

3.6. Normalización

El proceso de normalización en bases de datos de relaciones conlleva aplicar un conjunto de técnicas para convertir un esquema relacional en otro que, representando la misma información, es más eficiente al contener menos redundancia y evitar anomalías en los procesos de alta, baja y actualización de los datos.

Estas técnicas parten de un modelo formal basado en la teoría de conjunto. Una base de datos relacional es un conjunto de relaciones representadas como $R(a_1, a_2, \dots, a_n)$, donde R es el nombre de dicha relación o tabla, que se define por una serie de atributos a_i . La integridad referencial indica que una clave ajena solo debe contener valores que existan en la otra relación a la que se referencia, o sean nulos.

El proceso de normalización conlleva realizar comprobaciones secuenciales del esquema original con el fin de determinar si cumplen ciertas formas normales atendiendo a las dependencias funcionales que existan en cada paso.

La siguiente relación llamada Empleados tiene como atributos el dni del empleado, el número de la Seguridad Social, el cargo que ocupa en la empresa, el sueldo que actualmente gana y sus teléfonos de contacto. Los datos que inicialmente tiene la tabla Empleados están recogidos en la Figura 3.57.


 Dni	numSegSoc	Nombre	Cargo	Sueldo	Telefonos
29992092	49009123	Adolfo Cantos	Técnico	1700	618443392; 954332219
44200101	789458341	Ana Sánchez	Jefe de Servicio	2600	699098743; 91900300; 619097568
84105768	48998410923	Carlos Gómez	Técnico	1700	690450985

Figura 3.57. Esta tabla no está en tercera forma normal. Tiene celdas en las que hay más de un valor y existe dependencia entre el sueldo y el cargo.

Se dice que una tabla está en primera forma normal, 1FN, cuando sus atributos contienen valores atómicos, es decir, en una columna solo hay una ocurrencia. El campo correo no cumple la primera forma normal. Para resolver esta violación se puede duplicar los registros con valores repetidos eliminando el atributo que hace que la tabla no esté en primera forma normal, incluyéndolo como clave primaria de la nueva tabla (Figura 3.58).




 Dni	numSegSoc	Nombre	Cargo	Sueldo	 Telefonos
29992092	49009123	Adolfo Cantos	Técnico	1700	618443392
29992092	49009123	Adolfo Cantos	Técnico	1700	954332219
44200101	789458341	Ana Sánchez	Jefe de Servicio	2600	699098743
44200101	789458341	Ana Sánchez	Jefe de Servicio	2600	91900300
44200101	789458341	Ana Sánchez	Jefe de Servicio	2600	619097568
84105768	48998410923	Carlos Gómez	Técnico	1700	690450985

Figura 3.58. Se resuelve la primera forma normal. En cada celda solo hay un valor.

Otra solución consiste en sustituir la relación por una nueva que no contiene dicha columna. Se crea una nueva relación cuya clave ajena es la columna que ha sido eliminada y referenciada a la nueva relación recién creada. Las tablas resultantes se recogen en la Figura 3.59.

 Dni	numSegSoc	Nombre	Cargo	Sueldo
29992092	49009123	Adolfo Cantos	Técnico	1700
44200101	789458341	Ana Sánchez	Jefe de Servicio	2600
84105768	48998410923	Carlos Gómez	Técnico	1700



  Dni	Telefonos
29992092	618443392
29992092	954332219
44200101	699098743
44200101	91900300
44200101	619097568
84105768	690450985

Figura 3.59. Las tablas se separan para no repetir los registros en la tabla de la izquierda. La redundancia ahora es nula.



El resto de los atributos cumplen la primera forma normal. En cuanto a la segunda, se dice que está en segunda forma normal, 2FN, si lo está en la primera y todos los atributos que no forman parte de la clave primaria tienen dependencias funcionales completas con respecto a todas las claves existentes en el esquema. La segunda forma normal se aplica a las relaciones que tienen claves primarias compuestas por más de un atributo.

to. Si una relación está en primera forma normal y su clave primaria es simple, entonces también está en segunda forma normal.

Una relación está en tercera forma normal si está en segunda y, además, cada atributo que no participa en la clave primaria no depende transitivamente de esta. Esta dependencia se denomina **dependencia funcional**. Existe una dependencia funcional en un esquema relacional cuando un valor de una columna siempre está asociado al mismo valor en otra columna. Por ejemplo, una dependencia funcional sería si el sueldo siempre es el mismo cuando se tiene el mismo cargo.

Cada dependencia funcional queda expresada en una nueva tabla con un nuevo campo clave como identificador de cada uno de los posibles valores. Esta columna será también clave ajena en la tabla primera.

Siguiendo el ejemplo y atendiendo a que se ha detectado una dependencia funcional entre el cargo y el sueldo, la tabla queda dividida como se ve en la Figura 3.60.

 Dni	numSegSoc	Nombre	 Cargo
29992092	49009123	Adolfo Cantos	Técnico
44200101	789458341	Ana Sánchez	Jefe de Servicio
84105768	48998410923	Carlos Gómez	Técnico

 Cargo	Sueldo
Técnico	1700
Jefe de Servicio	2600

Figura 3.60. Se resuelve la dependencia que existe entre cargo y sueldo. Aparece como resultado una nueva tabla.