

## FORMAS NORMALES DE UNA BASE DE DATOS

### Qué es la normalización en una base de datos

Normalización es un conjunto de reglas que sirven para ayudar a los diseñadores a desarrollar un esquema que minimice los problemas de lógica. Cada regla está basada en la que le antecede. La normalización se adoptó porque el viejo estilo de poner todos los datos en un solo lugar, como un archivo o una tabla de la base de datos, era ineficiente y conducía a errores de lógica cuando se trataba de manipular los datos.

Existen varios niveles de normalización: Primera Forma Normal, Segunda Forma Normal, Tercera Forma Normal, Forma Normal Boyce-Codd, Cuarta Forma Normal, Quinta Forma Normal 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.

Enfocare las 5 principales formas:

- **Primera forma normal:** La regla de la Primera Forma Normal establece que las columnas repetidas deben eliminarse y colocarse en tablas separadas.

La primera forma normal, impone un requisito muy elemental a las relaciones; a diferencia de las demás formas normales, no exige información adicional como las dependencias funcionales.

Un dominio es atómico si se considera que los elementos del dominio son unidades indivisibles. Se dice que el esquema de una relación R esta en la primera forma normal (1FN) si los dominios de todos los atributos de R son atómicos.

Un conjunto de nombres es un ejemplo de valor no atómico. Por ejemplo, si el esquema de la relación empleado incluyera el atributo hijos, los elementos de cuyo dominio son conjuntos de nombres, el esquema no se hallaría en la primera forma normal.

Los atributos compuestos, como el atributo dirección con sus atributos componentes calle y ciudad, tienen también dominios no atómicos.

Se da por supuesto que los enteros son atómicos, por lo que el conjunto de enteros es un dominio atómico; el conjunto de todos los conjuntos de enteros es un dominio no atómico. La diferencia estriba en que normalmente no se considera que los enteros tengan subpartes, pero si se considera que los tienen los conjuntos de enteros., es decir, los enteros que componen el conjunto. Pero lo importante no es lo que sea el propio dominio, si no el modo en que se utilizan los elementos del dominio en la base de datos.

El dominio de todos los enteros no sería atómico si se considerara que cada entero es una lista ordenada de cifras.

Poner la base de datos en la Primera Forma Normal resuelve el problema de los encabezados de columna múltiples.

La normalización ayuda a clarificar la base de datos ya organizarla en partes más pequeñas y más fáciles de entender. En lugar de tener que entender una tabla gigantesca y monolítica que tiene muchos diferentes aspectos, usted sólo tiene que entender objetos pequeños y más tangibles, así como las relaciones que guardan con otros objetos también pequeños.

- **Segunda forma normal (2FN):** La regla de la Segunda Forma Normal establece que todas las dependencias parciales se deben eliminar y separar dentro de sus propias tablas. Una dependencia parcial es un término que describe a aquellos datos que no dependen de la clave de la tabla para identificarlos.

En Otras palabras pudiésemos decir que la segunda forma normal está basada en el concepto de dependencia completamente funcional. Una dependencia funcional  $X \rightarrow Y$  es completamente funcional si al eliminar los atributos  $A$  de  $X$  significa que la dependencia no es mantenida, esto es que  $A \in X$ ,  $(X - \tilde{A}) \not\rightarrow Y$ . Una dependencia funcional  $X \rightarrow Y$  es una dependencia parcial si hay algunos atributos  $A \in X$  que pueden ser removidos de  $X$  y la dependencia todavía se mantiene, esto es  $A \in X$ ,  $(X - \tilde{A}) \rightarrow Y$ . Por ejemplo  $\{SSN, PNUMBER\} \rightarrow HOURS$  es completamente dependencia dado que ni  $SSN \rightarrow HOURS$  ni  $PNUMBER \rightarrow HOURS$  mantienen la dependencia. Sin embargo  $\{SSN, PNUMBER\} \rightarrow ENAME$  es parcialmente dependiente dado que  $SSN \rightarrow ENAME$  mantiene la dependencia.

Una de las mayores desventajas de la normalización es el tiempo que lleva hacerlo. La mayoría de la gente está demasiado ocupada, y emplear tiempo para asegurarse de que sus datos están normalizados cuando todo funciona más o menos bien, parece ser un desperdicio de tiempo. Pero no es así. Usted tendrá que emplear más tiempo arreglando una base de datos no normalizada que el que emplearía en una normalizada.

Al haber alcanzado la Segunda Forma Normal, usted puede disfrutar de algunas de las ventajas de las bases de datos relacionales. Por ejemplo, puede añadir

nuevas columnas a la tabla Clientes sin afectar a las tablas Productos y Pedidos. Lo mismo aplica para las otras tablas. Alcanzar este nivel de normalización permite que los datos se acomoden de una manera natural dentro de los límites esperados.

Una vez que ha alcanzado el nivel de la Segunda Forma Normal, se han controlado la mayoría de los problemas de lógica. Puede insertar un registro sin un exceso de datos en la mayoría de las tablas.

- **Tercera forma normal (3FN):** La regla de la Tercera Forma Normal señala que hay que eliminar y separar cualquier dato que no sea clave. El valor de esta columna debe depender de la clave. Todos los valores deben identificarse únicamente por la clave.

Un ejemplo de este concepto sería que, una dependencia funcional  $X \rightarrow Y$  en un esquema de relación  $R$  es una dependencia transitiva si hay un conjunto de atributos  $Z$  que no es un subconjunto de alguna clave de  $R$ , donde se mantiene  $X \rightarrow Z$  y  $Z \rightarrow Y$ . Por ejemplo, la dependencia  $SSN \rightarrow DMGRSSN$  es una dependencia transitiva en  $EMP\_DEPT$  de la siguiente figura. Decimos que la dependencia de  $DMGRSSN$  el atributo clave  $SSN$  es transitiva vía  $DNUMBER$  porque las dependencias  $SSN \rightarrow DNUMBER$  y  $DNUMBER \rightarrow DMGRSSN$  son mantenidas, y  $DNUMBER$  no es un subconjunto de la clave de  $EMP\_DEPT$ . Intuitivamente, podemos ver que la dependencia de  $DMGRSSN$  sobre  $DNUMBER$  es indeseable en  $EMP\_DEPT$  dado que  $DNUMBER$  no es una clave de  $EMP\_DEPT$ .

Esto le da más flexibilidad y previene errores de lógica cuando inserta o borra registros. Cada columna en la tabla está identificada de manera única por la clave, y no hay datos repetidos. Esto provee un esquema limpio y elegante, que es fácil de trabajar y expandir.

- **Forma normal de boyce-cood(FNBC) :** Es una de las formas normales mas deseables que se pueden obtener. Es un esquema de relación  $R$  esta en FNBC respecto a un conjunto de dependencias funcionales  $F$  si , para todas las dependencias funcionales de  $F^+$  de la forma  $\alpha \rightarrow \beta$ , donde  $\alpha \rightarrow R$  y  $B \rightarrow R$ , se cumple al menos una de las siguientes condiciones:

$\alpha \rightarrow$  es una dependencia funcional trivial  $\alpha$  es una superclave del esquema  $R$

Un diseño de base de datos esta en FNBC si cada miembro del conjunto de esquemas de relación que constituye el diseño esta en FNBC.

- **Cuarta forma normal (4FN):** Existe dependencia funcional multivalorada o de múltiples valores si, dados tres atributos de una tabla, si para cada valor del primer atributo existen múltiples valores en el segundo atributo y no hay ninguna relación entre el tercer atributo y el primero, a no ser a través del segundo atributo.

Una tabla está en Cuarta Forma Normal o 4FN si está en FNBC y las únicas dependencias funcionales multivaloradas que existen son las dependencias funcionales de la clave con los atributos que no forman parte de la misma. Estas dependencias multievaluadas de la clave con los atributos que no forman parte de la misma son dependencias triviales, por lo que algunos autores dicen que no existen dependencias multievaluadas en 4FN.

- **Quinta forma normal (5FN):** Se dice que hay dependencia de JOIN, de unión o de producto si una tabla tiene dependencia de \*unión con varias de sus \*proyecciones y se puede obtener la tabla por medio de la unión de dichas proyecciones.

**\*Proyección:** Creación de una tabla cuyos elementos forman un subconjunto de una tabla dada. Se incluyen todas las filas y algunas columnas.

**\*Unión:** Formar, a partir de dos tablas, una nueva con todos los campos de una de ellas y los registros de ambas, excepto los repetidos. Ambas tablas han de tener el mismo grado y las mismas columnas.

Una tabla esta en Quinta Forma Normal (5FN) o Forma Normal de Proyección-Unión si está en 4FN y las únicas dependencias que existen son las dependencias de unión de una tabla con sus proyecciones relacionándose entre las distintas proyecciones mediante la clave primaria o cualquier clave alternativa. La 5FN se emplea cuando en una misma tabla tenemos mucha información redundante, con pocos atributos o cuando una tabla posee una gran cantidad de atributos y se hace por ello inmanejable.

Para conseguir que una tabla 4FN con gran cantidad de atributos esté en 5FN, se parte la tabla original en tantas tablas como se desee, teniendo cada una de ellas en común con las demás los campos que forman la clave primaria en la tabla original.

## **CONCLUSIONES**

La normalización es una técnica que se utiliza para crear relaciones lógicas apropiadas entre tablas de una base de datos.

Ayuda a prevenir errores lógicos en la manipulación de datos. La normalización facilita también agregar nuevas columnas sin romper el esquema actual ni las relaciones.

Normalizar demasiado puede conducir a tener una base de datos ineficiente y hacer a su esquema demasiado complejo para trabajar.

Los datos redundantes desperdician el espacio de disco y crean problemas de mantenimiento.

## **BIBLIOGRAFIA**

- ✓ <http://www.slideshare.net/MonjeOneble/formas-normales>
- ✓ <http://www.mitecnologico.com/Main/FormasNormalesBasesDatos>
- ✓ [http://www.trucostecnicos.com/trucos/ver.php?id\\_art=278](http://www.trucostecnicos.com/trucos/ver.php?id_art=278)
- ✓ <http://support.microsoft.com/kb/283878/es>
- ✓ <http://www.scourdesign.com/articulos/BD-FN.php>