

- Si el atributo es compuesto y, en su caso, qué atributos simples lo forman.
- Si el atributo es derivado y, en su caso, cómo se calcula su valor.

El dominio de un atributo es el conjunto de valores que puede tomar el atributo, a esto se le llama restricción de dominio.

Actividad de aprendizaje 3.5

En equipo de dos o tres integrantes elaborar un diagrama de flujo que represente las etapas del diseño lógico.

3.4.4 Procesos de normalización

El proceso de normalización de bases de datos consiste en designar y aplicar una serie de reglas a las relaciones obtenidas tras el modelo entidad-relación.

Las bases de datos se normalizan para:

- Evitar la redundancia de datos.
- Disminuir problemas de actualización de los datos en las tablas.
- Proteger la integridad de los datos.

Edgar F. Codd originalmente definió las tres primeras formas normales (1FN, 2FN y 3FN) en 1970. La forma normal de Boyce-Codd (FNBC) fue introducida en 1974 por los dos autores de los que se desprende su nombre. Las cuarta y quinta formas normales (4FN y 5FN) se encargan específicamente de la representación de las relaciones muchos a muchos y uno a muchos, entre los atributos, y fueron introducidas por Fagin en 1977 y 1979, respectivamente. Cada forma normal incluye a las anteriores, como se muestra en la figura 3.9.

Antes de ahondar en los conceptos relativos a las formas normales, revisemos algunas definiciones previas:

- Dependencia funcional: $A \rightarrow B$. Representa que B es funcionalmente dependiente de A. Entonces, para un valor de A siempre aparece un valor de B. Ejemplo: si A es el ID, y B el Nombre, está claro que para un determinado número de ID, siempre aparecerá el mismo nombre de titular.
- Dependencia funcional completa: $A \twoheadrightarrow B$. Representa que B depende de A en su totalidad. Ejemplo: tiene sentido plantearse este tipo de dependencia cuando A está compuesto por más de un atributo. Por ejemplo, supongamos que A corresponde al atributo compuesto ID_Empleado + Cod_Dpto. y B es Nombre_Dpto. En este caso, B depende del Cod_Dpto., pero no del ID_Empleado. Por tanto, no habría dependencia funcional completa.
- Dependencia transitiva: $A \rightarrow B \rightarrow C$, si $A \rightarrow B$ y $B \rightarrow C$. Representa que C depende de forma transitiva de A. Ejemplo: Sea A el ID de un alumno, B la localidad en la que vive y C la provincia, entonces se trata de un caso de dependencia transitiva $A \rightarrow B \rightarrow C$.
- Determinante funcional: representa todo atributo, o conjunto de ellos, de los que depende algún otro atributo. Ejemplo: el ID es un determinante funcional, pues atributos como nombre, dirección, localidad, etc., dependen de él.

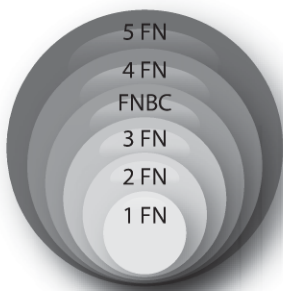


Figura 3.9 ►

- Dependencia multivaluada: $A \twoheadrightarrow B$. Es un tipo de dependencia en la que un determinante funcional no implica un único valor, sino un conjunto de ellos. El valor de A siempre implica varios valores de B. Ejemplo: $\text{CursoBachillerato} \twoheadrightarrow \text{Modalidad}$. Para el primer curso siempre aparecerá en el campo Modalidad uno de los siguientes valores: Ciencias, Humanidades/Ciencias Sociales o Artes. Igual para segundo curso.

Primera forma normal

Una relación está en 1FN si y solo si cada atributo es atómico.

Ejemplo: Supongamos que tenemos la siguiente tabla con los datos de los alumnos de una escuela secundaria.

Tabla 3.2

Alumnos							
ID	Nombre	Curso	Fecha Matricula	Tutor	Localidad	Provincia Alumno	Teléfonos
11111111A	Eva	1ESO-A	01-Julio-2016	Isabel	Ecatepec	México	660111222
22222222B	Ana	1ESO-A	09-Julio-2016	Isabel	Ecatepec	México	660222333 660333444 660444555
33333333C	Susana	1ESO-B	11-Julio-2016	Roberto	Ecatepec	México	
44444444D	Juan	2ESO-A	05-Julio-2016	Federico	Aragón	CDMX	
55555555E	José	2ESO-A	02-Julio-2016	Federico	Aragón	CDMX	661000111 661000222

Como se puede observar, esta tabla no está en 1FN puesto que el campo “Teléfonos” contiene varios datos dentro de una misma celda y, por tanto, no es un campo cuyos valores sean atómicos debido a que es un atributo multivaluado. La manera de solucionarlo sería la siguiente:

Tabla 3.3

Alumnos						
ID	Nombre	Curso	Fecha Matricula	Tutor	Localidad	Provincia Alumno
11111111A	Eva	1ESO-A	01-Julio-2016	Isabel	Ecatepec	México
22222222B	Ana	1ESO-A	09-Julio-2016	Isabel	Ecatepec	México
33333333C	Susana	1ESO-B	11-Julio-2016	Roberto	Ecatepec	México
44444444D	Juan	2ESO-A	05-Julio-2016	Federico	Aragón	CDMX
55555555E	José	2ESO-A	02-Julio-2016	Federico	Aragón	CDMX

Tabla 3.4

Teléfonos	
DNI	Teléfono
11111111A	660111222
22222222B	660222333
22222222B	660333444
22222222B	660444555
55555555E	661000111
55555555E	661000222

Segunda forma normal

Una relación está en 2FN si y solo si está en 1FN y todos los atributos que no forman parte de la clave principal tienen dependencia funcional completa de ella. Es decir, no existen referencias parciales.

Ejemplo: Volviendo al ejemplo anterior, observa la siguiente tabla:

Tabla 3.5

Alumnos						
ID	Nombre	Curso	Fecha Matriculación	Tutor	Localidad	Provincia Alumno
11111111A	Eva	1ESO-A	01-Julio-2016	Isabel	Ecatepec	México
22222222B	Ana	1ESO-A	09-Julio-2016	Isabel	Ecatepec	México
33333333C	Susana	1ESO-B	11-Julio-2016	Roberto	Ecatepec	México
44444444D	Juan	2ESO-A	05-Julio-2016	Federico	Aragón	CDMX
55555555E	José	2ESO-A	02-Julio-2016	Federico	Aragón	CDMX

Ahora, vamos a examinar las dependencias funcionales. El gráfico que las representa es el siguiente:

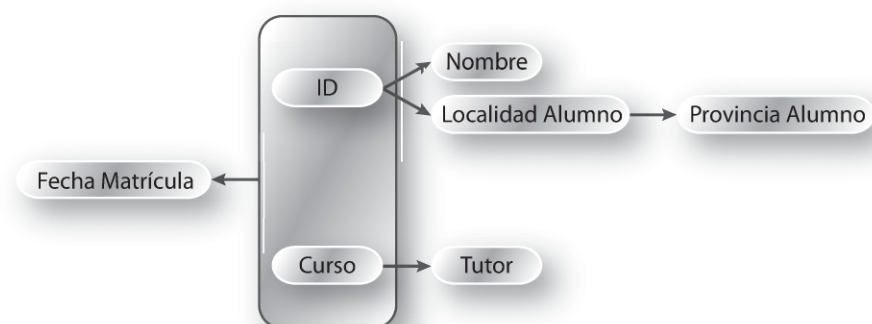


Figura 3.10 ►

- Siempre que aparece un ID aparecerá el Nombre correspondiente y la Localidad Alumno correspondiente. Por tanto, $ID \rightarrow Nombre$ e $ID \rightarrow Localidad\ Alumno$. Por otro lado, siempre que aparece un Curso aparecerá el Tutor correspondiente. Por tanto, $Curso \rightarrow Tutor$. Los atributos Nombre y Localidad Alumno no dependen funcionalmente de Curso, así como el atributo Tutor no depende funcionalmente de ID.
- El único atributo que sí depende de forma completa de la clave compuesta ID y Curso es Fecha Matriculación: $(ID, Curso) \rightarrow Fecha\ Matriculación$.

A la hora de establecer la clave primaria de una tabla debemos escoger un atributo, o conjunto de ellos, de los que dependan funcionalmente el resto de los atributos. Además debe tratarse de una dependencia funcional completa. Si escogemos ID como clave primaria, tenemos un atributo (Tutor) que no depende funcionalmente de ella. En tanto que si escogemos Curso como clave primaria, tenemos otros atributos que no dependen de ella.

Si escogemos la combinación (ID, Curso) como clave primaria, entonces el resto de los atributos sí tienen dependencia funcional respecto a esta clave. No obstante,

se trata de una dependencia parcial, no total (salvo Fecha Matrícula, donde sí existe dependencia completa). Por tanto, esta tabla no está en 2FN, por lo que la solución sería la siguiente:

Tabla 3.6

Alumnos			
ID	Nombre	Localidad	Provincia
11111111A	Eva	Ecatepec	México
22222222B	Ana	Ecatepec	México
33333333C	Susana	Ecatepec	México
44444444D	Juan	Aragón	CDMX
55555555E	José	Aragón	CDMX

Tabla 3.7

Matrículas		
ID	Curso	FechaMatrícula
11111111A	1ESO-A	01-Julio-2016
22222222B	1ESO-A	09-Julio-2016
33333333C	1ESO-B	11-Julio-2016
44444444D	2ESO-A	05-Julio-2016
55555555E	2ESO-A	02-Julio-2016

Tabla 3.8

Cursos	
Curso	Tutor
1ESO-A	Isabel
1ESO-B	Roberto
2ESO-A	Federico

Tercera forma normal. Forma normal Boyce-Cood

Una relación está en 3FN si y solo si está en 2FN y no existen dependencias transitivas. Todas las dependencias funcionales deben darse respecto a la clave principal.

Ejemplo: Retomemos el ejemplo de la secundaria, observa la siguiente tabla:

Tabla 3.9

Alumnos			
ID	Nombre	Localidad	Provincia
11111111A	Eva	Ecatepec	México
22222222B	Ana	Ecatepec	México
33333333C	Susana	Ecatepec	México
44444444D	Juan	Aragón	CDMX
55555555E	José	Aragón	CDMX

Como podemos observar, existe una dependencia funcional transitiva: ID \rightarrow Localidad \rightarrow Provincia.

Figura 3.11 ►



Para que la tabla esté en 3FN, no pueden existir dependencias funcionales transitivas. Entonces, para solucionar el problema debemos crear una nueva tabla. Observa cómo lo hicimos:

Tabla 3.10

Alumnos		
ID	Nombre	Localidad
11111111A	Eva	Ecatepec
22222222B	Ana	Ecatepec
33333333C	Susana	Ecatepec
44444444D	Juan	Aragón
55555555E	José	Aragón

Tabla 3.11

Localidades	
Localidad	Provincia
Ecatepec	México
Aragón	CDMX

Resultado final

Tabla 3.12

Alumnos		
ID	Nombre	Localidad
11111111A	Eva	Ecatepec
22222222B	Ana	Ecatepec
33333333C	Susana	Ecatepec
44444444D	Juan	Aragón
55555555E	José	Aragón

Tabla 3.13

Localidades	
Localidad	Provincia
Ecatepec	México
Aragón	CDMX

Tabla 3.14

Teléfonos	
ID	Teléfono
11111111A	660111222
22222222B	660222333
22222222B	660333444
22222222B	660444555
55555555E	661000111
55555555E	661000222

Tabla 3.15

Matrículas		
ID	Curso	Fecha Matrícula
11111111A	1ESO-A	01-Julio-2016
22222222B	1ESO-A	09-Julio-2016
33333333C	1ESO-B	11-Julio-2016
44444444D	2ESO-A	05-Julio-2016
55555555E	2ESO-A	02-Julio-2016

Tabla 3.16

Cursos	
Curso	Tutor
1ESO-A	Isabel
1ESO-B	Roberto
2ESO-A	Federico

Forma normal de Boyce-Codd: FNBC

Una relación está en FNBC si está en 3FN, solamente debemos tener en cuenta esta forma normal cuando tenemos varias claves candidatas compuestas, aunque se trata de una situación que ocurre en muy raras ocasiones.

Ejemplo: Tenemos una tabla con la siguiente información proporcionada por un grupo de proveedores: nombre del proveedor, códigos de piezas y la cantidad de piezas disponibles.

Tabla 3.17

Suministros			
CIF	Nombre	Código Pieza	Cantidad Piezas
S-11111111A	Ferroman	1	10
B-22222222B	Ferrotex	1	7
M-33333333C	Ferropet	3	4
S-11111111A	Ferroman	2	20
S-11111111A	Ferroman	3	15
B-22222222B	Ferrotex	2	8
B-22222222B	Ferrotex	3	4

El gráfico de dependencias funcionales correspondiente a esta tabla se muestra en la figura 3.12:

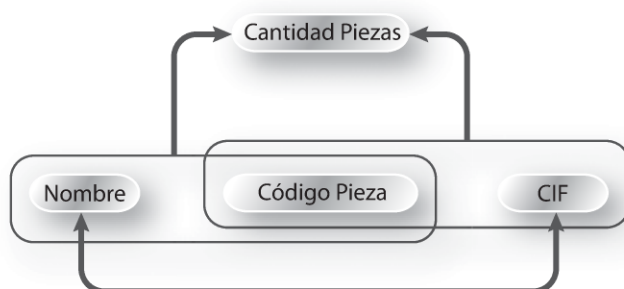


Figura 3.12 ►

Como podemos observar, el atributo CantidadPiezas tiene dependencia funcional de dos claves candidatas compuestas, que son:

(NombreProveedor, CódigoPieza)

(CIFProveedor, CódigoPieza)

Existe también una dependencia funcional en doble sentido (que no nos afecta): NombreProveedor <-> CIFProveedor.

Para esta tabla existe un solapamiento de dos claves candidatas compuestas. Para evitar esta situación dividimos la tabla de la siguiente manera:

Tabla 3.18

Proveedores	
CIF	Nombre
S-11111111A	Ferroman
B-22222222B	Ferrotex
M-33333333C	Ferropet

Tabla 3.19

Suministros		
CIF	CódigoPieza	CantidadPiezas
S-11111111A	1	10
B-22222222B	1	7
M-33333333C	3	4
S-11111111A	2	20
S-11111111A	3	15
B-22222222B	2	8
B-22222222B	3	4

Actividad de aprendizaje 3.6

Factura de compra-venta

☐

La empresa Colombian Systems te ha contratado como el ingeniero encargado de sistematizar la facturación. En la siguiente factura de compra-venta, analiza toda la información disponible y aplica el proceso de normalización hasta llegar a la tercera forma normal. Incluye una justificación detallada de cada uno de los pasos que seguiste para llegar al resultado final.

Factura(NUM_FAC, FECHA_FAC, NOM_CLIENTE, DIR_CLIENTE, RFC_CLIENTE, CIUDAD_CLIENTE, TELEF_CLIENTE, CATEGORIA, COD_PROD, DESP_PROD, VAL_UNIT, CANT_PROD)

Donde:

NUM_FAC: Número de la factura de compra-venta

FECHA_FAC: Fecha de la factura de compra-venta

NOM_CLIENTE: Nombre del cliente

DIR_CLIENTE: Dirección del cliente

RFC_CLIENTE: RFC del cliente

CIUDAD_CLIENTE: Ciudad del cliente

TELEF_CLIENTE: Teléfono del cliente

CATEGORIA: Categoría del producto

COD_PROD: Código del producto

DESCRIPCION: Descripción del producto

VAL_UNIT: Valor unitario del producto

CANT_PROD: Cantidad de productos que compra el cliente

La llave primaria es número de factura de venta: NUM_FAC.