

La **normalización** es una técnica para diseñar la estructura lógica de los datos de un sistema de información en el modelo relacional, desarrollada por E. F. Codd en 1972. Es una estrategia de diseño de abajo arriba: se parte de los atributos y éstos se van agrupando en relaciones (tablas) según su afinidad. Aquí no se utilizará la normalización como una técnica de diseño de bases de datos, sino como una etapa posterior a la correspondencia entre el esquema conceptual y el esquema lógico, que elimine las dependencias no deseadas entre atributos. Las ventajas de la normalización son las siguientes:

- Evita anomalías en inserciones, modificaciones y borrados.
- Mejora la independencia de datos.
- No establece restricciones artificiales en la estructura de los datos.

Uno de los conceptos fundamentales en la normalización es el de *dependencia funcional*. Una **dependencia funcional** es una relación entre atributos de una misma relación (tabla).

Si X e Y son atributos de la relación R , se dice que Y es funcionalmente dependiente de X (se denota por $X \rightarrow Y$) si cada valor de X tiene asociado un solo valor de Y (X e Y pueden constar de uno o varios atributos). A X se le denomina *determinante*, ya que X determina el valor de Y . Se dice que el atributo Y es *completamente dependiente* de X si depende funcionalmente de X y no depende de ningún subconjunto de X .

Esto es lo mismo que decir que si dos tuplas de R tienen el mismo valor para su atributo X , forzosamente han de tener el mismo valor para el atributo Y .

La dependencia funcional es una noción semántica. Si hay o no dependencias funcionales entre atributos no lo determina una serie abstracta de reglas, sino, más bien, los modelos mentales del usuario y las reglas de negocio de la organización o empresa para la que se desarrolla el sistema de información. Cada dependencia funcional es una clase especial de regla de integridad y representa una relación de uno a muchos.

En el proceso de normalización se debe ir comprobando que cada relación (tabla) cumple una serie de reglas que se basan en la clave primaria y las dependencias funcionales. Cada regla que se cumple aumenta el grado de normalización. Si una regla no se cumple, la relación se debe descomponer en varias relaciones que sí la cumplan.

La normalización se lleva a cabo en una serie de pasos. Cada paso corresponde a una forma normal que tiene unas propiedades. Conforme se va avanzando en la normalización, las relaciones tienen un formato más estricto (más fuerte) y, por lo tanto, son menos vulnerables a las anomalías de actualización. El modelo relacional sólo requiere un conjunto de relaciones en primera forma normal. Las restantes formas normales son opcionales. Sin embargo, para evitar las anomalías de actualización, es recomendable llegar al menos a la tercera forma normal.

CÁLCULO DE DEPENDENCIAS

Las **dependencias** son propiedades inherentes al contenido semántico de los datos formando parte de las restricciones de usuario del modelo relacional. Entre los atributos de una relación pueden existir dependencias de varios tipos:

- **Dependencias funcionales**

Son de primordial importancia a la hora de encontrar y eliminar la redundancia de los datos almacenados en las tablas de una base de datos relacional, se centran en el estudio de las dependencias que presenta cada atributo de una relación con respecto al resto de atributos de la misma.

Dada una relación R que contiene los atributos X e Y se dice que Y depende funcionalmente de X ($X \rightarrow Y$) si y sólo si en todo momento cada valor de X tiene asociado un solo valor de Y . Esto es lo mismo que decir que si dos tuplas de R tienen el mismo valor para su atributo X forzosamente han de tener el mismo valor para el atributo Y .

Tipos de dependencias:

- **Dependencias completa y parcial.** Dado un atributo compuesto X formado por los atributos X_1 y X_2 , se dice que el atributo Y tiene una dependencia funcional completa con respecto a X si: $X_1 \rightarrow Y$; $X_2 \rightarrow Y$; $Y \not\rightarrow X$ (Y no implica X) pero $X \rightarrow Y$.

Por el contrario, dado un atributo compuesto X formado por los atributos X_1 y X_2 , se dice que el atributo Y tiene una dependencia funcional parcial con respecto a X si: $X_1 \rightarrow Y$ o $X_2 \rightarrow Y$ y $X \rightarrow Y$

- **Dependencia transitiva.** Dados los atributos X , Y y Z de la relación R en la que existen las siguientes dependencias funcionales $X \rightarrow Y$; $Y \rightarrow Z$; se dice que Z tiene una dependencia transitiva respecto X a través de Y : $X \rightarrow Z$. Por ejemplo: Nombre de alumno Dirección; Dirección Ciudad; Nombre de alumno Ciudad.
- **Dependencia multivaluada o de valores múltiples.** Sean X e Y dos descriptores, X multidetermina a Y ($X \twoheadrightarrow Y$) si para cada valor de X existe un conjunto bien definido de valores posibles en Y , con independencia del resto de los atributos de la relación. Este tipo de dependencias produce redundancia de datos, como se puede apreciar en la Tabla 2.2, en donde las claves 34567-B y 89345-M tienen dos registros para mantener la serie de datos en forma independiente. Esto ocasiona que al realizarse una actualización se requiera de demasiadas operaciones para tal fin.

En la tabla siguiente, el atributo DNI multidetermina a Titulación; ver Tabla 2.2:

DNI	TITULACIÓN
34567-B	Licenciado en Exactas
34567-B	Licenciado en Ciencias Físicas
56667-C	Ingeniero de Telecomunicaciones
89345-M	Licenciado en Historia
89345-M	Licenciado en Filosofía

Tabla 2.2. Dependencia multivaluada.

- **Reglas de normalización**

Se dice que una relación está en una forma normal si satisface un cierto conjunto específico de restricciones impuestas por la regla de normalización correspondiente. La aplicación de una regla es una operación que toma como entrada una relación y da como resultado dos o más relaciones.

- **Primera forma normal. 1FN**

Se dice que una relación está en 1FN si y sólo si los valores que componen cada atributo de una tupla son atómicos, es decir, cada atributo de la relación toma un único valor del dominio correspondiente, o lo que es lo mismo no existen grupos repetitivos.

La tabla estará en 1FN si tiene un solo valor en la intersección de cada fila con cada columna. Un conjunto de relaciones se encuentra en 1FN si ninguna de ellas tiene grupos repetitivos.

Si una relación no está en 1FN, hay que eliminar de ella los grupos repetitivos. Un grupo repetitivo será el atributo o grupo de atributos que tiene múltiples valores para cada tupla de la relación. Hay dos formas de eliminar los grupos repetitivos:

- Repetir los atributos con un solo valor para cada valor del grupo repetitivo. De este modo, se introducen redundancias ya que se duplican valores, pero estas redundancias se eliminarán después mediante las restantes formas normales.
- La segunda forma de eliminar los grupos repetitivos consiste en poner cada uno de ellos en una relación aparte, heredando la clave primaria de la relación en la que se encontraban.



Caso práctico

- 10 La aplicación de esta regla es fácil. Este proceso se realiza casi siempre en el análisis del problema. Por ejemplo, consideramos la tabla ALUMNO, con clave primaria COD_ALUMNO, en la que el atributo TLF puede tomar varios valores: el móvil, el de casa, el del padre, el de la madre, etcétera. Ver Tabla 2.3:

COD_ALUMNO	NOMBRE	APELLIDO	TLF	DIRECCION
1111	PEPE	GARCÍA	678-900800 91-2233441 91-1231232	C/Las cañas 45
2222	MARÍA	SUÁREZ	91-7008001	C/Mayor 12
3333	JUAN	GIL	91-7562324 660-111222	C/La plaza
4444	FRANCISCO	MONTOYA	678-556443	C/La arboleda

Tabla 2.3. Tabla ALUMNO que no está en 1FN.

Esta tabla no está en 1FN, ya que hay dos alumnos que tienen varios teléfonos. Para que esta tabla esté en 1FN se puede:

- Definir como clave primaria de la tabla COD_ALUMNO, y el TLF, con el fin de que cada atributo tome un único valor en la tupla correspondiente. Ver Tabla 2.4.

COD_ALUMNO	TLF	NOMBRE	APELLIDO	DIRECCIÓN
1111	678-900800	PEPE	GARCÍA	C/Las cañas 45
1111	91-2233441	PEPE	GARCÍA	C/Las cañas 45
1111	91-1231232	PEPE	GARCÍA	C/Las cañas 45
2222	91-7008001	MARÍA	SUÁREZ	C/Mayor 12
3333	91-7562324	JUAN	GIL	C/La plaza
3333	660-111222	JUAN	GIL	C/La plaza
4444	678-556443	FRANCISCO	MONTOYA	C/La arboleda

Tabla 2.4. Tabla ALUMNO, primera transformación a 1FN.

- O también se eliminan los grupos repetitivos (TLF) y se crea una relación (tabla) junto con la clave inicial. Ver Tablas 2.5 y 2.6.

COD_ALUMNO	NOMBRE	APELLIDO	DIRECCION
1111	PEPE	GARCÍA	C/Las cañas 45
2222	MARÍA	SUÁREZ	C/Mayor 12
3333	JUAN	GIL	C/La plaza
4444	FRANCISCO	MONTOYA	C/La arboleda

Tabla 2.5. Tabla ALUMNO segunda transformación a 1FN.

La segunda tabla estará formada por el TLF, y el COD_ALUMNO. Los dos formarán la clave. Ver Tabla 2.6:

COD_ALUMNO(FK)	TLF
1111	678-900800
1111	91-2233441
1111	91-1231232
2222	91-7008001

Tabla 2.6. Tabla ALUM_TLF en 1FN.

- **Segunda forma normal. 2FN**

Se dice que una relación se encuentra en 2FN si y sólo si satisface la 1FN, y cada atributo de la relación que no está en la clave depende funcionalmente de forma completa de la clave primaria de la relación. La 2FN se aplica a las relaciones que tienen claves primarias compuestas por dos o más atributos.

Si una relación está en 1FN y su clave primaria es simple (tiene un solo atributo), entonces también está en 2FN. Las relaciones que no están en 2FN pueden sufrir anomalías cuando se realizan actualizaciones.

Si la clave primaria está formada por un solo atributo y la relación está en 1FN, ya está en 2FN.

Teorema de la 2FN: sea una relación formada por los atributos A, B, C, D con clave primaria compuesta por los atributos A y B. Si se cumple que (D depende funcionalmente de A): $A \rightarrow D$, entonces la relación puede descomponerse en dos relaciones R1 y R2 con los atributos respectivos: R1 (A, D) y R2 (A, B, C).

Para pasar una relación en 1FN a 2FN hay que eliminar las dependencias parciales de la clave primaria. Para ello, se eliminan los atributos, que son funcionalmente dependientes, y se ponen en una nueva relación con una copia de su determinante (los atributos de la clave primaria de los que dependen).

Se crearán dos tablas para eliminar las dependencias funcionales, una de ellas tendrá los atributos que dependen funcionalmente de la clave, y la otra los atributos que forman parte de la clave de la que dependen.



Caso práctico

- 11 Supongamos que tenemos una relación **ALUMNO** en la que representamos los datos de los alumnos y las notas en cada una de las asignaturas en que está matriculado. La clave es el número de matrícula **COD_ALUMNO** y la asignatura **ASIGNATURA**, ver Tabla 2.7:

COD_ALUMNO	NOM_ALUM	APE_ALUM	ASIGNATURA	NOTA	CURSO	AULA
1111	PEPE	GARCÍA	LENGUA I	5	1	15
1111	PEPE	GARCÍA	IDIOMA	5	2	16
2222	MARÍA	SUAREZ	IDIOMA	7	2	16
2222	MARÍA	SUAREZ	CIENCIAS	7	2	14
3333	JUAN	GIL	PLÁSTICA	6	1	18
3333	JUAN	GIL	MATEMÁTICAS I	6	1	12
4444	FRANCISCO	MONTOYA	LENGUA II	4	2	11
4444	FRANCISCO	MONTOYA	MATEMÁTICAS I	6	1	12
4444	FRANCISCO	MONTOYA	CIENCIAS	8	1	14

Tabla 2.7. Relación **ALUMNO** para transformar a 2FN.

Es obvio que todos los atributos no dependen de la clave completa (**COD_ALUMNO**, **ASIGNATURA**). En primer lugar, hay que ver las dependencias funcionales de cada uno de los atributos con respecto a los atributos de la clave y el resto de atributos:

- **NOM_ALUM** y **APE_ALUM**, sólo dependen de **COD_ALUMNO**.
- Los atributos **CURSO** y **AULA** están relacionados con la **ASIGNATURA**, es decir, existe una dependencia entre **ASIGNATURA** → **CURSO**, **ASIGNATURA** → **AULA**. Una asignatura pertenece a un curso y se imparte en un aula.
- El atributo **NOTA** depende funcionalmente de la clave, pues para que haya una nota tiene que haber una asignatura y un alumno.

Vistas las dependencias funcionales llegamos a la siguiente conclusión para que la relación **ALUMNO** esté en 2FN necesitamos crear tres relaciones: **ALUMNO**, **ASIGNATURAS** y **NOTAS** (ver Tablas 2.8, 2.9 y 2.10):

COD_ALUMNO	NOM_ALUM	APE_ALUM
1111	PEPE	GARCÍA
2222	MARÍA	SUAREZ
3333	JUAN	GIL
4444	FRANCISCO	MONTOYA

Tabla 2.8. Tabla **ALUMNO** en 2FN.

ASIGNATURA	CURSO	AULA
LENGUA I	1	15
IDIOMA	2	16
CIENCIAS	2	14
PLÁSTICA	1	18
MATEMÁTICAS I	1	12
LENGUA II	2	11

Tabla 2.9. Tabla **ASIGNATURAS** en 2FN.

COD_ALUMNO (FK)	ASIGNATURA (FK)	NOTA
1111	LENGUA I	5
1111	IDIOMA	5
2222	IDIOMA	7
2222	CIENCIAS	7
3333	PLÁSTICA	6
3333	MATEMÁTICAS I	6
4444	LENGUA II	4
4444	MATEMÁTICAS I	6
4444	CIENCIAS	8

Tabla 2.10. Tabla **NOTAS** en 2FN.

• Tercera forma normal. 3FN

Una relación está en tercera forma normal si, y sólo si, está en 2FN y, además, cada atributo que no está en la clave primaria no depende transitivamente de la clave primaria. Es decir, los atributos de la relación no dependen unos de otros, dependen únicamente de la clave, esté formada por uno o más atributos. La dependencia $X \rightarrow Z$ es transitiva si existen las dependencias $X \rightarrow Y$, $Y \rightarrow Z$, siendo X, Y , atributos o conjuntos de atributos de una misma relación.

Para pasar una relación de 2FN a 3FN hay que eliminar las dependencias transitivas. Para ello, se eliminan los atributos que dependen transitivamente y se ponen en una nueva relación con una copia de su determinante (el atributo o atributos no clave de los que dependen).

Teorema de la 3FN: sea una relación formada por los atributos A, B, C con clave primaria formada por el atributo A . Si se cumple que: $B \rightarrow C$, entonces la relación puede descomponerse en dos relaciones, R_1 y R_2 , con los atributos respectivos: $R_1 (A, B)$ y $R_2 (B, C)$.

Caso práctico



12 Supongamos que tenemos una relación **LIBROS** en la que representamos los datos de las editoriales de los mismos, ver Tabla 2.11:

COD LIBRO	TÍTULO	EDITORIAL	PAÍS
12345	DISEÑO DE BD RELACIONALES	RAMA	ESPAÑA
34562	INSTALACIÓN y MANTENIMIENTO DE EQUIPOS	MCGRAW-HILL	ESPAÑA
72224	FUNDAMENTOS DE PROGRAMACIÓN	SANTILLANA	ESPAÑA
34522	BASE DE DATOS OO	ADDISON	EEUU

Tabla 2.11. Relación **LIBROS** para transformar a 3FN.

Veamos las dependencias con respecto a la clave:

- TÍTULO y EDITORIAL dependen directamente del código del libro.
- El PAÍS, aunque en parte también depende del libro, está más ligado a la EDITORIAL a la que pertenece el libro. Por esta última razón, la relación libros no está en 3FN, la solución se muestra en las Tablas 2.12 y 2.13:

COD LIBRO	TÍTULO	EDITORIAL (FK)	EDITORIAL	PAÍS
12345	DISEÑO DE BD RELACIONALES	RAMA	RAMA	ESPAÑA
34562	INSTALACIÓN y MANTENIMIENTO DE EQUIPOS	MCGRAW-HILL	MCGRAW-HILL	ESPAÑA
72224	FUNDAMENTOS DE PROGRAMACIÓN	SANTILLANA	SANTILLANA	ESPAÑA
34522	BASE DE DATOS OO	ADDISON	ADDISON	EEUU



Actividades propuestas

2 Dada la tabla 2.14 normalizar hasta la 3FN.

DNI	NOMBRE	APELLIDOS	DIRECCIÓN	C_POST	POBLACIÓN	PROVINCIA
413245-B	JUAN	RAMOS	C/Las cañas 59 C/Pilón 12	19005 45589	GUADALAJARA CALERUELA	GUADALAJARA TOLEDO
23456-J	PEDRO	PÉREZ	C/Vitoria 3 C/El altozano	28804 10392	ALCALÁ DE HENARES BERROCALEJO	MADRID CÁCERES
34561-B	MARÍA	RODRÍGUEZ	C/Sanz Vázquez 2	19004	GUADALAJARA	GUADALAJARA
222346-J	JUAN	CABELLO	C/El ensanche 3 C/Los abedules 10	28802 10300	ALCALÁ DE HENARES NAVALMORAL DE LA MATA	MADRID CÁCERES

Tabla 2.14. Transformar a 3FN.

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

Se define determinante en una relación a un atributo del cual depende funcionalmente de manera completa cualquier otro atributo de la relación. Una relación está en la Forma Normal de Boyce - Codd (FNBC) si, y sólo si, todo determinante de ella es una clave candidata.

La 2FN y la 3FN eliminan las dependencias parciales y las dependencias transitivas de la clave primaria. Pero este tipo de dependencias todavía pueden existir sobre otras claves candidatas, si las hubiera. La BCFN es más fuerte que la 3FN, por lo tanto, toda relación en FNBC está en 3FN.

La violación de la FNBC es poco frecuente ya que se da bajo ciertas condiciones que raramente se presentan. Se debe comprobar si una relación viola la FNBC si tiene dos o más claves candidatas compuestas que tienen al menos un atributo en común.

Teorema de Boyce - Codd: sea una relación R formada por los atributos A, B, C, D con claves candidatas compuestas (A, B) y (B, C) tal que: A C, entonces la relación puede descomponerse en cualquiera de las dos siguientes maneras: R1 (A, C) y R2 (B, C, D) o bien, R1 (A, C) y R2 (A, B, D).



Caso práctico

- 13 Supongamos que tenemos una relación EMPLEADOS en la que representamos los datos de los empleados de una fábrica, ver Tabla 2.15:

DNI	NÚM_SEG_SOC	NOMBRE	APELLIDOS	DEPARTAMENTO	PUESTO	SALARIO
413245-B	28-1234566	JUAN	RAMOS	COMPRAS	GERENTE	2.300
23456-J	28-2345686	PEDRO	PÉREZ	NÓMINAS	AUXILIAR	1.200
123123-C	19-458766	MARÍA	GIL	ALMACÉN	CONSERJE	1.530
1234556-B	45-223344	ANTONIO	SANZ	COMPRAS	GESTIÓN	2.200

Tabla 2.15. Relación EMPLEADOS para transformar a FNBC.

Observando los atributos de esta relación podemos ver que NUM_SEG_SOC y el grupo NOMBRE-APELLIDOS son claves candidatas (determinantes). Esta relación se transforma en dos tablas: una contendrá la clave junto con las claves candidatas (EMPLEADOS) y la otra la clave con el resto de campos (EMPLE_TRABAJO), ver Tablas 2.16 y 2.17:

<u>DNI</u>	<u>NUM_SEG_SOC</u>	<u>NOMBRE</u>	<u>APELLIDOS</u>
413245-B	28-1234566	JUAN	RAMOS
23456-J	28-2345686	PEDRO	PÉREZ
123123-C	19-458766	MARÍA	GIL
1234556-B	45-223344	ANTONIO	SANZ

Tabla 2.16. Relación EMPLEADOS e FNBC.

<u>DNI (FK)</u>	<u>DEPARTAMENTO</u>	<u>PUESTO</u>	<u>SALARIO</u>
413245-B	COMPRAS	GERENTE	2.300
23456-J	NÓMINAS	AUXILIAR	1.200
123123-C	ALMACÉN	CONSERJE	1.530
1234556-B	COMPRAS	GESTIÓN	2.200

Tabla 2.17. Relación EMPLERTRABAJO en FNBC.

• Cuarta Forma Normal (4FN)

Una relación se encuentra en 4FN si, y sólo si, está en FNBC y no existen dependencias multivaluadas.

Teorema de Fagin: Dada la relación formada por los atributos X, Y, Z con las siguientes dependencias multivaluadas: $X \twoheadrightarrow Y$ y $X \twoheadrightarrow Z$, entonces la relación puede descomponerse en dos relaciones: R1 (X, Y) y R2 (X, Z).

Caso práctico



- 14 Consideramos la siguiente relación: GEOMETRÍA, con tres atributos con dependencias multivaluadas, los tres atributos forman la clave, ver Tabla 2.18:

<u>FIGURA</u>	<u>COLOR</u>	<u>TAMAÑO</u>
ESFERA	ROJO	GRANDE
ESFERA	VERDE	GRANDE
CUBO	BLANCO	GRANDE
CUBO	AZUL	GRANDE
PIRÁMIDE	BLANCO	MEDIANO
PIRÁMIDE	BLANCO	GRANDE
PIRÁMIDE	ROJO	GRANDE

Tabla 2.18. Relación GEOMETRÍA con dependencias multivaluadas.

En la Tabla 2.18, FIGURA determina valores múltiples de COLOR y TAMAÑO, pero COLOR y TAMAÑO son independientes entre sí. Así pues, las dependencias son FIGURA \twoheadrightarrow COLOR y FIGURA \twoheadrightarrow TAMAÑO. Observamos que se repiten ESFERA-GRANDE, o PIRÁMIDE-BLANCO, o PIRÁMIDE-GRANDE. Para eliminar la redundancia de los datos se deben eliminar las dependencias de valores múltiples. Esto se logra construyendo dos tablas, donde cada una almacena datos para solamente uno de los atributos de valores múltiples.

Estas repeticiones hacen que la tabla no esté en 4FN. Para pasarla a 4FN, creamos las Tablas 2.19 y 2.20:

<u>FIGURA</u>	<u>COLOR</u>
ESFERA	ROJO
ESFERA	VERDE
CUBO	BLANCO
CUBO	AZUL
PIRÁMIDE	BLANCO
PIRÁMIDE	ROJO

Tabla 2.19. Tabla COLOR en 4FN.

<u>FIGURA</u>	<u>TAMAÑO</u>
ESFERA	GRANDE
CUBO	GRANDE
CUBO	MEDIANO
PIRÁMIDE	GRANDE

Tabla 2.20. Tabla TAMAÑOS en 4FN.



Actividades propuestas

- 3 Normaliza la Tabla 2.21. Suponemos que los estudiantes de una academia pueden inscribirse en varias especialidades y en varios cursos. Por ejemplo, el número de matrícula 12345 tiene especialidades en Sistemas y Telemática y toma los cursos de Natación y Danza.

NUMATRÍCULA	ESPECIALIDAD	CURSO
12345	SISTEMAS	NATACIÓN
12345	TELEMÁTICA	DANZA
34567	SISTEMAS	PIANO
34567	DESARROLLO	NATACIÓN
67891	TELEMÁTICA	AERÓBIC

Tabla 2.21. Tabla ACADEMIA.

- **Quinta Forma Normal (5FN)**

Una relación se encuentra en 5FN si, y sólo si, toda dependencia de reunión en la relación es una consecuencia de las claves candidatas. Esto es, la relación estará en 5FN si está en 4FN y no existen restricciones impuestas por el creador de la base de datos. La 5FN se refiere a dependencias que son extrañas. Tiene que ver con tablas que pueden dividirse en subtablas, pero que no pueden reconstruirse. Su valor práctico es ambiguo ya que conduce a una gran división de tablas.