

1. Teoría de la normalización

La **normalización** es el proceso de **organizar datos en una base de datos**. Esto incluye la creación de tablas y el establecimiento de relaciones entre esas tablas de acuerdo con unas reglas diseñadas tanto para proteger los datos como para hacer que la base de datos sea más flexible al eliminar la redundancia y la dependencia inconsistente.

¿Qué son los datos redundantes?

Son aquellos que se repiten en distintos registros de la misma tabla o de tablas distintas y que desperdician espacio en disco y crean problemas de mantenimiento. Si tenemos que modificar en algún momento esos datos, deberemos hacerlo en todas las ubicaciones en las que se encuentren. En cambio, si, por ejemplo, debemos modificar la dirección de un/a alumno/a, sería mucho más fácil si esos datos estuviesen almacenados solo en la tabla ALUMNADO y en ningún otro lugar de la base de datos.

¿Qué es una dependencia inconsistente?

Significa que tenemos atributos en una tabla que no dependen del valor de la clave primaria de la misma. Si bien es intuitivo para un usuario buscar en la tabla ALUMNADO la dirección de un/a alumno/a en particular, puede que no tenga sentido buscar allí el número de horas de un módulo que cursa ese alumno/a. El número de horas del módulo está relacionado con el módulo o depende de él y, por lo tanto, debe trasladarse a la tabla MÓDULOS.

Tabla 5.1. Conceptos clave.

Las reglas formales de la teoría de la normalización son conocidas con el nombre de **formas normales**. Existen seis formas normales, de modo que cuando la base de datos cumple las reglas de la primera forma normal, se considera que está en primera forma normal (1FN); cuando pasan la segunda, que está en segunda forma normal (2FN), etcétera. También existe la forma normal de Boyce-Codd, que estudiaremos en el último apartado de esta unidad didáctica.

En adelante utilizaremos la notación 1FN, 2FN, 3FN y FNBC con las siguientes correspondencias:

1.ª forma normal	1FN
2.ª forma normal	2FN
3.ª forma normal	3FN
Forma normal de Boyce-Codd	FNBC

Tabla 5.2. Correspondencia de la notación a utilizar con el significado de esta.

Como ocurre con muchas reglas y especificaciones formales, los escenarios del mundo real no siempre permiten un cumplimiento perfecto. Además, la normalización requiere tablas adicionales. Debemos tener en cuenta que, **si no se cumplen las tres primeras reglas de normalización, tendremos**

que asegurarnos que la aplicación que utilice la base de datos anticipe cualquier problema que pueda ocurrir, como datos redundantes y dependencias inconsistentes.

De las seis formas normales, generalmente solo se aplican las tres primeras, considerando que **una base de datos que está en 3FN es una base de datos correctamente diseñada**.

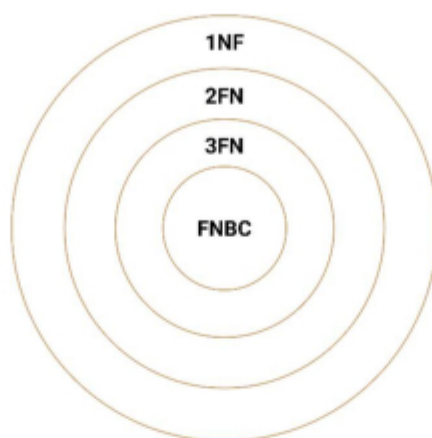


Fig. 5.1 Formas normales.

Las **anomalías de inserción, borrado y modificación** consisten en la aparición en la base de datos de estados no válidos en el mundo real. Por ejemplo, si cada profesor de una universidad tiene un departamento asociado a él, así como los grupos a los que imparte clase:

PROFESORADO			
Id_Profesor	Nombre	Departamento	Grupo
01054	José	Informática	1.º C
01078	Ana	Matemáticas	2.º A
01078	Ana	Matemáticas	2.º B
01069	Alejandro	Robótica	3.º A
01069	Alejandro	Robótica	1.º C

Fig. 5.2. Tabla PROFESORADO.

Anomalía de inserción	
	Es la imposibilidad de agregar datos a la base de datos debido a la ausencia de otros datos.
	Por ejemplo, suponemos que el campo Grupo está definido para que no se permitan valores nulos. Si se contrata a un nuevo/a profesor/a pero no se le asigna inmediatamente un grupo, este profesor no se podrá insertar en la base de datos. Esto da como resultado inconsistencias en la base de datos debido a omisiones.
Anomalía de borrado	
	Es la pérdida involuntaria de datos debido a la eliminación de otros datos.
	Por ejemplo, si el grupo de 1.º C se disolviera y fuera eliminado de la tabla anterior, el profesor José y el Departamento de Informática dejarían de existir. Esto da como resultado inconsistencias en la base de datos.


Anomalía de modificación	
	Es una inconsistencia de datos que resulta de la redundancia de datos y una actualización parcial.
	Si el departamento de Alejandro debe cambiar, debe actualizarse al menos 2 veces en la base de datos. Si el usuario que realiza la actualización no se percata de que los datos se almacenan de forma redundante, la actualización será incorrecta.

Tabla 5.3. Anomalías.

Las anomalías de actualización, eliminación e inserción son muy indeseables en cualquier base de datos. Las anomalías se evitan mediante el proceso de normalización.

