

TEMA 1. ALMACENAMIENTO DE LA INFORMACIÓN (Parte I)

1. INTRODUCCIÓN

2. SISTEMAS CLÁSICOS DE GESTIÓN DE FICHEROS

2.1. CONCEPTOS BÁSICOS

2.2. CLASIFICACIÓN DE LOS FICHEROS SEGÚN SU ORGANIZACIÓN

2.3. MODOS DE ACCESOS DE LOS FICHEROS SEGÚN SU ORGANIZACIÓN

2.4. CLASIFICACIÓN DE LOS FICHEROS SEGÚN SU USO

2.5. CLASIFICACIÓN DE LOS FICHEROS EN LA ACTUALIDAD

3. SISTEMAS GESTORES DE BASES DE DATOS

1. INTRODUCCIÓN

El **almacenamiento y manejo** de grandes cantidades de **datos** se hace necesario **en cualquier empresa para el logro de sus objetivos**. Por ejemplo, se necesitan los datos de los empleados, de los clientes, de los proveedores, de los productos almacenados, etc.

Normalmente la gestión de estos datos **se ha venido realizando de forma manual**. Se organizaban en forma de **fichas**, informes o expedientes, colocándolos en **carpetas** y almacenando estas en un **archivador**. Por ejemplo, se tiene un archivo de clientes en el que cada ficha contiene todos los datos correspondientes a un cliente. Cuando se necesita consultar o modificar los datos de un cliente concreto será necesario realizar la operación manualmente.

La utilización de los **ordenadores en la administración de las empresas** supuso una **revolución** respecto al almacenamiento y gestión de sus datos dando lugar a los denominados: **ficheros de datos informáticos y posteriormente a las bases de datos**.

Las ventajas se derivan de las características de los soportes y de los ordenadores:

- ✓ Gran **capacidad** de almacenamiento en un reducido espacio.(tamaño)
- ✓ **Rapidez** en el proceso de los datos. (velocidad)
- ✓ **Precisión** de los resultados obtenidos del proceso. (calidad)

Concretando, podemos decir que existen dos formas **básicas de almacenar los datos**:

Convencional. Se utilizan los ficheros planos y por tanto los Sistemas Clásicos de Gestión de Ficheros.

No Convencional. Se utilizan las Bases de datos y por consiguiente los Sistemas Gestores de Bases de Datos.

2. SISTEMAS CLÁSICOS DE GESTIÓN DE FICHEROS

2.1. CONCEPTOS BÁSICOS.

Un **archivo** o **fichero** de datos es una estructura de datos que reside en memoria secundaria, consistente en un conjunto de informaciones estructuradas en unidades de acceso denominadas **registros**, todos del mismo tipo y en número indeterminado.

Un **registro lógico** o simplemente registro, es cada uno de los componentes del archivo, conteniendo un conjunto de informaciones a la que se acceden y se tratan de manera unitaria. Está constituido por uno o más elementos denominadas **campos** que pueden ser de diferentes tipos.

Un registro tiene un **campo clave**, cuyo valor sirve para identificar de forma única el registro, y por tanto, dicho valor no puede aparecer repetido en otro registro diferente

Si un fichero contiene la información de un conjunto de individuos u objetos, entonces sus registros contienen la información de cada uno de ellos y los campos los datos de esta.

Un **registro físico** o **bloque** corresponde a la cantidad de información que se transfiere físicamente en cada operación de acceso (lectura o escritura).

El tamaño y formato del **registro lógico** los define el **programador**, mientras que el tamaño del registro físico viene dado por las características **físicas del ordenador**.

2.2. CLASIFICACIÓN DE LOS FICHEROS SEGÚN SU ORGANIZACIÓN

Al diseñar un archivo, **dependiendo del uso** que se va a hacer del mismo y del **soporte utilizado**, se pueden elegir diferentes maneras de organizar sus registros, siendo sus principales organizaciones las siguientes:

- ✓ Secuencial
- ✓ Directa o Aleatoria
- ✓ Secuencial Indexada

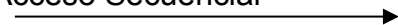
Organización Secuencial.

Es aquella en la cual los registros ocupan posiciones consecutivas de memoria, y sólo se puede acceder a ellos de uno en uno a partir de primero.

No se pueden hacer operaciones de escritura cuando se está leyendo, ni operaciones de lectura cuando se está escribiendo.

Para actualizarlos es preciso crear nuevos archivos donde se copien los registros que vayan a permanecer, modificados o no junto con los nuevos.

Acceso Secuencial



Reg 1º	Reg 2º	Reg 3º
--------	--------	--------	-------

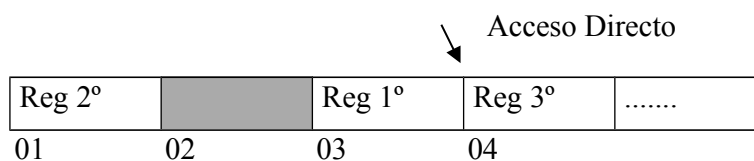
Organización Directa o Aleatoria

Esta organización también se denomina **relativa**. En estos archivos las informaciones se colocan y se acceden aleatoriamente mediante su posición, es decir, indicando el lugar relativo que ocupan dentro del conjunto de posiciones posibles.

En esta organización se pueden leer y escribir registros, en cualquier orden y cualquier lugar.

Presenta el *inconveniente* de que es tarea del programador establecer la relación entre la posición que ocupa el registro y su contenido. Esto se lleva a cabo mediante una función de aleatorización o hashing **(1)**. Además puede desaprovecharse parte del espacio destinado al archivo, ya que pueden quedar huecos libres entre unos registros y otros.

Su principal *ventaja* es su rapidez de acceso a un registro cualquiera, ya que para ello no es necesario pasar por los anteriores.



Organización Secuencial Indexada.

Un archivo de esta organización consta de tres áreas:

Área de Índices (fichero de índices).

Área Primaria (fichero de datos)

Área de Excedentes (fichero de excedentes u overflow)

El **Área Primaria** contendrá los registros de datos, clasificados en orden ascendentes por su clave.

El **Área de Índices** es un **archivo secuencial creado por el sistema**, en el que cada registro establece una división (segmento) en el área primaria, y contiene la dirección del comienzo del segmento y la clave más alta del mismo. De esta manera el sistema accede de forma directa a un segmento del área primaria a partir del área de índices, de forma similar a la búsqueda de un capítulo de un libro a partir de su índice.

Por último se reserva un espacio, llamado **área de excedentes**. Para añadir nuevos registros que no pueden ser colocados en el área primaria cuando se produce una actualización del archivo.

Ventajas:

Rápido acceso por medio de la clave del registro

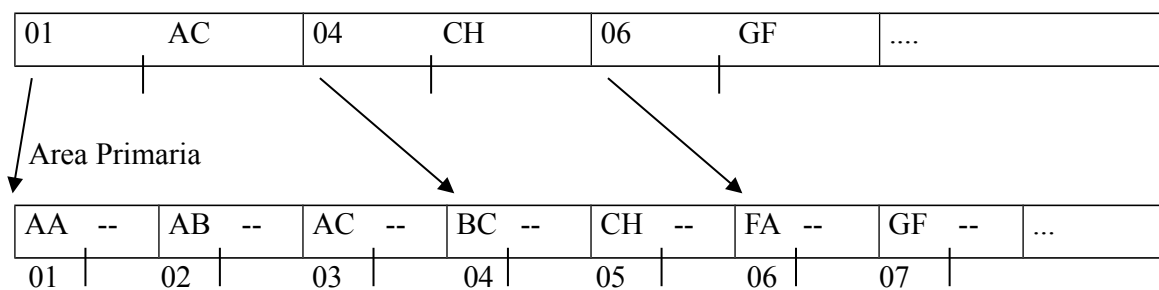
El sistema se encarga de relacionar la posición de cada registro con su contenido por medio del área de índices.

También es trabajo del sistema la gestión del área de índices y excedentes.

Inconvenientes:

Necesidad de espacio adicional para el área de índices

Área de Índices



Área de Excedentes

FH	--	AN	--
----	----	----	----	-------

2.3. MODOS DE ACCESOS DE LOS FICHEROS SEGÚN SU ORGANIZACIÓN

Se denomina **modo de acceso** a la forma en que un dispositivo que maneja un soporte de información, que contiene un archivo, se posiciona en un determinado lugar del mismo para realizar una operación de lectura o de escritura de un registro.

El modo de acceso lo decide el **programador** de la aplicación en función del tipo de organización.

Hay dos modos básicos: Secuencial y Directo.

Acceso Secuencial. Supone acceder inicialmente al primer registro del archivo y después, consecutivamente, a todos los sucesivos hasta llegar al registro deseado. Este modo de acceso se puede utilizar con cualquier organización.

Acceso directo. Consiste en el posicionamiento sobre cualquier registro sin necesidad de haber accedido antes a los anteriores.

Este modo de acceso se puede utilizar con la organización directa y con la indexada.

En los *archivos de organización directa*, este acceso se consigue proporcionando al dispositivo la posición del registro que se desea acceder. Normalmente **se programa una función de aleatorización o hashing**, la cual permite calcular la posición de cualquier registro a partir del valor de su clave.

En los *archivos indexados*, **el acceso directo lo realiza de forma automática el sistema a partir del valor del campo clave**; para ello utiliza, como se ha descrito en el apartado anterior, un conjunto auxiliar de índices que relacionan las claves con las posiciones de los registros.

(1) El programador puede utilizar una función de direccionamiento calculado (función de aleatorización o hashing) para calcular la posición del registro, a partir del valor de su clave, donde se va a leer o escribir.

Ejemplos:

Si la clave es de tipo entero

$H(K)=K$. La clave es del tipo 1, 2, 3, etc. y queremos almacenar de forma consecutiva.

$H(K)=K \text{ MOD } M$. Siendo M el número primo inmediatamente superior al número de registros.

Si la clave es alfanumérico

\sum (todos los ordinales) MOD M. Siendo M el número primo inmediatamente superior al número de registros.

En el caso de que la función nos de cómo resultado más de una vez la misma posición el programador debe resolver colisiones.

2.4. CLASIFICACIÓN DE LOS FICHEROS SEGÚN SU USO

La clasificación según el uso de los ficheros, es solo para los ficheros de organización secuencial. Los archivos se clasificaban según la utilización que se hace de ellos en tres grandes grupos:

ARCHIVOS PERMANENTES. Contiene información que varia poco a lo largo del tiempo. Pueden ser de tres clases:

- ✓ **Archivos Constantes**. Su información permanece prácticamente inamovible, utilizándose principalmente como **archivos de consulta**.

Ejemplo: Un archivo de este tipo puede ser el de la red de metro de una ciudad, que contiene la descripción, características, número de estaciones, numero de trenes, etc., de cada línea.

- ✓ **Archivos de Situación**. También denominados **archivos maestros**, contienen la información que **refleja el estado o situación de una empresa** en un determinado momento. Estos archivos **se actualizan periódicamente** para adaptarlo a cada nueva situación.

Ejemplo: Un archivo de personal con contrato temporal en una empresa, o el archivos de existencias de un almacén.

- ✓ **Archivos Históricos.** Se obtienen de los anteriores cuando se dejan fuera de uso para futuros estudios estadísticos o consultas.

Ejemplo: El archivo que contiene la información de libros adquiridos por una biblioteca en la década de los ochenta.

ARCHIVOS DE MOVIMIENTO. En ellos se almacena la información que **se utilizaban para actualizar los archivos maestros.** Sus registros, denominados movimientos o transacciones, son de tres clases: altas, baja y modificaciones.

Una vez realizado el proceso de actualización de un archivo de maestro por medio de un archivo de movimientos, éste pierde su validez y podemos deshacernos de él

Ejemplo: Un archivo de este tipo para actualizar el antes mencionado de personal contratado, es el que refleja las nuevas contrataciones, finalizaciones de contrato y modificaciones de los mismos producidas en la empresa durante el mes actual.

ARCHIVOS DE MANIOBRA O TRABAJO. Tienen una vida limitada, normalmente igual a la duración de la ejecución de un programa, y **se utilizan como auxiliares de los anteriores.**

Ejemplo: Se desea una lista alfabética de los nombres del personal contratado, se hará por medio de un archivo de maniobra en el que se almacene esta información a partir del archivo de personal. Este archivo desaparecerá una vez que tenga la lista impresa.

2.5. CLASIFICACIÓN DE LOS FICHEROS EN LA ACTUALIDAD

Las clasificaciones sobre los ficheros estudiadas anteriormente (clasificación según el uso o utilidad y según la organización) han quedado actualmente en desuso debido a la aparición de las bases de datos. No obstante, en este tema las estudiamos pues **a nivel físico una base de datos está formada por uno o más ficheros.**

Actualmente los ficheros se clasifican:

- Según su contenido (texto o datos binarios).
- Según su tipo (ejecutables, videos, imágenes, etc.)

Veamos la clasificación según su contenido:

Ficheros de Texto

Son ficheros que almacenan texto y se guardan en código ASCII (American Estándar Code for information Interchange).

Suelen tener extensión:

Los ficheros de configuración: .inf, .conf.

Los ficheros con código fuente : .c, .sql, .java

Los ficheros de página web: .html, .php

Ficheros Binarios

Son ficheros binarios todos los ficheros que no son de textos y requieren un formato para ser interpretados.

Tienen extensiones como:

Los ficheros de imagen: .jpg, .gif, bmp.

Los ficheros de videos: .mpg, .avi.

Los ficheros comprimidos: .zip, tar.

Los ficheros ejecutables: .exe, .com, .cgi.

Los ficheros que componen una base de datos son binarios puesto que la información que almacenan requieren una cierta estructura y se haría muy difícil almacenarla en ficheros de textos.