantidade	total	data	idCliente	nomeCliente	cidade
22546	P111345	silla	180	12	2160	11/11/2014	G948499	Pirrel	Ourense
23766	P748765	bolígrafo	0,9	30	27	12/07/2015	G948499	Pirrel	Ourense
17676	P847877	folios	3	20	60	23/08/2014	G948499	Pirrel	Ourense
34355	P111894	Silla	180	10	1800	25/08/2013	D482049	Xen	Lugo
77244	P111684	Silla	180	15	2700	02/01/2015	A124476	Carpazo	Pontevedra
86987	P748786	bolígrafo	0,9	33	29,7	12/12/2012	D482049	Xen	Lugo
55356	P748866	bolígrafo	0,9	22	19,8	07/09/2014	A124476	Carpazo	Pontevedra
56435	P847689	folios	3	21	63	15/04/2015	A124476	Carpazo	Pontevedra
12135	P135455	grapas	0,75	23	17,25	02/01/2015	J445532	Lana	A Coruña
90457	P748786	bolígrafo	0,9	45	40,5	01/01/2015	G948499	Pirrel	Ourense

Información redundante

Figura 2.5. Relación PEDIDO non normalizada que mostra anomalías de inserción

Ao engadir un novo pedido dun cliente seguiríase a introducira información redundante.

- ¿Qué ocorre se o introducimos información dun cliente que non fixo pedidos?.

PEDIDO									
CodPedido	codProducto	descripcion	prezoUnidade	cantidade	total	data	idCliente	nomeCliente	cidade
22546	P111345	silla	180	12	2160	11/11/2014	G948499	Pirrel	Ourense
23766	P748765	bolígrafo	0,9	30	27	12/07/2015	G948499	Pirrel	Ourense
17676	P847877	folios	3	20	60	23/08/2014	G948499	Pirrel	Ourense
34355	P111345	Silla	180	10	1800	25/08/2013	D482049	Xen	Lugo
77244	P111345	Silla	180	15	2700	02/01/2015	A124476	Carpazo	Pontevedra
86987	P748765	bolígrafo	0,9	33	29,7	12/12/2012	D482049	Xen	Lugo
55356	P748765	bolígrafo	0,9	22	19,8	07/09/2014	A124476	Carpazo	Pontevedra
56435	P847877	folios	3	21	63	15/04/2015	A124476	Carpazo	Pontevedra
12135	P135455	grapas	0,75	23	17,25	02/01/2015	J445532	Lana	A Coruña
90457	P748765	bolígrafo	0,9	45	40,5	01/01/2015	G948499	Pirrel	Ourense
44567	L948566	NULL	NULL	NULL	NULL	NULL	J554679	keppo	Madrid

Figura 2.6. Relación PEDIDO non normalizada que mostra anomalías de inserción.

Se queremos introducir os datos dun cliente que non fixo ningún pedido, teremos que xerar valores nulos.

O mesmo aconteceria nas seguintes situacións: cambiar un cliente de cidade, modificar o prezo dun produto.

- **Eliminación dunha tupla:** ¿Qué ocorre se eliminamos o produto P748765 da relación PEDIDO?.

PEDIDO									
CodPedido	codProducto	descripcion	prezoUnidade	cantidade	total	data	idCliente	nomeCliente	cidade
22546	P111345	silla	180	12	2160	11/11/2014	G948499	Pirrel	Ourense
23766	P748765	bolígrafo	0,9	30	27	12/07/2015	G948499	Pirrel	Ourense
17676	P847877	folios	3	20	60	23/08/2014	G948499	Pirrel	Ourense
34355	P111345	Silla	180	10	1800	25/08/2013	D482049	Xen	Lugo
77244	P111345	Silla	180	15	2700	02/01/2015	A124476	Carpazo	Pontevedra
86987	P748765	bolígrafo	0,9	33	29,7	12/12/2012	D482049	Xen	Lugo
55356	P748765	bolígrafo	0,9	22	19,8	07/09/2014	A124476	Carpazo	Pontevedra
56435	P847877	folios	3	21	63	15/04/2015	A124476	Carpazo	Pontevedra
12135	P135455	grapas	0,75	23	17,25	02/01/2015	J445532	Lana	A Coruña
90457	P748765	bolígrafo	0,9	45	40,5	01/01/2015	G948499	Pirrel	Ourense
44567	L948566	NULL	NULL	NULL	NULL	NULL	J554679	keppo	Madrid

Figura 2.7. Relación PEDIDO non normalizada que mostra anomalías de eliminación de tuplas

Ao eliminar a información do PEDIDO poderíamos perder todos os datos dos clientes con ese pedido. As inconsistencias afloran ao facer consultas: un cliente xa non aparece, un produto con diferente prezo, etc.

▪ Descomposición de reunión con perda (tuplas espúreas):

Ao dividir unha relación de forma incorrecta pode producir resultados inconsistentes, por exemplo supoñamos a seguinte división en dúas relacións do exemplo:

PEDIDOS					PEDIDO DE CLIENTE						
CodPedido	prezoUnidade	total	data	nomeCliente	idCliente	nomeCliente	cidade	telefonos	codProducto	descripcion	cantidade
22546	180	2160	11/11/2014	Pirrel	G948499	Pirrel	Ourense	988223344; 654327834	P111345	silla	12
23766	0,9	27	12/07/2015	Pirrel	G948499	Pirrel	Ourense	988123456; 654321648	P748765	bolígrafo	30
17676	3	60	23/08/2014	Pirrel	G948499	Pirrel	Ourense	988762311; 976534124	P847877	folios	20
34355	180	1800	25/08/2013	Xen	D482049	Xen	Lugo	982098765; 653876098	P111894	Silla	10
77244	180	2700	02/01/2015	Carpazo	F985980	Carpazo	Pontevedra	986341290; 645068234	P111684	Silla	15
86987	0,9	29,7	12/12/2012	Xen	D482049	Xen	Lugo	982009988; 618098345	P748786	bolígrafo	33
55356	0,9	19,8	07/09/2014	Carpazo	F985980	Carpazo	Pontevedra	986306187; 630982310	P748866	bolígrafo	22
56435	3	63	15/04/2015	Carpazo	F985980	Carpazo	Pontevedra	986976834; 639675143	P847689	folios	21
12135	0,75	17,25	02/01/2015	Lana	J445532	Lana	A Coruña	981675423; 623987407	P135455	grapas	23

Figura 2.8. Descomposición errónea da relación PEDIDO

Para recuperar posteriormente a información completa dos pedidos dun cliente xuntaríamos ambas táboas polo nome do cliente (o que se denomina operación de reunión ou join), obetendose o seguinte resultado:

PEDIDO DE CLIENTE							PEDIDOS				
idCliente	nomeCliente	cidade	telefonos	codProducto	descripcion	cantidade	CodPedido	prezoUnidade	total	data	nomeCliente
D482049	Xen	Lugo	982098765; 653876098	P111894	Silla	10	34355	180	1800	25/08/2013	Xen
D482049	Xen	Lugo	982098765; 653876098	P111894	Silla	10	86987	0,9	29,7	12/12/2012	Xen
D482049	Xen	Lugo	982009988; 618098345	P748786	bolígrafo	33	86987	180	1800	25/08/2013	Xen
D482049	Xen	Lugo	982009988; 618098345	P748786	bolígrafo	33	86987	0,9	29,7	12/12/2012	Xen

← Tuplas espúreas

Empregamos o atributo nome para localizar toda a información dos pedidos do cliente Xen

Figura 2.9. Reunión da descomposición da relación PEDIDO que xenera tuplas espúreas

O mezclarse cada unha das tuplas de PEDIDO DE CLIENTE con PEDIDOS empregando o nome do cliente (neste exemplo “Xen”) aparecen máis tuplas asociadas a “Xen” pero con menos información. Estas tuplas que non existen na relación inicial e que son incorrectas, reciben o nome de “espúreas”.

Para resolver estes problemas debemos realizar un bo deseño, seguindo os pasos de transformación a partir dun modelo conceptual entidade-interrelación e comprobando se as relacións resultantes están normalizadas, e no caso de que non o estean descompondo o esquema de relación en varios esquemas de relación. Os esquemas resultantes cumprirán certas formas normais (preservándose a unión ou combinación de relacións). Deste xeito, minimízase o grao de vulnerabilidade dunha táboa a inconsistencias e anomalías lóxicas.

1.2.4 Dependencias funcionais

O concepto de dependencia funcional é esencial para o deseño de esquemas de bases de datos sen redundancia.

- **Definimos a dependencia funcional (DF) da seguinte forma:** Dado un esquema relacional $R(a_1, a_2, \dots, a_N)$ e sexan α e β dos conxuntos diferentes de atributos de dito esquema. O feito de que α determine funcionalmente a β , e que β dependa funcionalmente de α se expresa: $\alpha \rightarrow \beta$.

Podese simplificar a definición dicindo que dados uns valores para os atributos de α , pódense determinar os correspondentes valores dos atributos β . Así, chámase X ao conxunto determinante e a Y o conxunto dependente.

En xeral, unha dependencia funcional (df) $\alpha \rightarrow \beta$ é unha restrición entre dous conxuntos de atributos (descriptores) α e β e é unha propiedade inherente ao contido semántico dos datos, que se debe cumprir para calquera extensión dunha relación r co esquema de relación R .

A inherencia ao contido semántico significa que unha dependencia funcional non pode ser demostrada, pero si afirmada por observación da realidade.

A súa utilidade é a de describir mellor un esquema de relación, mediante a especificación de restricións (sobre atributos) que deben cumprirse sempre.

Definición formal de dependencia funcional (DF)

Sexa R un esquema de relación, no que o seu conxunto de atributos é A , e sexan os descriptores α e β subconxuntos de A , dise que

- $\alpha \rightarrow \beta$ se " $t_1, t_2 \in r(R)$ e $t_1[\alpha] = t_2[\alpha]$ entón $t_1[\beta] = t_2[\beta]$ ".

β depende funcionalmente de α si dadas dous tuplas t_1 e t_2 (da extensión dunha relación $r(R)$) que teñan os mesmos valores nos atributos que forman α , tamén teñen os mesmos valores para os atributos de β (**Tamén se di que α determina funcionalmente a β**).

Dito doutro xeito que a un valor dado de α le corresponde un único valor de β e isto sucede independentemente do instante considerado, se coñecemos os valores do compoñente α dunha tupla se coñecerán os valores del compoñente β desa mesma tupla (os valores do compoñente β dunha tupla determinan funcionalmente dos valores do compoñente β).

D.F. é una aserción ou afirmación sobre todos os valores posibles e non sobre os valores actuais, polo tanto, caracteriza a intensión e non a extensión de un esquema de relación.

Débese clarexar que $\alpha \rightarrow \beta$ **non** significa que “dado α poida deducirse o valor de β ”, ademais que $\alpha \rightarrow \beta$, dous tuplas son iguais en α son iguais en β non indica nada acerca de si $\beta \rightarrow \alpha$, pode ocorrer que sendo diferentes en α sexan iguais en β .

Un esquema relacional utilizando dependencias funcionais podería definirse como $r(R, f)$, con R conxunto de atributo e f conxunto de dependencias funcionais. Por exemplo:

- $Actor = (\{idActor, nome, apleido1, apelido2\}, f)$.

Un atributo ou conxunto de atributos X dunha relación utilizando dependencias funcionais cumpre que $X \rightarrow A_1, A_2, \dots, A_n$.

Non existe ningún atributo ou conxunto de atributos Y , contido en X , tal que $Y \rightarrow A_1, A_2, \dots, A_n$, é dicir, a **clave é o conxunto mínimo de atributos que determina a tódolos demais**.

Por exemplo o nome dunha persoa depende funcionalmente do DNI; é dicir para un DNI concreto só hai un nome posible. Na relación do exemplo, o $idCliente$ non ten dependencia

funcional co codPedido, xa que para un mesmo idCliente pode haber máis dun pedido posible, eero o nomeCliente sí que depende do idCliente.

1.2.5 Proceso de normalización dunha relación

O proceso de normalización somete a un esquema de relación a unha serie de probas para "certificar" se pertence ou non a unha certa forma normal.

A normalización dos datos define o proceso no que os esquemas de relación que non cumpren as condicións se descompoñen, repartindo os seus atributos entre esquemas de relación máis pequenos que cumpren as condicións establecidas; deste xeito se garantízas que non ocorran anomalías de actualización.

A descomposición expresada máis formalmente **consiste na substitución da relación $R(A_1, A_2, \dots, A_n)$ por unha serie de relacións R_1, R_2, \dots, R_n obtidas mediante proxeccións de R e tales que a relación resultante das reunións de R_1, R_2, \dots, R_n teña o mesmo esquema que R .**

As propiedades que debe verificar unha boa descomposición son:

- **A propiedade de reunión sen perda**, é dicir, que a extensión das relacións normalizadas nos permitan obter todos os valores de R .
- **A propiedade de conservación das dependencias**, que asegura que todas as dependencias funcionais estean representadas nalgunha das relacións individuais resultantes.

Non todas as descomposición son válidas ou boas porque nalgunhas prodúcese perdas de información.

1.2.5.1 Primeira forma normal (1FN ou 1NF). Atributos multivaluados ou repetitivos

Un esquema R cumpre a primeira forma normal (1FN) se os dominios de todos os atributos son atómicos e univaluados, ningún atributo pode tomar valores compostos (dominio non atómico) nin multivaluados (mais dun valor ao mesmo tempo), estando libre de grupos repetitivos (táboas dentro de táboas). Esta forma normal é parte inherente do modelo relacional de datos.

Un atributo é atómico se os seus elementos pódense considerar como unidades indivisibles, exemplos de dominios non atómicos son:

- Nome completo: atributo composto polo nome máis os apelidos.
- Teléfonos: atributos multivaluado, xa que por regra xeral os contactos soen posuir máis dun número de teléfono (fixo, móbil, persoal, traballo, etc).

Moitas veces, a definición do que é un valor atómico non é clara, e pode depender dos usos que se vaian facer dos valores do atributo. Por exemplo, cando se trata de procesar datos de persoas pode ter sentido utilizar un atributo “nome completo” para almacenar o nome propio e os apelidos, se está información sempre se abordar de forma conxunta. En cambio, si se desexa acceder polo nome propio e/ou polos apelidos separadamente, para que o esquema de relación estea en 1FN debería ter atributos diferenciados para eles, “nomePropio”, “apelido1” “apelido2”.

Cando unha relación non está en primeira forma normal, dividise noutras relacións, repartindo os seus atributos entre as resultantes. Normalmente a idea é eliminar o atributo que incumpre a 1ª FN da relación orixinal e colocalo nunha relación diferente xunto coa clave primaria da relación de partida.

Pasos para converter unha relación en 1FN:

Cando unha relación non está en primeira forma normal, divídese noutras relacións, re-partindo os seus atributos entre as resultantes. Normalmente a idea é eliminar o atributo que incumpre a 1ª FN da relación orixinal e colocalo nunha relación diferente xunto coa clave primaria da relación de partida:

No exemplo o ounto de partida é o seguinte formulario:

PEDIDO

Nº Pedido
Fecha Pedido

Nº Cliente

Nombre

Enderezo

Codigo Producto	Descrición	PrecioUnidade	Cantidad	total

PedidoTotal

Figura 2.10. Formulario dun pedido

Pódese intuír a seguinte relación non normalizada que contén elementos repetidos e subtaboas (características non permitidas no modelo relacional). No esquema de relación marcáse como clave o atributo `codPedido`, que cumpre a condición de ser mínima e identifica univocamente cada unha das tuplas sen repetirse:

PEDIDO (*codPedido*, (*codProducto*, *descripcion*, *prezoUnidade*, *unidades*, *subtototal*), *data*, *idCliente*, *nomeCliente*, *cidade*, *telefono*, *totalPedido*)

PEDIDO											
CodPedido	codProducto	descripcion	prezoUnidade	cantidad	total	data	idCliente	nomeCliente	cidade	telefonos	totalPedido
22546	P111345	silla	180	12	2160	11/11/2014	G948499	Pirrel	Ourense	988223344 654327834	2247
	P748765	bolígrafo	0,9	30	27	11/11/2014	G948499	Pirrel	Ourense	988223344 654327834	
	P847877	folios	3	20	60	11/11/2014	G948499	Pirrel	Ourense	988223344 654327834	
34355	P111894	Silla	180	10	1800	25/08/2013	D482049	Xen	Lugo	982098765 653876098	1829,7
	P748786	bolígrafo	0,9	33	29,7	25/08/2013	D482049	Xen	Lugo	982098765 653876098	
77244	P111684	Silla	180	15	2700	15/04/2015	F985980	Carpazo	Pontevedra	986341290 645068234	2782,8
	P748866	bolígrafo	0,9	22	19,8	15/04/2015	F985980	Carpazo	Pontevedra	986341290 645068234	
	P847689	folios	3	21	63	15/04/2015	F985980	Carpazo	Pontevedra	986341290 645068234	
12135	P135455	grapap	0,75	23	17,25	02/01/2015	J445532	Lana	A Coruña	981675423 623987407	17,25

Figura 2.11. Relación PEDIDO non normalizada extraída do formulario PEDIDO

Pasos:

Todos os atributos deben ser atómicos e univaluados, ningún pode tomar valores compostos (dominio non atómico) nin multivaluados (mais dun valor ao mesmo tempo), estando libre de grupos repetitivos (táboas dentro de táboas).

- I) Crear unha nova relación cos atributos que incumpren ser atómico ou son táboas dentro de táboas, no exemplo comenzaremos coa subtaboa que se atopa dentro de cada tupla que representa un pedido completo:

PEDIDO (*codPedido*, (*codProducto*, *descripcion*, *prezoUnidade*, *unidades*, *total*), *data*, *idCliente*, *nomeCliente*, *cidade*, *totalPedido*)

PEDIDO										
CodPedido	codProducto	descripcion	prezoUnidade	cantidade	total	data	idCliente	nomeCliente	cidade	totalPedido
22546	P111345	silla	180	12	2160	11/11/2014	G948499	Pirrel	Ourense	2247
	P748765	bolígrafo	0,9	30	27	11/11/2014	G948499	Pirrel	Ourense	
	P847877	folios	3	20	60	11/11/2014	G948499	Pirrel	Ourense	
34355	P111894	Silla	180	10	1800	25/08/2013	D482049	Xen	Lugo	1829,7
	P748786	bolígrafo	0,9	33	29,7	25/08/2013	D482049	Xen	Lugo	
77244	P111684	Silla	180	15	2700	15/04/2015	F985980	Carpazo	Pontevedra	2782,8
	P748866	bolígrafo	0,9	22	19,8	15/04/2015	F985980	Carpazo	Pontevedra	
	P847689	folios	3	21	63	15/04/2015	F985980	Carpazo	Pontevedra	
12135	P135455	grapas	0,75	23	17,25	02/01/2015	J445532	Lana	A Coruña	17,25

Figura 2.12. Relación PEDIDO non normalizada que contén grupos repetitivos

- II) Engadir a esta nova relación a clave primaria da relación que orixinalmente a contiña.
- III) Darlle un nome á nova relación (opcionalmente no nome incluírase a FN na que a táboa aparece).

LIÑAPEDIDO_1					
codPedido	codProducto	descripcion	prezoUnidade	cantidade	total
22546	P111345	silla	180	12	2160
22546	P748765	bolígrafo	0,9	30	27
22546	P847877	folios	3	20	60
34355	P111894	Silla	180	10	1800
34355	P748786	bolígrafo	0,9	33	29,7
77244	P111684	Silla	180	15	2700
77244	P748866	bolígrafo	0,9	22	19,8
77244	P847689	folios	3	21	63
12135	P135455	grapas	0,75	23	17,25

Figura 2.13. Relación resultante do proceso de normalización a 1FN ao eliminar os elementos repetitivos

- IV) Determinar a clave primaria da nova relación: no exemplo os atributos que identifican unívocamente cada tupla da táboa teléfono sendo mínima, no exemplo a clave resultante é composta CodPedido e CodProducto.

LIÑAPEDIDO_1 (*codPedido*, *codProducto*, *descripcion*, *prezoUnidade*, *cantidad*, *total*)

LIÑAPEDIDO_1					
codPedido	codProducto	descripcion	prezoUnidade	cantidad	total
22546	P111345	silla	180	12	2160
22546	P748765	bolígrafo	0,9	30	27
22546	P847877	folios	3	20	60
34355	P111894	Silla	180	10	1800
34355	P748786	bolígrafo	0,9	33	29,7
77244	P111684	Silla	180	15	2700
77244	P748866	bolígrafo	0,9	22	19,8
77244	P847689	folios	3	21	63
12135	P135455	grapap	0,75	23	17,25

Figura 2.14. Relación resultante tralo proceso de detección de clave

- V) Repetir ata que non queden máis atributos non atómicos, no exemplo repítese o proceso para o tributo multivaluado teléfonos, xerando a seguinte táboa.



Tarefa 1: Normalizar a 1FN.

1.2.5.2 Segunda forma normal (2FN ou 2NF). Dependencia de clave

Unha relación R está en 2FN, se está 1FN e todo atributo que non sexa clave primaria, nin forme parte dela, ten dependencia funcional completa ou plena respecto da clave primaria desa relación, é dicir, o valor dos atributos non principais da relación ven determinado polo valor de todos os atributos da clave. Só os atributos dunha clave poden depender de partes dunha clave.

LIÑAPEDIDO_1					
codPedido	codProducto	descripcion	prezoUnidade	cantidad	total
22546	P111345	silla	180	12	2160
22546	P748765	bolígrafo	0,9	30	27
22546	P847877	folios	3	20	60
34355	P111894	Silla	180	10	1800
34355	P748786	bolígrafo	0,9	33	29,7
77244	P111684	Silla	180	15	2700
77244	P748866	bolígrafo	0,9	22	19,8
77244	P847689	folios	3	21	63
12135	P135455	grapap	0,75	23	17,25

Figura 2.16. Relación en 1FN con atributos (descripcion e prezoUnidade) que dependen de parte da clave cosProducto

No exemplo observase como certos atributos só dependen de parte da clave, coma descripcion e prezoUnidade que dependen funcionalmente do atributo codProducto pero non de codPedido polo que as dependencias funcionais (df) son:

- `codProduto` -> `descripcion`.
- `codProducto` -> `prezoUnidade`.

Esta forma normal só se aplica a relacións que teñen claves compostas, é dicir, que están formadas por máis dun atributo. Se un esquema de relación non está en 2FN se normaliza en varias relacións que si estean en 2FN, nas que os atributos que dependen dunha parte da clave formarán unha nova relación que terá esa parte da clave como clave primaria.

Proceso de descomposición formal:

Formalmente pódese definir o proceso de descomposición da seguinte maneira: Dada unha relación R con clave composta da forma $(R.x, R.y)$, e en esta relación existe una dependencia funcional incompleta de la forma $R.y \rightarrow R.z$, onde $R.z$ é un atributo que non pertence á clave) da relación R , o proceso de descomposición realizarase da forma seguinte:

- Da relación R elimínase o atributo $R.z$, quedando igual o resto.
- Constrúese unha nova relación $R1$, coa estrutura:
 - O atributo $R.z$ como atributo non clave da relación $R1$.
 - O atributo $R.y$ como atributo clave (clave primaria) da relación $R1$.
- Defínese a $R.y$ como clave foránea da clave da relación $R1$, $(R1.y)$.

A descomposición para a aplicación da 2FN débese facer en base á dependencia funcional incompleta, e nunca por outras dependencia entre atributos da relación. Neste proceso tanto $R.x$, $R.y$, $R.z$ poden ser a súa vez atributos simples ou agregados de datos.

Pasos para converter relacións en 1FN a 2FN

Normalizarase o esquema do exemplo seguindo os seguintes pasos:

- I) Eliminar os atributos que dependen parcialmente da clave primaria e crea con eles unha nova relación (sen elementos repetidos).
- II) Nomear á nova relación (engadir opcionalmente un 2 para indicar 2NF).
- III) Engadir a esta relación unha copia do atributo ou atributos do cales dependen (será á clave primaria da nova relación).

PRODUTO_2		
codProduto	descripcion	prezoUnidade
P111345	silla	180
P748765	bolígrafo	0,9
P847877	folios	3
P111894	silla	180
P748786	bolígrafo	0,9
P111684	silla	180
P748866	bolígrafo	0,9
P847689	folios	3
P135455	grapas	0,75

PRODUTO_2		
codProduto	descripcion	prezoUnidade
P111345	silla	180
P748765	bolígrafo	0,9
P847877	folios	3
P135455	grapas	0,75

Figura 2.17. Táboa resultante do proceso de normalización a 2FN

- IV) Renomear á relación orixinal (engadir un 2 para indicar 2FN)

LIÑAPEDIDO_2			
codPedido	codProducto	cantidade	total
22546	P111345	12	2160
22546	P748765	30	27
22546	P847877	20	60
34355	P111894	10	1800
34355	P748786	33	29,7
77244	P111684	15	2700
77244	P748866	22	19,8
77244	P847689	21	63
12135	P135455	23	17,25

PRODUTO_2		
codProducto	descripcion	prezoUnidade
P111345	silla	180
P748765	bolígrafo	0,9
P847877	folios	3
P135455	grapap	0,75

LIÑAPEDIDO_2 (*codPedido, codProducto, cantidade, total*)

PRODUTO_2 (*codProducto, descripcion, prezoUnidade*)

Figura 2.18. Táboas e grafo relacional obtido tras normalizar en 2FN

- V) Repetir o proceso ata que todas as táboas adopten a 2FN

A táboa orixinaria, o eliminar de ela os elementos repetitivos tamén se atopa en 2FN:

PEDIDO_2(*codPedido,, data, idCliente, nomeCliente, cidade, totalPedido*)

PEDIDO_2					
CodPedido	data	idCliente	nomeCliente	cidade	totalPedido
22546	11/11/2014	G948499	Pirrel	Ourense	2247
22546	11/11/2014	G948499	Pirrel	Ourense	2247
22546	11/11/2014	G948499	Pirrel	Ourense	2247
34355	25/08/2013	D482049	Xen	Lugo	1829,7
34355	25/08/2013	D482049	Xen	Lugo	1829,7
77244	15/04/2015	F985980	Carpazo	Pontevedra	2782,8
77244	15/04/2015	F985980	Carpazo	Pontevedra	2782,8
77244	15/04/2015	F985980	Carpazo	Pontevedra	2782,8
12135	02/01/2015	J445532	Lana	A Coruña	17,25

Figura 2.19. Táboa en 2FN tra a normalización



Tarefa 2: Normalizar relacións a 2FN

1.2.5.3 Terceira forma normal (3FN ou 3NF). Dependencias transitivas

Unha relación está en terceira forma normal se cumpre as seguintes condicións:

- A relación está en 2FN
- Todos os atributos da relación dependen funcionalmente só da clave, e non de ningún outro atributo non clave (todos os atributos non principais son independentes)
- Non existen dependencias funcionais transitivas respecto das claves. Un atributo depende transitivamente da clave primaria si depende doutro atributo que depende da clave

Resumindo, para toda DF: $X \rightarrow Y$, X é unha clave.

Proceso de descomposición formal:

Dada unha relación R de clave $R.x$ e dous atributos non clave $R.y$ e $R.z$, e nesta relación están presentes as seguintes dependencias funcionais:

$R.x \rightarrow R.y$, $R.y \rightarrow R.z$ e, por tanto a dependencia transitiva, $R.x \rightarrow R.z$, o proceso de descomposición realizarase da forma seguinte:

- Da relación R elimínase o atributo que mantén unha dependencia funcional transitiva coa clave da relación, é dicir, o atributo $R.z$, quedando igual ao resto da relación R .
- Constrúese unha nova relación $R1$, coa estrutura seguinte:
 - O atributo $R.z$ que mantña unha dependencia funcional transitiva.
 - O atributo $R.y$ co cal o atributo $R.z$ mantña unha dependencia funcional completa.
 - A clave da relación $R1$, será o atributo $R1.y$ e, nesta relación só existirá unha dependencia funcional completa da forma: $R1.y \rightarrow R1.z$.

Na relación R defínese $R.y$ como clave allea da relación $R1$ ($R1.y$).

Deste xeito, a descomposición que se debe realizar por aplicación da 3FN debe ser en base á dependencia funcional transitiva e nunca por outra dependencia entre os atributos da relación. Neste proceso tanto $R.x$, $R.y$, $R.z$ poden ser a súa vez atributos simples ou agregados de datos.

Pasos

No exemplo, a táboa PEDIDO_2 presenta anomalías por dependencias funcionais transitivas, seguimos os seguintes pasos para que adopte a 3FN:

- I) Comprobar que cada relación ten un número fixo de columnas e as variables son sinxelas (atómicas).
- II) Identifica a clave primaria.
- III) Comproba que todos os atributos (menos a clave primaria) depende de TODA a clave non de PARTE dela.
- IV) Si existe dependencia parcial dividir a relación en varias subrelacións.
- V) Comprobar que todos os atributos dependen da clave e non doutros atributos (o que se coñece coma dependencias transitivas): Observase como os atributos nomeCliente, cidade non dependen da clave ou de parte de ela senón do atributo idCliente que a súa vez si depende de codPedido, prodúcese unha dependencia transitiva:

PEDIDO_2					
CodPedido	data	idCliente	nomeCliente	cidade	totalPedido
22546	11/11/2014	G948499	Pirrel	Ourense	2247
22546	11/11/2014	G948499	Pirrel	Ourense	2247
22546	11/11/2014	G948499	Pirrel	Ourense	2247
34355	25/08/2013	D482049	Xen	Lugo	1829,7
34355	25/08/2013	D482049	Xen	Lugo	1829,7
77244	15/04/2015	F985980	Carpazo	Pontevedra	2782,8
77244	15/04/2015	F985980	Carpazo	Pontevedra	2782,8
77244	15/04/2015	F985980	Carpazo	Pontevedra	2782,8
12135	02/01/2015	J445532	Lana	A Coruña	17,25

Figura 2.20. Táboa con dependencias transitivas de CodPedido a idCliente e de idCliente a nomeCliente e cidade

- VI) Si existe dependencias non relacionadas con clave primaria dividir as relación.

CLIENTE_3		
idCliente	nomeCliente	cidade
G948499	Pirrel	Ourense
G948499	Pirrel	Ourense
G948499	Pirrel	Ourense
D482049	Xen	Lugo
D482049	Xen	Lugo
F985980	Carpazo	Pontevedra
F985980	Carpazo	Pontevedra
F985980	Carpazo	Pontevedra
J445532	Lana	A Coruña

PEDIDO_3			
CodPedido	data	idCliente	totalPedido
22546	11/11/2014	G948499	2247
22546	11/11/2014	G948499	2247
22546	11/11/2014	G948499	2247
34355	25/08/2013	D482049	1829,7
34355	25/08/2013	D482049	1829,7
77244	15/04/2015	F985980	2782,8
77244	15/04/2015	F985980	2782,8
77244	15/04/2015	F985980	2782,8
12135	02/01/2015	J445532	17,25

CLIENTE_3 (codCliente, nomeCliente, cidade) **PEDIDO_3** (codPedido, data, totalPedido)

Figura 2.21. Táboas e esquema obtidas tras a normalización a 3FN

A maioría das táboas en 3NF están libres de anomalías de actualización, inserción, e borrado. Certos tipos de táboas en 3NF, que na práctica raramente atópanse, son afectadas por tales anomalías; estas son táboas que non satisfán a forma normal de Boyce-Codd (FNBC ou BCNF) ou, si satisfán a FNBC, son insuficientes para satisfacer as formas normais máis altas 4NF ou 5NF, todas elas formas a estudar nos seguintes puntos.

O proceso de normalización ata a 3FN do exemplo mostra as seguintes dependencias funcionais entre atributos (DF):

- codProducto-> descricion.
- codProducto -> prezoUnidade.
- codCliente -> nomeCliente.
- codCliente -> cidade.
- CodPedido -> codCliente.
- codPedido -> cantidade.

- O esquema xeral xunto co seu contido en 3FN é:

Figura 2.22. Grafo relacional obtido o normalizar a 3FN

CLIENTE_3		
idCliente	nomeCliente	cidade
G948499	Pirrel	Ourense
G948499	Pirrel	Ourense
G948499	Pirrel	Ourense
D482049	Xen	Lugo
D482049	Xen	Lugo
F985980	Carpazo	Pontevedra
F985980	Carpazo	Pontevedra
F985980	Carpazo	Pontevedra
J445532	Lana	A Coruña

Figura 2.23. Táboas obtidas tras a normalización a 3FN



Tarefa 3: Normalizar relacións a 3FN

1.2.5.4 Forma normal de Boyce-Codd (FNBC ou BCNF). Dependencia de la clave

Un esquema R está en forma normal de Boyce_Codd, si para toda dependencia $X \rightarrow Y$ non trivial (se Y non está contida en X) X é unha superclave de R, dito doutro xeito, non pode haber máis dependencias que con superclaves (o resto de atributos non clave da relación dependen funcionalmente dela).

No noso exemplo se estudamos todas as relacións:

- PRODUTO_2(codProducto, descripcion,prezoUnidade).
 - Clave:codProducto.
 - DP: codProducto \rightarrow {descripcion,prezoUnidade}.
 - É superclaves atópase en FNBC.
- LIÑAPEDIDO_2 (codPedido,CodProducto, cantidade, total).
 - Clave: (codPedido, codProducto).
 - DP: (codPedido, codProducto) \rightarrow {cantidade, total}.
 - É superclaves atópase en FNBC.
- PEDIDO_3 (codPedido,codCliente,data,totalPedido).
 - Clave: (codPedido, codCliente).
 - DP: (codPedido, codCliente \rightarrow {data, totalPedido}.
 - É superclaves atópase en FNBC.
- CLIENTE_3(codCliente,nomeCliente,cidade).
 - Clave:codCliente.
 - DP: codCliente \rightarrow {nomeCliente,cidade}.
 - É superclaves atópase en FNBC.
- O exemplo xa se atopa en FNBC.

Estudemos agora un exemplo que non se atopa en FNBC pero si en 3FN.

Rúa	Numero	Piso	Municipio	Provincia	País	CP
Rúa do castelo	3	4	Maceda	Ourense	España	32700
Parque	8	1	Lalin	Pontevedra	España	36500
Angelo Colocci	1	3	Lugo	Lugo	España	27003
Avenida Castelao	64	2	Ferrol	A Coruña	España	15406

Figura 2.24. Táboa sen normalizar

O esquema relacional é ENDREZO (Rua, Numero, Piso, Municipio,Provincia,País, CP) onde:

- As claves poden ser: (Rua, Numero, Piso,CP) ou (Rua, Numero, Piso, Municipio, Provincia, País), non seleccionado esta última por non ser mínima.
- As dependencias funcionais serán:
 - (Rua, Numero, Piso, Municipio, Provincia, País), \rightarrow CP; non é superclave xa que existen atributos independentes de ela como numero e piso.
 - CP \rightarrow (Municipio, Provincia, País); non é superclave xa que existen atributos independentes de ela como numero, piso, e rúa.
- Polo tanto a relación non se atopa en FNBC pero si en 3FN xa que aunque non son superclaves , si forman parte da clave.

FNBC é unha definición máis estrita da 3FN, afecta so ás relación que teñen claves solapadas, si R está en 3FN e só ten unha clave, ou ben R ten varias claves pero ningunha delas é composta, ou ben R ten varias claves compostas pero non solapadas entón R está en FNBC.



Tarefa 4: Normalizar relación a FNBC

1.2.5.5 Cuarta forma normal (4FN ou 4NF) e as dependencias multivaluadas

Para o resto de formas normales (as deseñadas por Fagin, moito máis complexas), é importante definir as dependencias multivaluadas, que é distinta das funcionais. Si las funcionais eran a base da segunda e terceira forma normal (e da de Boyce-Codd), éstas son a base da cuarta forma normal.

Unha dependencia multivaluada de α sobre β (é dicir $\alpha \twoheadrightarrow \beta$), sendo α e β atributos da mesma táboa, ocorre cuando β tene un conxunto de valores ben definidos sobre calquera valor de α . É dicir, dado α saberemos os posibles valores que pode tomar β .

Refírese a posibles valores (en plural) e se trata de que os valores dese atributo sempre son os mesmos segundo o valor dun atributo e non doutro.

No seguinte exemplo mostrase o profesor e idioma no que se dará unha clase deportiva. Profesor e idioma son atributos con múltiples valores para deporte pero independentes entre si (esta é a dependencia multivaluada). A relación resultantes é CLASE(Deporte, Profesor, Idioma):

CLASE		
Deporte	Profesor	Idioma
Tenis	Xan	Galego
Tenis	Xan	Inglés
Tenis	Brais	Inglés
Tenis	Brais	Galego
Padel	Brais	Francés
Padel	Xonxa	Francés

Figura 2.25. relación ou táboa que mostra dependencias multivaluadas

A táboa presentaredundancias que poden acarrear erros de actualización a pesar de atoparse en FNBC, xa que non existen nela dependencias funcionais elementais posto que a súa clave está formada polo conxunto dos tres atributos.

Para que a táboa adopte a 4FN, evitando os inconvenientes de actualización, descomporase en dúas relacións cuxa unión sen perda (a unión de ambas táboas da como resultado as mesmas tuplas que a relación orixe).

Deporte	Profesor	Deporte	Idioma
Tenis	Xan	Tenis	Galego
Tenis	Xan	Tenis	Inglés
Tenis	Brais	Tenis	Inglés
Tenis	Brais	Tenis	Galego
Padel	Brais	Padel	Francés
Padel	Xonxa	Padel	Francés

Figura 2.26. Relacións obtidas tras a normalización en 4FN

O teorema de Fagin indica que cando existen tres pares de conxuntos de atributo X, Y, Z se $X \twoheadrightarrow Y$ e $X \twoheadrightarrow Z$ (tendo Y, Z dependencia multivaluada sobre X).

Un teorema de Fagin indica cuando hay tres pares de conjuntos de atributos X, Y e Z si ocorre $X \twoheadrightarrow Y$ e $X \twoheadrightarrow Z$ (Y y Z tienen dependencia multivaluada sobre X), entón as táboas X, Y e X, Z reproducen sen perda de información o que posuía la táboa orixinal. Este teorema é o que se aplica a hora de dividir as táboas e alcanzar a 4FN.

1.2.5.6 Quinta forma normal (5FN ou 5NF) dependencias de JOIN ou reunión

Ocorre cando unha táboa esta en 4FN e cada dependencia de unión ($JOIN\backslash$) nela é implicada por crávelas candidatas.

É a mais complexa e non está claro que o paso das relacións a 5FN mellore a base de datos, ademais, é raro atoparse este tipo de problemas cando a normalización chega a 4FN (ou incluso quedar na 3FN ou na FNBC). Débense a restricións semánticas especiais aplicadas sobre a táboa.

Operación JOIN ou de reunión

A operación JOIN procedente tamén da álgebra relacional, consiste en formar unha táboa coa unión de dúas táboas. A táboa resultante estará formada pola combinación de todas as columnas e filas de ambas, excepto as columnas e filas repetidas.

Operación de proxección

Unha proxección dunha táboa é a táboa resultante de tomar un subconxunto dos atributos dunha táboa (trátase dunha operación de álgebra relacional). É dicir unha táboa formada por unhas cantas columnas da táboa orixinal.

Dependencias de JOIN ou reunión

Dise que se ten unha táboa con dependencia de unión (ou de tipo $JOIN\backslash$) si se pode obter esa táboa como resultado de combinar mediante a operación JOIN varias proxeccións da mesma.

No seguinte exemplo móstrase a relación SUMINISTRADORES(Proveedor, Material, Proxecto), onde todos os atributos forman parte de clave. Esta relación mostra os códigos de material subministrado por un proveedor e utilizado nun determinado proxecto.

SUMINISTRADORES		
Provedor	Material	Proxecto
Prov1	Mat1	Prox2
Prov1	Mat2	Prox1
Prov2	Mat1	Prox1
Prov1	Mat1	Prox1

Figura 2.27. Relación con dependencia de JOIN

Así vista a relación, non permite ningunha proxección na que non perdamos datos.

Pero si ocorre unha restrición especial como por exemplo: Cando un proveedor subministra algunha vez un determinado material, si ese material aparece noutro proxecto, faremos que o proveedor anterior nos subministre tamén ese material para o proxecto.

Isto ocorre nos datos como o proveedor PROV1 forneceunos o material MAT1 para o PROX2 e no PROX1 utilizamos o MAT1, aparecerá a tupla PROV1, MAT1 e PROX1. Si un novo proxecto necesitase o MAT1, entón haberá que pedilo aos provedores PROV1 e PROV2 xa que xa participaron con anterioridade.

A dependencia de reunión que produce esta restrición é moi difícil de ver. Para esta restrición as táboas válidas sería:

SIMINISTRAMAT		MATPROXECTO		PROVEDORPROX	
Provedor	Material	Material	Proxecto	Provedor	Proxecto
Prov1	Mat1	Mat1	Prox2	Prov1	Prox2
Prov1	Mat2	Mat2	Prox1	Prov1	Prox1
Prov2	Mat1	Mat1	Prox1	Prov2	Prox1
Prov1	Mat1	Mat1	Prox1	Prov1	Prox1

Figura 2.28. Relacións obtidas tras a normalización a 5FN

Esa descomposición non perde valores neste caso, sabendo que si o proveedor fornécenos un material poderemos relacionarlle con todos os proxectos que utilizan ese material.

Resumindo, unha táboa non está en quinta forma normal si hai unha descomposición desa táboa que mostre a mesma información que a orixinal e esa descomposición non teña como clave a clave orixinal da táboa.

1.2.5.7 Forma normal de dominio clave (FNDC ou KDFN)

Trátase dunha forma normal enunciada por Fagin en 1981 ao darse conta dos problemas de redundancia que ocorren con algúns dominios.

Neste caso non se basa en dependencias entre os datos, senón que se basa en restricións de dominio e restricións de clave.

- **Restricións de dominio:** Trátase da restrición que fai que un determinado atributo obteña só certos valores.
- **Restrición de clave:** É a restrición que permite que un atributo ou un conxunto de atributos forme unha clave candidata.

Fagin di que unha táboa está en FNDC si toda restrición sobre a táboa é consecuencia lóxica de aplicar as restricións de dominio e clave sobre a mesma. Fagin demostra que esta situación ocorre incluso cando as táboas están normalizadas en 5FN.

No seguinte exemplo móstrase a seguinte relación ALUMNO(Nome, Nivel,Nota) con clave Nome:

ALUMNO		
Nome	Nivel	Nota
Antón	Moi alto e eficiente	9
Xulia	Moi alto pero con traballo pouco constante	9
Xenxo	Alto con traballo habitual	7
Lois	Medio e con traballo	5

Figura 2.29. Relación non normalizada

Observando os datos da táboa obsérvase que cando a nota é superior a 9, no nivel aparece a palabra alto, cando é un 7 ou un 8 medio, e un 5 ou 6 sería medio. É dicir temos restricións que non son nin de dominio nin de clave nesa táboa. O lóxico sería:

ALUMNO		
Nome	Nivel	Nota
Antón	Eficiente	9
Xulia	Traballo constante	9
Xenxo	Traballo habitual	7
Lois	Traballo medio	5

NOTA		
Nivel	Minima	Maxima
Moi alto	9	10
Alto	7	8,99
Medio	5	6,99
Baixo	0	4,99

Figura 2.30. Táboa normalizada ata a FNDC

Non se perde información ao deseñar as táboas desta forma e de feito é máis eficiente para a base de datos.

1.2.6 Outros criterios de deseño

Ademais da normalización (motivada polas anomalías de actualización) pódense ter en conta outros criterios para mellorar o deseño:

- Evitar valores NULOS:
 - –Crean problemas en operacións que implican comparacións, cálculos ou sumas.
 - –A proporción de NULOS sen un atributo pode ser un criterio para sacar o atributo a unha relación á parte.
- Semántica dos esquemas:
 - –A facilidade con que poden explicarse é unha medida informal da calidade do deseño.
- Eficiencia e desnormalización:
 - Táboas non normalizadas poden ser mais eficientes para algunhas consultas.

- Por motivos de eficiencia en ocasións compensa xuntar ou non descompoñer certas táboas: ceder espazo e robustez a cambio de eficiencia (o que afórrar operación de reunión).
- Dependendo da frecuencia de consulta, o tamaño das táboas e frecuencias de valores.

1.2.7 Desnormalización

A desnormalización é o proceso de procurar optimizar o deseño dunha base de datos por medio de agregar datos redundantes.

A pesar da existencia de algoritmos que permiten automatizar en parte o proceso, normalizar a partir dun esquema de BD “puro” ou de “laboratorio” pode ocasionar un notable esforzo. Agora ben, se aplicamos o modelo entidade-interrelación (MER) ben elaborado, é posible que a necesidade de normalización posterior se reduza a niveis mínimo ou incluso desaparece.

Por outra banda, na práctica a normalización en formas superiores á BCNF e na inmensa maioría dos casos innecesaria dada a dificultade de que se dean os requisitos para que aparezan dependencias multivaludadas.

Unha vez alcanzada a forma normal escollida, é posible que por motivos de eficiencia , a facilidade de implementación e/ou outros, se proceda a desnormalizar ou reestruturar os subesquemas resultantes.

No exemplo supoñamos unha relación CLIENTE (CodCliente, nomeCliente, rua, localidade, codPostal) onde se detectan as seguintes DF:

- DNICliente-> NomeCliente, codPostal.
- CodPostal -> localidade, provincia.

Para obter un esquema normalizado ata 3FN ou ata FNBC deberíanse crear dúas relacións do seguinte xeito:



Figura 2.31. Grafo relacional en FNBC

Pero esta solución máis normalizada empeorará o desempeño da BD o ter que mezcla (facer a operación JOIN) cada vez que se desexe obter a información completa dun cliente

1.3 Tarefas

As tarefas propostas son as seguintes:

- Tarefa 1. Normalizar relacións a 1FN.
- Tarefa 2. Normalizar relacións a 2FN.
- Tarefa 3. Normalizar relacións a 3FN.
- Tarefa 4. Normalizar relacións a FNBC.

1.3.1 Tarefa 1. Normalizar relacións a 1FN

- Tarefa 1.1: Xustifique se a seguinte relación atópase en 1FN ou non. De non atoparse nesta forma normal propoña unha descomposición válida.

ALUMNO							
dni	nome	apelidos	codPostal	provincia	modulo	profesor	nota
12345678	Uxia	Lameiro	32004	Ouren	Sistemas	A. Bonillo	9
12345678	Uxia	Lameiro	32004	Ourense	Programación	N. Jáimez	7
34567890	Xavier	Novoa	36002	Pontevedra	Programación	N. Jáimez	8
34567890	Xavier	Novoa	36002	Pontevedra	Robótica	S. López	9
23456789	Xurxo	Pazos	27080	Lugo	Robótica	S. López	9
23456789	Xurxo	Pazos	27080	Lugo	Sistemas	A. Bonillo	9

ALUMNO (dni, nome, apelido, codPostal, provincia, modulo, profesor, nota)

Solución

A relación ALUMNO atópase en 1FN xa que non hay ningún atributo que non cumpla ser atómico, e dicir, en cada tupla da relación todos os atributos toman un único valor

1.3.2 Tarefa 2. Normalizar relacións a 2FN

- Tarefa 2.1. : Xustifique se a seguinte relación atópase en 2FN ou non. De non atoparse nesta forma normal propoña unha descomposición válida

ALUMNO							
dni	nome	apelidos	codPostal	provincia	modulo	profesor	nota
12345678	Uxia	Lameiro	32004	Ouren	Sistemas	A. Bonillo	9
12345678	Uxia	Lameiro	32004	Ourense	Programación	N. Jáimez	7
34567890	Xavier	Novoa	36002	Pontevedra	Programación	N. Jáimez	8
34567890	Xavier	Novoa	36002	Pontevedra	Robótica	S. López	9
23456789	Xurxo	Pazos	27080	Lugo	Robótica	S. López	9
23456789	Xurxo	Pazos	27080	Lugo	Sistemas	A. Bonillo	9

ALUMNO (dni, nome, apelido, codPostal, provincia, modulo, profesor, nota)

Solución

Tras comprobar que a relación atópase en 1FN (todos os seus atributos son atómicos) comprobase si todos os atributos que non forman parte da clave primaria teñen dependencia completa de dita clave. Comenzarase polo tanto determinando a clave.

Para que cada tupla da relación sexa única a clave debe estar composta por dni e modulo.

Unha vez determinada a clave estudanse a DF onde todos os atributos non clave deben depender totalmente dos atributos clave

- dni -> (nome, apelido, codPostal): non depende de toda a clave.
- codPostal -> provincia: non depende de toda a clave.
- (dni, modulo) -> nota: non depende de toda a clave.
- Modulo -> profesor: non depende de toda a clave.

Polo tanto a relación non se atopa en 2FN.

Para normalizar descompondremos en táboas que si cumplan a condición 2FN (todos os atributos non clave dependen totalmente da clave).

ALUMNO				ENDEREZO	
dni	nome	apelidos	codPostal	codPostal	provincia
12345678	Uxia	Lameiro	32004	32004	Ouren
12345678	Uxia	Lameiro	32004	32004	Ourense
34567890	Xavier	Novoa	36002	36002	Pontevedra
34567890	Xavier	Novoa	36002	36002	Pontevedra
23456789	Xurxo	Pazos	27080	27080	Lugo
23456789	Xurxo	Pazos	27080	27080	Lugo

CALIFICACION			DOCENTE	
dni	profesor	nota	profesor	modulo
12345678	A. Bonillo	9	A. Bonillo	Sistemas
12345678	N. Jáimez	7	N. Jáimez	Programació
34567890	N. Jáimez	8	N. Jáimez	Programació
34567890	S. López	9	S. López	Robótica
23456789	S. López	9	S. López	Robótica
23456789	A. Bonillo	9	A. Bonillo	Sistemas

ALUMNO (dni, nome, apelido, codPostal)

ENDEREZO (codPostal, provincia)

CALIFICACION (dni, asignatura, nota)

DOCENTE (modulo, profesor)

1.3.3 Tarefa 3. Normalizar relacións a 3FN

- Tarefa 3.1: Xustifique o esquema atópase en 3Fn ou non. De non atoparse nesta forma normal propoña unha descomposición válida.

ALUMNO				ENDEREZO	
dni	nome	apelidos	codPostal	codPostal	provincia
12345678	Uxia	Lameiro	32004	32004	Ouren
12345678	Uxia	Lameiro	32004	32004	Ourense
34567890	Xavier	Novoa	36002	36002	Pontevedra
34567890	Xavier	Novoa	36002	36002	Pontevedra
23456789	Xurxo	Pazos	27080	27080	Lugo
23456789	Xurxo	Pazos	27080	27080	Lugo

CALIFICACION			DOCENTE	
dni	profesor	nota	profesor	modulo
12345678	A. Bonillo	9	A. Bonillo	Sistemas
12345678	N. Jáimez	7	N. Jáimez	Programación
34567890	N. Jáimez	8	N. Jáimez	Programación
34567890	S. López	9	S. López	Robótica
23456789	S. López	9	S. López	Robótica
23456789	A. Bonillo	9	A. Bonillo	Sistemas

ALUMNO (dni, nome, apelido, codPostal)

ENDEREZO (codPostal, provincia)

CALIFICACION (dni, asignatura, nota)

DOCENTE (modulo, profesor)

Solución

Todas as relacións atópanse en 2FN. Comprobarase se non existan dependencias funcionais transitivas en ningunha relación. As dependencias funcionais para cada unha das relacións son:

- Táboa ALUMNO: Dni-> (nome, apelido, codPostal)
- Táboa ENDEREZO: codPostal-> provincia.
- Táboa CALIFICACION: (dni, modulo)-> nota.
- Táboa DOCENTE: modulo-> profesor.

Non existe ningunh dependencia funcional polo que que si se atopa en 3FN

1.3.4 Tarefa 4 Normalizar relacións a FNBC

- Tarefa 4.1: Xustifique o esquema atópase en 3FN ou non. De non atoparse nesta forma normal propoña unha descomposición válida

ALUMNO				ENDEREZO	
dni	nome	apelidos	codPostal	codPostal	provincia
12345678	Uxia	Lameiro	32004	32004	Ouren
12345678	Uxia	Lameiro	32004	32004	Ourense
34567890	Xavier	Novoa	36002	36002	Pontevedra
34567890	Xavier	Novoa	36002	36002	Pontevedra
23456789	Xurxo	Pazos	27080	27080	Lugo
23456789	Xurxo	Pazos	27080	27080	Lugo

CALIFICACION			DOCENTE	
dni	profesor	nota	profesor	modulo
12345678	A. Bonillo	9	A. Bonillo	Sistemas
12345678	N. Jáimez	7	N. Jáimez	Programación
34567890	N. Jáimez	8	N. Jáimez	Programación
34567890	S. López	9	S. López	Robótica
23456789	S. López	9	S. López	Robótica
23456789	A. Bonillo	9	A. Bonillo	Sistemas

ALUMNO (dni, nome, apelido, codPostal)

ENDEREZO (codPostal, provincia)

CALIFICACION (dni, asignatura, nota)

DOCENTE (modulo, profesor)

Solución

Unha vez comprobado que as relación están en 3FN débese analizar si nas relacións todo determinante é clave candidata. Nas relacións desta base de datos todas as relacións teñen un único determinante (parte X da DF) que ademais conincide coa clave primaria da relación polo que todas elas están en FNBC

2. Materiais

2.1 Documentos de apoio ou referencia

- [EN 2002] ELMASRI, R.;NAVATHE, S.B.*Fundamentos de Sistemas de Bases de Datos* Addison-Wesley, 2002.
- [MPM 1999] DE MIGUEL, A; PIATTINI, M; MARCOS, E. *.Diseño de base de datos relacionales*. Ra-MA. 1999.
- CONNOLLY, T; BEGG, C; STRACHAN, A. *Database system: A practical aproach de-sing, implementation and magnagement*.Addisson-Wesley, 1998.
- SILBERSCHATZ,A; KORTH. H; SUDARSHAN, S; CONNOLLY, T; BEGG, C; STRACHAN, A. *Fundamentos de bases de datos*. McGraw-Hill, 1998.
- DATE,C.J. *Introducción a los sistemas de bases de datos*. Addisson-Wesley, 1992.
- DE MIGUEL, A; PIATTINI, M. *Concepción y diseño de bases de datos*. Ra-Ma, 1993.
- DE MIGUEL, A; PIATTINI, M. *Fundamentos y modelos de bases de datos*. Ra-Ma, 1993.
- Métrica:
 - http://administracionelectronica.gob.es/pae_Home/pae_Documentacion/pae_Metodolog/pae_Metrica_v3.html#.ViS6eGThBz8