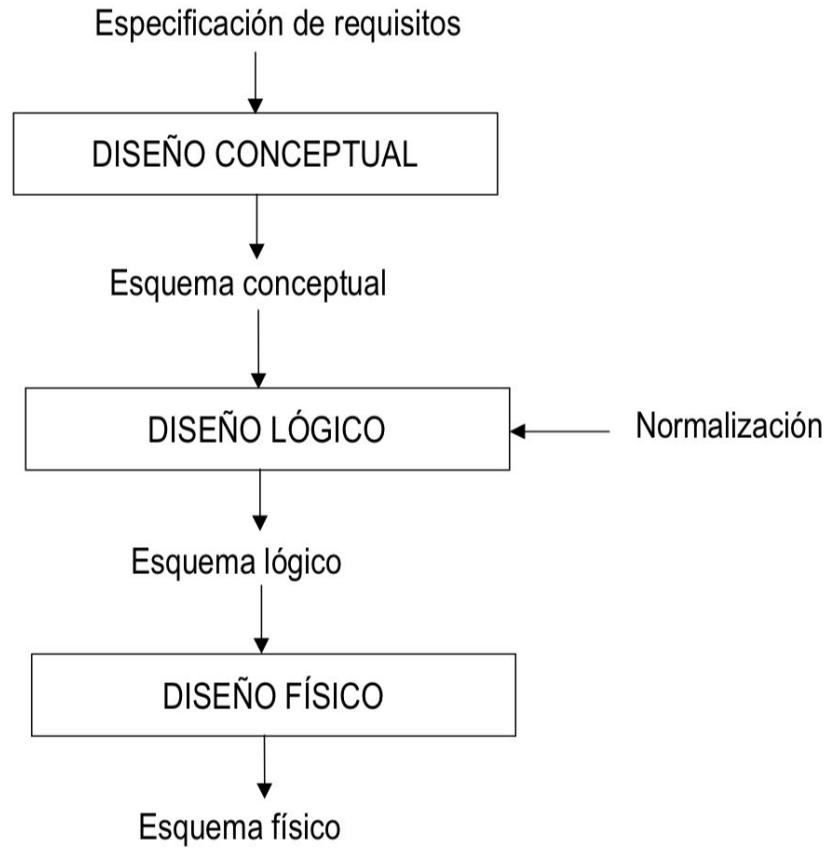


Bases de datos relacionales

Diseño Lógico

2. Metodología de diseño de bases de datos



Normalización

Cuando una tabla no está normalizada, cabe la posibilidad que alguna celda no esté llena, que los datos se repitan, que aparezcan dos datos o más en una celda, que algunos datos como atributos incorrectos, que los datos no sean valores simples, o que la misma tabla no cuente con claves primarias.

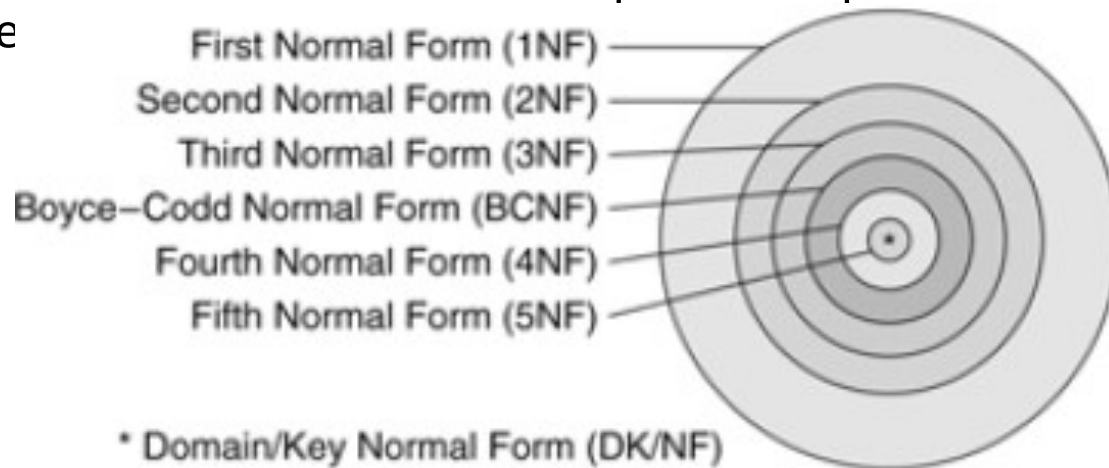
- Dependencia Funcional
- Primera Forma Normal
- Segunda Forma Normal
- Tercera Forma Normal

Normalización

- El proceso de normalización de bases de datos consiste en aplicar una serie de reglas a las tablas obtenidas tras el paso del modelo conceptual al modelo lógico.
- Las bases de datos relacionales se normalizan para:
 - Evitar la redundancia de los datos.
 - Evitar problemas de actualización de los datos en las tablas.
 - Proteger la integridad de los datos.

Normalización

- Según la propuesta original de Codd (1972), se somete un esquema de relación a una serie de pruebas para "certificar" si pertenece o no a una cierta forma normal.
- En un principio, Codd propuso tres formas normales, a las cuales llamó primera, segunda y tercera formas normales (1FN, 2FN, 3FN).
- Posteriormente, Boyce y Codd propusieron una definición más estricta de 3FN, a la que se conoce como forma normal de Boyce-Codd (FNBC).
- **Todas estas formas normales se basan en las dependencias funcionales entre los atributos de una tabla.**
- Más adelante se propusieron una cuarta forma normal (4FN) y una quinta (5FN), con fundamento en los conceptos de dependencias multivaluadas y depende



Dependencia Funcional



- Ocurre cuando el valor de un atributo o varios determina el valor de un segundo atributo o varios
- $A \rightarrow B$ (A determina funcionalmente a B)
- El atributo en la parte izquierda de la dependencia funcional es llamado el determinante
- Si conocemos el valor de A podemos conocer el valor de B.

Dependencia Funcional

- Por ejemplo, en una tabla

CLIENTE(Número_cliente, Nombre, Teléfono, Fecha_Nacimiento, Edad)

Existen las siguientes dependencias funcionales:

- DF: Número_cliente ~~Nombre~~
 - DF: Número_cliente ~~Teléfono~~
 - DF: Número_cliente  Fecha_Nacimiento
 - DF: Número_cliente  Edad
 - DF: Fecha_Nacimiento ~~Edad~~
-
- Para comenzar el proceso de normalización tenemos que estudiar las propiedades de todos los atributos de la tabla y analizar como están relacionados entre sí, buscando las posibles dependencias funcionales que existan.
 - Otro de los pasos previos al proceso de normalización es decidir cual es la clave primaria de la tabla.

Primera Forma Normal

- Una tabla está en 1FN si satisface las siguientes condiciones:
 1. Todos los atributos son “atómicos”. Por ejemplo, en el campo teléfono no tenemos varios teléfonos.
 2. La tabla contiene una clave primaria única. Por ejemplo el NIF para personas, la matrícula para vehículos o un simple id autoincremental. Si no tiene clave, no es 1FN.
 3. La clave primaria no contiene atributos nulos. No podemos tener filas para las que no haya clave (por ejemplo, personas sin NIF o vehículos sin matrícula).
 4. No debe existir variación en el número de columnas. Si algunas filas tienen 8 columnas y otras 3, pues no estamos en 1FN.

5. Los campos no clave deben identificarse por la clave. Es decir, que **los campos no clave dependen funcionalmente de la clave**. Esto es prácticamente lo mismo que decir que existe clave primaria.
6. Debe Existir una independencia del orden tanto de las filas como de las columnas, es decir, si los datos cambian de orden no deben cambiar sus significados. Por ejemplo, si en la tercera fila tenemos el tercer mejor expediente y en la quinta fila el quinto, no estamos en 1FN.

La primera forma normal se definió para prohibir los atributos multivaluados, compuestos y sus combinaciones.

1^{ra}. Forma Normal

- Indivisible: atomización de valores

Varios o distintos tipos de valores, en una misma celda de todos y cada uno de los registros, **no están permitidos**

Formadores
Diego Rodríguez Martín
Luz de León
Luis Ángel Pesce
Ricardo Balbín

Tabla: **Formadores**

ID	Nombre	Apellido
1	Diego	Rodríguez Martín ⚠
2	Luz	de León ⚠
3	Luis Ángel ⚠	Pesce
4	Ricardo	Balbín

Cliente			
ID Cliente	Nombre	Apellido	Teléfono
123	Rachel	Ingram	555-861-2025
456	James	Wright	555-403-1659 555-776-4100
789	Maria	Fernandez	555-808-9633

Cliente		
ID Cliente	Nombre	Apellido
123	Rachel	Ingram
456	James	Wright
789	Maria	Fernandez

Teléfono del cliente	
ID Cliente	Teléfono
123	555-861-2025
456	555-403-1659
456	555-776-4100
789	555-808-9633

- Cuando una tabla no está en primera forma normal, se divide en otras tablas, repartiendo sus atributos entre las resultantes. Normalmente la idea es eliminar el atributo que viola la 1FN de la relación original y colocarlo en una relación aparte junto con la clave primaria de la relación de partida.

Segunda Forma Normal

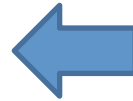
- Un esquema está en 2FN si:
 - Está en 1FN.
 - Y si todos los atributos no clave dependen funcionalmente de la clave **completa** y no sólo de una parte de esta.
- Este paso sólo se aplica a relaciones que tienen claves compuestas, es decir, que están formadas por mas de un atributo.

La Segunda Forma Normal (2NF) dice que todos los atributos no clave deben depender por completo de la clave primaria.

Segunda Forma Normal

Habilidades de los empleados

<u>Empleado</u>	<u>Habilidad</u>	Lugar actual de trabajo
Jones	Mecanografía	114 Main Street
Jones	Taquigrafía	114 Main Street
Jones	Tallado	114 Main Street
Bravo	Limpieza ligera	73 Industrial Way
Ellis	Alquimia	73 Industrial Way
Ellis	Malabarismo	73 Industrial Way
Harrison	Limpieza ligera	73 Industrial Way

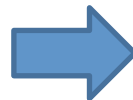


El atributo restante, Lugar actual de trabajo, es dependiente solo en parte de la clave candidata, llamada Empleado. Por lo tanto la tabla no está en 2NF.

Empleados

<u>Empleado</u>	Lugar actual de trabajo
Jones	114 Main Street
Bravo	73 Industrial Way
Ellis	73 Industrial Way
Harrison	73 Industrial Way

- Si una tabla no está en 2FN, se le puede normalizar a varias tablas en 2FN en las que los atributos que dependen de una parte de la clave formarán una nueva tabla que tendrá esa parte de la clave como clave primaria.



Habilidades de los empleados

<u>Empleado</u>	<u>Habilidad</u>
Jones	Mecanografía
Jones	Taquigrafía
Jones	Tallado
Bravo	Limpieza ligera
Ellis	Alquimia
Ellis	Malabarismo
Harrison	Limpieza ligera

ALUMNOS

<u>DNI</u>	<u>Cod Curso</u>	Nombre	Apellido1	Nota
12121219A	34	Pedro	Valiente	9
12121219A	25	Pedro	Valiente	8
3457775G	34	Ana	Fernández	6
5674378J	25	Sara	Crespo	7
5674378J	34	Sara	Crespo	6

La solución sería:

ALUMNOS

<u>DNI</u>	Nombre	Apellido1
12121219A	Pedro	Valiente
3457775G	Ana	Fernández
5674378J	Sara	Crespo

ASISTENCIA

<u>DNI</u>	<u>Cod Curso</u>	Nota
12121219A	34	9
12121219A	25	8
3457775G	34	6
5674378J	25	7
5674378J	34	6

Tercera Forma Normal

- Una tabla está en 3NF si y solo si las dos condiciones siguientes se mantienen:
 1. La tabla está en la segunda forma normal(2FN)
 2. Ningún atributo no--primario de la tabla es dependiente transitivamente de una clave candidata
- Es decir, una relación está en tercera forma normal si todos los atributos de la relación dependen funcionalmente sólo de la clave, y no de ningún otro atributo

Dependencia Transitiva - Sean X , Y , Z tres atributos (o grupos de atributos) de la misma entidad. Si Y depende funcionalmente de X y Z de Y , pero X no depende funcionalmente de Y , se dice que Z depende transitivamente de X . Simbólicamente sería:

$X \twoheadrightarrow Y \twoheadrightarrow Z$ entonces $X \twoheadrightarrow Z$

$FechaDeNacimiento \twoheadrightarrow Edad$

$Edad \twoheadrightarrow Conducir$

$FechaDeNacimiento \twoheadrightarrow Edad \twoheadrightarrow Conducir$



Tercera Forma Normal

Un ejemplo de una tabla 2NF que falla en satisfacer los requerimientos de la 3NF es:

Ganadores del torneo

<u>Torneo</u>	<u>Año</u>	Ganador	Fecha de nacimiento del ganador
Des Moines Masters	1998	Chip Masterson	14 de marzo de 1977
Indiana Invitational	1998	Al Fredrickson	21 de julio de 1975
Cleveland Open	1999	Bob Albertson	28 de septiembre de 1968
Des Moines Masters	1999	Al Fredrickson	21 de julio de 1975
Indiana Invitational	1999	Chip Masterson	14 de marzo de 1977

- La única clave candidata es {Torneo, Año}.
- **La violación de la 3NF ocurre porque el atributo no primario Fecha de nacimiento del ganador es dependiente transitivamente de {Torneo, Año} vía el atributo no primario Ganador.**
- El hecho de que la Fecha de nacimiento del ganador es funcionalmente dependiente en el Ganador hace la tabla vulnerable a inconsistencias lógicas, pues no hay nada que impida a la misma persona ser mostrada con diferentes fechas de nacimiento en diversos registros.
- Para expresar los mismos hechos sin violar la 3NF, es necesario dividir la tabla en dos:

Tercera Forma Normal

Ganadores del torneo

<u>Torneo</u>	<u>Año</u>	<u>Ganador</u>
Indiana Invitational	1998	Al Fredrickson
Cleveland Open	1999	Bob Albertson
Des Moines Masters	1999	Al Fredrickson
Indiana Invitational	1999	Chip Masterson

Fecha de nacimiento del jugador

<u>Jugador</u>	<u>Fecha de nacimiento</u>
Chip Masterson	14 de marzo de 1977
Al Fredrickson	21 de julio de 1975
Bob Albertson	28 de septiembre de 1968

Tercera Forma Normal

Puesto → Salario

IdEmpleado → Puesto → Salario

<u>idEmpleado</u>	Puesto	Salario
100	Secretaria	3100
150	Gerente	3200
200	Secretaria	3100
250	Subgerente	3200
300	Secretaria	3100

Relación que elimina la dependencia transitiva:

<u>idEmpleado</u>	Puesto
100	Secretaria
150	Gerente
200	Secretaria
250	Subgerente
300	Secretaria

<u>Puesto</u>	Salario
Secretaria	3100
Gerente	3200
Subgerente	3200

- La mayoría de las tablas 3NF están libres de anomalías de actualización, inserción, y borrado.

¿Qué tan lejos debe llevar la normalización?

La siguiente decisión es ¿qué tan lejos debe llevar la normalización? La normalización es una ciencia subjetiva. Determinar las necesidades de simplificación depende de nosotros. Si nuestra base de datos va a proveer información a un solo usuario para un propósito simple y existen pocas posibilidades de expansión, normalizar los datos hasta la 3FN quizá sea algo exagerado. Las reglas de normalización existen como guías para crear tablas que sean fáciles de manejar, así como flexibles y eficientes. A veces puede ocurrir que normalizar los datos hasta el nivel más alto no tenga sentido.

¿Se están dividiendo tablas sólo para seguir las reglas o estas divisiones son en verdad prácticas?. Éstas son el tipo de cosas que nosotros como diseñadores de la base de datos, necesitamos decidir, y la experiencia y el sentido común nos pueden auxiliar para tomar la decisión correcta. La normalización no es una ciencia exacta, más bien subjetiva.

Existen seis niveles más de normalización que no se han discutido aquí. Ellos son Forma Normal Boyce-Codd, Cuarta Forma Normal (4NF), Quinta Forma Normal (5NF) o Forma Normal de Proyección-Unión, Forma Normal de Proyección-Unión Fuerte, Forma Normal de Proyección-Unión Extra Fuerte y Forma Normal de Clave de Dominio. Estas formas de normalización pueden llevar las cosas más allá de lo que necesitamos. Éstas existen para hacer una base de datos realmente relacional. Tienen que ver principalmente con dependencias múltiples y claves relacionales.

Dada la relación Película(Título, Año, Duración, Tipo, Estudio, Actor) con las tuplas mostradas a continuación:

Título	Año	Duración	Tipo	Estudio	Dirección Estudio	Actor
Star Wars	1977	124	Color	Fox	Hollywood	Carrie Fisher
Star Wars	1977	124	Color	Fox	Hollywood	Mark Hamill
Star Wars	1977	124	Color	Fox	Hollywood	Harrison Ford
Mighty Ducks	1991	104	Color	Disney	Buena Vista	Emilio Estevez
Ben Hur	1959	212	Color	MGM	Hollywood	Charlton Heston
Ben Hur	1959	212	Color	MGM	Hollywood	Martha Scott
El retorno del Jedi	1983	124	Color	Fox	Hollywood	Carrie Fisher

Teniendo en cuenta, además, que:

- de un título se pueden haber realizado varias versiones en distintos años pero nunca con los mismos actores
- un determinado actor puede haber participado en varias películas durante un año
- no existe ningún estudio que esté ubicado en varias ciudades
- un actor puede trabajar con distintos estudios

ID_ORDEN	FECHA	ID_CLIENTE	NOM_CLIENTE	ESTADO	NUM_ITEM	DESC_ITEM	CANT	PRECIO
2301	2/23/03	101	MARTI	CA	3786	RED	3	35
2301	2/23/03	101	MARTI	CA	4011	RAQUETA	6	65
2301	2/23/03	101	MARTI	CA	9132	PAQ-3	8	4.75
2302	2/25/03	107	HERMAN	WI	5794	PAQ-6	4	5.0
2303	2/27/03	110	WE-SPORTS	MI	4011	RAQUETA	2	65
2303	2/27/03	110	WE-SPORTS	MI	3141	FUNDA	2	10