

TEMA 2

BASES DE DATOS

1. Sistemas de información

En la evolución de los sistemas de información se han distinguido 2 tipos de SI:

SI orientados a procesos

SI orientados a datos

1.1. SI orientados a procesos

En estos SI se crean diversas aplicaciones (software) para gestionar diferentes aspectos del sistema.

Los datos que utilizan las aplicaciones se almacenan en varios ficheros. Cada programa almacena y utiliza sus propios datos.

Así tenemos:



Ejemplo: dpto. Comercial dpto. contabilidad dpto. ventas

Datos

datos

datos

aplicación

aplicación

aplicación

La única ventaja es que los procesos son independientes por lo que la modificación de uno no afectaba al resto.

- Inconvenientes

- Datos redundantes (misma información en diferentes departamentos)
- Datos inconsistentes (En un sitio puede haber una información y en otro sitio otra distinta)

- **Dificultad de acceso a los datos** (De un departamento a otro porque **no son visibles**)
- **Datos y aplicaciones van unido**. (Para saber como se almacenan los datos hay que ver el código de la aplicación)

1.2. SI orientados a datos

En los **años 70** aparecieron los **SI orientados a datos**. En este tipo de sistemas los **datos se centralizan en una base de datos común** a todas las aplicaciones.

En esos sistemas los **datos se almacenan en una única estructura lógica** que es **utilizable por las aplicaciones**. A través de esa estructura **se accede a los datos que son comunes a todas las aplicaciones**.

Cuando una aplicación modifica un dato, la modificación de dicho dato será visible para el resto de aplicaciones.



- **Ventajas**
 - **Independencia de datos y programas**
 - **Menor redundancia**
 - **Más difícil que exista incoherencia de datos**
 - **Mayor seguridad en los datos** ya que podemos establecer niveles de acceso a los datos.
- **Inconvenientes**
 - **Instalación costosa**. **Cambiar equipos antiguos por nuevos que puedan trabajar con BD.**

- *Requiere personal cualificado*
- *Dependencia de los sistemas comerciales del mercado aunque muchos están aceptados como estándar de hecho.*

2. Bases de datos

Una BD es una *colección de datos relacionados lógicamente entre sí, con una definición y descripción comunes y que están estructurados de una determinada manera. Es un conjunto estructurado de datos que representa entidades y sus interrelaciones, almacenados con la mínima redundancia y posibilitando el acceso a ellos eficientemente por parte de varias aplicaciones y usuarios.*

Una base de datos constará de los siguientes elementos:

- ✓ *Entidades: objeto real o abstracto con características diferenciadoras de otros, del que se almacena información en la base de datos. En una base de datos de una clínica veterinaria, posibles entidades podrían ser: ejemplar, doctor, consulta, etc.*
- ✓ *Atributos: son los datos que se almacenan de la entidad. Cualquier propiedad o característica de una entidad puede ser atributo. Continuando con nuestro ejemplo, podrían ser atributos: raza, color, nombre, número de identificación, etc.*
- ✓ *Registros: donde se almacena la información de cada entidad. Es un conjunto de atributos que contienen los datos que pertenecen a una misma repetición de entidad. En nuestro ejemplo, un registro podría ser: 2123056, Sultán, Podenco, Gris, 23/03/2009.*
- ✓ *Campos: donde se almacenan los atributos de cada registro. Teniendo en cuenta el ejemplo anterior, un campo podría ser el valor Podenco.*

2.1. Usos

Ya sabemos lo que es una base de datos y sus características principales, pero es necesario conocer quien las usa y para qué.

¿Quién utiliza las bases de datos?

Existen cuatro tipos de personas que pueden hacer uso de una base de datos: el administrador, los diseñadores de la base de datos, los programadores de aplicaciones y los usuarios finales.

¿Quién utiliza las bases de datos?	
Tipo	Funciones y características
El administrador	Es la persona encargada de la creación o implementación física de la base de datos. Es quien escoge los tipos de ficheros, los índices que hay que crear, la ubicación de éstos, etc. En general, es quien toma las decisiones relacionadas con el funcionamiento físico del almacenamiento de información. Siempre teniendo en cuenta las posibilidades del sistema de información con el que trabaje. Junto a estas tareas, el administrador establecerá la política de seguridad y de acceso para garantizar el menor número de problemas.
Los diseñadores	Son las personas encargadas de diseñar cómo será la base de datos. Llevarán a cabo la identificación de los datos, las relaciones entre ellos, sus restricciones, etc. Para ello han de conocer a fondo los datos y procesos a representar en la base de datos. Si estamos hablando de una empresa, será necesario que conozcan las reglas de negocio en la que esta se mueve. Para obtener un buen resultado, el diseñador de la base de datos debe implicar en el proceso a todos los usuarios de la base de datos, tan pronto como sea posible
Los programadores de aplicaciones	Una vez diseñada y construida la base de datos, los programadores se encargarán de implementar los programas de aplicación que servirán a los usuarios finales. Estos programas de aplicación ofrecerán la posibilidad de realizar consultas de datos, inserción, actualización o eliminación de los mismos. Para desarrollar estos programas se utilizan lenguajes de tercera o cuarta generación.
Los usuarios finales	Son los clientes finales de la base de datos. Al diseñar, implementar y mantener la base de datos se busca cumplir los requisitos establecidos por el cliente para la gestión de su información.

¿Para qué se utilizan las bases de datos?

Enumerar todos y cada uno de los campos donde se utilizan las bases de datos es complejo, aunque seguro que quedarán muchos en el tintero, a continuación se recopilan algunos de los ámbitos donde se aplican.

- ✓ **Banca**: información de clientes, cuentas, transacciones, préstamos, etc.
- ✓ **Líneas aéreas**: información de clientes, horarios, vuelos, destinos, etc.
- ✓ **Universidades**: información de estudiantes, carreras, horarios, materias, etc.
- ✓ **Transacciones de tarjeta de crédito**: para comprar con tarjetas de crédito y la generación de los extractos mensuales.
- ✓ **Telecomunicaciones**: para guardar registros de llamadas realizadas, generar facturas mensuales, mantener el saldo de las tarjetas telefónicas de prepago y almacenar información sobre las redes.
- ✓ **Medicina**: información hospitalaria, biomedicina, genética, etc.
- ✓ **Justicia y Seguridad**: delincuentes, casos, sentencias, investigaciones, etc.
- ✓ **Legislación**: normativa, registros, etc.
- ✓ **Organismos públicos**: datos ciudadanos, certificados, etc.
- ✓ **Sistemas de posicionamiento geográfico**.
- ✓ **Hostelería y turismo**: reservas de hotel, vuelos, excursiones, etc.
- ✓ **Ocio digital**: juegos online, apuestas, etc.
- ✓ **Cultura**: gestión de bibliotecas, museos virtuales, etc.
- ✓ Etc.

3. Modelos de bases de datos.

La **clasificación tradicional** de las bases de datos establece tres modelos de bases de datos: **jerárquico**, **en red** y **relacional**. En la actualidad el modelo de bases de datos **más extendido es el relacional**. Aunque, hay que tener en cuenta que dos de sus variantes (modelo de bases de datos distribuidas y orientadas a objetos) son las que se más se están utilizando en los últimos tiempos.

En los siguientes epígrafes analizaremos cada uno de ellos, así como otros modelos de bases de datos existentes.

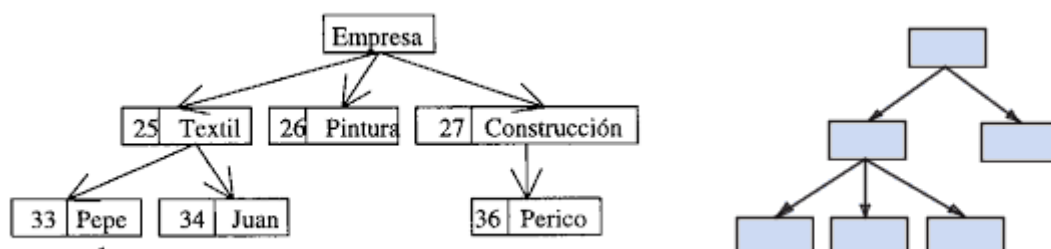
3.1. Modelo jerárquico

Cuando **IBM creó su Sistema Administrador de Información o IMS**, se **establecieron las bases** para que la gran mayoría de **sistemas de gestión de información** de los años setenta **utilizaran el modelo jerárquico**. También **recibe el nombre de modelo en árbol**, ya que utiliza una estructura en árbol invertido para la organización de los datos.

La información se organiza con una **jerarquía en la que la relación entre las entidades de este modelo siempre es del tipo padre/hijo**. De tal manera que existen nodos que contienen atributos o campos y que se relacionarán con sus nodos hijos, pudiendo tener cada nodo más de un hijo, pero un nodo siempre tendrá un sólo padre.

Los **datos de este modelo se almacenan en estructuras lógicas llamadas segmentos**. Los segmentos **se relacionan entre sí utilizando arcos**. La forma visual de este modelo es de árbol invertido, en la parte superior están los padres y en la inferior los hijos.

Hoy en día, debido a sus limitaciones, el modelo jerárquico está en desuso. En el siguiente gráfico puedes observar la estructura de almacenamiento del modelo jerárquico.

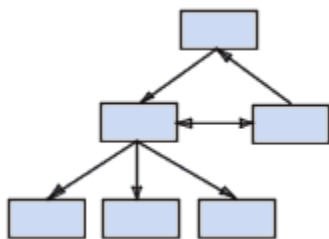


3.2. Modelo en red.

El modelo de datos en red aparece a mediados de los sesenta como respuesta a limitaciones del modelo jerárquico en cuanto a representación de relaciones más complejas. Podemos considerar a IDS (Integrated Data Store) de Bachman como el primer sistema de base de datos en red. Tras él se intentó crear un estándar de modelo de red por parte de CODASYL, siendo un modelo que tuvo gran aceptación a principios de los setenta.

El modelo en red organiza la información en registros (también llamados nodos) y enlaces. En los registros se almacenan los datos, mientras que los enlaces permiten relacionar estos datos. Las bases de datos en red son parecidas a las jerárquicas sólo que en ellas puede haber más de un padre.

En este modelo se pueden representar perfectamente cualquier tipo de relación entre los datos, pero hace muy complicado su manejo. Al no tener que duplicar la información se ahorra espacio de almacenamiento.

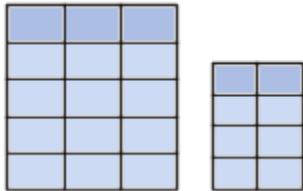
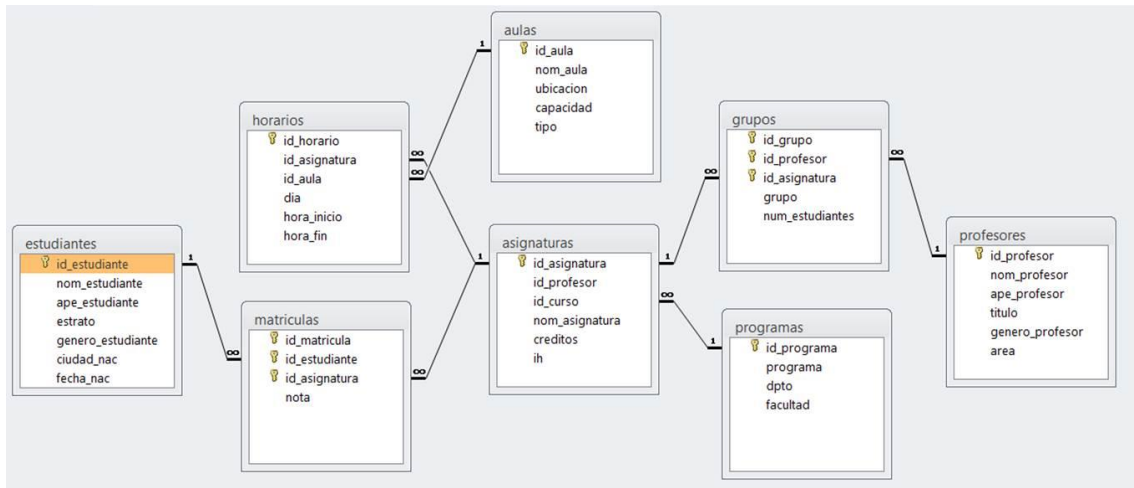


3.3. Modelo relacional

Este modelo es posterior a los dos anteriores y fue desarrollado por Codd en 1970. Hoy en día las bases de datos relacionales son las más utilizadas.

En el modelo relacional la base de datos es percibida por el usuario como un conjunto de tablas. Esta percepción es sólo a nivel lógico, ya que a nivel físico puede estar implementada mediante distintas estructuras de almacenamiento.

El modelo relacional utiliza tablas bidimensionales (relaciones) para la representación lógica de los datos y las relaciones entre ellos. Cada relación (tabla) posee un nombre que es único y contiene un conjunto de columnas.



3.4. Modelo orientado a objetos.

El modelo orientado a objetos define una base de datos en términos de objetos, sus propiedades y sus operaciones. Los objetos con la misma estructura y comportamiento pertenecen a una clase, y las clases se organizan en jerarquías. Las operaciones de cada clase se especifican en términos de procedimientos predefinidos denominados métodos. Algunos sistemas existentes en el mercado, basados en el modelo relacional, han sufrido evoluciones incorporando conceptos orientados a objetos. A estos modelos se les conoce como sistemas objeto-relacionales.

El objetivo del modelo orientado a objetos es cubrir las limitaciones del modelo relacional. Gracias a este modelo se incorporan mejoras como la herencia entre tablas, los tipos definidos por el usuario, disparadores almacenables en la base de datos (triggers), soporte multimedia, etc.

No han llegado a reemplazar a las BD relacionales.

4. Tipos de bases de datos.

Vamos a ver a continuación una detallada descripción de los tipos de bases de datos teniendo en cuenta varios criterios.

Diferentes clasificaciones de las bases de datos

❖ Bases de datos según su contenido

- **Bases de datos con información actual:** Contienen **información muy concreta y actualizada**, normalmente, de tipo numérico: **estadísticas**, **series históricas**, **resultados de encuestas**, convocatorias de becas o subvenciones, convocatorias de eventos, **ofertas de empleo**, ...
- **Directorios:** **recogen datos sobre personas o instituciones especializadas en una actividad o materia concreta**. Hay **directorios de profesionales**, **de investigadores**, de centros de investigación, de bibliotecas, de revistas científicas, **de empresas**, **de editoriales**, ...
- **Bases de datos documentales:** En éste último grupo, **cada registro se corresponde con un documento**, sea éste de cualquier tipo: una **publicación impresa**, un **documento audiovisual**, **gráfico**. Dependiendo de si incluyen o no el contenido completo de los documentos que describen, **podremos tener**:
 - **Bases de datos de texto completo:** **constituidas por los propios documentos en formato electrónico**, con un volcado completo de su texto.
 - **Archivos electrónicos de imágenes:** Constituidos por **referencias que permiten un enlace directo con la imagen del documento original**, sea éste un documento iconográfico (fotografías, imágenes de televisión, ...) o un documento impreso digitalizado en formato de imagen.
 - **Bases de datos referenciales:** Sus registros no contienen el texto original sino tan sólo la **información fundamental para describir y permitir la localización de documentos impresos, sonoros, iconográficos, audiovisuales o electrónicos**. En estos sistemas de información sólo se puede obtener referencias sobre documentos que habrá que localizar posteriormente en otro servicio (archivo, biblioteca, fonoteca, ...) o solicitar a un servicio de suministro de documentos.

❖ Bases de datos según su uso:

- **Base de datos individual:** Es una base de datos **utilizada** básicamente **por una persona**. El sistema **administrador** de la base de datos y los **datos son controlados por el mismo usuario**. Puede estar almacenada en la unidad de disco duro del usuario o en el servidor de archivos de una red de área local. Por ejemplo, un gerente de ventas podría contar con una base de datos para el control de sus vendedores y su desempeño.
- **Base de datos compartida:** Son bases de datos con **múltiples usuarios** y que muy probablemente pertenezcan a la **misma organización**, **como la base de datos de una compañía**. Se encuentra almacenada en una computadora potente y bajo el cuidado de un profesional en el área, el

administrador de la base de datos. Los usuarios tienen acceso a la base de datos mediante una red de área local o una red de área extensa.

- **Bases de datos de acceso público:** Son bases de datos accesibles por cualquier persona. Puede no ser necesario pagar un canon para hacer uso de los datos contenidos en ellas.
- **Bases de datos propietarias o bancos de datos:** Se trata en general de bases de datos de gran tamaño, desarrolladas por una organización y que contienen temas especializados o de carácter particular. El público general puede tener acceso a estas bases a veces de forma gratuita y otras mediante el pago de una cuota. Pueden ofrecer información que va desde negocios, economía, inversión, técnica y científica hasta servicios de entretenimiento. Permiten encontrar en minutos lo que tardaría horas ojeando revistas.
- ❖ **Bases de datos según la variabilidad de la información**
 - **Bases de datos estáticas:** Son bases de datos de sólo lectura. Se utilizan para el almacenamiento de datos históricos que pueden ser analizados y utilizados para el estudio del comportamiento de un conjunto de datos a través del tiempo. Permiten realizar proyecciones y toma de decisiones.
 - **Bases de datos dinámicas:** Son bases de datos donde la información almacenada se modifica con el tiempo, permitiendo operaciones como actualización y adición de datos, además de las operaciones fundamentales de consulta.
- ❖ **Bases de datos según la localización de la información**
 - **Bases de datos centralizadas:** Se trata de bases de datos ubicadas en un único lugar, un único computador. Pueden ser bases de datos monousuario que se ejecutan en ordenadores personales o sistemas de bases de datos de alto rendimiento que se ejecutan en grandes sistemas. Este tipo de organización facilita las labores de mantenimiento, sin embargo, hace que la información contenida en dicha base, sea más vulnerable a posibles fallos y limita su acceso. Este tipo de bases de datos puede ofrecer dentro de la arquitectura Cliente/Servidor dos configuraciones:
 - **Basada en anfitrión:** ocurre cuando la máquina cliente y la máquina servidor son la misma. Los usuarios se conectarán directamente a la máquina donde se encuentra la base de datos.
 - **Basada en Cliente/Servidor:** ocurrirá cuando la base de datos reside en una máquina servidor y los resultados acceden a la base de datos desde su máquina cliente a través de una red
 - **Bases de datos distribuidas:** Según la naturaleza de la organización es probable que los datos no se almacenen en un único punto, sino que se sitúen en un lugar o lugares diferentes a donde se encuentran los usuarios. Una base de datos distribuida es la unión de las bases de datos mediante redes. Los usuarios se vinculan a los servicios de bases de datos distantes mediante una amplia variedad de redes de comunicación. Puede imaginarse una compañía con diferentes oficinas regionales, donde se encuentra distribuida la base de datos. Si embargo, los ejecutivos pueden tener acceso a la información de todas las oficinas regionales.

❖ **Bases de datos según el organismo productor**

- **Bases de datos de organismos públicos y de la administración:** Las bibliotecas y centros de documentación de los ministerios, instituciones públicas, universidades y organismos públicos de investigación elaboran gran cantidad de recursos de información. Estos sistemas pueden ser:
 - Bases de datos de acceso público, sean gratuitas o no.
 - Bases de datos de uso interno, con información de acceso restringido.
- **Bases de datos de instituciones sin ánimo de lucro:** Fundaciones, asociaciones, sindicatos y organizaciones no gubernamentales elaboran frecuentemente sus propios sistemas de información especializados.
- **Bases de datos de entidades privadas o comerciales:** Los centros de documentación, bibliotecas y archivos de las empresas pueden elaborar distintos tipos de sistemas de información:
 - Bases de datos de uso interno para facilitar la circulación de información dentro de la empresa.
 - Bases de datos de uso interno que ocasionalmente ofrecen servicio hacia el exterior (usuarios particulares u otras instituciones)
 - Bases de datos comerciales, diseñadas específicamente para ser utilizadas por usuarios externos.
- **Bases de datos realizadas por cooperación en red:** Se trata de sistemas de información cuya elaboración es compartida por diversas instituciones. Bases de datos internacionales se elaboran a través de este sistema de trabajo, con diversos centros nacionales responsables de la información perteneciente a cada país.

❖ **Bases de datos según el modelo de acceso:**

- **Bases de datos de acceso local:** Para consultarlas es necesario acudir al organismo productor, a su biblioteca o centro de documentación. Pueden ser consultables en monopuesto o en varios puntos de una red local.
- **Bases de datos en CD-ROM:** Pueden adquirirse por compra o suscripción bien directamente por un particular o por una biblioteca o centro de documentación que permita su consulta a sus usuarios. En algunas instituciones se instalan diferentes CD-ROM en una red local para permitir su consulta desde cualquier ordenador conectado a la misma.
- **Bases de datos en línea:** Pueden consultarse desde cualquier ordenador conectado a Internet. La consulta puede ser libre (gratuita) o exigir la solicitud previa de una clave personal entrada (denominada comúnmente con el término inglés password). Para obtener un password puede exigir la firma de un contrato. Hay diferentes tipos de acceso en línea:
 - **Acceso via telnet o mediante línea de Internet:** el usuario realiza una conexión estable al host (gran ordenador) en donde se halla la base de datos, a través de Internet. La interfaz de usuario instalada en dicho ordenador remoto determinará si la interrogación debe realizarse por menús o por comandos o expresiones de un lenguaje determinado. Cuando un usuario entra en una base de datos vía telnet establece una sesión de trabajo interactiva con el programa que gestiona la base de datos, que le permite aplicar todas las

posibilidades de interrogación que tenga el sistema selección, combinación y visualización o impresión de resultados. En cualquier momento podrá visualizar todas las búsquedas realizadas hasta ese instante y establecer combinaciones entre ellas.

- **Acceso vía web:** conexión a través de un formulario existente en una página web de Internet, diseñado para lanzar preguntas a una base de datos.

Una misma base de datos puede tener acceso local y además una edición en CD-ROM y un sistema de acceso en línea. Sin embargo, puede haber diferencias en el contenido presente en cada uno de estos formatos o en el grado de actualización de la información. Por ejemplo, el productor de una base de datos puede ofrecer la conexión en línea a la base de datos completa con actualización diaria y, en cambio, editar un CD-ROM que tan sólo contenga los últimos cinco años de información y se actualice semestralmente.

❖ **Bases de datos según cobertura temática:**

- **Bases de datos científico-tecnológicas:** Contienen información destinada a los investigadores de cualquier ámbito científico o técnico. A su vez, este grupo puede dividirse en:
 - **Bases de datos multidisciplinarias:** abarcan varias disciplinas científicas o técnicas.
 - **Bases de datos especializadas:** recopilan y analizan documentos pertinentes para una disciplina o subdisciplina concreta: investigación biomédica, farmacéutica, química, agroalimentaria, social, humanística, etc
- **Bases de datos económico-empresariales:** Contienen información de interés para empresas, entidades financieras, ...
- **Bases de datos de medios de comunicación:** contienen información de interés para los profesionales de medios de comunicación de masas: prensa, radio, televisión, ...
- **Bases de datos de ámbito político-administrativo y jurídico:** Contienen información de interés para los organismos de la administración y los profesionales del Derecho: legislación, jurisprudencia,
- **Bases de datos de ámbito sanitario:** Además de las propias del primer grupo especializadas en ciencias de la salud, existen otros sistemas con información de interés sanitario: historiales médicos, archivos hospitalarios, ...
- **Bases de datos para el gran público:** Contiene información destinada a cubrir necesidades de información general, de interés para un gran número de usuarios.

5. Sistemas gestores de BD

Un SGBD es el software que permite a los usuarios procesar, describir, administrar y recuperar datos almacenados en una BD. En estos sistemas existen un conjunto de programas, procedimientos y lenguajes que permiten realizar las diferentes tareas sobre los datos.

Las ventajas del uso de SGBD son:

- ✓ Proporcionan al usuario una visión abstracta de los datos, ocultando parte de la complejidad relacionada con cómo se almacenan y mantienen los datos.
- ✓ Ofrecen Independencia física, es decir, la visión que tiene de la información el usuario, y la manipulación de los datos almacenados en la Base de Datos, es independiente de cómo estén almacenados físicamente.
- ✓ Disminuyen la redundancia y la inconsistencia de datos.
- ✓ Aseguran la integridad de los datos.
- ✓ Facilitan el acceso a los datos, aportando rapidez y evitando la pérdida de datos.
- ✓ Aumentan la seguridad y privacidad de los datos.
- ✓ Mejoran la eficiencia.
- ✓ Permiten compartir datos y accesos concurrentes.
- ✓ Facilitan el intercambio de datos entre distintos sistemas.
- ✓ Incorporan mecanismos de copias de seguridad y recuperación para restablecer la información en caso de fallos en el sistema.

El SGBD interacciona con otros elementos software existentes en el sistema, concretamente con el sistema operativo (SO). Los datos almacenados de forma estructurada en la base de datos son utilizados indistintamente por otras aplicaciones, será el SGBD quien ofrecerá una serie de facilidades a éstas para el acceso y manipulación de la información, basándose en las funciones y métodos propios del sistema operativo.

5.1. Funciones de un SGBD

Un SGBD desarrolla tres funciones fundamentales como son las de descripción, manipulación y utilización de los datos. A continuación se detallan cada una de ellas.

1. Función de descripción o definición: Permite al diseñador de la base de datos crear las estructuras apropiadas para integrar adecuadamente los datos.

Esta función se realiza mediante el lenguaje de descripción de datos o DDL. Mediante ese lenguaje: se definen las estructuras de datos, se

definen las relaciones entre los datos y se definen las reglas (restricciones) que han de cumplir los datos.

2. Función de manipulación: permite a los usuarios de la base buscar, añadir, suprimir o modificar los datos de la misma, siempre de acuerdo con las especificaciones y las normas de seguridad dictadas por el administrador. Se llevará a cabo por medio de un **lenguaje de manipulación de datos (DML)** que facilita los instrumentos necesarios para la realización de estas tareas.

También se encarga de definir la **vista externa** de todos los usuarios de la base de datos o vistas parciales que cada usuario tiene de los datos definidos con el DDL.

3. Función de control: permite al administrador de la base de datos establecer mecanismos de protección de las diferentes visiones de los datos asociadas a cada usuario, proporcionando elementos de creación y modificación de dichos usuarios. Adicionalmente, incorpora sistemas para la creación de copias de seguridad, carga de ficheros, auditoría, protección de ataques, configuración del sistema, etc. El lenguaje que implementa esta función es el **lenguaje de control de datos o DCL**.

¿Y a través de qué lenguaje podremos desarrollar estas funciones sobre la base de datos? Lo haremos utilizando el **Lenguaje Estructurado de Consultas (SQL: Structured Query Language)**. Este lenguaje proporciona sentencias para realizar operaciones de DDL, DML y DCL. SQL fue publicado por el ANSI en 1986 (American National Standard Institute) y ha ido evolucionando a lo largo del tiempo. Además, los SGBD suelen proporcionar otras herramientas que complementan a estos lenguajes como generadores de formularios, informes, interfaces gráficas, generadores de aplicaciones, etc.

5.2. Componentes

Una vez descritas las funciones que un SGBD debe llevar a cabo, imaginarás que un **SGBD es un paquete de software complejo que ha de proporcionar servicios relacionados con el almacenamiento y la explotación de los datos de forma eficiente**. Para ello, cuenta con una serie de componentes que se detallan a continuación:

1. Lenguajes de la base de datos. Cualquier sistema gestor de base de datos ofrece la posibilidad de utilizar lenguajes e interfaces adecuadas para sus diferentes tipos de usuarios. A través de los lenguajes se pueden especificar los datos que componen la BD, su estructura, relaciones, reglas de integridad, control de acceso, características físicas y vistas externas de los usuarios. Los lenguajes del SGBD son:

Lenguaje de Definición de los Datos (DDL), Lenguaje de Manejo de Datos (DML) y Lenguaje de Control de Datos (DCL).

2. El diccionario de datos. Descripción de los datos almacenados. Se trata de *información útil para los programadores de aplicaciones*. Es el lugar donde *se deposita la información sobre la totalidad de los datos que forman la base de datos*. Contiene las *características lógicas de las estructuras que almacenan los datos, su nombre, descripción, contenido y organización*. En una base de datos relacional, el diccionario de datos aportará información sobre:

- ✓ Estructura lógica y física de la BD.
- ✓ Definición de tablas, vistas, índices, disparadores, procedimientos, funciones, etc.
- ✓ Cantidad de espacio asignado y utilizado por los elementos de la BD.
- ✓ Descripción de las restricciones de integridad.
- ✓ Información sobre los permisos asociados a cada perfil de usuario.
- ✓ Auditoría de acceso a los datos, utilización, etc.

3. El gestor de la base de datos. Es la *parte de software encargada de garantizar el correcto, seguro, íntegro y eficiente acceso y almacenamiento de los datos*. Este componente es el encargado de proporcionar una interfaz entre los datos almacenados y los programas de aplicación que los manejan. Es un *intermediario entre el usuario y los datos*. Es el encargado de *garantizar la privacidad, seguridad e integridad de los datos, controlando los accesos concurrentes e interactuando con el sistema operativo*.

4. Usuarios de la base de datos. En los SGBD existen diferentes perfiles de usuario, cada uno de ellos con una serie de permisos sobre los objetos de la BD.

5. Herramientas de la base de datos. Son un *conjunto de aplicaciones que permiten a los administradores la gestión de la base de datos, de los usuarios y permisos, generadores de formularios, informes, interfaces gráficas, generadores de aplicaciones, etc.*

5.3. Arquitectura de las BD.

Todas las personas que utilizan la BD no tienen la misma visión de los datos. Por ejemplo un experto informático desea conocer la ubicación de los archivos físicos y un usuario realizara consultas sobre las nóminas de los empleados.

Se distinguen 3 niveles de abstracción

- *Nivel interno o físico*

En este nivel se describe la estructura física de la base de datos a través de un esquema interno encargado de detallar el sistema de almacenamiento de la base de datos y sus métodos de acceso. Es el nivel más cercano al almacenamiento físico. A través del esquema físico se indican, entre otros, los archivos que contienen la información, su organización, los métodos de acceso a los registros, los tipos de registros, la longitud, los campos que los componen, las unidades de almacenamiento, etc.

- *Nivel lógico o conceptual*

En este nivel se describe la estructura completa de la base de datos a través de un esquema que detalla las entidades, atributos, relaciones, operaciones de los usuarios y restricciones. Los detalles relacionados con las estructuras de almacenamiento se ocultan, permitiendo realizar una abstracción a más alto nivel.

- *Nivel externo o de visión del usuario*

En este nivel se describen las diferentes vistas que los usuarios percibirán de la base de datos. Cada tipo de usuario o grupo de ellos verá sólo la parte de la base de datos que le interesa, ocultando el resto.

Para una base de datos, sólo existirá un único esquema interno, un único esquema conceptual y podrían existir varios esquemas externos definidos para uno o varios usuarios.

*Gracias a esta arquitectura se consigue la **independencia de datos** a dos niveles:*

- **Independencia lógica:** *Podemos modificar el esquema conceptual sin alterar los esquemas externos ni los programas de aplicación.*
- **Independencia física:** *Podemos modificar el esquema interno sin necesidad de modificar el conceptual o el externo. Es decir, se puede cambiar el sistema de almacenamiento, reorganizar los ficheros, añadir nuevos, etc., sin que esto afecte al resto de esquemas.*

5.4. Tipos de SGBD

¿Qué tipos de SGBD existen? Para responder a esta pregunta podemos realizar la siguiente clasificación, atendiendo a diferentes criterios:

*a. El primer criterio que se suele utilizar es **por el modelo lógico en que se basan**. Actualmente, el modelo lógico que más se utiliza es el **relacional**. Los modelos en red y jerárquico han quedado obsoletos. Otro de los modelos que más extensión está teniendo es el **modelo orientado a objetos**. Por tanto, en esta primera clasificación tendremos:*

- *Modelo Jerárquico.*
- *Modelo de Red.*
- *Modelo Relacional.*
- *Modelo Orientado a Objetos.*

*b. El segundo criterio de clasificación se centra en el **número de usuarios** a los que da servicio el sistema:*

- **Monousuario:** *sólo atienden a un usuario a la vez, y su principal uso se da en los ordenadores personales.*
- **Multiusuario:** *entre los que se encuentran la mayor parte de los SGBD, atienden a varios usuarios al mismo tiempo.*

*c. El tercer criterio se basa en el **número de sitios en los que está distribuida la base de datos**:*

- **Centralizados:** *sus datos se almacenan en un solo computador. Los SGBD centralizados pueden atender a varios usuarios, pero el SGBD y la base de datos en sí residen por completo en una sola máquina.*
- **Distribuidos (Homogéneos, Heterogéneos):** *la base de datos real y el propio software del SGBD pueden estar distribuidos en varios sitios conectados por una red. Los sistemas homogéneos utilizan el mismo SGBD en múltiples sitios. Una tendencia reciente consiste en crear software para tener acceso a varias bases de datos autónomas preexistentes almacenadas en sistemas distribuidos heterogéneos. Esto da lugar a los SGBD federados o sistemas multibase de datos en los que los SGBD participantes tienen cierto grado de autonomía local.*

d. El cuarto criterio toma como referencia el coste. La mayor parte de los paquetes cuestan entre 10.000 y 100.000 euros. Los sistemas monousuario más económicos para microcomputadores cuestan entre 0 y 3.000 euros. En el otro extremo, los paquetes más completos cuestan más de 100.000 euros.

e. El quinto, y último, criterio establece su clasificación **según el propósito**:

- **Propósito General:** pueden ser utilizados para el tratamiento de cualquier tipo de base de datos y aplicación.
- **Propósito Específico:** Cuando el rendimiento es fundamental, se puede diseñar y construir un software de propósito especial para una aplicación específica, y este sistema no sirve para otras aplicaciones. Muchos sistemas de reservas de líneas aéreas son de propósito especial y pertenecen a la categoría de **sistemas de procesamiento de transacciones en línea**, que deben atender un gran número de transacciones concurrentes sin imponer excesivos retrasos.

6. SGBD comerciales

Actualmente, en el mercado de software existen multitud de sistemas gestores de bases de datos comerciales. En este epígrafe se desglosan las características fundamentales de los más importantes y extendidos hasta la fecha. Pero, como podrás observar, la elección de un SGBD es una decisión muy importante a la hora de desarrollar proyectos. A veces, el sistema más avanzado, "el mejor" según los entendidos, puede no serlo para el tipo de proyecto que estemos desarrollando. Hemos de tener en cuenta qué volumen de carga debe soportar la base de datos, qué sistema operativo utilizaremos como soporte, cuál es nuestro presupuesto, plazos de entrega, etc.

A través de la siguiente tabla se exponen los SGBD comerciales más utilizados y sus características más relevantes:

SGBD	Descripción
ORACLE	Reconocido como uno de los mejores a nivel mundial. Es multiplataforma, confiable y seguro. Es Cliente/Servidor. Basado en el modelo de datos Relacional. De gran potencia, aunque con un precio elevado hace que sólo se vea en empresas muy grandes y multinacionales. Ofrece una versión gratuita Oracle Database 10g Express Edition.
MYSQL	Sistema muy extendido que se ofrece bajo dos tipos de licencia, comercial o libre. Para aquellas empresas que deseen incorporarlo en productos privativos, deben comprar una licencia específica. Es Relacional, Multihilo, Multiusuario y Multiplataforma. Su gran velocidad lo hace

	<i>ideal para consulta de bases de datos y plataformas web.</i>
DB2	<i>Multiplataforma, el motor de base de datos relacional integra XML de manera nativa, lo que IBM ha llamado pureXML, que permite almacenar documentos completos para realizar operaciones y búsquedas de manera jerárquica dentro de éste, e integrarlo con búsquedas relacionales.</i>
INFORMIX	<i>Otra opción de IBM para el mundo empresarial que necesita un DBMS sencillo y confiable. Es un gestor de base de datos relacional basado en SQL. Multiplataforma. Consume menos recursos que Oracle, con utilidades muy avanzadas respecto a conectividad y funciones relacionadas con tecnologías de Internet/Intranet, XML, etc.</i>
Microsoft SQL SERVER	<i>Sistema Gestor de Base de Datos producido por Microsoft. Es relacional, sólo funciona bajo Microsoft Windows, utiliza arquitectura Cliente/Servidor. Constituye la alternativa a otros potentes SGBD como son Oracle, PostgreSQL o MySQL</i>

Otros SGBD comerciales importantes son: DBASE, ACCESS, INTERBASE y FOXPRO.

7. SGBD libres.

SGBD	Descripción
MySQL	<i>Es un sistema de gestión de base de datos relacional, multihilo y multiusuario con más de seis millones de instalaciones. Distribuido bajo dos tipos de licencias, comercial y libre. Multiplataforma, posee varios motores de almacenamiento, accesible a través de múltiples lenguajes de programación y muy ligado a aplicaciones web.</i>
PostgreSQL	<i>Sistema Relacional Orientado a Objetos. Considerado como la base de datos de código abierto más avanzada del mundo. Desarrollado por una comunidad de desarrolladores que trabajan de forma desinteresada, altruista, libre y/o apoyados</i>

	<i>por organizaciones comerciales. Es multiplataforma y accesible desde múltiples lenguajes de programación.</i>
Firebird	<i>Sistema Gestor de Base de Datos relacional, multiplataforma, con bajo consumo de recursos, excelente gestión de la concurrencia, alto rendimiento y potente soporte para diferentes lenguajes.</i>
Apache Derby	<i>Sistema Gestor escrito en Java, de reducido tamaño, con soporte multilenguaje, multiplataforma, altamente portable, puede funcionar embebido o en modo cliente/servidor.</i>
SQLite	<i>Sistema relacional, basado en una biblioteca escrita en C que interactúa directamente con los programas, reduce los tiempos de acceso siendo más rápido que MySQL o PostGreSQL, es multiplataforma y con soporte para varios lenguajes de programación.</i>