

Sigla Asignatura	MDY1131	Nombre de la Asignatura	Modelamiento de Base de datos	Tiempo	1 h.
Nombre del Recurso Didáctico	1.1.6 Material Complementario Lectura Enfoques de Bases de Datos				
Experiencia de Aprendizaje N° 1	Construyendo un Modelo Conceptual Simple				

Sistemas de Gestión de Bases de Datos.

Introducción: Justificación del uso de las Bases de Datos

Evolución en el tratamiento de la información.

En casi todos los ámbitos del mundo actual, es fundamental el procesamiento de datos para obtener información que permita tomar decisiones.

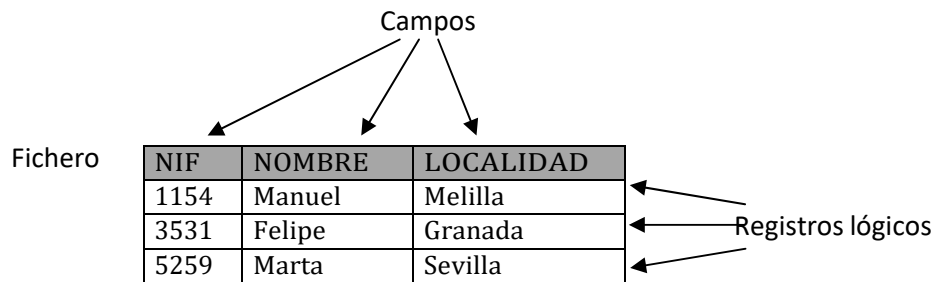
El mundo empresarial y administrativo cada vez utilizan más datos, lo cual dificulta su procesamiento. A partir de los años sesenta varias empresas empiezan a almacenar sus datos en soportes magnéticos para ser tratados directamente por computadores. Desde entonces el auge de los equipos informáticos y soportes de información ha ido aumentando a pasos agigantados, y de igual forma han ido disminuyendo su costo.

En la evolución del tratamiento automático de la información podemos ver varias etapas: Las primeras aplicaciones integraban los datos en los programas en forma de constantes y sólo se analizaban pequeñas parcelas del mundo científico. Posteriormente aparecen los ficheros, que representan pequeñas parcelas del mundo real. En un principio estos ficheros sólo podían ser accedidos de forma secuencial y posteriormente de forma aleatoria. Las aplicaciones estaban constituidas por programas independientes y cada uno de ellos procesaba sus propios ficheros. Por ello, algunos datos tenían que ser repetidos en distintos archivos, pudiendo producirse incoherencias entre los mismos. A finales de los años sesenta, para evitar estos problemas, surgen las bases de datos, que pretenden modelar grandes parcelas del mundo real mediante la interrelación de varios archivos, evitando las redundancias y posibles incoherencias.

Características de los ficheros convencionales

El computador no sólo trabaja con datos individuales y aislados, sino con conjuntos de datos lógicamente relacionados. Se denomina **registro lógico** a un conjunto de datos referentes a una misma entidad y que constituyen una unidad para el proceso. A cada uno de los datos que forman un registro lógico se le denomina **campo** o **atributo**.

Llamamos **archivo** a un conjunto de registros (lógicos) del mismo tipo, que posee una organización coherente y que se ubica en dispositivos de almacenamiento secundario.



Se llama **registro físico o bloque** al conjunto de datos transferidos en una sola operación de entrada/salida. Se conoce como **factor de bloqueo** al número de registros lógicos que hay en cada registro físico o bloque.

Los registros pueden ser de **longitud fija** o de **longitud variable**. La utilización de registros de longitud fija provoca pérdida de espacio, pero son más fáciles de manejar. Los registros de longitud variable provocan menos pérdida de espacio pero su utilización será más difícil al tener que utilizar marcas de principio y final de registro, o campos extra con la longitud del registro.

Asociado a los archivos de un soporte suele existir un **directorio de archivos**, se trata de un archivo, donde para cada archivo del soporte, existe un registro de directorio. Este registro contiene información acerca de la posición y el formato de los registros que forman el archivo. Según el tipo de organización el contenido de este registro será diferente. Normalmente el contenido de cada registro directorio suele ser el nombre del archivo, el propietario, el punto inicial, el punto final, la cantidad de espacio permitida y la utilizada actualmente. El propietario de este archivo directorio es el Sistema Operativo.

Se denomina **apertura de un fichero** al proceso de leer el registro directorio asociado al fichero y retener esta información para su consulta posterior. El proceso contrario, **cerrar el archivo**, consiste en la actualización del directorio de archivo correspondiente al fichero, en el caso que haya sufrido alguna modificación.

Tipos de acceso: Se denomina acceso al proceso físico de localizar la información contenida en el fichero. Podrá ser secuencial o directo:

- **Acceso secuencial:** Se accede a los registros siguiendo su orden físico de aparición en el soporte de entrada.
- **Acceso al azar o directo:** Se accede a los registros atendiendo a un orden aleatorio de solicitud de estos. Este orden es externo al programa y no predecible.

El tipo de acceso dependerá directamente del **tipo de soporte** que se utilice:

- **Soporte de acceso secuencial:** Para acceder a un registro físico estamos obligados a acceder a todos los registros físicos que le preceden. Por ejemplo: una cinta magnética.
- **Soporte de acceso directo:** Son los que permiten el acceso a cualquier registro físico una vez conocida su dirección. Por ejemplo: un disco magnético. Los soportes de acceso directo también admiten accesos secuenciales.

Tipo de proceso: Dependerá de la aplicación que procese el archivo. Podrá ser de dos tipos:

- **En lote o batch:** En general son los que no requieren la intervención del usuario. Un caso típico son las actualizaciones de un fichero maestro a partir de uno de movimientos. En estos procesos todos los movimientos (altas, bajas, y modificaciones) sobre un fichero denominado maestro son acumulados durante cierto tiempo en un fichero de movimientos, que posteriormente es procesado actualizando al fichero maestro. Los procesos en lote son eficaces siempre que no sea necesario que el fichero maestro esté continuamente actualizado.
- **En línea:** En los procesos en línea los movimientos o alteraciones del fichero son tratadas en el momento en que se producen. Estos procesos se caracterizan por permitir el diálogo con el usuario, se dice que son interactivos o conversacionales.

Clasificación de los ficheros según su uso.

Los archivos se clasifican según la utilización que se hace de ellos en tres grupos:

- **Archivos permanentes.** Contienen información que varía poco a lo largo del tiempo. Su contenido es esencial para la aplicación que los utiliza. Deben protegerse ante alteraciones accidentales y realizar copias de seguridad de forma periódica para garantizar la reconstrucción en caso de pérdida total o parcial. Pueden ser de tres clases:
 - **Archivos de constantes.** Su información permanece prácticamente inamovible, utilizándose principalmente como archivos de consulta. Un archivo de este tipo puede ser el archivo de datos personales de empleados, que suele sufrir escasas modificaciones.
 - **Archivos de situación o maestros.** Reflejan el estado o situación de una empresa, entidad o algún aspecto de ella en un determinado momento. Estos archivos se actualizan periódicamente para adaptarlos a cada nueva situación. Un ejemplo de este tipo de archivos sería: el archivo de existencias de almacén, que para cada entrada o salida de almacén deberá ser modificado.
 - **Archivos históricos.** Se obtienen de los anteriores cuando quedan fuera de uso. Aunque la información que contienen no es de actualidad, reflejan la evolución de esta. Suelen ser de gran tamaño, y no es frecuente su consulta al azar, su uso habitual es para elaborar estudios estadísticos. Un archivo histórico puede ser el que contiene la información de libros adquiridos por una biblioteca en la década de los ochenta. Otro podría ser, la facturación de una empresa en el año 1.998.

Sistemas Gestores de Bases de Datos (SGBD)

Definición de Base de Datos

Una definición breve e informal podría ser la siguiente: “Una colección de datos que están lógicamente relacionados entre sí, que tienen una definición y una descripción comunes y que están estructurados de forma particular”. Una base de datos es también, un **modelo** de una parcela del mundo real, por tanto debe poder servir para una amplia gama de aplicaciones.

Objetivos de los sistemas gestores de bases de datos.

Los objetivos que pretenden cubrir las bases de datos podrían resumirse en los siguientes:

- **Independencia entre datos y aplicaciones:** En una base de datos el programa accede a los datos por medio de un software intermedio que oculta al programa de aplicación la estructura concreta de los datos. Cuando se tenga que cambiar la descripción de un fichero o el formato de sus datos, no habrá que cambiar los programas de aplicación.
- Evitar **redundancias** de datos, es decir, ficheros distintos con información común destinados a diferentes aplicaciones.
- Evitar la **inconsistencia** o **incoherencia** de datos, para ello es fundamental evitar la redundancia de datos en distintos ficheros. La actualización de un dato se hará una única vez en un único fichero físico al que se accede desde cualquier aplicación, y por lo tanto no habrá riesgo de valores inconsistentes para el mismo dato.
- **Centralización del control** de la base de datos mediante un software de control **SGBD** (sistema gestor de la base de datos) y la figura del **administrador de la base de datos**.
- Posibilidad de obtener información de la base de datos sin necesidad de tener que construir programas “a propósito”. Esto puede realizarse gracias a los **lenguajes de consulta** (*Query Language*).
- Preservar la **integridad** de los datos mediante sistemas automáticos de validación, respaldos sistemáticos y procedimientos automáticos de reconstrucción.
- Garantizar la **confidencialidad y seguridad** de los datos contra accesos incorrectos o no autorizados.
- Permitir los **accesos concurrentes** a los datos.

Elementos de un sistema gestor de bases de datos.

1. **Los datos.** Se pueden clasificar en
 - Datos Simples. Ej. : nombre, dirección, teléfono, etc.
 - Datos Compuestos. Ej. : tablas, etc.
2. **El diccionario de datos.** Es una base de datos que contiene datos de interés sobre los datos de la base de datos a la que pertenece. Por ejemplo, puede recoger información sobre la representación física de los datos, asignación a dispositivos, formas de acceso, índices..., además posee el repertorio de todos los campos y variables que existen en la BD, así como descripción, longitud, posibles valores de cada uno de ellos, etc.
3. **El sistema de gestión de la Base de Datos (SGBD).** Conjunto de programas que gestionan la Base de Datos. Las funciones principales son crear, actualizar, consultar, proteger la Base de Datos.
4. **El administrador de la Base de Datos (DBA).** Es una función que pueden llevar a cabo una o varias personas. Entre las responsabilidades del administrador están la de creación, gestión y mantenimiento de la Base de Datos.
5. **Los usuarios.** Todos aquellos que utilizan la Base de Datos. Distinguiendo entre usuarios informáticos, aquellos a los que el administrador le dará todas las herramientas para que puedan realizar su trabajo, y los no informáticos (usuarios finales) a los que va dirigida la Base de Datos.

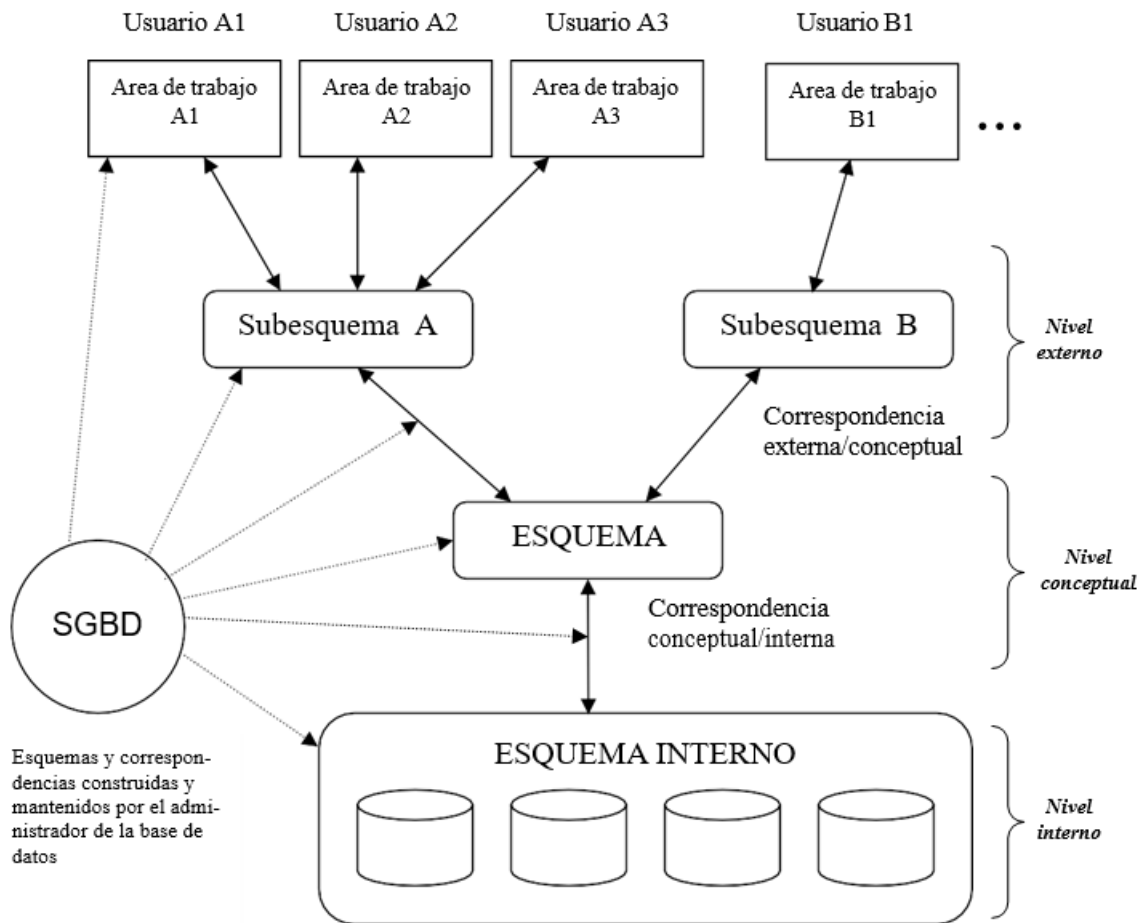
Arquitectura en tres niveles (ANSI – SPARC)

Uno de los objetivos de un sistema de bases de datos es proporcionar a los usuarios una visión abstracta de la información, ocultando ciertos detalles acerca de cómo se almacenan los datos, pero permitiendo una recuperación eficaz de la información.

Existen varios paquetes de sistemas de bases de datos en el mercado con diferente arquitectura. Una de las arquitecturas más estandarizada es la especificada por la normativa ANSI/X3/SPARC (*Standard Planning and Requirements Committee of the American National Standards Institute – Comité de Planificación y Requerimientos de estándares del Instituto Nacional de Estándares de Estados Unidos, división X3*).

Según esta norma la arquitectura de una base de datos debe tener tres niveles: **externo**, **conceptual** e **interno**. En la siguiente figura se muestra la organización de esta arquitectura.

Arquitectura ANSI/X3/SPARC:



Nivel externo.

Es el más cercano a los usuarios. Representa la visión de cada usuario o programador de la base de datos. En el nivel externo se describe de forma individual un subconjunto de datos para un usuario o conjunto de usuarios. Los usuarios pueden trabajar con los archivos externos como si existiesen físicamente, aunque éstos en realidad no existen tal como son vistos por el usuario.

Por cada tipo de usuario es necesario especificar un *esquema externo*, **subesquema** o *vista externa*, que describe un subconjunto de datos de la base de datos. Por lo tanto, existirán tantos subesquemas como tipos de usuarios tenga la base de datos (uno por cada uno de ellos). Habrá usuarios que podrán acceder a más de un esquema externo, y un esquema externo podrá ser compartido por varios usuarios.

Nivel conceptual.

Es una representación de toda la información contenida en la base de datos en una forma más abstracta que la contenida en la estructura interna. Se puede considerar como un modelo teórico de la base de datos sobre el cual estarán asentados los submodelos externos. Se trata de tener una visión global de los datos “como realmente son”, sin estar forzados a definirlos por medio de las restricciones de un lenguaje de programación o un dispositivo hardware concreto.

El administrador es el único que trabaja a nivel conceptual, ya que los usuarios trabajan a nivel externo utilizando subconjuntos de la estructura conceptual.

El objetivo de este nivel es proporcionar una descripción real de la organización independiente de la estructura física de los datos. El nivel conceptual se describe por medio de un **esquema** o *vista conceptual* de la base de datos. El esquema conceptual podría contener:

- Las **entidades** del mundo real (Por ejemplo: empleados, departamentos, componentes, etc.)
- Los **atributos** de las distintas entidades (Por ejemplo para la entidad empleado: ID Empleado, nombre, dirección, RUN, nº hijos, sueldo etc.)
- **Relaciones** entre las distintas entidades (Por ejemplo: Un empleado pertenece a un departamento, un departamento contiene varios empleados, etc.)
- Podrían incluirse verificaciones de seguridad e **integridad** (Por ejemplo: un empleado no podría pertenecer a un departamento que no existe).

Nivel interno

Es el más cercano al almacenamiento físico. Se ocupa de cómo los datos se almacenan en el soporte:

- La estructura de almacenamiento físico de los datos.
- La organización de los ficheros.
- Los modos de acceso.
- Los índices y apuntadores.
- El bloqueo de registros.

El nivel interno se describe por medio de un **esquema o vista internos**. Si el nivel interno y conceptual estuviesen totalmente separados, no se necesitaría de este esquema, ya que podría diseñarlo automáticamente el propio sistema a partir del esquema conceptual. Como no suele ser así, el administrador deberá comunicar al sistema gestor de la base de datos ciertas características respecto a la estructura interna que tiendan a conseguir una mayor eficiencia en el almacenamiento y recuperación de la base de datos.

El administrador es el único que trabaja a nivel interno. Al diseñar el esquema interno se intenta conseguir los siguientes objetivos:

- Disminuir al máximo el tiempo de respuesta del sistema.
- Minimizar el espacio de almacenamiento.
- Evitar en lo posible la redundancia de información.
- Optimización de los recursos de los que se disponen.

SGBD (Sistema Gestor de Bases de Datos)

El SGBD es un conjunto coordinado de programas, procedimientos y lenguajes, que suministra, tanto a los usuarios informáticos, como no informáticos y al administrador, los medios necesarios para describir, recuperar y manipular los datos integrados en la base de datos, asegurando su confidencialidad y seguridad.

Principales funciones del SGBD:

- Facilitar al administrador la descripción de los elementos de datos que integran la base de datos, su estructura y las relaciones que existen entre ellos, así como los controles a efectuar antes de realizar el acceso a la base de datos. Esta función es llevada a cabo por medio de Lenguajes de Definición de Datos (LDD).
- Facilitar la manipulación de los datos, permitiendo a los usuarios de la base de datos añadir, buscar, suprimir o modificar los datos de esta. Se lleva a cabo por medio de Lenguajes de Manipulación de Datos (LMD).
- Garantizar la seguridad y confidencialidad de la base de datos. Por ejemplo: controlando el acceso de los usuarios a la base de datos mediante palabras de paso y limitando el acceso a los datos mediante un sistema de derechos o privilegios.
- Garantizar la fiabilidad e integridad de los datos. Por ejemplo: No permitir la introducción de fechas erróneas, valores no permitidos, valores incoherentes, relaciones inconsistentes, etc.
- Garantizar la compartibilidad de los datos. Permitir que varios usuarios puedan utilizar los mismos datos sin que se produzcan problemas en los accesos simultáneos.
- Facilitar estadísticas sobre la marcha de la explotación de la base de datos, donde se pueda observar su rendimiento.