

Sistemas Gestores de Bases de Datos

Gestión y diseño de bases de datos



1. Datos y Archivos.

1.1 La necesidad de gestionar datos

En el mundo actual existe una cada vez mayor demanda de datos. Esta demanda siempre ha sido patente en empresas y sociedades, pero en estos años la demanda todavía de ha disparado más debido al acceso multitudinario a Internet.

El propio nombre **Informática** hace referencia al hecho de ser una ciencia que trabaja con información. Desde los albores de la creación de ordenadores, la información se ha considerado como uno de los pilares de las computadoras digitales. Por ello las bases de datos son una de las aplicaciones más antiguas de la informática.

En informática se conoce como **dato** a cualquier elemento informativo que tenga relevancia para el sistema. Desde el inicio de la informática se ha reconocido al dato como al elemento fundamental de trabajo en un ordenador. Por ello se han realizado numerosos estudios y aplicaciones para mejorar la gestión que desde las computadoras se realiza de los datos.

Inicialmente los datos que se necesitaba almacenar y gestionar eran pocos, pero poco a poco han ido creciendo. En la actualidad las numerosas aplicaciones de Internet han producido enormes sistemas de información que incluso para poder gestionarles requieren decenas de máquinas haciendo la información accesible desde cualquier parte del planeta y en un tiempo rápido. Eso ha requerido que la ciencia de las bases de datos esté en continua renovación para hacer frente a esas enormes necesidades.

Pero incluso podemos remontarnos más al hablar de datos. El ser humano desde siempre ha necesitado gestionar datos; de esta forma se controlaban almacenes de alimentos, controles de inventario y otras muchos sistemas de datos. Como herramienta el ser humano al principio sólo poseía su memoria y cálculo y como mucho la ayuda de sus dedos.

La escritura fue la herramienta que permitió al ser humano poder gestionar bases cada vez más grandes de datos. Además de permitir compartir esa información entre diferentes personas, también permitió que los datos se guardaran de manera continua e incluso estuvieran disponibles para las siguientes generaciones. Los problemas actuales con la privacidad ya aparecieron con la propia escritura y así el cifrado de datos es una técnica tan antigua como la propia escritura para conseguir uno de los todavía requisitos fundamentales de la gestión de datos, la **seguridad**.

Para poder almacenar datos y cada vez más datos, el ser humano ideó nuevas herramientas archivos, cajones, carpetas y fichas en las que se almacenaban los datos.



Tableta de arcilla del año 3000/3500 antes de Cristo. British Museum
Imagen de BabelStone:
<http://commons.wikimedia.org/wiki/User:BabelStone>

Antes de la aparición del ordenador, el tiempo requerido para manipular estos datos era enorme. Sin embargo el proceso de aprendizaje era relativamente sencillo ya que se usaban elementos que el usuario reconocía perfectamente.

Por esa razón, la informática adaptó sus herramientas para que los elementos que el usuario maneja en el ordenador se parezcan a los que utilizaba manualmente. Así en informática se sigue hablando de ficheros, formularios, carpetas, directorios,....



Oficina de patentes de Estados Unidos, 1940
Imagen de la Librería del Congreso de Estados Unidos:
<http://www.loc.gov/pictures/item/hec2009015033/>

2. Sistemas de Información.

2.1 La empresa como sistema

Según la RAE, la definición de sistema es **"Conjunto de cosas que ordenadamente relacionadas entre sí contribuyen a un determinado objeto"**.

La clientela fundamental del profesional de la informática es la empresa. La empresa se puede entender como un sistema formado por diversos objetos: el **capital**, los **recursos humanos**, los **inmuebles**, los **servicios** que presta, etc.



Oficina central de Archivos
MTM Business Systems

El sistema completo que forma la empresa, por otra parte, se suele dividir en los siguientes subsistemas:

- ✓ **Subsistema productivo.** También llamado subsistema real o físico. Representa la parte de la empresa encargada de gestionar la producción de la misma.
- ✓ **Subsistema financiero.** Encargado de la gestión de los bienes económicos de la empresa
- ✓ **Subsistema directivo.** Encargado de la gestión organizativa de la empresa

Hay que hacer notar que cada subsistema se asocia a un departamento concreto de la empresa.

2.2 Sistemas de información.

Los sistemas que aglutinan los elementos que intervienen para gestionar la información que manejan los subsistemas empresariales es lo que se conoce como Sistemas de Información. Se suele utilizar las siglas **SI** o **IS** (de **Information Server**) para referirse a ello).

Realmente un sistema de información sólo incluye la información que nos interesa de la empresa y los elementos necesarios para gestionar esa información.

Un sistema de información genérico está formado por los siguientes elementos:

- ✓ **Recursos físicos.** Carpetas, documentos, equipamiento, discos,...
- ✓ **Recursos humanos.** Personal que maneja la información
- ✓ **Protocolo.** Normas que debe cumplir la información para que sea manejada (formato de la información, modelo para los documentos,...)

Las empresas necesitan implantar estos sistemas de información debido a la competencia que las obliga a gestionar de la forma más eficiente sus datos para una mayor calidad en la organización de las actividades de los subsistemas empresariales.



2.3 Componentes de un sistema de información electrónico

En el caso de una **gestión electrónica de la información** (lo que actualmente se considera un **sistema de información electrónico**), los componentes son:

- ✓ **Datos.** Se trata de la información relevante que almacena y gestiona el sistema de información. Ejemplos de datos son: *Sánchez, 12764569F, Calle Mayo 5, Azul...*
- ✓ **Hardware.** Equipamiento físico que se utiliza para gestionar los datos. cada uno de los dispositivos electrónicos que permiten el funcionamiento del sistema de información.
- ✓ **Software.** Aplicaciones informáticas que se encargan de la gestión de la base de datos y de las herramientas que facilitan su uso.
- ✓ **Recursos humanos.** Personal que maneja el sistema de información.

3. Archivos.

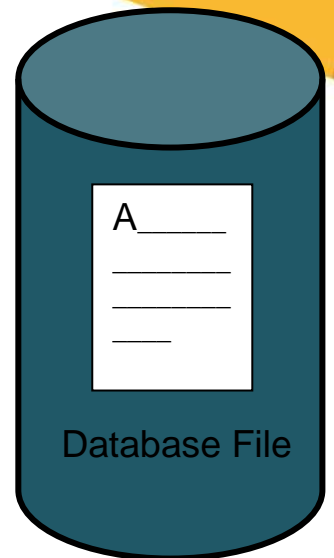
Los ficheros o archivos son la herramienta fundamental de trabajo en una computadora todavía a día de hoy. Las computadoras siguen almacenando la información en ficheros, eso sí de estructura cada vez más compleja.

Los datos deben de ser almacenados en componentes de almacenamiento permanente, lo que se conoce como **memoria secundaria** (discos duros u otras unidades de disco). En esas memorias, los datos se estructuran en archivos (también llamados ficheros).

Un fichero es una secuencia de números binarios que organiza información relacionada a un mismo aspecto.

En general sobre los archivos se pueden realizar las siguientes operaciones:

- ✓ **Abrir (*open*)**. Prepara el fichero para su proceso.
- ✓ **Cerrar (*close*)**. Cierra el fichero impidiendo su proceso inmediato.
- ✓ **Leer (*read*)**. Obtiene información del fichero.
- ✓ **Escribir (*write*)**. Graba información en el fichero.
- ✓ **Posicionarse (*seek*)**. Coloca el puntero de lectura en una posición concreta del mismo (no se puede realizar en todos los tipos de ficheros).
- ✓ **Fin de fichero (*eof*)**. Indica si hemos llegado al final del fichero.



Cuando los ficheros almacenan datos, se dice que constan de **registros**. Cada registro contiene datos relativos a un mismo elemento u objeto. Por ejemplo en un fichero de personas, cada registro contiene datos de una persona. Si el archivo contiene datos de 1000 personas, constará de 1000 registros. A continuación se explican los tipos más habituales de ficheros.

3.1 Ficheros secuenciales.

En estos ficheros, los datos se organizan secuencialmente en el orden en el que fueron grabados. Para leer los últimos datos hay que leer los anteriores. Es decir leer el registro número nueve, implica leer previamente los ocho anteriores.

Ventajas

- ✓ Rápidos para obtener registros contiguos de una base de datos
- ✓ No hay huecos en el archivo al grabarse los datos seguidos, datos más compactos.

Desventajas

- ✓ Consultas muy lentas al tener que leer todos los datos anteriores al dato que queremos leer
- ✓ Algoritmos de lectura y escritura más complejos
- ✓ No se pueden eliminar registros del fichero (se pueden marcar de manera especial para que no sean tenidos en cuenta, pero no se pueden borrar)
- ✓ El borrado provoca archivos que no son compactos
- ✓ La ordenación de los datos requiere volver a crearle de nuevo



3.2 Ficheros de acceso directo o aleatorio.

Se puede leer una posición concreta del fichero, con saber la posición (normalmente en bytes) del dato a leer. Cuando se almacenan registros, posicionarnos en el quinto registro se haría de golpe, lo único necesitamos saber el tamaño del registro, que en este tipo de ficheros debe de ser el mismo. Suponiendo que cada registro ocupa 100 bytes, el quinto registro comienza en la posición 400. Lo que se hace es colocar el llamado **puntero de archivo** en esa posición y después leer.

Ficheros Aleatorios		
Nº de registro de la posición en byte		Tamaño de cada registro 100 bytes
1	0	
2	100	
3	200	
4	300	
5	400	
6	500	
7	600	
8	700	
9	800	
10	...	

Acceso directo al registro Nº 5

Ventajas

- ✓ Acceso rápido al no tener que leer los datos anteriores
- ✓ La modificación de datos es más sencilla
- ✓ Permiten acceso secuencial
- ✓ Permiten leer y escribir a la vez
- ✓ Aptos para organizaciones **relativas directas**, en las que la clave del registro se relaciona con su posición en el archivo

Desventajas

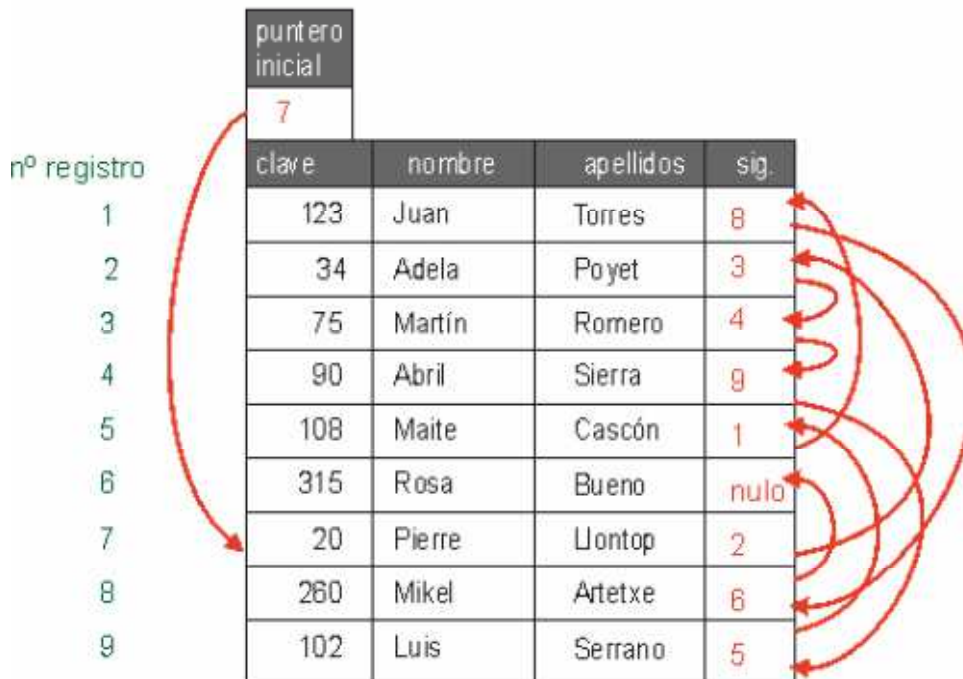
- ✓ Salvo en archivos relativos directos, no es apto por sí mismo para usar en bases de datos, ya que los datos se organizan en base a una clave
- ✓ No se pueden borrar datos (sí marcar para borrado, pero generarán huecos)
- ✓ Las consultas sobre multitud de registros son más lentas que en el caso anterior.

3.3 Ficheros secuenciales encadenados.

Son ficheros secuenciales gestionados mediante punteros, datos especiales que contienen la dirección de cada registro del fichero. Cada registro posee ese puntero que indica la dirección del siguiente registro y que se puede modificar en cualquier momento. El puntero permite recorrer los datos en un orden concreto.

Cuando aparece un nuevo registro, se añade al final del archivo, pero los punteros se reordenan para que se mantenga el orden.

Ficheros secuenciales encadenados



Ventajas

- ✓ El fichero mantiene el orden en el que se añadieron los registros y un segundo orden en base a una clave
- ✓ La ordenación no requiere reorganizar todo el fichero, sino sólo modificar los punteros
- ✓ Las mismas ventajas que el acceso secuencial
- ✓ En esta caso sí se borran los registros y al reorganizar, se perderán definitivamente

Desventajas

- ✓ No se borran los registros, sino que se marcan para ser ignorados. Por lo que se malgasta espacio
- ✓ Añadir registros o modificar las claves son operaciones que requieren recalcular los punteros

3.4 Ficheros secuenciales indexados.

Se utilizan dos ficheros para los datos, uno posee los registros almacenados de forma secuencial, pero que permite su acceso aleatorio. El otro posee una tabla con punteros a la posición ordenada de los registros. Ese segundo fichero es el **índice**, una tabla con la ordenación deseada para los registros y la posición que ocupan en el archivo.

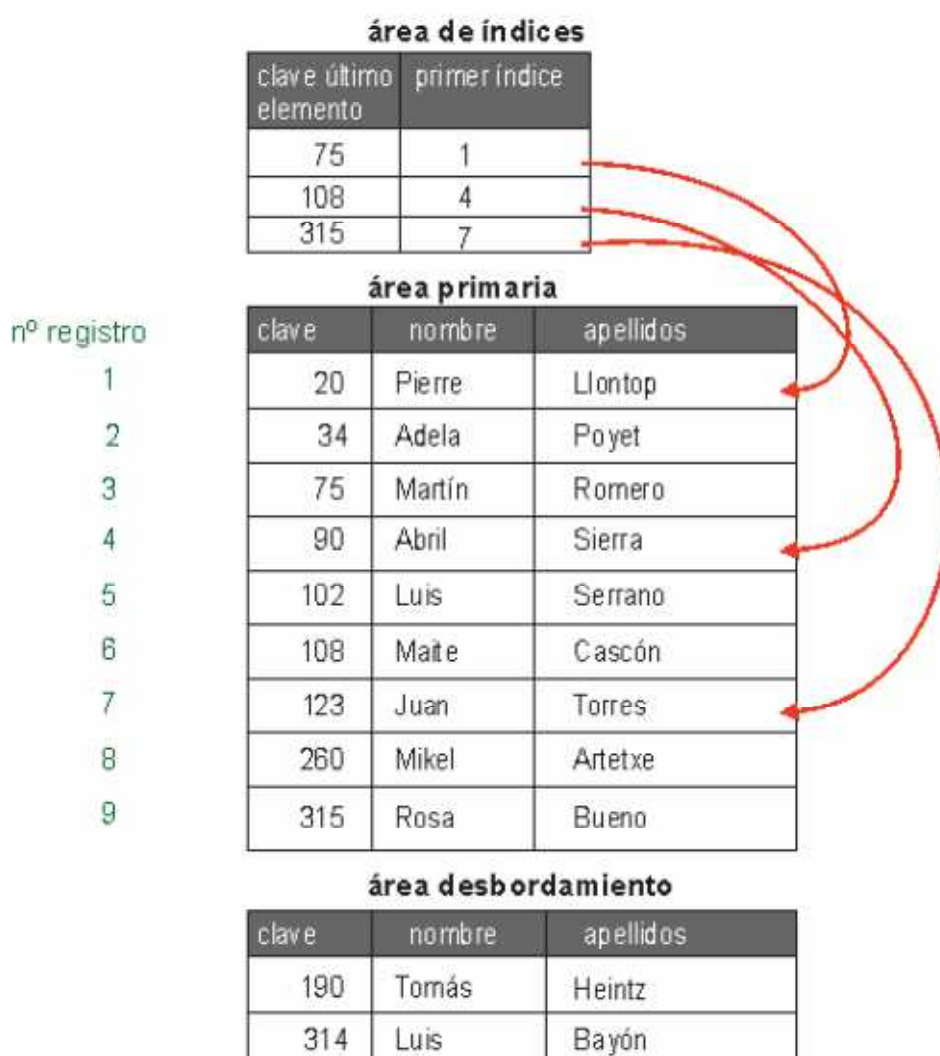
El archivo de índices posee unas cuantas entradas sólo en las que se indica la posición de ciertos valores claves en el archivo (cada 10, 15, 20, ... registros del archivo principal se añade una entrada en el de índices). El archivo principal tiene que estar siempre ordenado y así cuando se busca un registro, se busca su valor clave en la tabla de índices, la cual poseerá la posición del registro buscado. Desde esa posición se busca secuencialmente el registro hasta encontrarlo.

Existe un archivo llamado de **desbordamiento** u **overflow** en el que se colocan los nuevos registros que se van añadiendo (para no tener que ordenar el archivo principal cada vez que se añade un nuevo registro) este archivo está desordenado. Se utiliza sólo si se busca un registro y no se encuentra en el archivo principal. En ese caso se recorre todo el archivo de overflow hasta encontrarlo.

Para no tener demasiados archivos en overflow (lo que restaría velocidad), cada cierto tiempo se reorganiza el archivo principal.

Ejemplo:

Ficheros secuenciales indexados



Ventajas

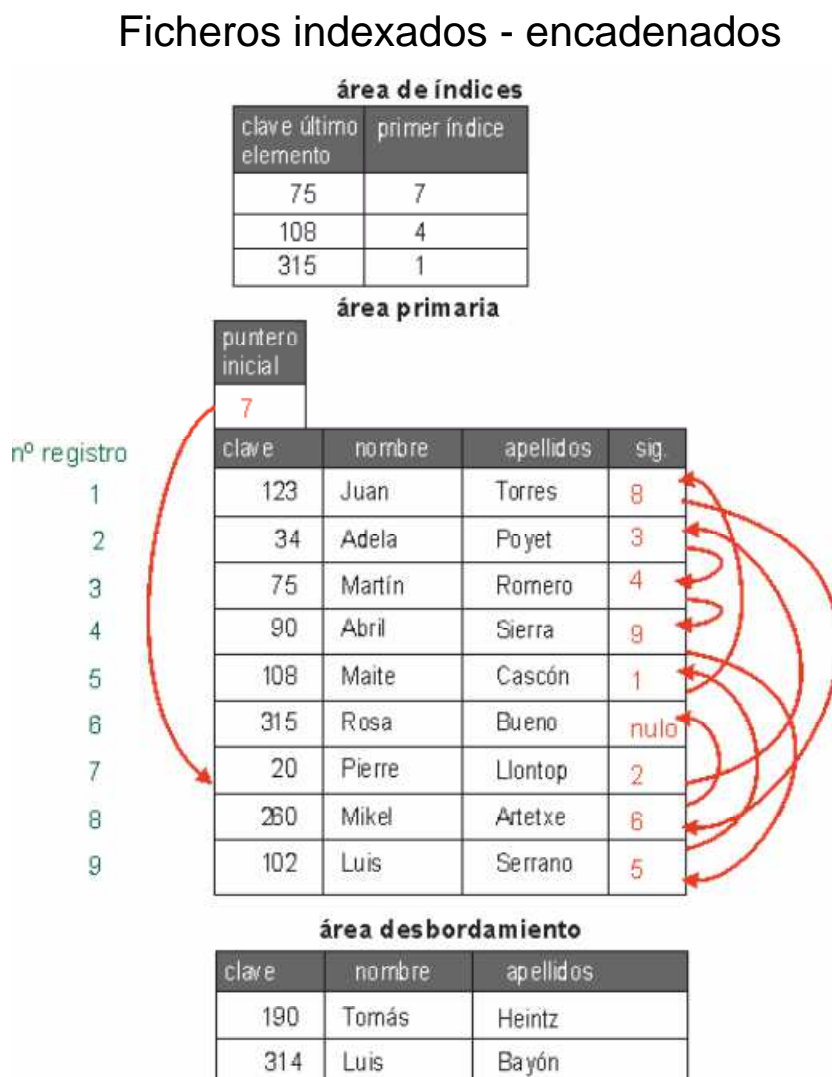
- ✓ El archivo está siempre ordenado en base a una clave
- ✓ La búsqueda de datos es rapidísima
- ✓ Permite la lectura secuencial (que además será en el orden de la clave)
- ✓ El borrado de registros es posible (aunque más problemático que en el caso anterior)

Desventajas

- ✓ Para un uso óptimo hay que reorganizar el archivo principal y esta operación es muy costosa ya que hay que reescribir de nuevo y de forma ordenada todo el archivo.
- ✓ La adición de registros requiere más tiempo que en los casos anteriores al tener que reordenar los índices

3.5 Ficheros indexado – encadenados.

Utiliza punteros e índices, es una variante encadenada del caso anterior. Hay un fichero de índices equivalente al comentado en el caso anterior y otro fichero de tipo encadenado con punteros a los siguientes registros. Cuando se añaden registros se añaden en un tercer registro llamado de desbordamiento u **overflow**. En ese archivo los datos se almacenan secuencialmente, se accede a ellos si se busca un dato y no se encuentra en la tabla de índices.



Ventajas

- ✓ Posee las mismas ventajas que los archivos secuenciales indexados, además de una mayor rapidez al reorganizar el fichero (sólo se modifican los punteros)

Desventajas

- ✓ Requieren compactar los datos a menudo para reorganizar índices y quitar el fichero de desbordamiento.

4. Operaciones relacionadas con uso de ficheros en bases de datos.

Borrado y recuperación de registros

Algunos de los tipos de ficheros vistos anteriormente no admiten el borrado real de datos, sino que sólo permiten añadir un dato que indica si el registro está borrado o no. Esto es interesante ya que permite anular una operación de borrado. Por ello esta técnica de marcar registros, se utiliza casi siempre en todos los tipos de archivos.

En otros casos los datos antes de ser eliminados del todo pasan a un fichero especial (conocido como **papelera**) en el que se mantienen durante cierto tiempo para su posible recuperación.

Fragmentación y compactación de datos

La fragmentación en un archivo hace referencia a la posibilidad de que éste tenga huecos interiores debido a borrado de datos u a otras causas. Causa los siguientes problemas:

- ✓ Mayor espacio de almacenamiento
- ✓ Lentitud en las operaciones de lectura y escritura del fichero

Por ello se requiere **compactar** los datos. Esta técnica permite eliminar los huecos interiores a un archivo. Las formas de realizarla son:

- ✓ **Reescribir el archivo para eliminar los huecos.** Es la mejor, pero lógicamente es la más lenta al requerir releer y reorganizar todo el contenido del fichero.
- ✓ **Aprovechar huecos.** De forma que los nuevos registros se inserten en esos huecos. Esta técnica suele requerir un paso previo para reorganizar esos huecos.

Compresión de datos

En muchos casos para ahorrar espacio de almacenamiento, se utilizan técnicas de compresión de datos. La ventaja es que los datos ocupan menos espacio y la desventaja es que al manipular los datos hay que descomprimirlos lo que hace que la manipulación de los datos sea lenta.

Cifrado de datos

Otra de las opciones habituales sobre ficheros de datos es utilizar técnicas de cifrado para proteger los ficheros en caso de que alguien no autorizado se haga con el fichero. Para leer un fichero de datos, haría falta descifrar el fichero. Para descifrar necesitamos una clave o bien aplicar métodos de descifrado; lógicamente cuanto mejor sea la técnica de cifrado, más difícil será descifrar los datos mediante la fuerza bruta.