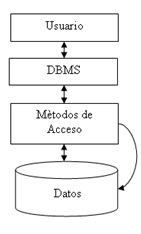
**Bases de datos**

Una base de datos es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso. En este sentido; una biblioteca puede considerarse una base de datos compuesta en su mayoría por documentos y textos impresos en papel e indexados para su consulta. Actualmente, y debido al desarrollo tecnológico de campos como la informática y la electrónica, la mayoría de las bases de datos están en formato digital, siendo este un componente electrónico, por tanto, se ha desarrollado y se ofrece un amplio rango de soluciones al problema del almacenamiento de datos.

Hay programas denominados sistemas gestores de bases de datos, abreviado SGBD (del inglés Database Management System o DBMS), que permiten almacenar y posteriormente acceder a los datos de forma rápida y estructurada. Las propiedades de estos DBMS, así como su utilización y administración, se estudian dentro del ámbito de la informática.

Las aplicaciones más usuales son para la gestión de empresas e instituciones públicas; También son ampliamente utilizadas en entornos científicos con el objeto de almacenar la información experimental.



Aunque las bases de datos pueden contener muchos tipos de datos, algunos de ellos se encuentran mutuamente protegidos por las leyes de varios países. Por ejemplo, en España, los datos personales se encuentran protegidos por la Ley Orgánica de Protección de Datos de Carácter Personal (LOPD), en México por la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental y en Argentina la Ley de Protección de Datos Personales.1​

En Argentina el Código Penal sanciona ciertas conductas relacionadas con una base de datos: acceder ilegítimamente a un banco de datos personales, proporcionar o revelar información registrada en un archivo o en un banco de datos personales cuyo secreto estuviere obligado a guardar por ley o insertar o hacer insertar datos en un archivo de datos personales. Si el autor es funcionario público, sufre además pena de inhabilitación especial.

**Clasificación de bases de datos**

**Base de datos de particulares**

Las bases de datos pueden clasificarse de varias maneras, de acuerdo con el contexto que se esté manejando, la utilidad de estas o las necesidades que satisfagan.

**Según la variabilidad de la base de datos**

**Bases de datos estáticas**

Son bases de datos únicamente de lectura, utilizadas principalmente para almacenar datos históricos que posteriormente se pueden utilizar para estudiar el comportamiento de un conjunto de datos a través del tiempo, realizar proyecciones, tomar decisiones y realizar análisis de datos para inteligencia empresarial.

**Bases de datos dinámicas**

Son bases de datos donde la información almacenada se modifica con el tiempo, permitiendo operaciones como actualización, borrado y edición de datos, además de las operaciones fundamentales de consulta. Un ejemplo, puede ser la base de datos utilizada en un sistema de información de un supermercado. 2​

**Según el contenido**

**Bases de datos bibliográficas**

Solo contienen una subrogante (representante) de la fuente primaria, que permite localizarla. Un registro típico de una base de datos bibliográfica contiene información sobre el autor, fecha de publicación, editorial, título, edición, de una determinada publicación, etc. Puede contener un resumen o extracto de la publicación original, pero nunca el texto completo, porque si no, estaríamos en presencia de una base de datos a texto completo (o de fuentes primarias —ver más abajo). Como su nombre lo indica, el contenido son cifras o números. Por ejemplo, una colección de resultados de análisis de laboratorio ayuda mucho a la redundancia de datos.

**Bases de datos de texto completo**

Almacenan las fuentes primarias, como, por ejemplo, todo el contenido de todas las ediciones de una colección de revistas científicas.

**Directorios**

Un ejemplo son las guías telefónicas en formato electrónico.

Estos directorios se pueden clasificar en dos grandes tipos dependiendo de si son personales o empresariales (llamadas páginas blancas o amarillas respectivamente).

Los directorios empresariales hay de tres tipos:

* Tienen nombre de la empresa y dirección.
* Contienen teléfono y los más avanzados contienen correo electrónico.
* Contienen datos como facturación o número de empleados además de códigos nacionales que ayudan a su distinción.
* Los directorios personales solo hay de un tipo, ya que leyes como la LOPD en España protege la privacidad de los usuarios pertenecientes al directorio.

La búsqueda inversa está prohibida en los directorios personales (a partir de un número de teléfono saber el titular de la línea).

**Bases de datos o “bibliotecas” de información química o biológica**

Son bases de datos que almacenan diferentes tipos de información proveniente de la química, las ciencias de la vida o médicas. Se pueden considerar en varios subtipos:

* Las que almacenan secuencias de nucleótidos o proteínas.
* Las bases de datos de rutas metabólicas.
* Bases de datos de estructura, comprende los registros de datos experimentales sobre estructuras 3D de biomoléculas.
* Bases de datos clínicas.
* Bases de datos bibliográficas (biológicas, químicas, médicas y de otros campos): PubChem, Medline, EBSCOhost.

**Sistema de Administración de Bases de Datos (DBMS)**

**Bases de datos**

Las bases de datos requieren de un software que permita la administración de dicha base de datos. Estos programas especializados sirven como interfaz para que los usuarios puedan, administrar como se estructura y optimiza toda la información recopilada. Un sistema de administración de bases de datos también permite un gran número de operaciones relacionadas con la administración, tal como, supervisar la productividad, ajustes, backups y restauración de los datos.

Entre los gestores de bases de datos o DBMS más conocidos se encuentran: Microsoft SQL Server, MySQL,Oracle Database, Microsoft Access, FileMaker Pro, y dBASE.3​

**Diferencias entre bases de datos y hojas de cálculo**

Las bases de datos y las hojas de cálculo (por ejemplo, las hojas de cálculo de los paquetes de ofimática) son formas convenientes de almacenar información. Las principales diferencias entre las ambas son:

* La forma de manipular y guardar la información.
* La cantidad de datos que se pueden almacenar.
* La accesibilidad a esos datos almacenados.

Las hojas de cálculo desde sus comienzos fueron diseñadas para un usuario, y puede ser observado en sus características. Son excelentes para uno o un pequeño número de usuarios que no necesitan utilizar un gran volumen de datos complejos. Las bases de datos, por otro lado, fueron creadas para almacenar gran cantidad de información organizada, enormes cantidades en ocasiones. Las bases de datos permiten consultas multiusuario, que permite acceder y consultar los datos de forma rápida y segura a muchos usuarios al mismo tiempo, utilizando una lógica y un lenguaje altamente complejos.

**Modelos de bases de datos**

Además de la clasificación por la función de las bases de datos, estas también se pueden clasificar de acuerdo con su modelo de administración de datos.

Un modelo de datos es básicamente una "descripción" de algo conocido como contenedor de datos (algo en donde se guardan los datos), así como de los métodos para almacenar y recuperar datos de esos contenedores. Los modelos de datos no son cosas físicas: son abstracciones que permiten la implementación de un sistema eficiente de base de datos; por lo general se refieren a algoritmos, y conceptos matemáticos.

Algunos modelos con frecuencia utilizados en las bases de datos:

**Bases de datos jerárquicas**

En este modelo los datos se organizan en forma de árbol invertido (algunos dicen raíz), en donde un nodo padre de información puede tener varios hijos. El nodo que no tiene padres es llamado raíz, y a los nodos que no tienen hijos se los conoce como hojas.

Las bases de datos jerárquicas son especialmente útiles en el caso de aplicaciones que manejan un gran volumen de información y datos muy compartidos permitiendo crear estructuras estables y de gran rendimiento.

Una de las principales limitaciones de este modelo es su incapacidad de representar eficientemente la redundancia de datos.

**Base de datos de red**

Este es un modelo ligeramente distinto del jerárquico; su diferencia fundamental es la modificación del concepto de nodo: se permite que un mismo nodo tenga varios padres (posibilidad no permitida en el modelo jerárquico).

Fue una gran mejora con respecto al modelo jerárquico, ya que ofrecía una solución eficiente al problema de redundancia de datos; pero, aun así, la dificultad que significa administrar la información en una base de datos de red ha significado que sea un modelo utilizado en su mayoría por programadores más que por usuarios finales.

**Bases de datos transaccionales**

Son bases de datos cuyo único fin es el envío y recepción de datos a grandes velocidades, estas bases son muy poco comunes y están dirigidas por lo general al entorno de análisis de calidad, datos de producción e industrial, es importante entender que su fin único es recolectar y recuperar los datos a la mayor velocidad posible, por lo tanto la redundancia y duplicación de información no es un problema como con las demás bases de datos, por lo general para poderlas aprovechar al máximo permiten algún tipo de conectividad a bases de datos relacionales.

Un ejemplo habitual de transacción es el traspaso de una cantidad de dinero entre cuentas bancarias. Normalmente se realiza mediante dos operaciones distintas, una en la que se debita el saldo de la cuenta origen y otra en la que acreditamos el saldo de la cuenta destino. Para garantizar la atomicidad del sistema (es decir, para que no aparezca o desaparezca dinero), las dos operaciones deben ser atómicas, es decir, el sistema debe garantizar que, bajo cualquier circunstancia (incluso una caída del sistema), el resultado final es que, o bien se han realizado las dos operaciones, o bien no se ha realizado ninguna.

**Bases de datos relacionales**

Este es el modelo utilizado en la actualidad para representar problemas reales y administrar datos dinámicamente. Tras ser postulados sus fundamentos en 1970 por Edgar Frank Codd,5​ de los laboratorios IBM en San José (California), no tardó en consolidarse como un nuevo paradigma en los modelos de base de datos. Su idea fundamental es el uso de "relaciones". Estas relaciones podrían considerarse en forma lógica como conjuntos de datos llamados "tuplas". Pese a que esta es la teoría de las bases de datos relacionales creadas por Codd, la mayoría de las veces se conceptualiza de una manera más fácil de imaginar. Esto es pensando en cada relación como si fuese una tabla que está compuesta por registros (las filas de una tabla), que representarían las tuplas, y campos (las columnas de una tabla).

En este modelo, el lugar y la forma en que se almacenen los datos no tienen relevancia (a diferencia de otros modelos como el jerárquico y el de red). Esto tiene la considerable ventaja de que es más fácil de entender y de utilizar para un usuario esporádico de la base de datos. La información puede ser recuperada o almacenada mediante "consultas" que ofrecen una amplia flexibilidad y poder para administrar la información.

El lenguaje más habitual para construir las consultas a bases de datos relacionales es SQL, Structured Query Language o Lenguaje Estructurado de Consultas, un estándar implementado por los principales motores o sistemas de gestión de bases de datos relacionales.

Durante su diseño, una base de datos relacional pasa por un proceso al que se le conoce como normalización de una base de datos.

**Bases de datos multidimensionales**

Son bases de datos ideadas para desarrollar aplicaciones muy concretas, como creación de Cubos OLAP. Básicamente no se diferencian demasiado de las bases de datos relacionales (una tabla en una base de datos relacional podría serlo también en una base de datos multidimensional), la diferencia está más bien a nivel conceptual; en las bases de datos multidimensionales los campos o atributos de una tabla pueden ser de dos tipos, o bien representan dimensiones de la tabla, o bien representan métricas que se desean aprender.

**Bases de datos orientadas a objeto**

Este modelo, bastante reciente, y propio de los modelos informáticos orientados a objetos, trata de almacenar en la base de datos los objetos completos (estado y comportamiento).

Una base de datos orientada a objetos es una base de datos que incorpora todos los conceptos importantes del paradigma de objetos:

**Encapsulación** - Propiedad que permite ocultar la información al resto de los objetos, impidiendo así accesos incorrectos o conflictos.

**Herencia** - Propiedad a través de la cual los objetos heredan comportamiento dentro de una jerarquía de clases.

**Polimorfismo** - Propiedad de una operación mediante la cual puede ser aplicada a distintos tipos de objetos.

En bases de datos orientadas a objetos, los usuarios pueden definir operaciones sobre los datos como parte de la definición de la base de datos. Una operación (llamada función) se especifica en dos partes. La interfaz (o signatura) de una operación incluye el nombre de la operación y los tipos de datos de sus argumentos (o parámetros). La implementación (o método) de la operación se especifica separadamente y puede modificarse sin afectar la interfaz. Los programas de aplicación de los usuarios pueden operar sobre los datos invocando a dichas operaciones a través de sus nombres y argumentos, sea cual sea la forma en la que se han implementado. Esto podría denominarse independencia entre programas y operaciones.

**Bases de datos documentales**

Permiten la indexación a texto completo, y en líneas generales realizar búsquedas más potentes, sirven para almacenar grandes volúmenes de información de antecedentes históricos. Tesaurus es un sistema de índices optimizado para este tipo de bases de datos.

**Bases de datos deductivas**

Un sistema de base de datos deductiva es un sistema de base de datos, pero con la diferencia de que permite hacer deducciones a través de inferencias. Se basa principalmente en reglas y hechos que son almacenados en la base de datos. Las bases de datos deductivas son también llamadas bases de datos lógicas, a raíz de que se basa en lógica matemática. Este tipo de base de datos surge debido a las limitaciones de la base de datos relacional de responder a consultas recursivas y de deducir relaciones indirectas de los datos almacenados en la base de datos.

**Lenguaje**

Utiliza un subconjunto del lenguaje Prolog llamado Datalog el cual es declarativo y permite al ordenador hacer deducciones para contestar a consultas basándose en los hechos y reglas almacenados.

**Ventajas**

* Uso de reglas lógicas para expresar las consultas.
* Permite responder consultas recursivas.
* Cuenta con negaciones estratificadas
* Capacidad de obtener nueva información a través de la ya almacenada en la base de datos mediante inferencia.
* Uso de algoritmos que optimizan las consultas.
* Soporta objetos y conjuntos complejos.

**Fases**

**Fase de Interrogación**: se encarga de buscar en la base de datos informaciones deducibles implícitas. Las reglas de esta fase se denominan reglas de derivación.

**Fase de Modificación**: se encarga de añadir a la base de datos nuevas informaciones deducibles. Las reglas de esta fase se denominan reglas de generación.

**Interpretación**

Encontramos dos teorías de interpretación de las bases de datos deductiva por lo cual consideramos las reglas y los hechos como axiomas. Los hechos son axiomas base que se consideran como verdaderos y no contienen variables. Las reglas son axiomas deductivos ya que se utilizan para deducir nuevos hechos.

**Teoría de Modelos**: una interpretación es llamada modelo cuando para un conjunto específico de reglas, estas se cumplen siempre para esa interpretación. Consiste en asignar a un predicado todas las combinaciones de valores y argumentos de un dominio de valores constantes dado. A continuación, se debe verificar si ese predicado es verdadero o falso.

**Mecanismos**

Existen dos mecanismos de inferencia:

* **Ascendente**: donde se parte de los hechos y se obtiene nuevos aplicando reglas de inferencia.
* **Descendente**: donde se parte del predicado (objetivo de la consulta realizada) e intenta encontrar similitudes entre las variables que nos lleven a hechos correctos almacenados en la base de datos.

**Sistema de Gestión de bases de datos distribuida (SGBD)**

La base de datos y el software SGBD pueden estar distribuidos en múltiples sitios conectados por una red. Hay de dos tipos:

1. Distribuidos homogéneos: utilizan el mismo SGBD en múltiples sitios.

2. Distribuidos heterogéneos: Da lugar a los SGBD federados o sistemas multibase de datos en los que los SGBD participantes tienen cierto grado de autonomía local y tienen acceso a varias bases de datos autónomas preexistentes almacenados en los SGBD, muchos de estos emplean una arquitectura cliente-servidor.

Estas surgen debido a la existencia física de organismos descentralizados. Esto les da la capacidad de unir las bases de datos de cada localidad y acceder así a distintas universidades, sucursales de tiendas, etc.

**Base de datos orientada a grafos**

Una base de datos orientada a grafos (BDOG) representa la información como nodos de un grafo y sus relaciones con las aristas de este, de manera que se pueda usar teoría de grafos para recorrer la base de datos ya que esta puede describir atributos de los nodos (entidades) y las aristas (relaciones).

Una BDOG debe estar absolutamente normalizada, esto quiere decir que cada tabla tendría una sola columna y cada relación tan solo dos, con esto se consigue que cualquier cambio en la estructura de la información tenga un efecto solamente local.

**Otros tipos de base de datos**

En la actualidad, existen muchos tipos de bases de datos, algunas menos comunes adaptadas a funciones financieras, funciones científicas y otras funciones altamente específicas, todo dependiendo de cómo va avanzando la tecnología. Entre algunas de ellas se pueden citar:

* Bases de datos en la nube. Una base de datos en la nube es una colección de datos, pueden ser estructurados o no estructurados, que se localiza en una plataforma de computación en la nube privada, pública o una combinación de as dos anteriores (híbrida). Existen dos modelos de base de datos en la nube: tradicional y database as a service (DBaaS). Con DBaaS, donde las gestiones administrativas y el mantenimiento son realizados por un proveedor de servicios.
* De código abierto. Un sistema de base de datos de código abierto es aquel cuyo código fuente es de código abierto; podrían ser bases de datos SQL o NoSQL.
* Base de datos documental/JSON. para gestionar información basada en documentos, las bases de datos documentales son una forma moderna de almacenar datos en formato JSON en lugar de filas y columnas.
* Base de datos multimodelo. combinan diferentes tipos de modelos de base de datos en un único back-end integrado. de este modo varios tipos de datos pueden convivir en una misma base de datos.
* Bases de datos independientes. Las bases de datos independientes, es lo nuevo en base de datos, (conocidas como bases de datos autónomas), se basan en la nube y utilizan el aprendizaje automático para automatizar el ajuste, la seguridad, las copias de seguridad, las actualizaciones y otras tareas de administración de rutina de las bases de datos que tradicionalmente realizan los administradores de bases de datos.

**Consulta a base de datos**

Una consulta es el método para acceder a la información en las bases de datos. Con las consultas se puede modificar, borrar, mostrar y agregar datos en una base de datos, también pueden utilizarse como origen de registro para formularios. Para esto se utiliza un Lenguaje de consulta.

Las consultas a la base de datos se realizan a través de un Lenguaje de manipulación de datos, el lenguaje de consultas a base de datos más utilizado es SQL.