

NORMALIZACIÓN

1.- INTRODUCCIÓN A LAS FORMAS NORMALES

- 1.1.- Necesidad de la Normalización.
- 1.2.- Tipos de Formas Normales.
- 1.3.- Noción intuitiva de las F.N.

2.- CONCEPTOS PREVIOS

- 2.1.- Dependencia funcional.
- 2.2.- Dependencia funcional completa.
- 2.3.- Dependencia funcional trivial.
- 2.4.- Dependencia funcional elemental.
- 2.5.- Dependencia funcional transitiva.
- 2.6.- Representación gráfica de las dependencias funcionales.
- 2.7.- Revisión del concepto de clave.

3.- TEORIA FORMAL DE LA NORMALIZACIÓN

- 3.1.- Conservación de la información

4.- FORMAS NORMALES

- 3.1.- Primera Forma Normal.
- 3.2.- Segunda Forma Normal.
- 3.3.- Tercera Forma Normal.
- 3.4.- Forma Normal de Boyce Codd.

5.- EJERCICIOS

1.- INTRODUCCIÓN A LAS FORMAS NORMALES

La Teoría de la Normalización la definió Codd en el año 1970, se basa en la idea de que "los datos son independientes de los programas" y se trata de una serie de reglas o normas a seguir en el diseño de la estructura de datos que va a utilizar el sistema de información.

Las ventajas derivadas de la normalización son las siguientes:

- *Facilidad de uso.* Los datos se agrupan en tablas que identifican claramente cada entidad o relación.
- *Precisión.* Se consigue mantener información diferente relacionada con toda exactitud.
- *Seguridad.* Los controles de acceso resultan sencillos de implementar tanto a nivel de relaciones como de atributos.
- *Independencia de datos.* Los programas no están unidos a las estructuras, y se pueden añadir nuevas tablas o nuevos atributos sin que afecte a los programas que los utilizan.
- *Claridad.* La representación de la información es clara y sencilla para el usuario.
- *Facilidad de gestión.* Los lenguajes manipulan la información de forma sencilla.
- *Mínima redundancia.* La información no se repite de forma innecesaria.
- *Máximo rendimiento de las aplicaciones.* Cada aplicación trata solamente los datos que le son de utilidad.

El objetivo último de la normalización es obtener una estructura optima en el diseño lógico de los datos. Es decir que cada tabla esté constituida por los atributos imprescindibles para representar la entidad o relación de la que se deriva.

Desde el punto de vista práctico, puede considerarse como un "*refinamiento*", *último y necesario*, del *esquema relacional* obtenido del esquema conceptual (esquema E-R).

1.1.- NECESIDAD DE LA NORMALIZACIÓN

Tanto en el modelo relacional como en los demás modelos, el diseño de una base de datos puede abordarse de dos formas distintas:

- a) Obteniendo es esquema relacional directamente a partir de la observación del universo del discurso.
- b) Realizando el proceso de diseño en las dos fases ya vistas:
 1. Diseño conceptual, en nuestro caso en el modelo E/R, que permite obtener el esquema conceptual.
 2. Transformación del esquema conceptual en un esquema relacional.

Estas relaciones obtenidas por uno u otro proceso, de hecho el modelo E/R de Chen apareció seis años después que la Teoría de la Normalización, pueden presentar algunos problemas derivados de fallos en la percepción del UD o errores en la transformación al modelo relacional.

Los problemas más destacables son los siguientes:

- Incapacidad para almacenar ciertos hechos.
- Redundancias, y por tanto, posibilidad de incoherencias.
- Ambigüedades.
- Pérdida de información (aparición de tuplas espúreas).
- Pérdida de determinadas restricciones que dan lugar a interdependencia entre los datos. Esto se denomina pérdida de dependencias funcionales (el concepto de dependencia funcional se verá más adelante en esta unidad).
- Aparición entre los datos situaciones que no son válidos en el mundo real; es decir, anomalías de inserción, borrado y modificación.

Por tanto, el esquema relacional debe ser analizado para comprobar que no presenta los problemas anteriormente citados.

Veamos un ejemplo de diseño incorrecto:

Sea la relación ALQUILERES de una base de datos de una empresa de alquiler de coches, que se muestra a continuación.

ALQUILERES

Matrícula	Modelo	Marca	FechaAlquiler	Kms	NIFCliente	Nombre
BU-3456-AB	R 16	Renault	12/12/1999	3.000	13098456	José Martínez Peña
BU-4405-CA	Fiat Uno	Fiat	16/12/1999	5.550	21095465	Arturo Ortiz Majín
BU-3456-AB	R 16	Renault	18/12/1999	1.024	33097358	Ana Marín Osés
BU-3456-AB	R 16	Renault	23/12/1999	2.404	70794569	Andrés Delgado Pí
SO-0234-K	A3	Audi	23/12/1999	4432	33097358	Ana Marín Osés
BU-4405-CA	Fiat Uno	Fiat	26/12/1999	1053	13098456	José Martínez Peña

Contiene datos sobre vehículos (Matrícula, Modelo y Marca), sobre los clientes (NIFCliente y Nombre) y sobre el alquiler de los vehículos (FechaAlquiler y Kms). La clave estaría formada por (Matrícula, NIFCliente y FechaAlquiler).

Analizando la relación ALQUILERES, vemos que presenta varios de los problemas enumerados anteriormente:

- Redundancias. Se repiten las matrículas, los modelos, los NIF y los nombres de los clientes.
- Anomalías de modificación. Es consecuencia del apartado anterior. Si por error nos equivocamos al introducir el NIF del cliente o el nombre, tendríamos NIF distintos asociados a nombres iguales o al contrario. Lo mismo podría ocurrir con los coches y las matrículas. Evidentemente, esto produce incoherencias en los datos.
- Anomalías de inserción. Ya que si se deseara introducir información sobre un nuevo coche adquirido por la empresa, al formar parte de la clave el atributo Matrícula, tendríamos que esperar a que el vehículo en cuestión fuese alquilado, ya que la clave la forman también los atributos NIFCliente y FechaAlquiler.
- Anomalías de borrado. Si deseáramos dar de baja un vehículo, se perderían los datos de los clientes que lo hubieran alquilado y viceversa.

Esta relación presenta todos estos problemas, y alguno más, debido a que atenta contra el principio básico en todo diseño:

"hechos distintos se deben almacenar en objetos distintos".

En nuestro caso, las relaciones resultantes de aplicar dicho principio son las siguientes:

VEHICULOS (Matrícula, Modelo)

CLIENTES (NIFCliente, Nombre)

ALQUILER (Matrícula, NIFCliente, FechaAlquiler, Kms)

La pregunta que surge de inmediato: ¿Un diseñador poco experimentado puede llegar al diseño adecuado, en cualquiera de los múltiples UD que se puedan presentar, con una regla tan genérica como la enunciada?

La respuesta se encuentra en la primera modelización, que se debe realizar estableciendo los esquemas conceptuales por medio del modelo E/R, la posterior traducción a los esquemas relacionales y, por último, seguir las reglas que establecidas en la teoría de la normalización.

La teoría de la normalización puede definirse como una técnica formal para organizar datos, ya que nos ayuda a determinar aquello que está equivocado y nos enseña la manera de corregirlo.

EJERCICIO 1

Determinar los atributos que forman la clave de la relación BIBLIOTECA, que se da a continuación, y después determina el esquema relacional que consideras correcto.

BIBLIOTECA

Autor	Nacionalidad	CodLibro	Título	Editorial	AñoPub
Canosa, José	Española	98765	Curso lenguaje C	Marcombo	1992
Canosa, José	Española	98777	C++ Práctico	Marcombo	1994
García, Fco.	Española	98789	Prog. Aplic. Windows	McGrawHill	1994
Miguel, A. de	Española	88123	Concepción y Diseño de BD	ra-ma	1993
Piattini, Mario	Española	88123	Concepción y Diseño de BD	ra-ma	1993
Warnier J.D.	Francesa	78976	Programación Lógica	eta,s.a.	1973
Flanagan B.M.	Francesa	78976	Programación Lógica	eta,s.a.	1973
Codd, E. F.	Norteamericana	88345	The Relational Model for DM	Addison-W	1990

1.2.- TIPOS DE FORMAS NORMALES

La teoría de la normalización consta de seis reglas o *formas normales*, enunciadas en orden creciente de restricción, de forma que para que una relación cumpla con una de ellas, es necesario que además de cumplir las especificaciones de dicha forma normal, debe cumplir con los requisitos de todas las anteriores.

En consecuencia la aplicación debe hacerse siempre de forma progresivamente creciente partiendo de la primera (1FN).

Codd fue el que definió las tres primeras formas normales (1FN, 2FN y 3FN). Posteriormente, en 1971, Heath volvió a formular la 3FN definida por Codd, ya que algunas relaciones en 3FN presentan problemas que no se evitan con la definición de Codd. Esta nueva forma normal, redefinida en 1974 por Boyce y Codd, se conoce en la actualidad con el nombre de *Forma Normal de Boyce/Codd* (FNBC).

En 1977, Fagin definió la cuarta forma normal (4FN), denominada así porque en ese momento se conocía la FNBC como la 3FN. Por último, en 1979, este mismo investigador junto con Zaniolo (1981) define la quinta forma normal (5FN), que constituye la última posible.

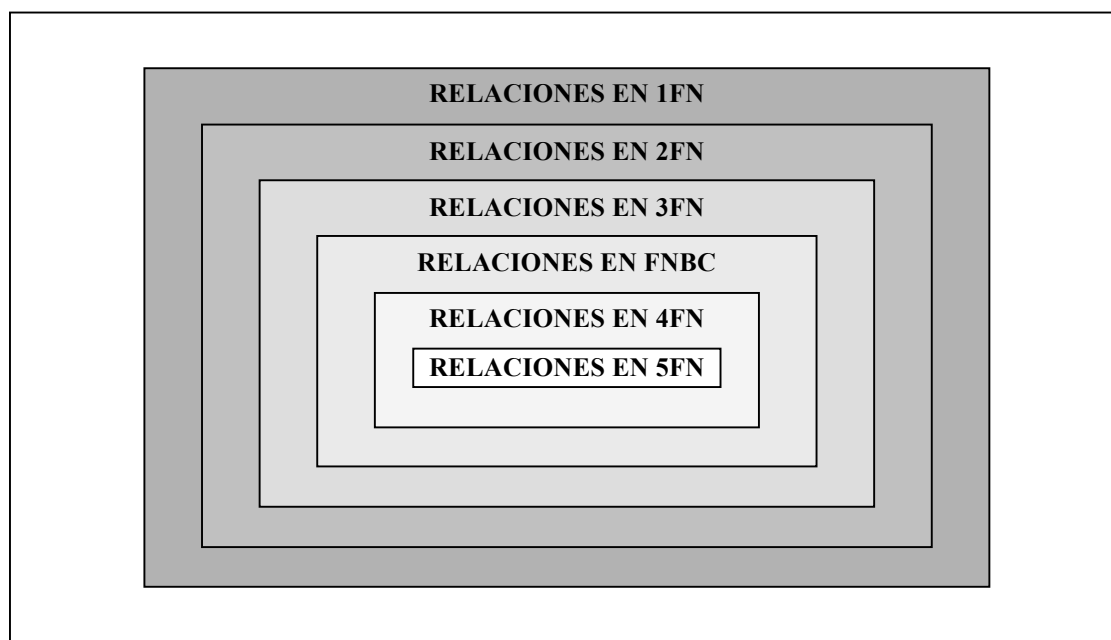


Fig. 1. Representación de las Formas Normales

Como se indicó anteriormente y se refleja en la figura, todas las relaciones que están en 5FN, también están en 4FN, en FNBC, etc. Sin embargo, aquellas relaciones que estén en 1FN no están en 2FN ni en 3FN, etc.

A fin de evitar las anomalías descritas anteriormente, es preferible una relación en 2FN que en 1FN, una relación en 3FN que en 2FN, etc.

La teoría de la normalización se basa en ciertas restricciones definidas sobre los atributos de una relación, estas restricciones se conocen con el nombre de dependencias. Existen varios tipos de dependencias:

- Dependencias funcionales. Están relacionadas con la 2FN, 3FN y FNBC.
- Dependencias multivaluadas. Relacionadas con la 4FN.
- Dependencias de proyección/combinación. Relacionadas con la 5FN.

En la práctica se normaliza hasta la FNBC, ya que un exceso de normalización produce un exceso de tablas que puede resultar poco operativo, de hecho algunas veces después de normalizar hay que recorrer el camino inverso *desnormalizar* para hacer operativa la BD, por ello en esta unidad se verán sólo las dependencias funcionales.

1.3.- NOCIÓN INTUITIVA DE LAS F.N.

La aproximación intuitiva a las formas normales nos da idea de la finalidad que se persigue con cada una de ellas y pretende ser el primer paso en la organización en la forma de pensar del diseñador.

Primera forma normal

Una relación se encuentra en 1FN si no existen grupos repetitivos; es decir, que para una determinada tupla, un atributo tome más de un valor de su dominio.

Por ejemplo, la relación ALQUILER que se da a continuación, no está en 1FN.

ALQUILER

Matrícula	Modelo	Marca	NIFCliente	Nombre
BU-3456-AB	R 16	Renault	13098456 33097358 70794569	José Martínez Peña Ana Marín Osés Andrés Delgado Pí
BU-4405-CA	Corsa	Opel	13098456 21095465	José Martínez Peña Arturo Ortiz Majín
SO-0234-KO	A3	Audi	33097358	Ana Marín Osés

Si embargo la relación ALQUILER1 sí está en 1FN.

ALQUILER1

Matrícula	Modelo	Marca	NIFCliente	Nombre
BU-3456-AB	R 16	Renault	13098456	José Martínez Peña
BU-3456-AB	R 16	Renault	33097358	Ana Marín Osés
BU-3456-AB	R 16	Renault	70794569	Andrés Delgado Pí
BU-4405-CA	Corsa	Opel	13098456	José Martínez Peña
BU-4405-CA	Corsa	Opel	21095465	Arturo Ortiz Majín
SO-0234-KO	A3	Audi	33097358	Ana Marín Osés

En realidad, si una tabla se encuentra en 1FN, se considera que está normalizada, aunque se debe procurar que el nivel de normalización sea el óptimo.

Segunda forma normal

Una relación se encuentra en 2FN si, además de estar en 1FN, todos los atributos que no forman parte de ninguna clave candidata suministran información acerca de la clave completa.

Analicemos la relación ALQUILER, de primer ejemplo:

ALQUILER (Matrícula, Modelo, Marca, Nombre, NIFCliente, FechaAlquiler, Kms)

Las claves candidatas son:

(Matrícula, NIFCliente, FechaAlquiler)
(Matrícula, Nombre, FechaAlquiler)

La segunda clave es válida si suponemos que no existen clientes con el mismo nombre.

Se puede observar que ciertos atributos, que no forman parte de ninguna clave candidata, no aportan información sobre la clave completa, tal es el caso de Modelo y Marca. Luego la relación ALQUILER no se encuentra en 2FN.

La solución es crear una nueva relación con dichos atributos y la clave a la que aportan información. Las relaciones resultantes son las siguientes:

ALQUILER1 (Matrícula, Nombre, NIFCliente, FechaAlquiler, Kms)
VEHICULOS (Matrícula, Modelo, Marca)

En ALQUILER1, el único atributo que no forma parte de ninguna clave es Kms pero aporta información sobre la totalidad de ambas claves candidatas.

En VEHICULOS, la clave es Matrícula y tanto Modelo como Marca aportan información sobre la clave completa.

Por tanto ambas relaciones se encuentran en 2FN.

Tercera forma normal

Una relación se encuentra en 3FN si, además de estar en 1FN y 2FN, los atributos que no forman parte de ninguna clave candidata facilitan información sólo acerca de la/s clave/s candidata/s y no acerca de otros atributos.

Se dice que no existen dependencias transitivas.

En la relación ALQUILER1, el atributo Kms sólo facilita información acerca de las claves, luego si está en 3FN.

En cambio, en la relación VEHICULOS, el atributo Marca facilita información acerca de la casa que ha fabricado el modelo. En realidad, Marca facilita información a Matrícula a través del atributo Modelo. En consecuencia, VEHICULOS no se encuentra en 3FN.

Para solucionarlo realizamos una nueva descomposición. Las relaciones resultantes, son las siguientes:

VEHICULOS1 (Matrícula, Modelo)
MODELOS (Modelo, Marca)

Las dos relaciones resultantes sí están en 3FN, ya que en cada una de ellas el atributo no clave sólo aporta información sobre la clave. Evidentemente, en la relación MODELOS la clave es Modelo.

Considerando de nuevo la relación ALQUILER1, observamos que a pesar de encontrarse en 3FN, los atributos Nombre y NIFCliente se repiten innecesariamente. Para solucionar este problema disponemos de la FNBC.

Forma normal de Boyce/Codd

Puede considerarse como una redefinición más estricta de la 3FN. Se centra en los atributos de la clave.

Una relación está en FNBC si las claves candidatas son los únicos descriptores sobre los que se facilita información por cualquier otro atributo. Es decir, que cualquier atributo de una clave candidata no aporte información sobre cualquier otro atributo de la relación.

En nuestro ejemplo, la relación

ALQUILER1 (Matrícula, Nombre, NIFCliente, FechaAlquiler, Kms)

está en 3FN, pero no está en FNBC, porque NIFCliente es una información acerca de Nombre y viceversa, pero ninguno de estos atributos son clave (aunque formen parte de la clave).

Para solucionarlo, descomponemos la relación en dos y nos quedan las siguientes relaciones:

CLIENTES (NIFCliente, Nombre)

ALQUILER2 (Matrícula, NIFCliente, FechaAlquiler, Kms)

Las cuales sí están en FNBC. Las relaciones VEHICULOS1 y MODELOS al tener la clave formada por un sólo atributo también están en FNBC.

En relación con esta forma normal también hay que tener presente las siguientes ideas:

*Si una relación tiene varias claves candidatas y éstas tienen atributos comunes se dice que dicha relación tiene **claves solapadas**.*

Si una relación tiene varias claves candidatas que no están solapadas y está en 3FN está también en FNBC.

Si una relación está 3FN con claves candidatas solapadas y los atributos que las forman son los únicos de la relación también está en FNBC.

Por tanto, la relación ALQUILER después del proceso de normalización quedaría convertida en cuatro relaciones, todas ellas en FNBC.

VEHICULOS1 (Matrícula, Modelo)

MODELOS (Modelo, Marca)

CLIENTES (NIFCliente, Nombre)

ALQUILER2 (Matrícula, NIFCliente, FechaAlquiler, Kms)

EJERCICIO 2

Analiza la relación PRESTAMOS(CodLibro, Título, Editorial, País, FechaPréstamo, NumSocio, NombreSocio) y contesta las siguientes cuestiones:

1. Determinar sus claves candidatas.
2. Indicar la forma normal en que se encuentra
3. Realizar las transformaciones adecuadas para que cumpla con la FNBC.

2.- CONCEPTOS PREVIOS

Antes de afrontar el estudio formal de las formas normales es conveniente conocer las ideas fundamentales de la teoría de las dependencias, sobre todo aquellas que hacen referencia a las dependencias funcionales, y profundizar en algunos aspectos del concepto de clave.

2.1.- DEPENDENCIA FUNCIONAL

Como se indicó en el apartado 1.2 las dependencias funcionales son un determinado tipo de dependencias, que es el más extendido y que está relacionado con las tres formas normales más relevantes desde el punto de vista del diseño: 2FN, 3FN y FNBC.

Dada una relación cuyo esquema es $R(X, Y)$, se dice que un atributo (o conjunto de atributos) Y depende funcionalmente del atributo (o conjunto de atributos) X si, y solo si, cada valor de X determina o implica un único valor de Y .

Se representa por: $X \longrightarrow Y$

Al atributo o conjunto de atributos X e Y se les denomina **descriptores**.

Al descriptor que se encuentra a la izquierda del símbolo de implicación, en este caso X , se denomina **determinante** o **implicante**.

Por ejemplo, entre DNI y Nombre existe una dependencia funcional ya que un DNI determina un, y solo un, Nombre.

Lo representamos por: $\text{DNI} \longrightarrow \text{Nombre}$

Y lo interpretamos como:

DNI determina Nombre, o bien, Nombre depende funcionalmente de DNI.

Si suponemos que los nombres no se repiten tendríamos que:

$\text{Nombre} \longrightarrow \text{DNI}$ (Nombre determina DNI)

En este caso tenemos que la dependencia funcional se produce de forma biunívoca, entonces se dice que los **descriptores son equivalentes** y se puede representar por :

$\text{DNI} \longleftrightarrow \text{Nombre}$

Veamos otras situaciones:

Supongamos que una persona vive en una sola casa, considerando los atributos DNI y CALLE tendríamos:

$\text{DNI} \longrightarrow \text{Calle}$ (DNI determina Calle)

Pero Calle no determina DNI, porque en una misma calle viven muchas personas.

Esta situación la representamos: $\text{Calle} \not\longrightarrow \text{DNI}$.

Hay atributos entre los que no existe ninguna dependencia funcional.

Por ejemplo, si consideramos CodAutor y TítuloLibro, tendremos:

CodAutor \nrightarrow TítuloLibro, ya que un autor puede haber escrito muchos libros.

TítuloLibro \nrightarrow CodAutor, puesto que hay libros escritos por varios autores.

Este caso lo podemos reflejar con la siguiente notación:

CodAutor \longleftrightarrow TítuloLibro

Determinante o implicante compuesto

Hay situaciones en las que no basta un atributo para determinar un único valor de otro atributo. Consideremos la relación EMPLEADOS (DNI, Empresa, Sueldo) en la que una misma persona puede trabajar en varias empresas y dependiendo de la empresa percibe diferente sueldo:

DNI, Empresa \longrightarrow Sueldo

También se da la situación contraria, un único atributo determina varios atributos:

DNI \longrightarrow Nombre | Calle

El proceso de normalización se basa en distintas variantes de la dependencia funcional. Por tanto, es importante conocer las dependencias funcionales que relacionan atributos de distintas entidades.

2.2.- DEPENDENCIA FUNCIONAL COMPLETA

Para un determinante X compuesto: X (X1, X2), se dice que Y tiene una dependencia funcional completa o total de X, si depende funcionalmente de X, pero no depende de ningún subconjunto de X. Es decir:

$$\left. \begin{array}{l} X \longrightarrow Y \\ X1 \nrightarrow Y \\ X2 \nrightarrow Y \end{array} \right\} \text{ Se representa por } X \Rightarrow Y$$

Por ejemplo. En la relación ARTICULOS (Artículo, Revista, Fecha, Página)
Se dan las dependencias funcionales siguientes:

Artículo, Revista, Fecha \longrightarrow Página
 Artículo \nrightarrow Página
 Revista \nrightarrow Página
 Número \nrightarrow Página

Por tanto Página depende funcionalmente, de forma completa, de Artículo, Revista, Fecha.
Es decir:

Artículo, Revista, Fecha \Rightarrow Página

Por el contrario, en la relación EMPLEADOS (DNI, NombreEmpl, Empresa, Sueldo), con las mismas condiciones que el caso anterior: una misma persona puede trabajar en varias empresas y dependiendo de la empresa percibe diferente sueldo.

Si estableciéramos: $\text{DNI, Empresa} \longrightarrow \text{NombreEmpl}$

Tendríamos una dependencia no completa, ya que nombre queda determinado sólo con DNI. Decimos entonces, que Empresa es un atributo *redundante*, también llamado **ajeno o extraño**.

Sin embargo: $\text{DNI, Empresa} \longrightarrow \text{Sueldo}$

sí es una dependencia funcional completa ya que para conocer el sueldo necesitamos conocer la persona y la empresa.

2.3.- DEPENDENCIA FUNCIONAL TRIVIAL

Una dependencia funcional $X \longrightarrow Y$ se dice que es trivial si Y es un subconjunto de X.

Por ejemplo, son triviales las siguientes dependencias:

$\text{CodEmpleado} \longrightarrow \text{CodEmpleado}$
 $\text{CodEmpleado, Delegación} \longrightarrow \text{Delegación}$

2.4.- DEPENDENCIA FUNCIONAL ELEMENTAL

Decimos que una dependencia funcional $X \longrightarrow Y$ es elemental si Y es un atributo único no incluido en X, y no existe X' incluido en X tal que $X' \longrightarrow Y$; es decir, *la dependencia funcional elemental es una dependencia funcional completa no trivial en la que el implicado es un atributo único.*

Las dependencias funcionales elementales son las únicas que son útiles para la teoría de la normalización.

EJERCICIO 3

Indica las dependencias funcionales de las siguientes relaciones:

1. PROFESORES (NombreProfesor, Especialidad, Instituto, horas)
Un profesor sólo tiene una especialidad, pero puede impartir en varios institutos.
2. PRESTAMOS (CodLibro, Título, Editorial, FechaPresta, NumSocio, NombreSocio)
En la BD sólo figuran los préstamos en vigor.

2.5.- DEPENDENCIA FUNCIONAL TRANSITIVA

Se produce cuando un atributo tiene dependencia funcional de otro a través de un tercero.

Sean tres subconjuntos distintos de atributos X, Y y Z, pertenecientes a la misma relación, en la que se cumplen las dependencias funcionales reflejadas en la siguiente figura:

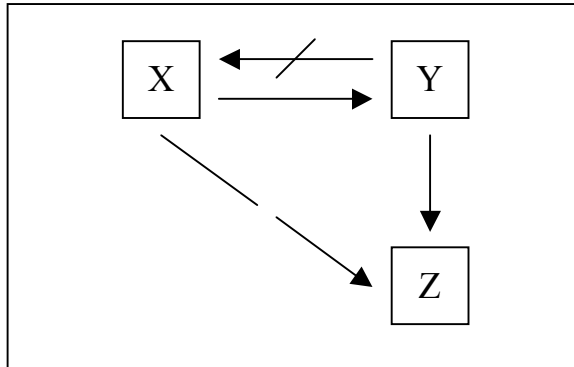


Fig. 2. Definición de dependencia funcional transitiva

Es decir, si las dependencias entre dichos atributos son:

$$\left. \begin{array}{l} X \longrightarrow Y \\ Y \longrightarrow Z \\ Y \not\longrightarrow X \end{array} \right\} \quad \text{Se dice que Z tiene una dependencia funcional transitiva respecto a X a través de Y.}$$

Se representa por: $X \text{ --- } \longrightarrow Z$

Si además, $Z \not\longrightarrow Y$ se dice que la dependencia transitiva es **estricta**.

Por ejemplo, consideremos la relación: VEHICULOS (Matrícula, Modelo, Marca)

Las dependencias funcionales son las siguientes:

Matrícula \longrightarrow Modelo
 Modelo \longrightarrow Marca
 Modelo $\not\longrightarrow$ Matrícula

Por tanto, la dependencia funcional entre Matrícula y Marca es una dependencia transitiva a través de modelo, y la representamos por:

Matrícula $\text{--- } \longrightarrow$ Marca

Es decir, Marca aporta información a Matrícula pero indirectamente, ya que completa la información de Modelo.

Al ser Matrícula el atributo clave de la relación, nos pone de manifiesto la definición intuitiva que hacíamos de la 3FN, donde decíamos que los atributos secundarios deben facilitar información sólo acerca de la clave candidata y no acerca de otros atributos.

Además, como $\text{Marca} \not\longrightarrow \text{Modelo}$, estamos ante una dependencia transitiva estricta.

EJERCICIO 4

Indica las dependencias funcionales de las siguientes relaciones:

1. ALUMNOS (NumMatrícula, NombreAlumno, Grupo, Aula)
Cada alumno se asigna a un grupo y cada grupo tiene un aula.
Pueden existir alumnos con el mismo nombre.
2. EDITORES (CodLibro, Título, Editorial, País)
No existen títulos iguales y cada editorial publica en un solo país.
3. LIBROS (CodLibro, Título, Idioma, Autor)
Pueden existir títulos iguales.

2.6.- REPRESENTACIÓN GRÁFICA DE LA DEPENDENCIAS FUNCIONALES

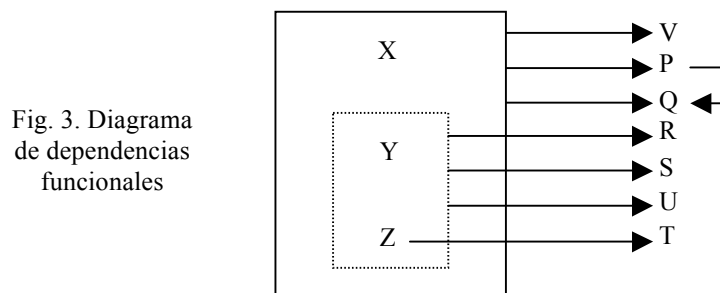
Los atributos de una relación y sus dependencias funcionales se pueden expresar gráficamente por medio de un diagrama de dependencias funcionales.

Este grafo o diagrama está formado por los nombres de los atributos unidos por las flechas de las dependencias. Cuando el implicante de una dependencia está formado por varios atributos, estos se encierran en un recuadro. Si existe un implicante subconjunto de otro, que es compuesto o implica a varios atributos secundarios, se encierra con trazo discontinuo para facilitar la interpretación del diagrama.

Por ejemplo, sea la relación: R (X, Y, Z, P, Q, R, S, T, U, V), en la que se dan las siguientes correspondencias:

$$\begin{aligned} X,Y,Z &\longrightarrow P|Q|V \\ P &\longrightarrow Q \\ Y,Z &\longrightarrow R|S|U \\ Z &\longrightarrow T \end{aligned}$$

El gráfico asociado es el siguiente:



EJERCICIO 5

Realiza el diagrama de dependencias funcionales de las siguientes relaciones:

1. ALUMNOS (NumMatrícula, NombreAlumno, Grupo, Aula)
2. PUBLICA (Artículo, Revista, Número, Página)
3. ALQUILER (Matrícula, Modelo, Marca, Nombre, NIFCliente, FechaAlquiler, Kms)

2.7.- REVISIÓN DEL CONCEPTO DE CLAVE

Concepto de clave

En una primera aproximación y como ya se vio en unidades anteriores, podemos definir clave como el atributo o conjunto mínimo de atributos que permiten identificar de forma unívoca cada tupla de la relación.

Superclave

Se denomina superclave (SK) al conjunto de atributos de una relación que determinan el resto de los atributos de la relación.

Clave candidata

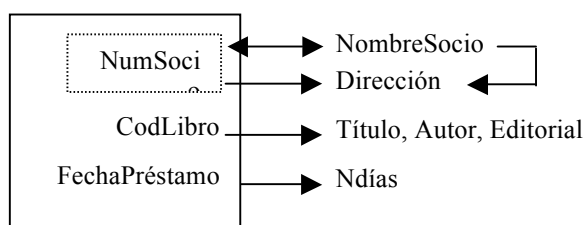
Una clave es candidata, cuando además de ser superclave, en el conjunto de atributos que la forman no existe ningún subconjunto de ellos que también determine los atributos de la relación (condición de minimalidad).

La clave candidata, es por tanto, un caso particular de superclave.

Por ejemplo, sea la relación:

BIBLIOTECA (NumSocio, NombreSocio, Dirección, CodLibro, Título, Autor, Editorial
FechaPréstamo, Ndías)

El diagrama de dependencias funcionales es el siguiente:



Existen dos superclaves: NumSocio, CodLibro, FechaPréstamo y NombreSocio, CodLibro, FechaPréstamo. Como ambos casos son paralelos, realizamos el estudio con la primera.

Vamos a verificar que el descriptor de la superclave: {NumSocio, CodLibro, FechaPréstamo}, constituye el conjunto minimal.

Es decir:

- NumSocio sólo determina NombreSocio y Dirección
- CodLibro sólo determina Título, Autor y Editorial
- FechaPréstamo por sí solo no determina a ningún atributo.
- NumSocio, CodLibro no determina a FechaPréstamo, ya que un mismo socio puede tomar prestado el mismo libro en diferentes fechas. Por tanto, no determina a la totalidad de los atributos de BIBLIOTECA.
- NumSocio, FechaPréstamo no determinan a ningún atributo.
- CodLibro, FechaPréstamo no determinan a ningún atributo. (Podría darse el caso de que el mismo día un mismo libro sea prestado a más de un socio).

En consecuencia, ninguno de los subconjuntos de la clave determinan la totalidad de los atributos de la relación, luego el descriptor implicante formado por NumSocio, CodLibro y FechaPréstamo es un conjunto minimal y por tanto es clave candidata.

Por las mismas razones, también lo es NombreSocio, CodLibro, FechaPréstamo.

Atributos principales

Se denominan atributos principales de una relación, aquellos que forman parte de las claves candidatas, mientras que al resto se les denomina atributos secundarios o no principales.

3.- TEORÍA FORMAL DE LA NORMALIZACIÓN

Normalizar consiste en convertir un esquema relacional de partida, que podemos considerar formado por un conjunto de atributos y un conjunto de dependencias funcionales, en una serie de relaciones que cumplan determinadas condiciones, fundamentalmente exentas de redundancias, y sean equivalentes al esquema inicial.

El proceso propuesto por Codd consiste en partir de una única relación constituida por todos los atributos que intervienen en el universo del discurso, denominada **relación universal**, y realizar descomposiciones sucesivas basadas en el operador de proyección teniendo en cuenta las dependencias entre los atributos.

Se trata por tanto, de buscar un conjunto de esquemas, libres de redundancias en la medida de lo posible, que sean equivalentes al esquema que soporta la relación universal.

Estos esquemas resultantes deben tener tres propiedades:

1. **Conservación de la información.** Ver apartado 3.1.
2. **Conservación de las dependencias.** Las dependencias son restricciones que recogen la semántica del universo del discurso, y por tanto, deben quedar reflejadas en el esquema de nuestra base de datos.

En consecuencia, el conjunto de dependencias de las relaciones resultantes del proceso de normalización debe ser equivalente al conjunto de dependencias de la relación de partida. Es lo que se denomina **equivalencia de dependencias**. Nosotros, nos centraremos en las dependencias funcionales.

3. **Mínima redundancia de los datos.** Es el objetivo fundamental de la normalización de las relaciones.

Además de las condiciones anteriores, Kambayashi (1982) indica que un buen diseño debe cumplir además los siguientes objetivos:

- **Minimización de las dependencias.** Este objetivo implica minimizar el número de dependencias y el número de atributos contenidos en ellas.
- **Minimización de los esquemas resultantes.** Minimizar el número de relaciones y el número de sus atributos.

Evidentemente, no siempre es posible conseguir todos los objetivos; por tanto, habrá que optar por la solución más conveniente desde el punto de vista de la utilización que se va a dar a la base de datos.

3.1.- CONSERVACIÓN DE LA INFORMACIÓN

El proceso de normalización debe llevarse a cabo sin pérdida de la información que recoge la relación de partida, es lo que a veces se llama *equivalencia de datos*.

Esta propiedad lleva implícitas dos condiciones:

1. **Conservación de los atributos.** El conjunto de los atributos de los esquemas resultantes debe ser igual que el conjunto de atributos del esquema de origen.
2. **Conservación del contenido de las tuplas.** La combinación natural (JOIN) de las relaciones resultantes debe producir la relación de origen.

Si el proceso no se lleva adecuadamente, la combinación de las relaciones resultantes puede producir tuplas que no estaban en la relación original, son las que se denominan **tuplas espúreas**. Por ejemplo, consideremos la relación VEHICULOS:

VEHICULOS

Matrícula	Modelo	Marca
BU-3460-AB	R 16	Renault
BU-4070-AC	R 5	Renault
BU-4490-CA	Xantia	Citroen
SO-0240-KO	A3	Audi

Tratamos de normalizarla creando las dos relaciones siguientes:

COCHES

Matrícula	Marca
BU-3460-AB	Renault
BU-4070-AC	Renault
BU-4490-CA	Citroen
SO-0240-KO	Audi

MARCAS

Modelo	Marca
R 16	Renault
R 5	Renault
Xantia	Citroen
A3	Audi

Verificamos con la operación JOIN:

COCHES*MARCAS

Matrícula	Modelo	Marca
BU-3460-AB	R 16	Renault
BU-3460-AB	R 5	Renault
BU-4070-AC	R 16	Renault
BU-4070-AC	R 5	Renault
BU-4490-CA	Xantia	Citroen
SO-0240-KO	A3	Audi

} Tuplas espúreas

Observamos que hay pérdida de información ya que han aparecido tuplas espúreas. La pretendida normalización se ha realizado mal.

4.- FORMAS NORMALES

Una vez vistos los conceptos básicos sobre las dependencias funcionales vamos a formalizar la definición de las tres primeras formas normales y la de Boyce y Codd que, como ya se ha indicado, son las que se apoyan en las dependencias funcionales.

4.1.- PRIMERA FORMA NORMAL

Como ya se indicó en la aproximación intuitiva a las formas normales, una relación se encuentra en 1FN si no existen grupos repetitivos. Esta es la condición básica para que una tabla pueda ser considerada una relación.

Para pasar a 1FN, la solución típica es repetir los valores del resto de los atributos de la tupla para cada uno de los valores del grupo repetitivo.

4.2.- SEGUNDA FORMA NORMAL

Una relación se dice que está en segunda forma normal si cumple las siguientes condiciones:

1. *Está en 1FN.*
2. *Todo atributo secundario, tiene una dependencia funcional completa respecto de cada una de las claves candidatas.*

Si todas las claves candidatas están formadas por un solo atributo, ya se encuentra en segunda forma normal.

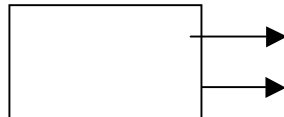
Por ejemplo, consideremos la siguiente relación:

EXISTENCIAS (CodAlmacén, Dirección, CodPieza, Cantidad)

¡Error! Marcador no definido. CodAlmacén	Dirección	CodPieza	Cantidad
A1	SAN PABLO, 7	P1	100
A1	SAN PABLO, 7	P2	500
A2	MADRID, 15	P1	300

Está en Primera Forma Normal

1.- Diagrama de dependencias funcionales



2.- Clave candidata: CodAlmacén, CodPieza

El atributo secundario Dirección no tiene una dependencia funcional completa de la clave candidata. Por tanto, EXISTENCIAS no está en 2FN y debemos crear una nueva relación.

ALMACENES (CodAlmacén, Dirección)

EXISTENCIAS1 (CodAlmacén, CodPieza, CANTIDAD)

Ahora tendríamos dos tablas en segunda forma normal

EJERCICIO 6

Estudiar la relación EMPLEADOS, estableciendo la forma normal en que se encuentra, el diagrama de dependencias funcionales, la clave o claves candidatas y la descomposición oportuna para que cumpla con 2FN.

EMPLEADOS

DNI	Nombre	Dirección	Empresa	Sueldo
13056782	José Machón Lara	Calzadas, 7	ANSA, S.A. CAMPOFRÍO Colegio "Sta. Maria"	70.000 200.000 80.000
14098456	María Ostiz Pita	Santander, 3	Solaris, S.A. Informática, S.A.	150.000 200.000
23456231	Irene Boch Tello	San Pablo, 12	Delegación Fiat	230.000

4.3.- TERCERA FORMA NORMAL

Una relación (tabla) se dice que está en tercera forma normal si y solo si se cumplen las siguientes condiciones:

1. *Está en 2FN*
2. *Ningún atributo secundario depende transitivamente de ninguna clave de la relación.*

Esto quiere decir que un atributo secundario (los que no son posibles claves) solo se puede conocer a través de las claves candidatas nunca mediante otro atributo secundario.

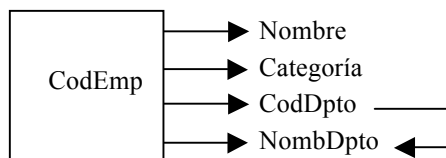
Por ejemplo:

Sea la relación EMPLEADOS (CodEmp, Nombre, Categoría, CodDpto, NombDpto)

CodEmp	Nombre	Categoría	CodDpto	NombDpto
EC1912	José Peña Ortiz	Auxiliar	AD1	Administración
TC2345	Arturo Osés Más	Técnico	MT2	Mantenimiento
TC2367	Mario Pérez Tuy	Maquinista	PR1	Producción

Está en primera forma normal ya que los atributos de todas las tuplas son monovalentes.

1.- Diagrama de dependencias funcionales



2.- Clave candidata: CodEmp

Todos los atributos secundarios tienen una dependencia funcional completa de la clave candidata, luego la relación sí está en 2FN.

No está en tercera forma normal porque el atributo NombDpto tiene una dependencia transitiva respecto de la clave: CodEmp — \longrightarrow NombDpto

Por tanto, crearemos una nueva relación:

DEPARTAMENTOS (CodDpto, NombDpto)
 EMPLEADOS1 (CodEmp, Nombre, Categoría, CodDpto)

Estas relaciones sí están en 3FN.

EJERCICIO 7

Estudiar la relación PRESTAMOS que se da a continuación, en la que no se almacenan datos históricos, ya sólo se utiliza para determinar los libros prestados en un momento determinado, para establecer:

1. La forma normal en que se encuentra.
2. Diagrama de dependencias funcionales.
3. Las descomposiciones que procedan para situarla en 3FN.

PRESTAMOS

CodSocio	NombreSocio	FechaPréstamo	CodLibro	Editorial	País
091234	Azucena Martín Arce	12/1/2000	912345	Marcombo	España
091289	Mario Pérez Sopena	4/2/2000	834568	Planeta	España
049872	María Artés Miñón	7/2/2000	994563	Penguin	Gran Bretaña

EJERCICIO 8

Convertir los siguientes esquemas relacionales para que cumplan con la 3FN

1. ALQUILER (Matrícula, Modelo, Marca, Nombre, NIFCliente, FechaAlquiler, Kms)
 Puede haber clientes con el mismo nombre.
2. FACTURAS (Nfactura, FechaFactura, CodCli, NombreCli, DirCli, CodArticulo, Descripción, Cantidad, Precio). Puede haber clientes con el mismo nombre.

4.4.- FORMA NORMAL DE BOYCE CODD

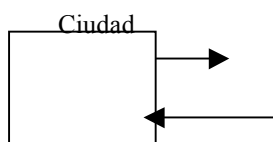
Como se indicó anteriormente, es una mejora de la tercera forma normal.

Una relación está en forma normal de Boyce-Codd, si y solo si todo determinante es una clave candidata.

Supongamos
la siguiente tabla:

CALLEJERO		
¡Error! Marcador no definido. Dirección	Ciudad	CodPostal
Estación, 2	Miranda de Ebro	09200
Rivera, 5	Miranda de Ebro	09200
Calzadas, 4	Burgos	09006
Briviesca, 4	Burgos	09006
Madrid, 7	Burgos	09001
Rivera, 16	Aranda de Duero	09194

El diagrama de dependencias
funcionales es el siguiente:



Luego las claves candidatas son Dirección, Ciudad y Dirección, CodPostal.

La relación está en 3FN ya que todos los atributos son principales. Sin embargo, se observan redundancias, ya que para cada código postal se repite la ciudad.

Además, se observan anomalías de inserción, puesto que es posible introducir un código postal ya existente y asignarlo a una ciudad diferente, al no estar implementada la dependencia $\text{CodPostal} \rightarrow \text{Ciudad}$.

Por último, en el diagrama de dependencias se observa que CodPostal es determinante de Ciudad y dicho determinante no es clave candidata (aunque forme parte de una de las claves), luego dicha relación no está en la FNBC.

La descomposición sería la siguiente:

CIUDADES (CodPostal, Ciudad)
DIRECCIONES (Dirección, CodPostal)

Las dos se encuentran ahora en la FNBC. Sin embargo, se ha perdido la dependencia funcional $\text{Dirección, Ciudad} \rightarrow \text{CodPostal}$, pero se puede recuperar por medio de la una operación de *combinación natural* (JOIN) de las dos relaciones resultantes:

Estudiando las dependencias también se verifica:

DIRECCIONES * CIUDADES		
¡Error! Marcador no definido. Dirección	CodPostal	Ciudad
Estación, 2	09200	Miranda de Ebro
Rivera, 5	09200	Miranda de Ebro
Calzadas, 4	09006	Burgos
Briviesca, 4	09006	Burgos
Madrid, 7	09001	Burgos

De DIRECCIONES:
Dirección, CodPostal \rightarrow CodPostal
(Creamos una dependencia funcional trivial)
Como en CIUDADES tenemos: CodPostal \rightarrow Ciudad,
podemos determinar:
Dirección, Ciudad \rightarrow CodPostal

Rivera, 16	09194	Aranda de Duero
------------	-------	-----------------

Si una relación está en 3FN y sus claves candidatas no están solapadas (no tienen atributos comunes), esta relación se encuentra también en FNBC.

5.- EJERCICIOS

8.- Define los siguientes términos:

1. Dependencia funcional.
2. Dependencia transitiva.
3. Atributos principales.
4. 1FN, 2FN, 3FN, FNBC.

9.- Indica si las siguientes afirmaciones son ciertas o falsas, razonando las respuestas:

1. Toda relación binaria (compuesta por dos atributos) está en tercera forma normal.
2. Toda relación binaria está en forma normal de Boyce_Codd.
3. Toda relación ternaria con una única clave candidata compuesta de dos atributos está en la FNBC.

10.- Razonar cuál es la forma normal más alta que cumple la siguiente relación, con las condiciones que se indican, y convertirla a FNBC.

TAREAS (NombTrabajo, Lugar, CodEmpleado, FechaComienzo)

- Cada trabajo se desarrolla en un lugar y ocupa a varias personas.
- Cada persona, solo tiene una fecha de comienzo.

11.- Examinar la relación PEDIDOS y la semántica que la acompaña y desarrollar los siguientes apartados:

1. Las dependencias funcionales.
2. El diagrama de dependencias.
3. Las claves candidatas.
4. La forma normal en que se encuentra.
5. Transformarla a la FNBC, si no lo estuviera.
6. Diseñar el diagrama conceptual en el modelo E/R, obtener el esquema relacional y verificar el grado de coincidencia con las relaciones resultantes del apartado 5.

PEDIDOS (NumPedido, NumPieza, Precio, Cantidad, FechaPedido)

- Cada pedido puede abarcar varias piezas.
- Cada pieza tiene su propio precio.

12.- Realizar las mismas tareas que las indicadas en el ejercicio anterior con la siguiente relación:

SUBASTAS (NumLote, Descripción, Comprador, Fecha, Dirección)

- Cada lote se venden al comprador con mejor oferta.
- Cada lote sólo puede ser vendido una vez en un día.
- Cada lote tiene una descripción y cada comprador tiene una dirección.
- Un lote puede ser vendido más de una vez pero en fechas diferentes.

13.- Estudiar las dependencias funcionales y convertir los siguientes esquemas relacionales para que cumplan con la FNBC

- ARTÍCULOS (CodArtículo, Título, Revista, Número, Página)
- VINOS (CodVino, Denominación, País, Región)
Una región siempre pertenece a un país.
Cada denominación puede pertenecer a más de una región o país.
- ALQUILERES (Matrícula, NIFCliente, Nombre, FechaAlquiler, Kms)
No existen nombres de clientes repetidos.

14.- Razonar cuál es la forma normal más alta que cumple la siguiente relación, con las condiciones que se indican, y convertirla a FNBC.

SOLICITUDES (NumSolicitud, NombreSolicitante, DirecciónSolicitante, TipoPrestamo)

- Cada solicitante se identifica por NumSolicitud.
- Cada solicitud permite un tipo de préstamo.
- Cada solicitud tiene una dirección del solicitante
- Un solicitante puede hacer varias solicitudes.

15.- Examinar la relación EMPLEADOS y la semántica que la acompaña y desarrollar los siguientes apartados:

1. Las dependencias funcionales.
2. El diagrama de dependencias.
3. Las claves candidatas.
4. La forma normal en que se encuentra.
5. Transformarla a la FNBC, si no lo estuviera.
6. Diseñar el diagrama conceptual en el modelo E/R, obtener el esquema relacional y verificar el grado de coincidencia con las relaciones resultantes del apartado 5.

EMPLEADOS (NumEmpl, NomEmpl, NumDep, NumDivisión, Administrador)

- Un empleado pertenece a un solo departamento.
- Cada departamento pertenece a una división.
- Cada división tiene un administrador.

16.- Realizar las mismas tareas que las indicadas en el ejercicio anterior con la siguiente relación:

ESTUDIANTES (Nmatrícula, NombreEst, Curso, Centro, Profesor, Texto)

- Un estudiante puede estar matriculado en varios cursos.
- Un estudiante tiene un número de matrícula distinto por cada curso en el que está matriculado.
- Un curso se imparte en un solo centro.
- El número de matrícula identifica al curso y al centro en que se imparte.
- Un curso es impartido por un solo profesor, pero un profesor puede impartir varios cursos.
- Un curso se apoya en distintos textos, y un mismo texto puede servir de soporte para varios cursos.

17.- Indica la FN en que se encuentran y realiza las descomposiciones oportunas para determinar la mejor opción posible para la FNBC, en las siguientes relaciones:

1. NOTAS ({DniProfesor, NombreProf, DniAlumno, Calificación},
{ DniProfesor \longleftrightarrow NombreProf, (DniProfesor, DniAlumno) \rightarrow Calificación})

2. LIBROS ({Título, Colección, Editorial},
{Título, Editorial \rightarrow Colección \rightarrow Editorial})

Un libro puede ser publicado por varias editoriales distintas, pero una colección sólo se publica en una editorial.

18. - Normalizar la siguiente relación:

VENTAS

Ncliente	Nombre	Localidad	NAartículo	Denominación	Cantidad	Precio	Fecha
111	Andrés	Burgos	A123	Discos	200	100	12/1/2000
111	Andrés	Burgos	A321	Papel	10	2000	10/2/2000
222	María	Madrid	B543	Grapas	200	20	14/12/1999
444	Ana	Burgos	A321	Papel	20	2000	14/2/2000