**Resumen Normalización**

**Introducción**

El proceso de normalización de una BD consiste en aplicar una serie de reglas a las tablas obtenidas a partir del diseño físico. Una BD debe normalizarse para evitar la redundancia de los datos, dado que cuando se repite innecesariamente la información, aumentan los costos de mantenerla actualizada. Además, la normalización ayuda a proteger la integridad de la información de la BD. Entonces, la normalización es un mecanismo que permite que un conjunto de tablas cumpla una serie de propiedades deseables, las mismas consisten en evitar:

* Redundancia de datos.
* Anomalías de actualización.
* Pérdida de integridad de datos.

El proceso de normalización es un proceso deseado pero no estrictamente necesario. Durante dicho proceso, se toman decisiones que producen cambios en las tablas, los cuales pueden afectar los tiempos de acceso a los datos almacenados en esas tablas por lo cual en determinadas situaciones la solución normalizada no es conveniente para la operatoria sobre la BD, pues el tiempo de consulta podría resultar muy elevado debido a que si bien evita la redundancia de datos, al mismo tiempo segmenta más la información.

**Redundancia y anomalías de actualización**

Si bien el proceso de normalización tiene como principal objetivo eliminar la redundancia, existe redundancia necesaria que es llamada *redundancia deseada*, la misma se genera, por ejemplo, cuando una tabla con las CP de las dos entidades que relaciona, ésta redundancia (información que se repite) es necesaria para representar el modelo físico; éste tipo de redundancia no afecta al proceso de normalización. Los tipos de anomalías que pueden producirse son:

* Anomalías de inserción.
* Anomalías de borrado.
* Anomalías de modificación.

**Dependencias funcionales**

Una *Dependencia Funcional (DF)* representa una restricción entre atributos de una tabla de la BD. Se dice que un atributo Y depende funcionalmente de un atributo X (denotado X -> Y) cuándo para un valor de X siempre se encuentra el mismo valor para el atributo Y. Esto permite deducir que cualquier atributo depende funcionalmente de la CP y de las claves candidatas. Es conjunto mínimo de DF tiene las siguientes propiedades:

* El consecuente de la DF está formado por un solo atributo. Si se define la DF X -> Y, Y debe ser un sólo atributo.
* Si se define la DF X -> Y, no es posible encontrar otra dependencia Z -> Y donde > sea un subconjunto de atributos de Z. En este caso se define a la DF como completa.
* No se puede quitar de un conjunto de DF alguna de ellas, y seguir teniendo un conjunto mínimo.

Una DF X -> Y se denomina *parcial* cuando, además, existe otra DF Z -> Y, siendo Z un subconjunto de X. Èste tipo de DF contradicen la definición de conjunto mínimo de DF lo que trae aparejado repetición de informaciòn y, consecuentemente, las anomalías de actualización. Una DF X -> Y se denomina *transitiva* cuando existe un atributo Z tal que X -> Z y Z -> Y. Este tipo de DF contradice la definición de conjunto mínimo de DF ocasionando anomalías de actualización.

**Dependencia funcional de Boyce-Codd**

Una DF X -> Y se denomina Dependencia Funcional de Boyce-Codd (DFBC) cuándo X no es una CP o CC, e Y es una CP o CC, o parte de ella.

**Formas normales**

**Primera Forma Normal**

La Primera Forma Normal plantea que todos los atributos que conforman una tabla deben ser monovalentes, es decir, la cardinalidad de los atributos debe ser cero o uno. La solución es quitas el atributo polivalente y convertirlo en una tabla relacionada a la entidad dueña del atributo polivalente. Éste proceso se realiza en el esquema lógico asegurando que el esquema físico obtenido tendrá todas sus tablas en primera forma normal.

**Segunda Forma Normal**

Un modelo está en Segunda Forma Normal si está en 1FN y, además, no existe en ninguna tabla del modelo una DF parcial. La solución para tratar una DF parcial es quitar de la tabla el atributo o los atributos que generan ésta DF, y situarlos en una nueva tabla, lo que significa que la DF planteada debe mantenerse en el modelo, pero en una ubicación donde ya no sea una DF parcial.

**Tercera Forma Normal**

Un modelo está en Tercera Forma Normal si está en 2FN y, además, no existe en ninguna del modelo una DF transitiva.

**Forma Normal de Boyce-Codd**

La Forma Normal de Boyce-Codd fue propuesta con posterioridad a la 3FN como una simplificación de éste, pero en definitiva generó una forma normal más restrictiva. Un modelo está en Forma Normal de Boyce-Codd si está en 3FN y, además, no existe en ninguna tabla del modelo una DF de BC. Entonces, para generar un modelo en FNBC se debe garantizar que los determinantes de cada DF sean siempre CC. La regla de normalización plantea que las DF de BC deben ser quitadas de la tabla donde se generan, y llevadas a una nueva tabla.

**Dependencia multivaluada y Cuarta Forma Normal**

Una \*\*Dependencia Multivaluada (DM)\*, denotada como X ->> Y siendo X e Y conjuntos de atributos de una tabla, indica que para un valor determinado de X es posible determinar múltiples valores para el atributo Y. El problema de la multideterminación de datos comienza cuando en una tabla cualquiera tiene tres atributos X, Z, Y y se analizan siguientes DM:  
(X, Y) ->> Z  
(X, Z) ->> Y  
Pero en realidad tanto Y como Z solamente dependen de X en cada una de las DF. La solución en éste caso es tratar de generar un modelo dónde las DM que se definen no presenten ésta anomalía. Ésta solución se plantea como la Cuarta Forma Normal.

Un modelo está en Cuarta Forma Normal si está en FNBC y as DM que se puedan encontrar son triviales.

Una DM X ->> Y se considera trivial si Y está multideterminado por el atributo X y no por un subconjunto de X.

**Nota sobre más formas normales.**

Las formas normales posteriores a 4FN representan situaciones muy particulares y específicas.