

## Objetivo

Definir los criterios generales para el diseño de las bases de datos de los sistemas de información de la SEMARNAT.

## Descripción

La utilización de bases de datos como plataforma para el desarrollo de sistemas de información en las organizaciones se ha generalizado en los últimos años. Lo anterior se debe a las ventajas que ofrece su utilización, algunas de las cuales son:

- Globalización de la información.- Permite a los diferentes usuarios considerar la información como un recurso institucional que carece de dueños específicos.
- Eliminación de información inconsistente: Si existen dos o más archivos con la misma información, los cambios que se hagan a éstos deberán hacerse a todas las copias existentes en la institución.
- Permite compartir información.
- Permite mantener la integridad en la información: La integridad de la información es una de sus cualidades deseables necesarias para crear confiabilidad respecto de los datos y tiene por objetivo que sólo se almacene la información correcta.
- Independencia de datos: Para los desarrolladores de sistemas tiene múltiples ventajas de diseño ya implica una clara división entre programas y datos.
- Seguridad de la información: Adicionalmente a considerar la información como un activo institucional, el uso de bases de datos proporciona mecanismos de respaldo y recuperación que difícilmente podrán proporcionar esquemas de almacenamiento orientados exclusivamente a la creación de contenidos.

Una de las fases críticas en la construcción de un sistema de información que maneje una base de datos, es sin duda, su propio diseño. Si las tablas no son definidas apropiadamente, pueden ocasionar múltiples problemas al momento de ejecutar, consultar y manipular la información.

El presente documento permitirá clarificar los elementos que deberán tomarse en cuenta para el diseño de las bases de datos utilizadas por los sistemas de información de la Secretaría y que permitan fortalecer paulatinamente el uso de convenciones, tanto en la nomenclatura, como en los criterios de normalización y documentación utilizados.

## Objetivos del Diseño de Bases de Datos

- Satisfacer los requisitos de contenido e información de los usuarios y aplicaciones.
- Proporcionar una estructura de la información, natural y fácil de entender.
- Soportar los requisitos de procesamiento y cualesquier otros objetivos de rendimiento y el espacio de almacenamiento.

El diseño de una base de datos deberá descomponerse en tres etapas:

Diseño conceptual  
Diseño lógico y;  
Diseño físico.

### Diseño Conceptual

El diseño conceptual parte de las especificaciones de los requisitos del usuario y su resultado es el esquema conceptual de la base de datos, el cual se define como una descripción de alto nivel de la estructura que puede tomar la base de datos.

En esta etapa se deberá construir un mapa de la información que se utilizará en el universo de datos que se especifique para cada aplicación, independientemente de cualquier consideración física. Al desarrollar el esquema, los diseñadores definirán el significado de los datos de la Organización. Encontrarán entidades, atributos y relaciones. El objetivo de esta fase será comprender los siguientes aspectos:

- La perspectiva que cada usuario tiene de los datos.
- La naturaleza de los datos, independientemente de su representación física.
- El uso de los datos a través de las áreas de aplicación.

El producto de esta etapa puede ser utilizada para que el diseñador transmita al usuario lo que ha entendido sobre la información que maneja. Para ello, ambas partes deben estar familiarizadas con la notación utilizada en el esquema. La notación a utilizar es la referente al modelo entidad-relación.

### Diseño lógico

En esta etapa, se transforma el esquema conceptual en un esquema lógico que utilizará las estructuras de datos del modelo de base de datos en el que se basa el DBMS que se vaya a utilizar, como puede ser el modelo relacional, el modelo de red, el modelo jerárquico o el modelo orientado a objetos. Conforme se va desarrollando el esquema lógico, éste se probará y validará contra los requisitos del usuario. En el caso de los sistemas de información transaccionales de la Secretaría, se utilizará el modelo relacional.

El esquema lógico es una fuente de información para el diseño físico. Además, juega un papel importante durante la etapa de mantenimiento del sistema de información, ya que permite que los futuros cambios que se realicen sobre los programas de aplicación o sobre los datos, se representen correctamente en la base de datos.

Tanto el diseño conceptual, como el diseño lógico, son procesos iterativos, tendrán un punto de inicio y se deberán refinar continuamente manteniendo un estricto control de cambios.

### **Diseño físico**

El diseño físico es el proceso de producir la descripción de la implementación de la base de datos en estructuras de almacenamiento y métodos de acceso que garanticen un desempeño eficiente en la manipulación de los datos.

Para llevar a cabo esta etapa, ya debió haberse tomado la decisión respecto a cuál será el DBMS que se va a utilizar, ya que el esquema físico se adapta en gran parte a él. Entre el diseño físico y el diseño lógico existe un proceso constante de realimentación, ya que algunas de las decisiones que se tomen durante el diseño físico para mejorar el potencial relacionado al manejador, pueden afectar a la estructura del esquema lógico.

### **Consideraciones del Diseño de Bases de Datos**

#### **Nomenclatura**

- Utilizar caracteres alfanuméricos.
- Limitar los nombres a menos de 64 caracteres.
- Utilizar el guión bajo (\_) para separar palabras.
- Utilizar palabras en minúsculas.
- Los nombres de las tablas deberían ir en plural y los nombres de las columnas en singular.
- Utilizar las letras ID en las columnas de llave primaria y foránea.
- En una tabla, colocar primero la llave primaria seguida de las llaves foráneas.
- Los nombres de los campos deben ser descriptivos de su contenido.
- Los nombres de los campos deben ser unívocos entre tablas, excepción hecha de las llaves.

Los puntos señalados, permitirán que paulatinamente la nomenclatura utilizada en las bases de datos de la Secretaría sea coherente y consistente con el fin de minimizar la posibilidad de errores al momento de crear interfases que permitan el intercambio de información entre aplicaciones.

#### **Definición de Catálogos o Datos Maestros**

Durante el diseño de la base de datos deberá procurarse eliminar al máximo el uso y aplicación de datos almacenados en texto libre. Será importante que a criterio del diseñador se promueva el uso de información catalogada, la cual permitirá incluso facilitar la localización y consulta de información plenamente identificada.

Por otra parte, es necesario considerar todas aquellas entidades y atributos que se encuentran normadas por alguna institución externa, o bien por la propia Secretaría. Tal es el caso de las Entidades Federativas, Municipios y Localidades que pertenecen a catálogos normados y emitidos por el INEGI.

Para todos los casos anteriores, se hará obligatorio el uso y definición de los mismos, a fin de garantizar el traslado de la información incluso entre entidades ajenas a la SEMARNAT.

### **Normalización**

La normalización de los datos deberá considerarse como un proceso de análisis de los esquemas de relación obtenidos en la etapa de diseño lógico de la Base de Datos. La forma de efectuar esto a través de la optimización de las dependencias determinadas dentro de cada relación.

Los lineamientos para el diseño de relaciones de bases de datos:

1. Cada entidad de datos separada debe crear un archivo maestro. No es correcto combinar dos entidades distintas en un solo archivo.
2. Un campo de datos específicos debe existir solamente en un archivo maestro.
3. Cada archivo maestro o relación de base de datos debe tener programas para crear, leer, actualizar y borrar registros, lo ideal es que sólo un programa añada registros y otro elimine.

El proceso de normalización proporcionará a los diseñadores los siguientes aspectos importantes:

- Un marco formal para analizar los esquemas de relación basándose en sus llaves y en las dependencias funcionales entre sus atributos.
- Una serie de pruebas de formas normales que puedan afectarse sobre esquemas de relación individuales de modo que la base de datos pueda normalizarse hasta el grado deseado.

En lo general deberá buscarse que el esquema de datos definido en su conjunto alcance la tercera forma normal, en cuyo caso una relación no debería tener un atributo no clave determinado funcionalmente por otro atributo no clave (o por un conjunto de ellos).

### **Almacenamiento de información**

Durante el diseño, se deberá decidir qué información es realmente necesaria. Frecuentemente se pueden generar algunos datos sobre la marcha sin tener que almacenarlos en una tabla de la base de datos.

### **Uso de tablas temporales**

El diseñador podrá considerar el uso de tablas temporales para la manipulación de información que no necesariamente deba ser almacenada en forma definitiva.

Generalmente, una tabla temporal podrá existir mientras dure la conexión a la base de datos. Cuando esta conexión se interrumpe, el DBMS deberá remover automáticamente la tabla y liberar el espacio que utilizaba.

12

### **Criterios para la creación de índices**

Cuando existen búsquedas frecuentes sobre un atributo o conjunto de ellos, lo recomendable es tener esa columna indexada para obtener más rápido la información.

Los criterios a seguir son muy simples, si la información de la base de datos se consultará con frecuencia sobre un atributo, ese es el candidato ideal para ser indexado.

El almacenamiento de grandes textos en una base de datos puede resultar erróneo, así como crear índices sobre campos de texto. En todos los casos, si se tiene visualizado que algún dato es candidato a convertirse en criterio de búsqueda, no deberá ser definido del tipo texto libre y por el contrario, se buscará que pueda ser catalogado en un archivo maestro.

### **Características Adicionales a considerar en la etapa del Diseño**

**Independencia de los Datos:** Durante el diseño deberá considerarse al menos la división entre las capas de definición de reglas de negocio o transacciones y las del almacenamiento de datos.

**Crecimiento:** Mantenimiento constante al modelo de datos. Conforme crezca la base de datos para incorporar nuevos tipos de información, así también deberá crecer la definición de la base de datos.

**Capacidad de Auditoria:** El diseño de la base de datos siempre deberá considerar el manejo y registro de usuarios y permitir la rastreabilidad de las transacciones. Para todas aquellas entidades que estén diseñadas para soportar el intercambio de información en procesamiento batch, deberá ser diseñada vista la generación y almacenamiento de bitácoras o logs, que si bien, no necesariamente deberán alojarse en forma definitiva en la base, es recomendable su permanencia al menos por tres meses posteriores a su ejecución. Este concepto es otra forma de asegurar la calidad de la información contenida en el sistema y coadyuva para asegurar su confiabilidad.

**Desempeño:** El diseño de la base de datos siempre deberá responder a las necesidades de desempeño. Es válido que durante la fase de diseño físico se realicen los ajustes apropiado al manejador que se haya elegido. O bien ante cambios en los requerimientos tanto funcionales como técnicos.