

Introducción a las Bases de Datos

Este capítulo está dedicado a introducir el concepto de base de datos. Sus orígenes desde los primitivos sistemas de almacenamiento basados en ficheros hasta las actuales bases de datos. Se especificarán las características que debe tener una base de datos, junto con los niveles de descripción que hay. Posteriormente se introducirá el concepto de sistema gestor de base de datos como el software que se encarga de gestionar una base de datos

Introducción a las bases de datos

Copyright © 2024 by Rafael Lozano Luján.

Este documento está sujeto a derechos de autor. Todos los derechos están reservados por el Autor de la obra, ya sea en su totalidad o en parte de ella, específicamente los derechos de:

- La reproducción total o parcial mediante fotocopia, escaneo o descarga.
- La distribución o publicación de copias a terceros.
- La transformación mediante la modificación, traducción o adaptación que altere la forma de la obra hasta obtener una diferente a la original.

Tabla de contenido

1	Introducción.....	1
1.1	Sistema de Información.....	1
1.2	Sistemas de Información Automatizado.....	2
1.2.1	SIA basado en ficheros.....	3
1.3	Tipos de archivos.....	3
1.3.1	Archivo secuencial.....	4
1.3.2	Archivo relativo.....	5
1.3.3	Archivo secuencial indexado.....	6
1.4	Inconvenientes de las BD basadas en ficheros.....	8
2	Las bases de datos.....	9
2.1	Ventajas de las BD.....	9
2.2	Características de las BD.....	10
2.3	Tipos de bases de datos.....	11
2.3.1	Relacional.....	11
2.3.2	Distribuida.....	12
2.3.3	Orientada a objeto.....	12
2.3.4	Jerárquicas.....	12
2.4	Niveles de la base de datos.....	13
3	Sistemas gestores de bases de datos.....	15
3.1	Funciones de un SGBD.....	15
3.1.1	Definición o descripción de los datos.....	16
3.1.2	Manipulación de los datos.....	16
3.1.3	Control de la BD.....	16
3.2	Componentes de un SGBD.....	16
3.2.1	Lenguaje de definición de datos DDL.....	17
3.2.2	Lenguaje de manipulación de datos DML.....	17
3.2.3	El diccionario de datos.....	17
3.2.4	El gestor de la base de datos.....	18
3.3	La estructura operacional de un SGBD.....	18
4	Bases de datos distribuidas.....	19
4.1	Tipos de bases de datos distribuidas.....	20
4.2	Fragmentación.....	20
5	Protección de datos.....	21
6	Big Data y Bussiness Intelligence.....	23
7	Bibliografía.....	25

Introducción a las Bases de Datos

1 Introducción

Actualmente cualquier organización con un sistema de información automatizado dispone de una Base de Datos (BD) para almacenar y gestionar toda la información de dicha organización. Sin embargo, antes de la aparición de las BD, las aplicaciones de usuario se encargaban de gestionar su propia información, lo que acarreaba una serie de inconvenientes que dificultaba la actividad de las organizaciones.

Vamos a describir a continuación los primitivos sistemas de almacenamiento de la información, los inconvenientes que sufrían y como las BD se crearon para superar estas dificultades.

1.1 Sistema de Información

Toda organización, grande o pequeña, necesita una infraestructura para poder desarrollar sus actividades. Esta estructura organizativa está basada en un conjunto de funciones que hay que desarrollar y que, entre otras, incluye las siguientes:

- ✓ Controlar y gestionar los recursos financieros a través de la función contable y gestión económica.
- ✓ Comercializar los productos o servicios en los que la empresa basa su negocio: la actividad comercial y de ventas.
- ✓ Fabricar productos o crear servicios que venden en el mercado: departamento de producción.

Todas estas funciones tienen que coordinarse entre sí mediante la gestión y la

comunicación de información de calidad. Por ello, las organizaciones incluyen una infraestructura para coordinar el flujo y registro de información necesaria para desarrollar sus actividades de acuerdo a una estrategia de negocio. El sistema dedicado a este cometido es el que se denomina sistema de información (SI) de la organización.

Una definición formal de Sistema de Información sería:

Un conjunto formal de procesos, que operando sobre una colección de datos estructurada según las necesidades de la organización, recopilan, transforman y distribuyen la información, o parte de ella, necesaria para las operaciones de dicha empresa y para las actividades de dirección y control correspondientes (toma de decisiones) con objeto de desempeñar su actividad de acuerdo a su estrategia de negocio.

Los componentes que forman un SI son:

- ✓ La información → Se debe adaptar a las personas que la manejan y al equipo disponible, según los procedimientos de trabajo definidos en la organización para la realización de su actividad.
- ✓ Los procedimientos → Prácticas habituales de trabajo que se siguen al realizar toda clase de actividades necesarias para el buen funcionamiento de la organización. Estos procedimientos son para coordinar a los distintos elementos de la organización. El SI da soporte a la gestión de la información que hay que proporcionar en función de estos procedimientos.
- ✓ Usuarios → Personas o grupos de la organización que introducen, manejan o usan la información para realizar sus actividades en función de los procedimientos de trabajo establecido.
- ✓ Recursos → El equipo de soporte para la comunicación, procesamiento y almacenamiento de información. Constituye la parte más visible del SI. Puede incluir elementos de los más variados niveles tecnológicos.

1.2 Sistemas de Información Automatizado

Las organizaciones han ido incorporando nuevas tecnologías desde el siglo XX hasta la actualidad para mejorar el rendimiento y la eficacia de los SI. Se comenzó con máquinas de escribir, calculadoras mecánicas, teléfonos, cintas de papel perforado, etc. y se ha llegado ya a utilizar las tecnologías de la información y la comunicación.

Cuando el SI emplea las TIC en el procesamiento, comunicación y almacenamiento de la información se denomina Sistema de Información Automatizado (SIA). La automatización de un SI debe contemplar la elección del hardware y la configuración más adecuada de software de base (sistema operativo), y por supuesto, las aplicaciones software que permitan cubrir las necesidades de información que marca la estructura del SI.

1.2.1 SIA basado en ficheros

Si analizamos la situación pasada de algunas organizaciones nos encontramos con sistemas de información compuestos por diversas aplicaciones software para gestionar la información de la organización. Cada aplicación realiza unas determinadas operaciones. Los datos de dichas aplicaciones se almacenan en ficheros o archivos digitales dentro de las unidades de almacenamiento del ordenador (a veces en archivos binarios, o en hojas de cálculo, o incluso en archivos de texto). Este tipo de SI se denomina SI orientado a procesos.

En esta situación, existe una proliferación de ficheros, específicos cada uno de ellos de una determinada aplicación. Por tanto, los datos que gestionaba una organización se almacenaban en un conjunto de ficheros, es decir, la BD de la organización comprendía todos los ficheros que utilizaba para almacenar dicha información.




Cada fichero era generado por la propia aplicación que gestionaba los datos almacenados en el propio fichero, es decir, nos encontramos una fuerte dependencia de los datos con la aplicación que los manipula y gestiona. Una definición formal de archivo o fichero sería:

Conjunto de información relacionada que se almacena como una secuencia de bytes en un dispositivo de almacenamiento y es tratada por el sistema operativo como una unidad lógica única.

1.3 Tipos de archivos

Existen muchos tipos de archivos, pero aparte de los habituales utilizados en aplicaciones ofimáticas de propósito general, existen otros exclusivos de ciertas aplicaciones de uso extendido en los SIA. Estos archivos están formados por un conjunto de registros, cada uno de los cuales almacenaba la información sobre una entidad concreta (cliente, producto, etc.). Cada registro estaba formado por un conjunto de campos, cada uno de los cuales representaba un elemento simple de información generalmente denominado campo (nif, nombre, teléfono, referencia, descripción, precio, etc.). Finalmente, si un SIA utiliza varias aplicaciones que emplean información de diversas fuentes, entonces era necesario varios archivos independientes que se almacenaban en una carpeta o directorio.

Los ficheros podían ser de distinto tipo en función de la organización de los registros dentro de cada fichero, cada uno de ellos con ventajas e inconvenientes. Esta organización del archivo determina la forma en que los registros con la información se crean, eliminan o actualizan dentro del archivo.

- **Campo**  Unidad mínima.
- **Registro**  Conjunto de campos relacionados.
- **Fichero**  Conjunto de registros relacionados.
- **Carpeta**  Conjunto de ficheros relacionados.

Los tipos de archivo según su organización más habituales son los siguientes:

1.3.1 Archivo secuencial

Consiste en un fichero en el que los registros se añaden uno detrás de otro en secuencia. Un archivo secuencial es la forma más simple de almacenar y recuperar registros de un archivo. En un archivo secuencial, se almacenan los registros uno tras otro. Siempre se almacenan los registros al final del archivo, después del último. Los registros tienen la misma estructura y por tanto todos son del mismo tamaño.

Dirección de memoria	Marca de borrado	Nombre	Apellidos	Teléfono
1200		Alfredo	Bárcena	768334472
1300	*	Isabel	De los Ríos	987335612
1400		Carmen	Sierra	955347612
1500		Fernando	Ruiz	674992455
1600		Juan Carlos	Abad	573982277

Figura 1: Archivo secuencial

Las operaciones de actualización son muy costosas. Debido a ello para el borrado de registros se utiliza un campo especial que contiene una marca de registro borrado para los registros que han sido eliminados. Sin embargo, las actualizaciones son más costosas ya que implica el uso de un archivo temporal para copiar todos los registros más el registro modificado.

Este tipo de archivos es el único que admite un sistema de almacenamiento secuencial (cinta magnética), ya en desuso. Entre sus ventajas están:

- ✓ Las inserciones de nuevos registros son muy rápidas.
- ✓ Proporcionan la mejor utilización de espacio en disco.
- ✓ Los tiempos de acceso son bajos cuando es necesario acceder en lectura a todos los

registros.

- ✓ Son muy sencillos de gestionar.

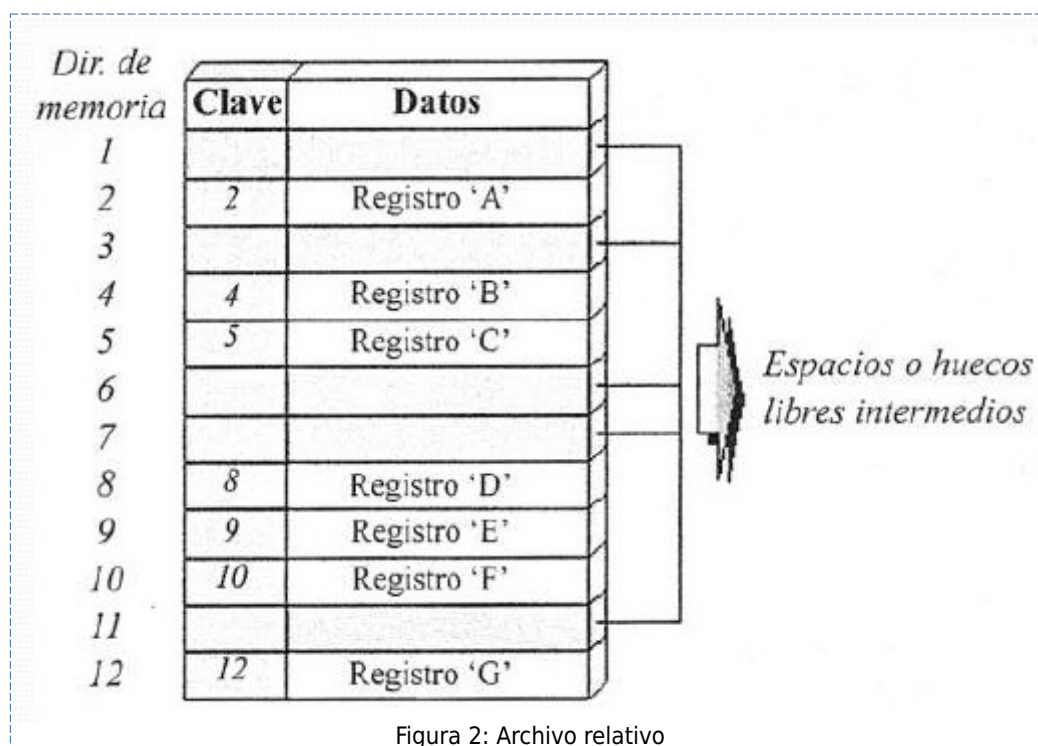
Entre sus desventajas están:

- ✓ Si el acceso a los datos debe ser aleatorio u según un orden diferente al orden de inserción de los registros, entonces los tiempos de acceso son muy altos.
- ✓ En dispositivos de almacenamiento de acceso aleatorio son muy ineficientes.
- ✓ Las operaciones de modificación de registros son muy lentas ya que necesitan generar otro archivo temporal que refleje la operación realizada.
- ✓ Son archivos muy vulnerables a fallos del sistema.

1.3.2 Archivo relativo

En un archivo con organización relativa el orden de inserción del registro es independiente de la posición en la que se almacena el registro en el archivo. Cuando se crea el archivo se dimensiona para albergar un número limitado de registros, por tanto tienen un tamaño definido que siempre es el mismo. Todo el espacio disponible para el archivo se divide en posiciones numeradas cada una de las cuales es capaz de albergar un registro.

Las celdas están numeradas desde el 0 y en orden creciente. A cada registro le corresponde una única celda cuyo número se obtiene por una operación realizada sobre la clave del fichero. El caso más favorable es cuando las claves son numéricas y permiten utilizar el direccionamiento directo.



Entre sus ventajas están:

- ✓ Permiten acceder de la forma mas rápida posible a un registro cualquiera del fichero. Por tanto es la organización ideal para los procesos en los que el tiempo de acceso a los registros es lo más importante, siempre y cuando el acceso se haga siempre por el campo clave.
- ✓ Permite la actualización de los registros en el mismo fichero, sin necesidad de utilizar ficheros temporales para copiar el fichero.
- ✓ Las inserciones, modificaciones y borrados son muy rápidos.

Entre sus desventajas están:

- ✓ Se desaprovecha mucho espacio si hay pocos registros y el fichero durante su creación se dimensionó para albergar muchos registros.
- ✓ El algoritmo de direccionamiento debe garantizar que no se producirán dos direcciones iguales para dos valores de clave distintos.
- ✓ No permite el acceso secuencial a los registros por otro orden que no sea el orden en que están grabados en el soporte.
- ✓ Sólo se puede utilizar un campo clave para acceder a los registros.

1.3.3 Archivo secuencial indexado

Es una variante de la organización secuencial que permite acceso directo o aleatorio a los registros, además de su acceso secuencial. De nuevo los registros tienen la misma

estructura y consecuentemente el mismo tamaño. Sin embargo, en el diseño del registro se indica un campo o combinación de campos que permite identificar a cada registro de forma única, es decir, no hay dos registros con el mismo valor de campo identificador. A este campo se le llama clave.

En un archivo secuencial indexado está dividido en varios archivos separados:

- ✓ Área de datos → Con los datos del archivo y organización secuencial. Los datos se almacenan en un soporte de almacenamiento directo, en secuencia ascendente, de acuerdo con los valores de la clave y en páginas o bloques de tamaño fijo.
- ✓ Área índice → Se crea por el sistema al mismo tiempo que se van almacenando los datos. Contiene una tabla que asocia las claves con las direcciones de los registros en el área de datos. Cada entrada del área de índices está formada por el valor más alto de la clave de cada grupo de registros y un puntero con la dirección del primer registro del grupo.
- ✓ Área de excedentes (overflow) → Donde se graban los registros que no tienen sitio en el área de datos. Los nuevos registros se insertan y quedan enlazados entre sí mediante punteros conservando el orden lógico que marca la clave o índice principal.

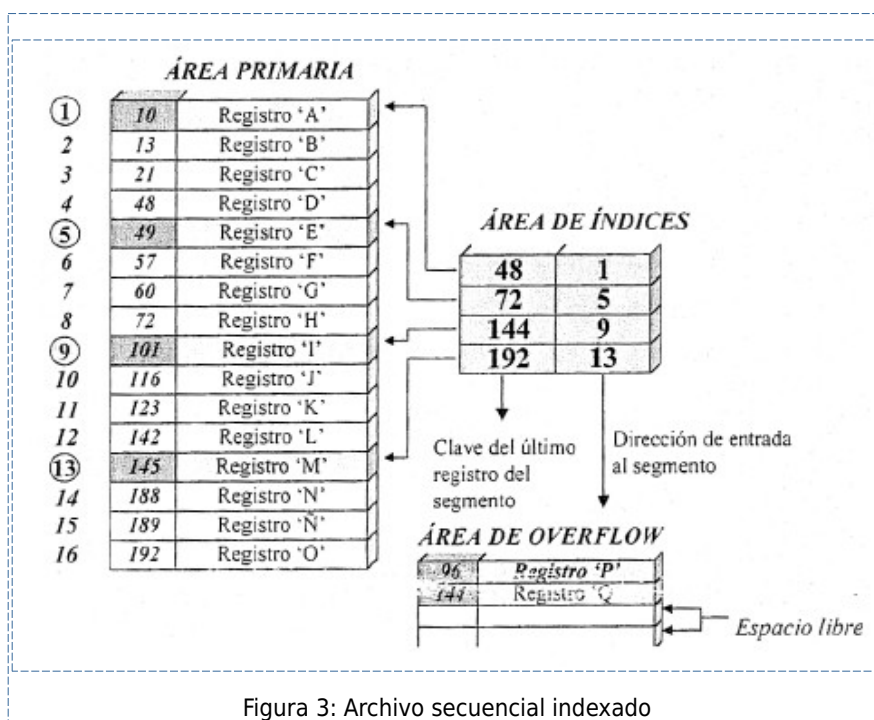


Figura 3: Archivo secuencial indexado

Entre sus ventajas están:

- ✓ Se puede acceder a todos los registros del archivo secuencialmente ordenados por el campo clave.
- ✓ Se permite el acceso directo a registros individuales a través de su campo clave.
- ✓ Se pueden actualizar los registros en el mismo fichero, sin necesidad de crear un

fichero nuevo de copia en el proceso de actualización.

Entre sus desventajas están:

- ✓ Ocupa más espacio en el disco que los ficheros secuenciales, debido al uso del área de índices.
- ✓ Tiene tendencia a degradarse los tiempos de acceso ya que cuando se producen muchas inserciones de registros nuevos con claves que hay que intercalar entre las existentes, aumenta el área de excedentes.
- ✓ Este tipo de archivo no puede existir en un soporte de almacenamiento no direccionable.

1.4 Inconvenientes de las BD basadas en ficheros

Independientemente del tipo de archivo que utilizara, la BD formada por un conjunto de ficheros tenía una fuerte dependencia de la aplicación que lo gestiona. Esto supone una serie de inconvenientes:

- ✓ Redundancia de datos.- Cuando más de una aplicación necesita utilizar los datos, estos se recogen y almacenan varias veces, estando repetidos en los distintos archivos. Esto es debido a que los ficheros solamente pueden ser accedidos por una aplicación, quedando bloqueados para el resto de aplicaciones. Esta redundancia, además de malgastar recursos, origina a menudo divergencias en los resultados. Las aplicaciones se analizan e implantan con independencia unas de otras, y los datos no se suelen transferir entre ellas, sino que se duplican siempre que los correspondientes trabajos los necesitan.
- ✓ Aumento del espacio de almacenamiento.- La redundancia de datos produce una ocupación inútil de memoria secundaria.
- ✓ Degradación del rendimiento.- Se hace necesario un aumento de los tiempos de proceso, al repetirse los mismos controles y operaciones en los distintos ficheros.
- ✓ Inconsistencia de la información.- A menudo se presentan incongruencias en los datos, debido a que la actualización de los mismos datos, cuando se encuentran en más de un fichero, no se suele realizar de forma simultánea en todos ellos.
- ✓ No se puede cambiar la estructura de los datos.- Ya que reside dentro del programa. Para cambiar la estructura de un fichero hay que modificar el programa.
- ✓ Los datos no están protegidos de accesos no autorizados.- Solo es necesario poseer el fichero para acceder a los datos.

Otra cuestión importante respecto a este tipo de BD era que los ficheros no estaban relacionados entre sí, y los datos almacenados representaban las relaciones existentes en la información mediante referencias simbólicas y/o físicas.

Se deduce claramente la necesidad de una gestión más racional del conjunto de datos, surgiendo así un nuevo enfoque que se apoya sobre una base de datos, en la cual los

datos son recogidos y almacenados una sola vez, con independencia de las aplicaciones que los maneja. Por tanto, la solución de los problemas asociados al tratamiento de los datos en los sistemas tradicionales con ficheros lleva a un cambio radical, en el cual los datos se organizan y se mantienen en un conjunto estructurado que no está diseñado para una aplicación concreta, sino que, por el contrario, tiende a satisfacer las necesidades de información de toda la organización; necesidades cuya diversidad se acentuó en el transcurso del tiempo. A este tipo de SI se le denomina SIA orientado a datos.

2 Las bases de datos

Al sistema de almacenamiento de información tan primitivo visto en el apartado anterior se le denominó BD, aunque carecía de una serie de características inherentes al término empleado. Para que una BD pueda denominarse como tal, tiene que acreditar una serie de propiedades, las cuales fueron incorporándose a estos sistemas a medida que el software de administración de la información que se desarrolló fue haciéndose más eficaz. Una definición formal de base de datos sería

Una base de datos es un conjunto de archivos relacionados que almacena tanto los datos que forma el sistema de información de una organización como una representación de estos datos. Tanto los datos como la representación de los mismos están sujetos a una serie de restricciones inherentes a los mismos y la cual también es almacenada en estos archivos. La información se organiza y estructura de un modo específico para que su contenido pueda ser accedido y actualizado de manera rápida y sencilla.

Para cumplir esta definición, es imprescindible independizar la estructura de la información de las aplicaciones encargadas de su tratamiento. Si se desea que cualquier modificación en la cantidad, contenido o estructura de la información no afecte a las aplicaciones desarrollados previamente para el mantenimiento de la misma, es necesario tener en cuenta que existe una independencia de los datos respecto a las aplicaciones.

2.1 Ventajas de las BD

Las BD presentan una multitud de ventajas frente a los sistemas clásicos de ficheros, que son:

- ✓ Independencia de los datos respecto a las aplicaciones y viceversa.- Un cambio en el tratamiento de los datos no imponen rediseñar la base de datos. Por otra parte al cambiar en la BD la estructura de los datos no obliga a alterar los programas. La independencia en la práctica nunca es absoluta, pero permite adaptar el sistema de información de las organizaciones a su propia evolución sin excesivas dificultades ni costes.
- ✓ Coherencia de los resultados.- Como la información sólo se recoge y almacena una vez, en las aplicaciones se utilizan siempre los mismos datos, por lo que los resultados son coherentes. Al desaparecer la redundancia de los datos, desaparece el problema de las actualizaciones redundantes.

- ✓ Mejor disponibilidad de los datos para el conjunto de los usuarios.- Cada usuario ya no es propietario de los datos, puesto que están compartidos por el conjunto de aplicaciones.
- ✓ Mayor valor informativo.- Además de representar los datos, en una base de datos también se representan las relaciones entre los datos, con lo que el conjunto de información es superior a la suma de los elementos individuales que lo forman.
- ✓ Mejor documentación de la información.- En los ficheros clásicos los datos están almacenados, pero su descripción se hacen en los programas. Este problema se atenúa en las bases de datos ya que los datos y su semántica están almacenados en la misma.
- ✓ Mayor eficiencia en la recogida, validación e introducción de los datos en el sistema.- Los datos sólo se recogen y validan una vez, aumentando el rendimiento de todo el proceso previo al almacenamiento.
- ✓ Reducción del espacio de almacenamiento.- Al desaparecer la redundancia de los datos, se reduce la ocupación de almacenamiento secundario, aunque la información referente a la estructura de los datos y su semántica ocupan espacio.

2.2 Características de las BD

Para que una BD se considere como tal, tiene que cumplir una serie de características:

- ✓ Versatilidad para la representación de la información.- La organización de la información en la BD debe permitir que diferentes programas puedan construir diferentes registros a partir de la información existente en la base de datos. Estos registros lógicos están formados por ítems de datos derivados del conjunto de los ítems de datos existentes en los registros físicos.
- ✓ Desempeño.- Hay que asegurar un tiempo de respuesta adecuado, permitiendo el acceso simultáneo al mismo o distinto conjunto de ítems de datos por la misma o distinta aplicación.
- ✓ Mínima redundancia.- La tecnología de BD surgió por diversas razones, pero una de las principales fue evitar la redundancia de información presente en sistemas de procesamiento que usaban archivos planos. Las BD no evitan totalmente la redundancia ya que es necesario representar relaciones que existen entre las entidades que forman parte de la información. La existencia de redundancia provoca problemas de inconsistencia de la información y necesidades de almacenamiento mayores.
- ✓ Integridad.- Veracidad de los datos almacenados respecto a la información que trata la BD. Muchos usuarios manejan los datos haciendo uso de muchos procedimientos que tratan los mismos datos de muchas formas, siendo necesario garantizar que estos datos no sean modificados ni destruidos de forma anómala. Los problemas que podrían poner en riesgo esta integridad pueden ser de dos tipos:
 - ✗ Durante el procesamiento pueden producirse fallos de muy diversa naturaleza:

errores del sistema general, del hardware, software, etc. Así, la BD debe garantizar la integridad de la información a pesar de los errores que se puedan producir a causa de los fallos con independencia de su naturaleza.

- ✗ La integridad debe garantizarse con respecto a la veracidad de los ítems de datos y sus relaciones. En una BD deben establecerse procedimientos que verifiquen que los valores de los datos se ajustan a los requerimientos y restricciones extraídos del análisis del problema.
- ✓ Seguridad y Privacidad.- La seguridad consiste en la capacidad de una BD para proteger los datos contra su pérdida total o parcial por fallos del sistema o por accesos accidentales o intencionados a los mismos. La privacidad de una base de datos hace referencia a la reserva de la información a personas no autorizadas. La información de una organización es un bien muy valioso que hay que proteger por lo que una base de datos deberá satisfacer los siguientes requisitos:
 - ✗ Seguridad contra la destrucción de los datos causada por el entorno: fuego, robo, inundaciones, etc.
 - ✗ Seguridad contra la destrucción de los datos causada por fallos del sistema (hardware o software) de forma que los datos puedan reconstruirse.
 - ✗ Seguridad contra accesos no autorizados a la base de datos.
 - ✗ Seguridad contra accesos indebidos a los datos.

En una BD existirán procedimientos de recuperación de la información por pérdida total o parcialmente, debido a cualquier causa, como procedimientos que supervisen el acceso a los datos por los usuarios de la base de datos.

2.3 Tipos de bases de datos

Dependiendo de la forma de estructurar y almacenar la información existen cuatro tipos de bases de datos. Son las siguientes:

2.3.1 Relacional

Este tipo de BD es el de uso más extendido debido a su parecido con los antiguos sistemas de ficheros. Se basa en el modelo ideado por Codd en 1970 en los laboratorios de IBM. El elemento fundamental de una BD relacional es la relación, la cual está formado por un conjunto de datos llamados *tuplas* las cuales a su vez contienen atributos como elementos simples de información. 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).

La información puede ser gestionada mediante consultas realizadas con un lenguaje estándar de BD relacionales denominado SQL (*Structured Query Language*). El usuario o las aplicaciones pueden insertar, recuperar o actualizar la información de la BD de datos

mediante consultas SQL que se envían a la BD.

2.3.2 Distribuida

Una base de datos distribuida (BDD) es un conjunto de múltiples bases de datos lógicamente relacionadas las cuales se encuentran distribuidas en diferentes localizaciones físicas e interconectados por una red de comunicaciones. Las bases de datos distribuidas se verán en un apartado posterior.

2.3.3 Orientada 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, lo que incluye sus propiedades y sus métodos (estado y comportamiento). Una BD orientada a objetos incorpora todos los conceptos importantes del paradigma de objetos:

- ✓ Encapsulación - Característica que permite ocultar la información de un objeto al resto de los objetos, impidiendo así accesos indebidos.
- ✓ Herencia – Característica a través de la cual los objetos heredan tanto propiedades como métodos de otros objetos, creando una jerarquía de clases.
- ✓ Polimorfismo – Característica por la cual una operación puede aplicarse a distintos objetos.

En BD orientadas a objetos, los usuarios pueden definir funciones sobre los datos como parte de la definición de la BD. Una función se especifica en dos partes:

- ✓ La interfaz → Incluye el nombre de la función y los tipos de datos de sus parámetros o argumentos.
- ✓ La implementación → Se especifica separadamente y puede modificarse sin afectar la interfaz. Consiste en el código ejecutable de la función que opera sobre los datos.

Los programas de aplicación de los usuarios pueden operar sobre los datos invocando a dichas funciones a través de interfaz, sea cual sea la forma en la que se han implementado.

2.3.4 Jerárquicas

Una base de datos jerárquica almacena la información en una estructura jerárquica en la que los datos se distribuyen en una estructura de árbol invertido, donde un nodo padre de puede tener varios nodos hijo, y estos a su vez, más nodos hijo, así sucesivamente.

Esta relación jerárquica no es estrictamente obligatoria, de manera que pueden establecerse relaciones entre nodos hermanos, y en este caso, la estructura en forma de árbol se convierte en una estructura en forma de grafo dirigido (esta variante se denomina BD de red).

Las principales características de las bases de datos jerárquicas son las siguientes:

- ✓ Se organizan en forma de árbol invertido, con un nodo raíz, nodos padre e hijos.

- ✓ El árbol se organiza en un conjunto de niveles, siendo el nivel 0 el nodo raíz y es el nivel más alto de la jerarquía.
- ✓ Los arcos (enlaces) representan las asociaciones jerárquicas entre dos nodos, carecen de nombre porque entre dos conjuntos de datos solo puede existir una interrelación.
- ✓ Uno nodo padre puede tener un número ilimitado de nodos hijos, pero a un nodo hijo solo le puede corresponder un padre.
- ✓ Todo nodo debe tener un padre, a excepción del nodo raíz.
- ✓ Los nodos sin descendientes se llaman «hojas».
- ✓ Los niveles de la estructura jerárquica se denominan «altura».
- ✓ El árbol siempre se recorre en un orden prefijado.
- ✓ La estructura del árbol no se puede modificar cuando ha quedado establecida.

La principal ventaja que presenta este tipo de base de datos es la rapidez en las consultas de información ya que la propia estructura jerárquica es ideal para este tipo de operaciones, no así para operaciones de actualización.

2.4 Niveles de la base de datos

Dependiendo de quien acceda o use la BD, ésta debe presentarle una visión de los datos que sea capaz de reconocer, interpretar y manejar. El organismo ANSI (*American National Standards Institute*) ha marcado la referencia para la construcción de un sistema gestor de base de datos (SGBD). El modelo definido por el grupo de trabajo SPARC se basa en estudios anteriores en los que se definían tres niveles de abstracción necesarios para gestionar una base de datos. ANSI profundiza más en esta idea y define cómo debe ser el proceso de creación y utilización de estos niveles.

Existen, pues tres visiones de los datos en una BD en función del tipo de usuario que la maneja y que permite describir la información de una BD en tres niveles:

- ✓ Visión externa.- Es la visión de los datos que tienen los usuarios finales de una base de datos. Un usuario trata sólo una visión parcial de la información, sólo aquella que interviene en su actividad. Este usuario verá la información como un registro, con independencia de a qué entidad pertenecen los ítems de datos, correspondientes a ese registro. Otro usuario verá también su registro particular de información cuyos ítems de datos podrán ser comunes o no al de otros registros particulares de otros usuarios. Estas visiones particulares de los usuarios son proporcionadas por los procedimientos o programas de aplicación que sólo manejan parte de la información de la base de datos.
- ✓ Visión conceptual.- Representación del problema de gestión de información tal y como se presenta en el mundo real y que se trata de resolver mediante una BD. Una BD representa la información de un problema del mundo real. Se determinan las

entidades, sus propiedades o características y las relaciones o dependencias que existen entre ellas. La visión conceptual de una base de datos es una representación abstracta del problema e independiente de cómo va a ser tratada esta información, de qué visiones externas pueda tener y de cómo está información pueda ser almacenada físicamente. Esta visión conceptual no cambia a no ser que cambie la naturaleza del problema.

- ✓ Visión física.- Representación de cómo la información es almacenada en los dispositivos de almacenamiento. Esta visión describe las estructuras de datos u organizaciones físicas, dispositivos, ficheros, tipos de datos, etc., estructuras que representan el problema entendible para el sistema informático.

Estas tres visiones, en realidad suponen tres representaciones del mismo problema a diferentes niveles. En cada nivel se describen aquellos objetos que pueden ser entendidos por los usuarios de ese nivel de la base de datos. El usuario final sólo entiende de registros o campos que maneja la organización (visión externa). El diseñador de la BD o analista de sistemas sólo entiende de tipos de entidades o clases de objetos que intervienen en el problema que la organización desea, de sus relaciones y los procedimientos que lleva a cabo la organización para solucionar el problema (visión conceptual). El administrador de la base de datos es el encargado de describir el nivel físico para determinar aquella organización física que pueda garantizar el desempeño óptimo del sistema (visión física).

Esta descripción de los datos a diferentes niveles garantiza, en principio, la independencia de los datos:

- ✓ Que pueda ser modificada la organización física sin que haya cambiado la descripción conceptual y sin que haya que modificar los programas de aplicación que manipula esa información.
- ✓ Que pueda ser modificada la representación conceptual del problema sin que por ello tenga que ser modificada la estructura física de la información, ni los programas de aplicación, siempre y cuando no se eliminen de la representación conceptual objetos necesarios o requeridos en estos otros niveles.
- ✓ Que las visiones externas puedan cambiar conforme a nuevos requerimientos funcionales sin que por ello se tenga que modificar ninguna descripción de datos a ningún nivel de abstracción.

Estos tres niveles deben ser independientes, pero en la mayoría de las BD no es del todo cierto. Sin embargo, puede conseguirse una buena independencia de los datos si:

- ✓ Una misma representación conceptual de los datos puede representarse de varias formas físicamente. La representación física de los datos vendrá determinada por requerimientos funcionales o de desempeño.
- ✓ El número de ítems y disposición de los mismos en una representación externa debe ser independiente de cómo estos ítems han sido representados en el nivel conceptual y de las relaciones que mantienen en el mismo.

El nivel más importante es el nivel conceptual, ya que los otros dos niveles se apoyan

en este y en base a su calidad, se garantiza que la base de datos solucione el problema de información de la organización. Una descripción conceptual de la organización incluirá todas las entidades que intervienen en el problema, sus atributos y propiedades, así como las relaciones existentes entre ellas. Se describen los ítems de datos que intervienen en el comportamiento del sistema y cuya información es necesario considerar.

Sin embargo, hay muchas formas, y con diversas técnicas, de representar un sistema conceptualmente. Debido a esto se puede hablar de un cuarto nivel en la representación de la información en una BD, el nivel lógico. Mientras que la descripción conceptual del problema es independiente de las descripciones externa y física, la descripción lógica si lo es. La descripción lógica deriva de la descripción conceptual en base a la aplicación de una serie de reglas y restricciones que tienen en cuenta cómo la información representada puede ser tratada por los procedimientos que van a manejar y definir la información en base a las otras representaciones.

3 Sistemas gestores de bases de datos

Es importante conocer la diferencia entre lo que es una base de datos y lo que es un Sistema de Gestión de Base de Datos, términos que se confunden muy a menudo cuando se está trabajando con la información haciendo uso de esta tecnología. Cuando se habla de base de datos se habla de información almacenada cumpliendo una serie de características y restricciones, pero para que la información pueda ser almacenada y el acceso a la misma satisfaga las características exigidas a una base de datos, es necesario que exista una serie de procedimientos, un sistema software, que sea capaz de llevar a cabo tal labor. A este sistema software es a lo que se le denomina Sistema de Gestión de Bases de Datos (SGBD).

Un SGBD es una colección de aplicaciones que proporcionan al usuario los medios necesarios para definir, manipular, mantener la integridad y la seguridad de los datos en una base de datos.

Un SGBD debe permitir realizar las siguientes tareas:

- ✓ Definición de los datos a los distintos niveles de abstracción (físico, lógico y externo)
- ✓ Manipulación de los datos en la base de datos. Es decir, la inserción, modificación, borrado y acceso o consulta a los mismos.
- ✓ Mantenimiento de la integridad de la base de datos. Integridad en cuanto a los datos en sí, sus valores y las relaciones entre ellos.
- ✓ Control de la privacidad y seguridad de los datos en la base de datos.

3.1 Funciones de un SGBD

Las funciones esenciales de un SGBD son las de descripción, manipulación y control.

3.1.1 Definición o descripción de los datos

Debe permitir al diseñador de la BD especificar los elementos de datos que la integran, su estructura y relaciones que existen entre ellos, las reglas de integridad semántica, etc. así como las características de tipo físico y las vistas externas de los usuarios.

Esta función la realiza el lenguaje de definición de datos (DDL) propio de cada SGBD, el cual debe de suministrar los medios para definir las tres estructuras de datos (externa, lógica e física), especificando las características de los datos a cada uno de estos niveles.

3.1.2 Manipulación de los datos

Una vez se han descrito los datos, es preciso cargar los datos en las estructuras previamente creadas, con lo que la BD estará ya dispuesta para su utilización. Los usuarios tendrán necesidad de recuperar la información o de actualizarla mediante consultas. La consulta a la base de datos podrá ser de dos tipos:

- ✓ Sobre la totalidad de los datos en la que se recuperan todos los datos de la BD o todos los de un determinado tipo.
- ✓ Consulta selectiva, en la que se tendrán que localizar los registros que cumplen una determinada condición (criterio de selección).

La actualización de los datos en una base de datos supondrá tres tipos de operaciones distintas:

- ✓ Inserción de nuevos registros.
- ✓ Borrado o eliminación de registros.
- ✓ Modificación de los datos de aquellos registros en los cuales se hayan producido cambios.

La función de manipulación la realiza el lenguaje de manipulación de datos (DML) el cual facilita los instrumentos para realizar estas tareas.

3.1.3 Control de la BD

Debe integrar una serie de instrumentos para facilitar la tarea de administración de la BD por parte del administrador. En la mayoría de los SGBD existen funciones de servicio, como cambiar la capacidad de los ficheros, obtener estadísticas de utilización, etc. y principalmente las relacionadas con la seguridad (copias de seguridad, re arranque en caso de caída del sistema, etc.) y de protección frente a accesos no autorizados.

3.2 Componentes de un SGBD

Para realizar las funciones anteriores y algunas más es necesario que el SGBD cuente con una serie de componentes, cada uno de los cuales puede desempeñar una o varias de sus funciones, satisfaciendo los requerimientos impuestos para estos sistemas.

3.2.1 Lenguaje de definición de datos DDL

Permite definir la representación lógica y física de los datos, denominada Esquema de la BD. En el esquema están definidas:

- ✓ Descripción de la BD nivel lógico → Las clases de objetos con sus propiedades, relaciones entre objetos y sus propiedades, restricciones tanto de las clases de objetos como de las relaciones.
- ✓ Descripción de la BD a nivel físico → Las unidades físicas en las cuales los datos se almacenarán, volúmenes y archivos utilizados, características físicas de los medios de almacenamiento y métodos de acceso a la información.

Además, el DDL cuenta con un sublenguaje encargado del control y seguridad de los datos. Este sublenguaje se denomina lenguaje de control de datos (DCL) y permite el control de acceso a la información almacenada en el diccionario de datos.

3.2.2 Lenguaje de manipulación de datos DML

Mediante este lenguaje se realizan dos funciones en la gestión de los datos:

- ✓ La definición del nivel externo o de usuario de los datos.
- ✓ La manipulación de los datos; es decir, la inserción, borrado, modificación y recuperación de los datos almacenados en la base de datos.

Dependiendo del SGBD y del modelo de datos empleado hay dos tipos de DML:

- ✓ Procedimentales.- Las sentencias del lenguaje requieren que se especifique qué datos se van a manipular y qué operaciones deben realizarse para ello.
- ✓ No procedimentales.- Solo requieren que en las sentencias se especifique qué datos se van a manipular, siendo el propio DML el encargado de determinar los procedimientos más efectivos para ello.

Estos últimos son los más fáciles para el usuario no experto, pero puede requerir modificar el código que generan estos lenguajes para asegurar un buen desempeño.

El DML tiene la función de describir la visión externa de los datos, mediante las vistas o visiones parciales que los usuarios tienen del esquema de la BD definido mediante el DDL. Estas vistas, también denominadas subesquemas pueden realizarse de dos formas:

- ✓ Usando solamente el DML
- ✓ Haciendo uso de un lenguaje de programación, mediante el cual se realizan los programas de aplicación que permiten al usuario manipular los datos de la base de datos. En el código fuente de estos programas están presentes sentencias del DML.

3.2.3 El diccionario de datos

Conjunto de archivos que contienen información acerca de los datos que pueden ser almacenados en la base de datos. Se almacenan todas las definiciones realizadas por el

DDL sobre el problema y algunas de las realizadas por el DML: el esquema lógico, físico y subesquemas. Además, el diccionario de datos también almacena:

- ✓ Restricciones de privacidad y acceso a los datos almacenados en la BD.
- ✓ Reglas, normas o restricciones referentes a la seguridad de los datos.
- ✓ Información para garantizar la integridad de los datos almacenados en la base de datos.

3.2.4 El gestor de la base de datos

Es un componente software encargado de garantizar el correcto, seguro, íntegro y eficiente acceso y almacenamiento de los datos. Proporciona una interfaz entre los datos almacenados y los programas de aplicación. Todas las operaciones realizadas en la base de datos deben contar con su aprobación. El gestor las interpreta y valida, devolviendo el resultado al usuario o programa de aplicación que la solicitó, o bien la rechaza. Es el responsable de:

- ✓ Garantizar la privacidad de los datos, permitiendo sólo acceso a los usuarios autorizados.
- ✓ Garantizar la seguridad de los datos, realizando procedimientos que puedan recuperar datos después de un fallo que ocasione pérdida o deterioro temporal de los mismos.
- ✓ Garantizar la integridad de los datos, gestionando que los datos que se almacenan en la base de datos satisfacen las restricciones definidas en el esquema de la misma.
- ✓ Garantizar el acceso concurrente a la base de datos de forma que varios usuarios puedan acceder al mismo o distinto dato sin que se provoque una pérdida de la integridad de los datos.
- ✓ Interaccionar con el sistema operativo, especialmente con el sistema de archivos, para que los procedimientos DML puedan ser entendidos por el sistema operativo para el correcto almacenamiento y recuperación de la información. Para ello el gestor dispone del procesador de consultas.

3.3 La estructura operacional de un SGBD

Actualmente casi todos los SGBD poseen la misma forma de funcionar. Siguen una arquitectura Cliente/Servidor en la que un servidor de base de datos ejecuta el SGBD y almacena la BD. Este servidor permanece en espera a la escucha de peticiones desde un conjunto de aplicaciones cliente para acceder a los datos. Dentro de este enfoque de operar tenemos las siguientes posibilidades:

- ✓ Cliente-Servidor en 2 capas.- Estructura clásica, la base de datos y su SGBD están en un servidor al cual acceden los clientes. El cliente posee software que permite al usuario enviar instrucciones al SGBD en el servidor y recibir los resultados de estas instrucciones. Para ello el software cliente y el servidor deben utilizar software de

comunicaciones en red.

- ✓ Cliente-Servidor en 3 capas.- Se trata de una forma de conexión por el que los clientes no conectan directamente con el SGBD sino con un intermediario (normalmente un servidor web o un servidor de aplicaciones) que tiene una mayor facilidad para comunicarse con los usuarios. Ese servidor se encarga de traducir lo que el cliente realiza a una forma entendible por la base de datos.
- ✓ Cliente multi-servidor.- Ocurre cuando los clientes acceden a datos situados en más de un servidor. Esta es la forma de operar de las BD distribuidas. El cliente no sabe si los datos están en uno o más servidores, ya que el resultado es el mismo independientemente de dónde se almacenan los datos. En esta estructura hay un servidor de aplicaciones que es el que recibe las peticiones y el encargado de traducirlas a los distintos servidores de datos para obtener los resultados.

4 Bases de datos distribuidas

Como vimos en un apartado anterior, una base de datos distribuida o *Distributed Database Management System* (DDBMS) consiste en varias bases de datos instaladas en diferentes servidores los cuales están ubicados en diferentes localizaciones físicas y conectados entre sí por un sistema de comunicaciones. Dichas BD tienen la capacidad de realizar procesamiento autónomo, esto permite realizar operaciones locales o distribuidas.

Los usuarios acceden a la información de forma transparente, sin conocer la ubicación física de los datos con los que trabajan. El administrador de la base de datos es el responsable de la gestión individualizada de cada servidor para proveer a usuarios y aplicaciones la información almacenada, con controles de seguridad, acceso y concurrencia.

Una BD distribuida se caracteriza por:

- ✓ Hay múltiples servidores, llamados sitios o nodos.
- ✓ El conjunto de todos los servidores que forman la BD distribuida se conoce como cluster.
- ✓ Estos nodos deben de estar comunicados por medio de algún tipo de red de comunicaciones para transmitir datos y órdenes entre los sitios.

Una arquitectura de DDBMS consiste en:

- ✓ Varias bases de datos, las cuales pueden ser gestionadas por diferentes SGBD (sistema heterogéneo) o ejecutar el mismo SGBD (sistema homogéneo).
- ✓ Los servidores están físicamente ubicados en diferentes lugares.
- ✓ Los servidores pueden disponer de diferentes plataformas (Windows, Linux, Solaris, ...).
- ✓ Las redes que comunican los servidores de bases de datos pueden tener distintos protocolos y arquitecturas.

Las razones para implementar BD distribuidas:

- ✓ **Compartición de la información y autonomía local** → Una ventaja de compartir los datos mediante la distribución consiste en que desde cada nodo se puede controlar la administración de los datos almacenados localmente.
- ✓ **Fiabilidad y disponibilidad** → Si se produce un fallo en algún nodo de un sistema distribuido o en las comunicaciones con este, es posible que los otros nodos puedan continuar trabajando y prestando servicio ya que puede haber redundancia de datos.
- ✓ **Agilización del procesamiento de consultas** → Cuando una consulta necesita acceder a datos almacenados en diferentes nodos, puede ser posible dividir la consulta en diferentes subconsultas que se ejecuten en los nodos respectivos.

4.1 Tipos de bases de datos distribuidas

En función de cómo se distribuye la información en los distintos nodos, se pueden encontrar los siguientes tipos de bases de datos:

- ✓ **Centralizada** → La base de datos se encuentra en un único servidor centralizado.
- ✓ **Replicada** → Cada nodo almacena una copia completa de la BD. Tiene un alto costo en el almacenamiento de la información. Debido a que la actualización de los datos debe ser realizada en todas las copias, también tiene un alto costo en operaciones de escritura, pero todo esto vale la pena si tenemos un sistema en el que se va a actualizar los datos pocas veces y leer muchas, y dónde la disponibilidad y fiabilidad de los datos sea de máxima importancia.
- ✓ **Particionada** → La información de la BD está distribuida por todos los nodos sin replicación. En cada nodo se aloja uno o más fragmentos disjuntos de la base de datos. Como los fragmentos no se replican se disminuye el costo de almacenamiento, pero también sacrifica la disponibilidad y fiabilidad de los datos. Algo que se debe tomar en cuenta cuando se desea implementar este modelo es la granularidad de la fragmentación. En un apartado posterior se verá los modelos de fragmentación que existen.
- ✓ **Híbrida o Mixta** → Es una combinación de los esquemas particionado y replicada. En este esquema, la información se encuentra particionada entre los distintos servidores y también se almacenan replicas.

Las bases de datos distribuidas suelen utilizar un modelo híbrido. Estos sistemas almacenan la información de manera particionada entre los distintos nodos generando además réplicas de los datos. Esto permite tener la información disponible incluso tras fallo en alguno de los nodos.

4.2 Fragmentación

En los modelos de BDD particionada e híbrida cada BD almacena fragmentos de la información. La fragmentación de los datos permite dividir un objeto en dos o más segmentos o fragmentos. El objeto podría ser una base de datos completa o tablas

individuales. Cada fragmento puede guardarse en cualquier servidor en la red. La información de la fragmentación de los datos se guarda en un catálogo de datos distribuidos, desde donde se accede por el procesador de transacciones para procesar las solicitudes de los usuarios.

Las estrategias de fragmentación de los datos, como se analizan en esta sección, están basadas a nivel de tabla y consisten en dividir una tabla en fragmentos lógicos. Vamos a explorar tres tipos de fragmentación de datos: horizontal, vertical y mixta o mezclada.

- ✓ La fragmentación horizontal se refiere a la división de una tabla en subconjuntos (fragmentos) de tuplas (filas o registros). Cada fragmento se guarda en un servidor diferente y cada uno de ellos tiene filas únicas que tienen los mismos atributos (columnas).
- ✓ La fragmentación vertical se refiere a la división de una tabla en subconjuntos de atributos (columna); cada subconjunto (fragmento) se guarda en un servidor diferente y cada fragmento tiene columnas únicas, con la excepción de la columna clave, la cual es común a todos los fragmentos.
- ✓ La fragmentación mixta se refiere a una combinación de estrategias horizontales y verticales. En otras palabras, una tabla puede dividirse en varios subconjuntos horizontales (tuplas), y cada una tiene un subconjunto de los atributos (columnas).

5 Protección de datos

El RGPD (Reglamento General de Protección de datos) es la normativa europea de privacidad que entró en vigor en mayo de 2018 y que en España fue adaptada por la LOPDGDD (Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales) en diciembre de ese mismo año.

El RGPD y la LOPDGDD ofrecen un marco de protección al tratamiento de datos personales, que debe ser cumplido por todas las organizaciones y empresas en el ámbito de la UE que traten este tipo de datos, incluyendo también las empresas que se encuentren fuera de la UE y ofrezcan bienes o servicios a personas ubicadas en la UE o que monitoricen su comportamiento.

La normativa de protección de datos se basa en la premisa de que todos los datos personales de los usuarios deben ser protegidos por los legisladores europeos. Se considera “dato personal” a todo dato informativo capaz de hacer identificada o identificable a una persona física. Es decir, cualquier dato que, por sí mismo o en consonancia a otros datos que pudieran cotejarse, pueda llevar a la identificación de una persona física deberá ser considerado dato personal.

Por otra parte, existen una serie de principios generales que las empresas deberán tener en cuenta a la hora de aplicar esta regulación:

- ✓ Minimización de los datos personales, debiéndose tratar los datos exclusivamente necesarios para el tratamiento de los mismos.

- ✓ Limitación de la finalidad del tratamiento. Es decir, que los tratamientos se realicen para unas finalidades concretas.
- ✓ Limitación del plazo de conservación de los datos personales. Es decir, bloquear y/o destruir los datos en plazo de tiempo determinado.
- ✓ Transparencia en cuanto a la información proporcionada a los interesados sobre el tratamiento de sus datos personales.
- ✓ Responsabilidad proactiva en el uso de medidas de seguridad y políticas antes brechas de seguridad en los sistemas utilizados por las empresas en el tratamiento de los datos personales.
- ✓ Exactitud y veracidad de los datos personales cuando los mismos no sean proporcionados por los interesados.
- ✓ Obtención de consentimiento como base de legitimación para el tratamiento de datos personales, cuando proceda. Este consentimiento deberá ser expreso.

Las bases de datos y su uso por las empresas y otras instituciones están reguladas por las normativas anteriores, las cuales establecen las obligaciones con las que deben cumplir los responsables del tratamiento de ficheros con datos personales para fines comerciales, como es el caso de las empresas de marketing que utilizan bases de datos para la gestión de sus campañas.

La esencia de esta normativa es que los titulares de los datos tengan conocimiento de cuáles son los datos que ha recabado una empresa y qué uso tiene previsto hacer de ellos. Y, por supuesto, este uso solo se puede hacer si el titular de los datos lo permite.

De acuerdo con la normativa, un responsable de la gestión de una base de datos debe cumplir con las siguientes obligaciones:

- ✓ Mantener las BBDD actualizadas, para reflejar la situación actual del titular de los datos.
- ✓ No utilizar los datos de manera que se excedan los fines para los que fueron recabados.
- ✓ Obtener el consentimiento previo al tratamiento de los datos. Para esto, se debe informar al usuario de sus derechos, para qué fines serán usados sus datos, quiénes serán los responsables de la gestión, y cuáles serán los plazos en los que serán usados.
- ✓ Si la base de datos es pública, el titular de los datos tiene derecho a oponerse al tratamiento de sus datos.
- ✓ El titular debe tener garantizado el acceso, rectificación, cancelación y derecho a la oposición al uso de sus datos para los fines establecidos por la empresa.
- ✓ Cuando surge una reclamación por violación de estos derechos, es el responsable del tratamiento (la empresa) quien tiene la carga de la prueba, es decir, quien debe

constatar su inocencia mediante la aportación de pruebas, y no es el titular de los datos quien tiene que demostrar que sus derechos han sido violados.

- ✓ Las bases de datos están regidas por el secreto profesional, por lo tanto, la información debe ser guardada tanto durante el tiempo en el que el titular autorizó su uso, como posteriormente.
- ✓ El secreto profesional afecta a cualquier persona que haga uso de los datos, incluidos terceros a los que se les cedan estos datos para fines determinados.
- ✓ El responsable de los datos deberá inscribir los ficheros en la *Agencia Española de Protección de Datos*. Previamente, es necesario notificar de este procedimiento al titular de los datos.

6 Big Data y Bussiness Intelligence

En el mundo empresarial han surgido en los últimos años dos conceptos íntimamente relacionados que se emplean en la toma de decisiones estratégicas de las empresas.

Cuando hablamos de Big Data (BD) nos referimos a conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales dentro del tiempo necesario para que sean útiles.

La naturaleza compleja del BD se debe principalmente a la naturaleza no estructurada de gran parte de los datos generados por las tecnologías modernas, como los web logs, los sensores incorporados en dispositivos, la maquinaria, los vehículos, las búsquedas en Internet, las redes sociales, ordenadores portátiles, smartphones y dispositivos GPS.

El *Business Intelligence* (BI), término también conocido como Inteligencia Empresarial o de Negocios, se refiere a una serie de técnicas y herramientas informáticas que reúnen aplicaciones, metodologías y una infraestructura, diseñadas especialmente para analizar la información recopilada por una empresa.

El BI funciona desde la premisa de fusionar todos los datos empresariales en un único servidor. Una vez que estos datos se evalúan offline, se guardan en una plataforma conocida como *Data Warehouse*.

Mientras que el BD está diseñado para capturar y almacenar datos de gran volumen, las herramientas de BI son las encargadas de llevar a cabo el análisis de toda esta cantidad de información, siguiendo unos patrones predictivos de comportamiento, basados en los datos.

Lo que sí es cierto es que ambos son complementarios y ayudan a las organizaciones a lograr una visión 360 del negocio y, consecuentemente, a tomar decisiones acertadas en base a unos datos. Además, cuando estos se transforman en conocimiento, les permiten ofrecer servicios y experiencias innovadores al cliente, surgiendo un modelo de negocio de éxito.

La recopilación de grandes cantidades de datos y la búsqueda de tendencias dentro de los datos permiten que las empresas reaccionen mucho más rápidamente, sin problemas y de manera eficiente. También les permite eliminar las áreas problemáticas antes de que los problemas acaben con sus beneficios o su reputación.

El análisis de realizado con BI utilizando los datos obtenidos en BD ayuda a las organizaciones a aprovechar sus datos y utilizarlos para identificar nuevas oportunidades. Eso, a su vez, conduce a la toma de decisiones estratégicas más inteligentes, operaciones más eficientes, mayores beneficios y clientes más satisfechos. Las empresas consiguen valor de las siguientes formas:

- ✓ Reducción de costes.
- ✓ Mejora en la toma de decisiones.
- ✓ Nuevos productos y servicios. Con la capacidad de medir las necesidades de los clientes y la satisfacción a través de análisis viene el poder de dar a los clientes lo que quieren.

7 Bibliografía

GÓMEZ NIETO, Miguel Ángel y LUQUE RUIZ, Irene, *Diseño y uso de Bases de Datos Relacionales*. 1997 RA-MA ISBN 9788478972791

ISDI, *Big Data y Business Intelligence: diferencias y similitudes* (17 abril 2023)
<https://www.isdi.education/es/blog/big-data-y-business-intelligence-diferencias-y-similitudes>