

Normalización

El objetivo de diseñar una BD es generar un conjunto de esquemas sin redundancia pero que a la vez faciliten la recuperación de información. Un enfoque es diseñar esquemas en una forma normal apropiada.

Defectos de una BD mal diseñada :

- * Repetición de la información.
- * Pérdida de la información.
- * Incapacidad para representar cierta información.

Para solucionar el problema se debe descomponer el esquema de relación con problemas en varios esquemas de relaciones.

Un buen principio de diseño podría ser “cada dato en un lugar” y evitar la redundancia si es posible.

Se dice que una relación está en cierta forma normal si cumple con un conjunto de restricciones.

Codd definió la primera, segunda y tercera formas normales. El objetivo del diseñador debe ser obtener relaciones en, por lo menos, 3NF.

Las formas normales son :

1NF : Se dice que una relación está en 1NF si cada dominio simple subyacente contiene solo valores atómicos.

Un atributo se dice primo si es parte de la clave primaria.

$X \rightarrow Y$ es una DF completa si Y depende de X y de ningún subconjunto de X.

2NF : Una relación está en 2NF si está en 1NF y cada atributo no primo DF de manera completa de cada clave candidata en la relación.

3NF : Una relación está en 3NF si está en 2NF y además cada atributo no primo DF de manera no transitiva de la clave primaria.

El proceso de normalización permite llevar una relación en cierta forma normal (por ejemplo 1NF) a un conjunto de relaciones en una forma más deseable (2NF). El proceso es reversible pues dada la salida (conjunto de relaciones en 2NF), es posible obtener nuevamente la entrada (relación en 1NF). La operación de descomposición es la proyección y el de recomposición la reunión natural.

Dada una relación mal diseñada con demasiados atributos es posible descomponerla en 2 o más relaciones con menos atributos. Es esencial que la descomposición sea sin pérdida de información, pues la recomposición podría dar como resultado un superconjunto de la relación original y no hay manera de saber que tuplas son espurias. Por lo tanto una buena descomposición asegura que la recomposición producirá exactamente la relación original.

En general, los motivos para normalizar son :

- * Eliminar ciertos tipos de redundancia.
- * Evitar anomalías de actualización.
- * Obtener un diseño que sea una buena representación del mundo real que se está modelando.

A veces hay buenos motivos para no normalizar : Datos muy consultados y poco actualizados no conviene normalizar.