

ESCUELA DE PREPARACIÓN DE OPOSITORES
E. P. O.

C/. La Merced, 8 – Bajo A Telf.: 968 24 85 54
30001 MURCIA

INF14 - SAI14

Utilización de ficheros según su organización.

Esquema.

1	INTRODUCCIÓN.....	1
2	CONCEPTOS BÁSICOS.....	3
2.1	ALMACENAMIENTO DE LA INFORMACIÓN.....	3
2.2	EL GESTOR DE FICHEROS.....	3
2.3	RENDIMIENTO DE LA TRANSFERENCIA DE LA INFORMACIÓN.....	3
3	TIPOS DE ARCHIVOS.....	3
3.1	ARCHIVOS PERMANENTES.....	3
3.2	ARCHIVOS TEMPORALES.....	3
4	OPERACIONES SOBRE ARCHIVOS.....	3
4.1	CREACIÓN DE UN ARCHIVO.....	4
4.2	CONSULTA DE UN ARCHIVO.....	4
4.3	ACTUALIZACIÓN DE UN ARCHIVO.....	4
4.4	CLASIFICACIÓN DE UN ARCHIVO.....	5
4.5	REORGANIZACIÓN DE UN ARCHIVO.....	5
4.6	DESTRUCCIÓN DE UN ARCHIVO.....	6
4.7	REUNIÓN, FUSIÓN DE UN ARCHIVO.....	6
4.8	ROTURA/ESTALLIDO DE UN ARCHIVO.....	6
5	OPERACIONES SEGÚN ORGANIZACIONES.....	6
5.1	ORGANIZACIONES CONSECUTIVAS.....	7
5.2	ORGANIZACIONES DIRECCIONADAS.....	9
5.2.1	<i>Métodos de direccionamiento.....</i>	<i>10</i>
	Direccionadas directas.....	10
	Direccionadas dispersas.....	10
5.2.2	<i>Algoritmos de direccionamiento.....</i>	<i>10</i>
	Direccionamiento directo.....	10
	Direccionamiento asociado.....	10
	Direccionamiento calculado (hashing).....	10
5.2.3	<i>Operaciones básicas sobre un archivo con organización directa.....</i>	<i>10</i>
5.3	ORGANIZACIONES INDEXADAS.....	11
6	PARÁMETROS DE UTILIZACIÓN DE UN ARCHIVO.....	14
7	CONCLUSIONES.....	15

1 Introducción.

Los sistemas informáticos requieren, para la manipulación de la información, que ésta se encuentre previamente almacenada en alguno de los dispositivos de almacenamiento que estos sistemas reconocen. Si esta información puede ser obtenida

en tiempo de computación, bien mediante un cálculo operacional, o bien mediante la entrada desde un terminal de entrada/salida y, este proceso tiene un bajo coste computacional y, además, puede ser reproducido, esta información puede ser almacenada en la memoria central de la computadora con el conocimiento de que será perdida una vez el proceso que manipula esa información abandone la memoria central de la computadora.

Pero son muy pocos los problemas en los que el dominio de la información satisface estas características. En la mayoría de los casos el volumen de información es tan elevado, o los procesos de cómputo para obtenerla son tan costosos, que el alimentar a los procesos de manipulación de esa información con ésta en cada una de sus ejecuciones resulta algo inconcebible. Es en estos casos cuando el usuario de los sistemas informáticos requiere de los ficheros para el almacenamiento de la información que forma parte del dominio del problema que está tratando con un sistema informático.

Un fichero es una agregación de datos estructurados que es vista por el sistema operativo de la computadora como un todo, como una unidad estructural sobre la cual se pueden realizar una serie de operaciones dependiendo del dispositivo físico en el cual se encuentre y de la propia estructura interna de los datos en el mismo.

El conjunto de datos que forman parte de los ficheros se agrupan en unidades denominadas registros cuya función es representar a unidades de información del problema que se está tratando. El número y estructura de los datos que forman parte de los registros es determinado en la fase de análisis y diseño lógico del problema, mientras que la disposición de los registros y los datos en los ficheros (su estructura) es determinada en el diseño físico.

La estructura de los ficheros, la disposición de la información en los mismos, va a estar delimitada por el dispositivo físico en el cual se encuentren, de forma que un mismo dominio de información podrá almacenarse con estructuras diferentes dependiendo del dispositivo en el cual se encuentre y, en función, por supuesto, de los requerimientos funcionales exigidos a los procedimientos que la manipulan.

Existen, por tanto, múltiples estructuras diferentes en las cuales puede almacenarse la información dando lugar a lo que se conoce como organizaciones de la información y, en base a esas organizaciones los procedimientos que manipulan esa información son diferentes.

La vida de todo archivo comienza cuando se crea y acaba cuando se borra. Durante la vida del archivo se suelen realizar sobre él determinadas operaciones de recuperación o consulta y de mantenimiento o actualización. Estas operaciones las realizan programas específicos, que actúan a nivel de registro. Las operaciones de actualización que se pueden realizar sobre el archivo son:

- Modificación de un registro
- Eliminación o borrado de un registro.
- Inserción de un registro.

La mayor parte de las operaciones de recuperación y actualización implican la realización de una localización o búsqueda de un registro concreto para luego actuar sobre él (leerlo, escribir o cambiar parte de él, borrarlo, etc.).

2 Conceptos básicos.

Un fichero es una agregación de datos estructurados que es vista por el sistema operativo de la computadora como un todo, como una unidad estructural sobre la cual se pueden realizar una serie de operaciones dependiendo del dispositivo físico en el cual se encuentre y de la propia estructura interna de los datos en el mismo.

El conjunto de datos que forman parte de los ficheros se agrupan en unidades denominadas registros cuya función es representar a unidades de información del problema que se está tratando.

2.1 Almacenamiento de la información.

En este apartado debe explicarse el almacenamiento de la información que hemos estudiado en profundidad en el tema 13, página 2, apartado 2.1.

2.2 El gestor de ficheros.

En este apartado debe explicarse el gestor de ficheros que hemos estudiado en profundidad en el tema 13, página 3, apartado 2.2.

2.3 Rendimiento de la transferencia de la información.

En este apartado debe explicarse el rendimiento de la transferencia de la información que hemos estudiado en profundidad en el tema 13, página 4, apartado 2.3.

3 Tipos de archivos.

En una aplicación informática se pueden utilizar archivos para realizar funciones diversas. Conocer la función que va a desempeñar un archivo concreto es fundamental a la hora de decidir como se debe organizar éste. Podemos clasificar los archivos como archivos permanentes y archivos temporales.

3.1 Archivos permanentes.

En este apartado deben explicarse los archivos permanentes que hemos estudiado en profundidad en el tema 13, página 5, apartado 3.1.

3.2 Archivos temporales.

En este apartado deben explicarse los archivos temporales que hemos estudiado en profundidad en el tema 13, página 5, apartado 3.2.

4 Operaciones sobre archivos.

Las distintas operaciones que se pueden realizar son:

- Creación.
- Consulta.
- Actualización.
- Clasificación.
- Reorganización.
- Destrucción.
- Reunión, fusión.

- Rotura, estallido.

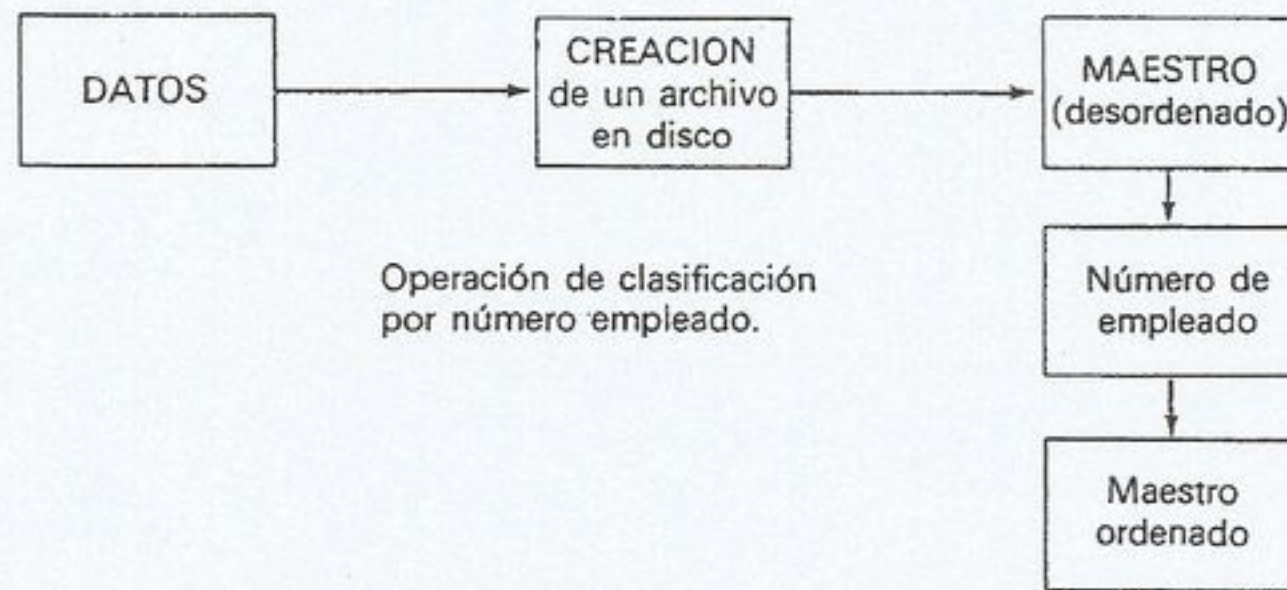
4.1 Creación de un archivo.

Es la primera operación que sufrirá el archivo de datos. Implica la elección de un entorno descriptivo que permita un ágil, rápido y eficaz tratamiento del archivo.

Para utilizar un archivo, éste tiene que existir, es decir, las informaciones de este archivo tienen que haber sido almacenados sobre un soporte y ser utilizables. La creación exige organización, estructura, localizar o reservar espacio en el soporte de almacenamiento, transferencia del archivo del soporte antiguo al nuevo.

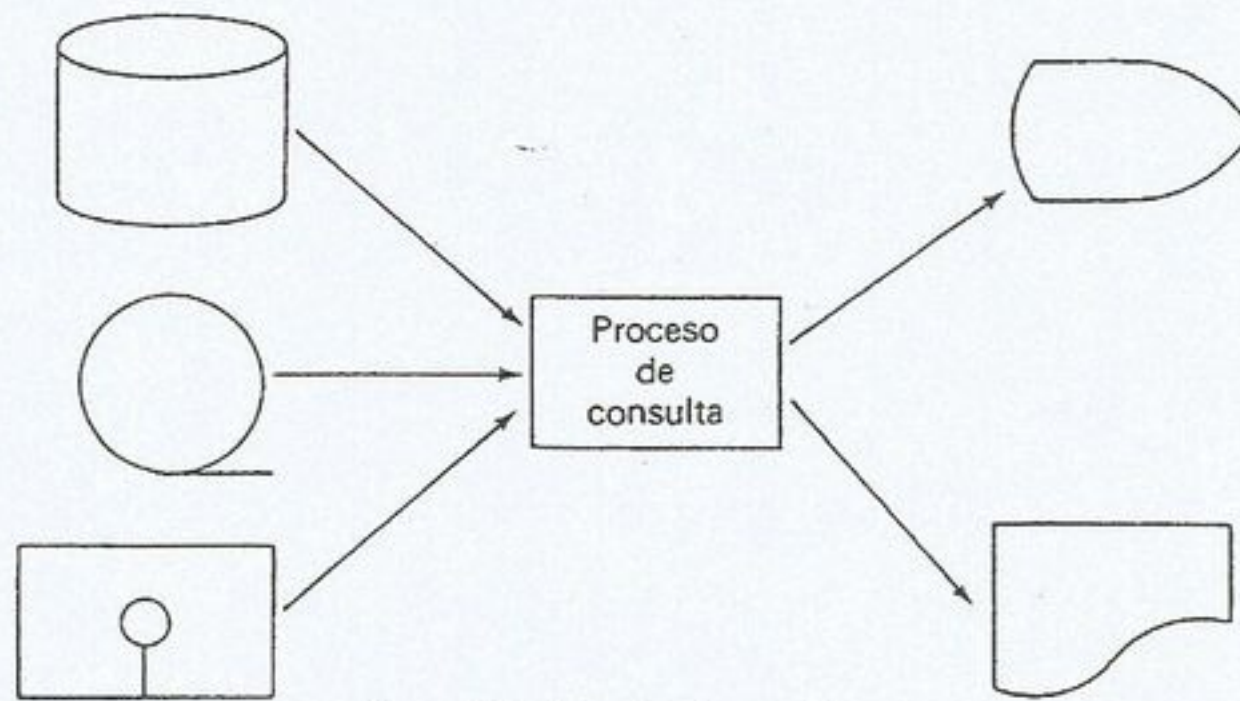
Un archivo puede ser creado por primera vez en un soporte, proceder de otro previamente existente en el mismo o diferente soporte, ser el resultado de un cálculo o ambas cosas a la vez.

La siguiente figura muestra un organigrama de la creación de un archivo ordenado.



4.2 Consulta de un archivo.

Es la operación que permite al usuario acceder al archivo de datos para conocer el contenido de uno, varios o todos los requisitos.

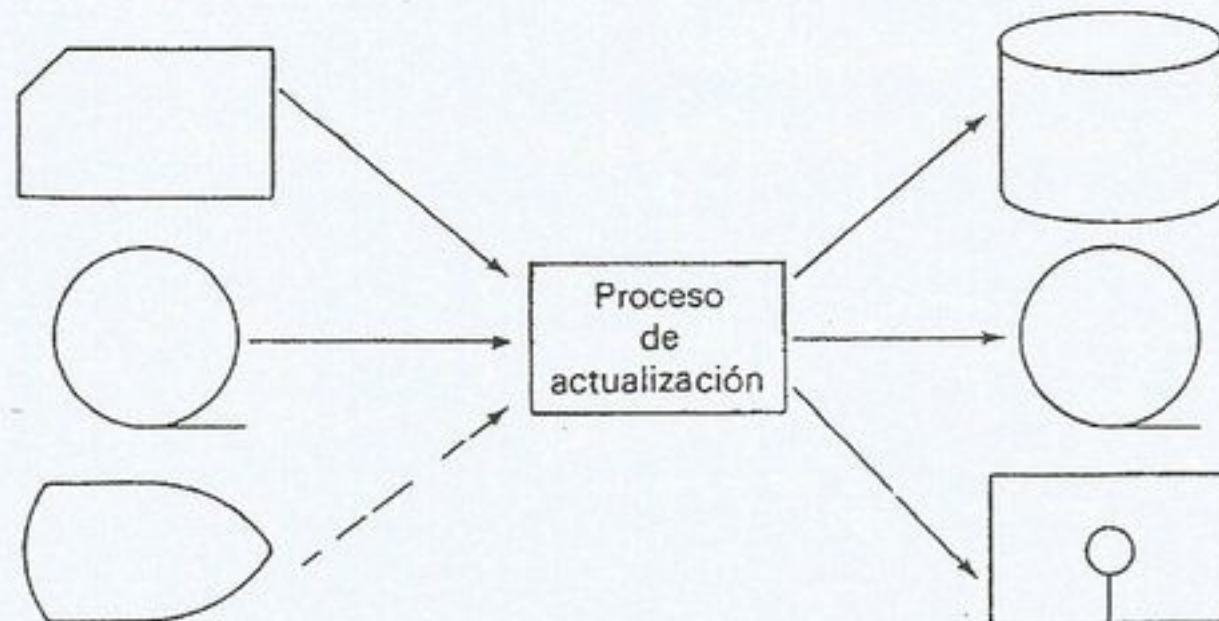


4.3 Actualización de un archivo.

Es la operación que permite tener actualizado (puesto al día) el archivo, de tal modo que sea posible realizar las siguientes operaciones con sus registros:

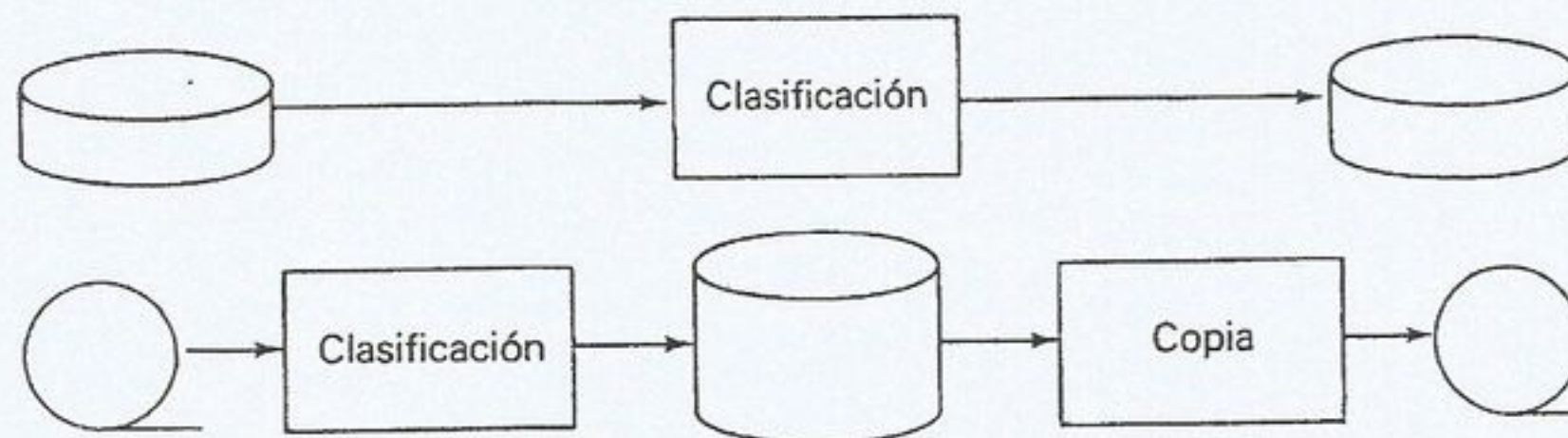
- Consulta del contenido de un registro.

- Inserción de un registro nuevo en el archivo.
- Supresión de un registro existente.
- Modificación de un registro.



4.4 Clasificación de un archivo.

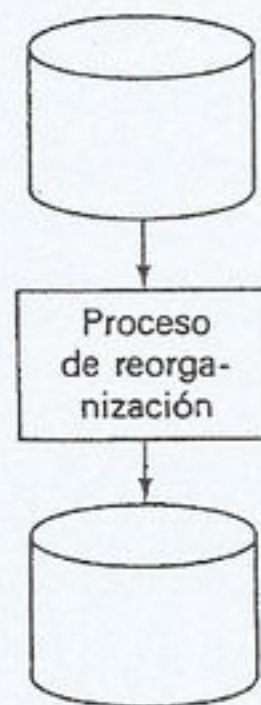
Una operación muy importante en un archivo es la clasificación u ordenación. Esta clasificación se realizará de acuerdo con el valor de un campo específico, pudiendo ser ascendente o descendente: alfabética o numérica.



4.5 Reorganización de un archivo.

Las operaciones sobre archivos modifican la estructura inicial o la óptima de un archivo. Los índices, enlaces (punteros), zonas de sinónimos, zonas de desbordamiento, etc., se modifican con el paso del tiempo, lo que hace a la operación de acceso al registro cada vez más lenta.

La reorganización suele consistir en la copia de un nuevo archivo a partir del archivo modificado, a fin de obtener una nueva estructura lo más óptima posible.

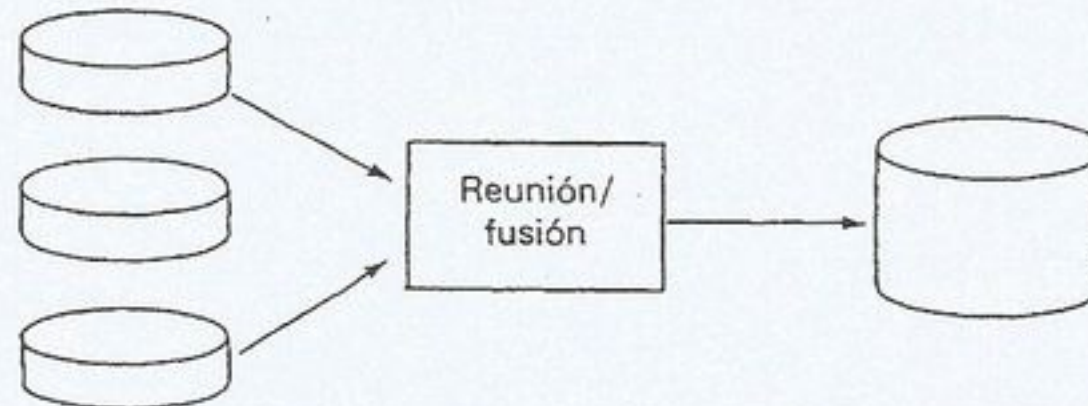


4.6 Destrucción de un archivo.

Es la operación inversa a la creación de un archivo. Cuando se destruye un archivo, éste ya no se puede utilizar y por consiguiente no se podrá acceder a ninguno de sus registros.

4.7 Reunión, fusión de un archivo.

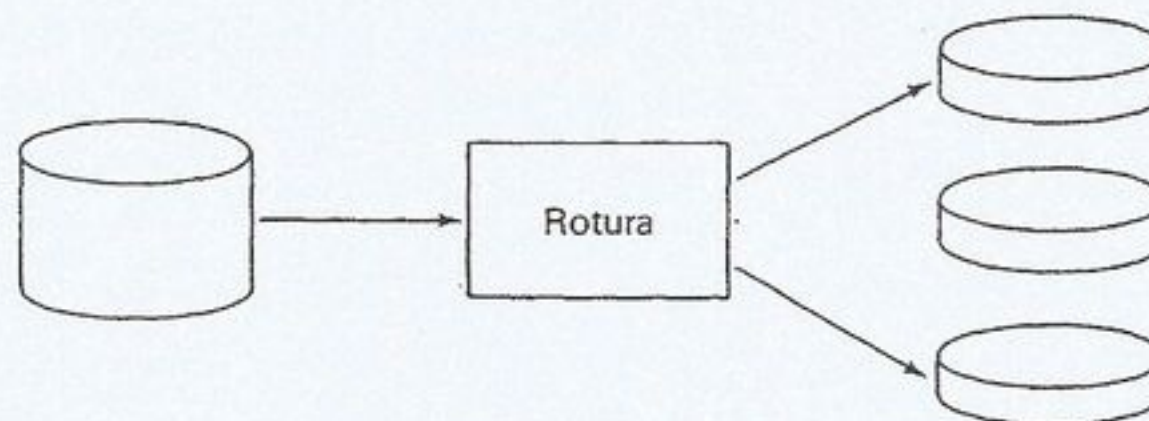
La reunión permite obtener un archivo a partir de otros varios.



Se realiza una fusión cuando se reúnen varios archivos en uno solo, intercalándose unos en otros, siguiendo unos criterios determinados.

4.8 Rotura/estallido de un archivo.

Es la operación de obtener varios archivos a partir de un mismo archivo inicial.



5 Operaciones según organizaciones.

La organización de un fichero determina el modo de estructurar los registros pertenecientes a un fichero en almacenamiento secundario. Los métodos de acceso nos permitirán localizar dichos registros. La elección de una determinada organización permitirá realizar distintos tipos de acceso a los registros del fichero y, viceversa, si se desea acceder a un registro o conjunto de registros de una determinada forma, se deberá elegir una organización adecuada para ello.

La organización de los ficheros y los métodos de acceso repercuten fuertemente en los tiempos de respuesta. Los métodos básicos de acceso son los siguientes:

- Acceso secuencial. Para localizar un registro se necesita haber accedido al registro anterior. Puede resultar el modo de acceso más eficiente cuando se tiene que procesar el fichero completo (o un porcentaje del mismo).
- Acceso directo. Se localiza un registro por su dirección, obtenida a partir del valor de una clave de direccionamiento o por la posición relativa que ocupa el registro en el fichero.

5.1 Organizaciones consecutivas.

Las organizaciones consecutivas sitúan los registros físicamente uno a continuación de otro. Si la colocación de los registros no tiene un orden lógico, el tipo de organización se denomina consecutiva serial; si, por el contrario, tienen un orden, según una clave de ordenación, se denomina consecutiva secuencial.

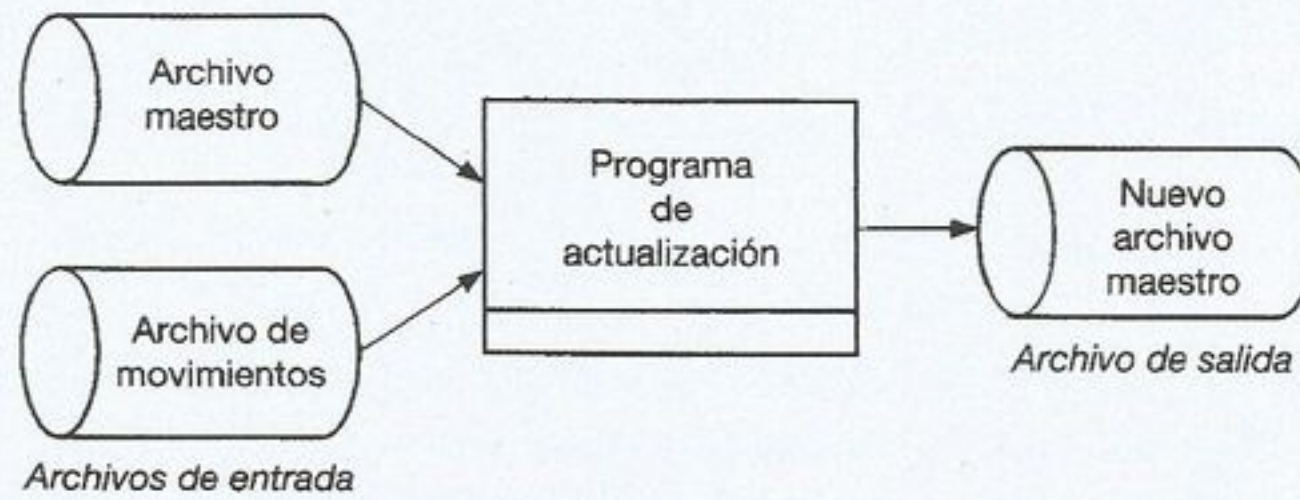
El método de acceso que se utiliza para localizar los registros almacenados es el secuencial si la organización es serial; si es secuencial puede utilizarse, además, un acceso directo por posición relativa del registro (búsqueda binaria o dicotómica).

Es adecuada para recuperar ficheros en procesos por lotes, pues optimiza al máximo el espacio de almacenamiento, es muy fácil de programar, puede utilizar registros de longitud variable y heterogéneos, y no está limitada al tipo de soporte directo. Es conveniente utilizar este tipo de organización cuando existe una carga masiva de datos, las tablas son pequeñas, o cuando, en el proceso del fichero se accede normalmente a casi todas las filas.

Esta es la única organización de archivos susceptible de ser gestionada en un dispositivo no direccionable (soportes secuenciales tales como cinta magnética, cinta de papel o tarjeta perforada). La secuencia en que aparecen los registros en el archivo puede estar determinada por el valor de algún campo o ser simplemente temporal.

Las distintas operaciones o acciones que se pueden realizar sobre archivos con esta organización son:

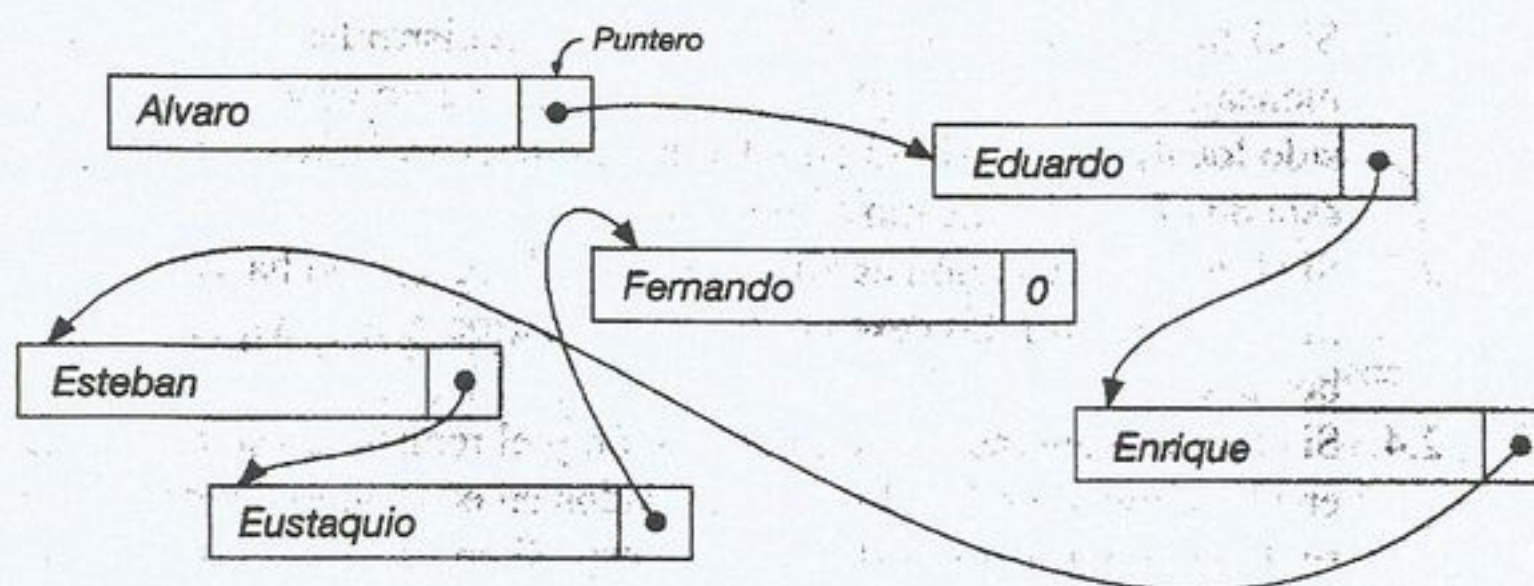
- Añadir. Sólo es posible escribir al final del archivo. La información se graba en el archivo escribiendo los registros uno a continuación de otro, en el orden en que se desea que estén en el archivo.
- Consulta o recuperación. La consulta se realiza en orden secuencial. Esto es, para leer el registro que ocupa la posición n en el archivo es necesario leer previamente los $n-1$ anteriores.
- Inserción, modificación y eliminación. No es posible realizar fácilmente ninguna de estas operaciones sobre un archivo secuencial. Si se necesita actualizar un archivo con organización secuencial se debe crear de nuevo el archivo. Esto se realiza por medio de un programa, escrito para tal fin, que utilizará la versión a modificar del archivo y un archivo intermedio, también secuencial, de los que leerá registros que escribirá en el nuevo archivo. Para esto es necesario que los registros aparezcan en el archivo ordenados según el valor de algún campo. Las modificaciones a realizar sobre el archivo se almacenarán previamente en el archivo intermedio antes mencionado (llamado usualmente archivo de movimientos) en la misma secuencia en que aparecen los registros en el archivo a modificar. Los registros del archivo de movimientos tienen la misma estructura que los del archivo a actualizar, más un campo, que normalmente se coloca al principio del registro, que contiene el código de la operación a realizar con el registro: modificar, eliminar o insertar.



Si el archivo se encuentra sobre un soporte direccionable (disco por ejemplo) es posible realizar algunas actualizaciones sobre el archivo secuencial, sin necesidad de crear otro archivo maestro. En este supuesto se podrán realizar las siguientes operaciones sobre el archivo, además de las ya descritas:

- **Modificación.** Una vez localizado un registro se puede reescribir éste en el propio archivo, siempre que al modificar el registro no aumente su longitud.
- **Borrado.** No es posible eliminar un registro del archivo. No obstante, es posible marcarlo de tal forma que al leer se identifique como no válido, esto es lo que se llama borrado lógico. El registro se marca reescribiéndolo con un valor determinado en un campo especial que indica si el registro está borrado o bien con un valor no válido en un campo de información del registro. En cualquier caso, el registro sigue estando en el archivo y por tanto, ocupa espacio.
- **Consulta.** Si los registros son de longitud fija es posible determinar la dirección de comienzo de cada uno a partir de su posición relativa en el archivo. Se puede, pues, acceder a un registro conociendo su posición relativa en el archivo. En este caso, y si el archivo está ordenado según el valor de algún campo, es posible consultar un registro conocido el valor del campo que se usa como clave, sin necesidad de leer el archivo secuencialmente. Para ello se deben usar algoritmos de búsqueda especiales.

Si el archivo se encuentra sobre un soporte direccionable (disco por ejemplo) es posible almacenar junto a cada registro un puntero con la dirección del registro siguiente, según el orden lógico del archivo. Estructuralmente un archivo secuencial encadenado es equivalente a una lista lineal de registros.



Las distintas operaciones se realizan sobre el archivo del siguiente modo:

- **Recuperación o consulta.** La consulta es secuencial, al igual que en un archivo con organización secuencial pura. Cada vez que se lee un registro se

lee la posición del siguiente, lo que permite seguir la secuencia lógica del archivo.

- **Inserción.** Para insertar un registro es necesario, en primer lugar, localizar la posición en que se desea insertar; esto es, entre qué dos registros se quiere que aparezca al leer el archivo. Físicamente el registro se escribe en una zona libre, con el mismo valor de puntero que el registro que le va a preceder. Por último, se modifica el registro anterior para actualizar el valor de su puntero, de forma que contenga la dirección del nuevo registro.



- **Borrado.** Un registro deja de estar en la secuencia de lectura del archivo cuando se elimina su dirección del puntero del registro anterior. Para borrar un registro se asigna al puntero del registro anterior la dirección del registro siguiente al que se desea borrar. El sistema operativo puede o no liberar el espacio ocupado por el registro. En el primer caso, dicho espacio puede ser ocupado por información perteneciente a otro archivo. En ambos casos, ese espacio podrá ser ocupado por otro registro del mismo archivo.
- **Modificación.** Si la modificación no implica un aumento de longitud del registro, éste puede reescribirse en el mismo espacio. En el caso de que el registro aumente de longitud se debe insertar el registro y, posteriormente, borrar la versión anterior a la modificación.

Si la longitud de los registros es mucho menor que el tamaño del bloque del dispositivo (factor de bloqueo mayor que uno), en cada acceso a disco se podrán leer o escribir varios registros. Si en cada nodo de la cadena hay un solo registro se ralentiza el acceso, ya que la lectura de cada registro implicará un acceso al dispositivo. Para reducir el tiempo de acceso se deben almacenar en cada bloque tantos registros como sea posible, encadenando los bloques, en lugar de encadenar los registros. Los registros contenidos en cada bloque serán consecutivos según el orden lógico del archivo.

5.2 Organizaciones direccionadas.

Este tipo de organización almacena los registros según una relación establecida entre el valor de la clave de direccionamiento y la dirección física del registro, siendo posible realizar un acceso directo a un registro si se conoce la clave de direccionamiento, aunque puede impedir (o al menos dificultar) el acceso secuencial.

La ubicación de un registro se determinará aplicando un algoritmo de transformación a la clave de direccionamiento, lo que dará como resultado la dirección base donde se debería almacenar dicho registro. La dirección base estará comprendida dentro de un rango de valores, que representan las direcciones relativas, denominado

espacio de direccionamiento, el cual se fija previamente en función del volumen de datos que hay que almacenar. La organización direccionada resulta muy adecuada en procesos de tipo selectivo donde hay que acceder a un registro por el valor de la clave de direccionamiento o por un rango de la misma.

5.2.1 Métodos de direccionamiento.

Dentro de este tipo de organización se puede distinguir entre organizaciones direccionadas directas y dispersas.

Direccionadas directas.

En este apartado debe explicarse la organización direccionada directa que hemos estudiado en profundidad en el tema 13, página 8, apartado 4.2.1.1.

Direccionadas dispersas.

En este apartado debe explicarse la organización direccionada dispersa que hemos estudiado en profundidad en el tema 13, página 8, apartado 4.2.1.2.

5.2.2 Algoritmos de direccionamiento.

Hay tres métodos básicos de direccionamiento para los archivos de organización directa:

- Directo. Se utiliza como dirección la propia llave.
- Asociado. A cada llave se le asocia una dirección mediante una tabla.
- Calculado o por transformación de llave. La dirección se obtiene realizando operaciones y transformaciones con la llave.

A continuación se describen brevemente estos tres tipos de direccionamientos para archivos de organización directa.

Direccionamiento directo.

En este apartado debe explicarse el direccionamiento directo que hemos estudiado en profundidad en el tema 13, página 9, apartado 4.2.2.1.

Direccionamiento asociado.

En este apartado debe explicarse el direccionamiento asociado que hemos estudiado en profundidad en el tema 13, página 10, apartado 4.2.2.2.

Direccionamiento calculado (hashing).

En este apartado debe explicarse el direccionamiento calculado que hemos estudiado en profundidad en el tema 13, página 10, apartado 4.2.2.3.

5.2.3 Operaciones básicas sobre un archivo con organización directa.

Con un archivo con organización directa se pueden realizar todas las operaciones descritas. A diferencia de otras organizaciones, es necesario realizar un proceso especial de creación del archivo. Esta implica la elección de uno de los procedimientos de direccionamiento descritos en los epígrafes precedentes y de un método de gestión de sinónimos, así como reservar espacio en disco para la zona principal.

- Consulta. La consulta se realiza por clave. Para leer un registro debe aplicarse a la clave el algoritmo de transformación y posteriormente leer el registro en la dirección resultante. Si el registro con la clave buscada no se encuentra allí se procederá, en función de cómo se haya resuelto gestionar los sinónimos. La consulta secuencial no siempre es factible.
- Borrado. Siempre se realiza un borrado lógico, pudiéndose reutilizar el espacio del registro eliminado.
- Modificación e inserción. Siempre se puede modificar o insertar un nuevo registro, realizando la transformación de clave correspondiente.

La organización directa es útil para archivos donde los accesos deben realizarse por clave, accediéndose siempre a registros concretos. Si la información se va a procesar en conjunto, con frecuencia puede ser más rentable una organización secuencial indexada.

5.3 Organizaciones indexadas.

Sobre las organizaciones básicas (consecutivas y directas) es posible superponer nuevas estructuras que consiguen mejorar la eficiencia en el acceso a los registros que cumplan ciertas condiciones de búsqueda; se trata de las organizaciones indexadas. Puede ocurrir también que la estructura indexada esté unida a los datos, de forma que éstos no tengan una organización básica previa a la construcción del índice.

Una estructura de índice se define sobre uno o más campos de un fichero, llamados clave de indexación; de forma que cada uno de los valores de la clave de indexación se asocia a la dirección del registro al que le corresponde dicho valor.



Con las organizaciones secuenciales indexadas se pueden realizar las siguientes operaciones a nivel de registro:

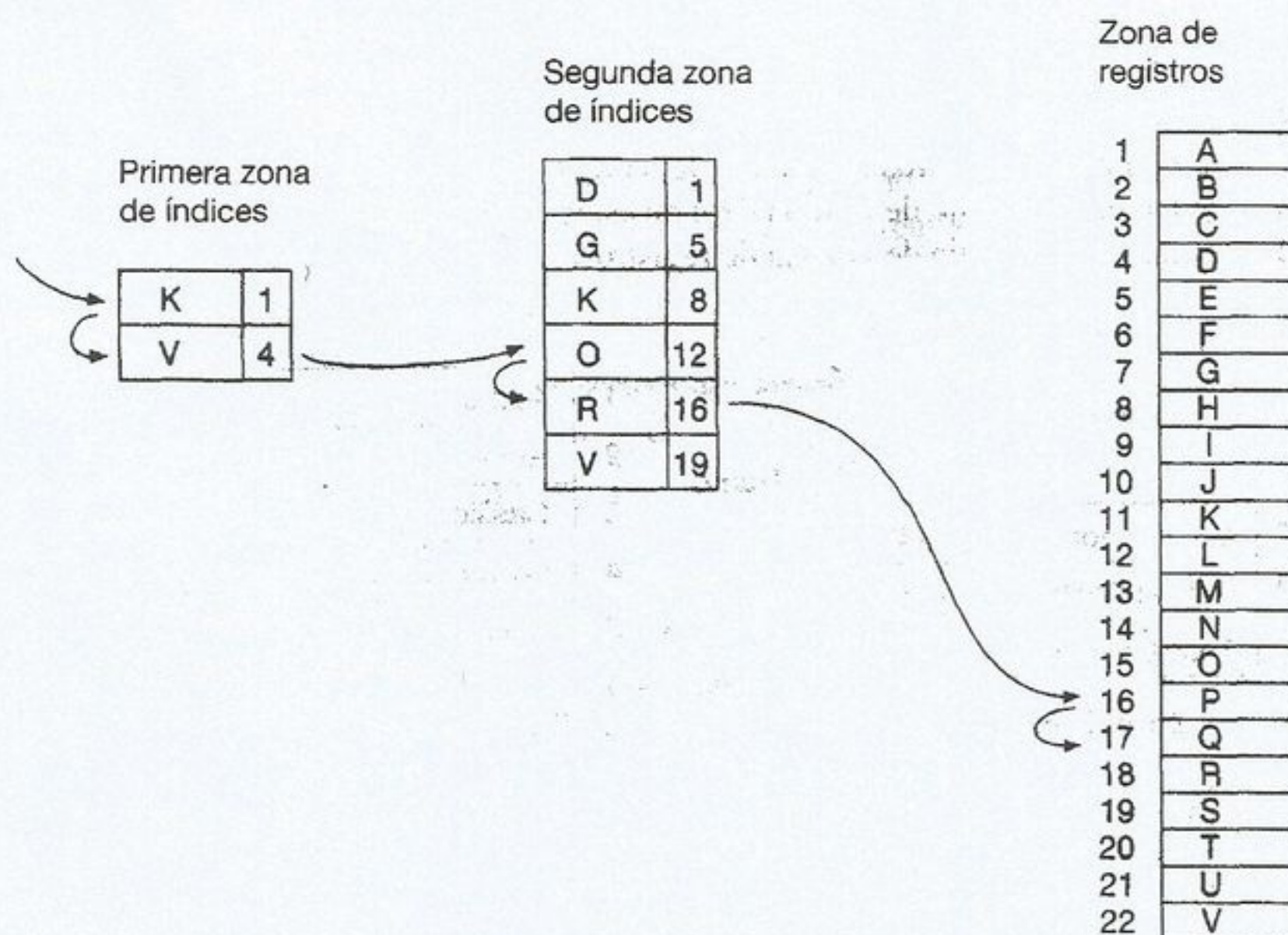
- Consulta. Con este tipo de archivos se pueden realizar consultas secuenciales. Además, esta organización de archivo permite realizar consultas por clave (esto es, localizar un registro conocida su clave) sin necesidad de leer los registros que le anteceden en el archivo. El procedimiento a seguir para realizar una consulta por clave es:
 1. Leer secuencialmente las claves en la zona de índices hasta encontrar una mayor o igual a la del registro buscado. Obtener la dirección de comienzo del tramo donde se encuentra el registro.

2. Leer secuencialmente en la zona de registros a partir de la dirección obtenida en la zona de índices hasta encontrar el registro buscado o uno con valor de clave mayor que el buscado. En este último caso el registro no se encuentra en el archivo.
- Inserción. Dado que ambas zonas son secuenciales, no es posible insertar un registro en archivos con esta organización. En algunos casos se permite la escritura de nuevos registros al final de la zona de registros. Estos registros, como es lógico, no podrán ser consultados por clave con el procedimiento antes descrito.
 - Eliminación y modificación. Al estar los registros escritos en secuencia no es posible borrar un registro. La única forma de eliminar la información contenida en un registro es marcándolo, lo que se conoce como borrado lógico. Las modificaciones son posibles tan solo si el registro no aumenta de longitud al modificarlo y no se altera el valor de la clave de éste.

Un valor de la clave de direccionamiento con la dirección asociada constituye una entrada al índice. El acceso al registro de datos se hace por el valor de la clave de indexación, siendo dicha clave el criterio de búsqueda en consultas selectivas. La clave de indexación puede ser un identificador del fichero de datos (índice primario), o un conjunto de campos no identificadores (índice secundario).

Sobre un mismo fichero de datos se pueden definir varios índices, cada uno sobre una clave de indexación. Cualquier búsqueda selectiva de un registro en una organización indexada obliga a una consulta previa al índice (la cual proporciona la dirección del registro) y un posterior acceso directo al fichero de datos por medio de la dirección obtenida. Por tanto, para optimizar la eficiencia en los accesos, será conveniente mantener el índice, siempre que sea posible, en memoria principal.

Este tipo de organización admite que no exista una entrada al índice por cada registro del fichero (índice no denso), en cuyo caso el fichero de datos tiene que estar ordenado por la clave de direccionamiento. Si existe una entrada por cada registro (índice denso) el fichero no necesita estar ordenado. Los índices no densos admiten la creación de índices sobre los índices, teniendo así estructuras indexadas a varios niveles (índices multinivel).



Esta organización resulta especialmente útil cuando se deben combinar consultas a registros concretos y el procesamiento secuencial de todo archivo. Su principal inconveniente es la imposibilidad de realizar actualizaciones.

Para permitir actualizaciones es necesario incluir en la estructura una zona de desbordamientos. En la zona de desbordamientos (o zona de *overflow*) los registros están desordenados, cada nuevo registro se escribe al final de ésta. Esto complica la búsqueda por llave, pues, si el registro no es encontrado en la zona de registros por el algoritmo antes descrito, es necesario buscarlo secuencialmente en la zona de desbordamientos. Además, se imposibilita la consulta secuencial del archivo, ya que los registros no aparecen ordenados por llave.

Otra mejora de la estructura anterior es incluir punteros entre los registros, de forma que éstos mantengan el orden lógico de los registros. A esta organización se le llama secuencial indexada encadenada. Cuando se crea un archivo con esta organización su estructura es igual a la de un archivo secuencial indexado, salvo que se ha previsto un campo en cada registro para albergar un puntero.

Para insertar un nuevo registro es necesario encontrar el que le sigue en la zona de registros. Se escribe el nuevo registro en la zona de desbordamientos y se reescribe el siguiente según el orden lógico, para incluir el puntero al registro recién grabado.

La eliminación de un registro debe realizarse por marcas. Esto no implica en ningún caso la realización de modificaciones en el índice, pero degrada al archivo. La consulta es semejante a la realizada para el caso no encadenado. Esta organización reúne las ventajas de la organización secuencial indexada y de la organización secuencial encadenada. Cuando el número de registros borrados es grande o las cadenas de desbordamiento son largas el archivo se degrada, esto es, su estructura deja de ser eficaz. Cuando se llega a este punto es necesario reorganizar el archivo; esto es, volver a crearlo introduciendo los registros secuencialmente según el valor de clave.

De entre todas las organizaciones basadas en índices destacan las secuencias indexadas (ISAM), que añaden a las ventajas de las secuenciales (rapidez en el acceso secuencial al fichero), la posibilidad de acceder directamente a ciertos registros (acceso directo) apoyándose en el índice. En este tipo de organizaciones se penalizan las actuaciones respecto a las de tipo secuencial, por la necesidad de actualizar el índice; además, se deterioran cuando se producen muchas actualizaciones.

Las técnicas de indexación difieren a causa del tipo de índice (primario/secundario, denso/no denso), la estructura del índice (ordenado/desordenado, mononivel/multinivel), la organización del fichero de datos (serial/secuencial/disperso), y/o la forma de tratar los desbordamientos

Los índices multinivel, pueden no tener un número fijo de niveles a fin de evitar desbordamientos y/o reorganizaciones del índice, así como de los datos, cuando se insertan registros; en este caso se fija previamente el número de entradas por cada nivel, y se va incrementando el número de niveles, cuando las inserciones obligan a ello. Surgen así distintos tipos de índices (binarios, equilibrados, etc.), basados, en general, en estructuras de tipo árbol. Entre todas las estructuras en árbol destacan la familia de árboles B.

Los nodos del árbol son bloques del espacio de almacenamiento que contienen los valores de las claves de búsqueda y punteros a los nodos hijos; además han de contener, bien los datos, bien los punteros a los mismos. Se asegura una determinada ocupación de los bloques no inferior al 50%, que se estabiliza en un promedio del 69%.

El árbol B es un árbol especial de búsqueda con restricciones adicionales, que garantizan que siempre estará equilibrado y que su densidad es siempre superior a un determinado valor (el espacio de almacenamiento que se “desperdicia” nunca será excesivo). Existen diversas variedades de árboles B (B^+ , B^* , con la información asociada o separada de la estructura del árbol, etc.), siendo estas estructuras el fundamento de las organizaciones físicas en los sistemas relacionales. Los árboles B^+ soportan, al igual que las estructuras ISAM, tanto acceso secuencial como directo, con la ventaja adicional de que crecen dinámicamente cuando lo hace la tabla, manteniendo además el orden de la clave de acceso.

La utilización de índices mejora los tiempos de respuesta ante consultas que impliquen a los atributos indexados, pero disminuye el rendimiento de la base de datos, ya que se debe actualizar el índice cuando se actualizan los atributos sobre los que está definido, además de aumentar el espacio de almacenamiento.

Por estas razones suele ser conveniente indexar la clave primaria (mediante un índice único) en el caso de que el producto no lo haga, las claves alternativas que se utilicen frecuentemente (también mediante un índice único), y aquellas claves ajenas que se utilicen en combinaciones con otras tablas. Sin embargo, en tablas pequeñas, o en aquellas en que prácticamente se recuperan todas las filas, no suele ser conveniente, ya que es mejor una búsqueda secuencial. También se deberá tener en cuenta a la hora de indexar el tipo de datos de los atributos afectados, ya que no es conveniente indexar datos de tipo carácter muy largos.

6 Parámetros de utilización de un archivo.

La organización de un archivo debe elegirse teniendo presente la utilización que va a hacerse del mismo. Ninguna organización es siempre mejor que otra, pero para cada caso concreto hay una ideal. Para caracterizar la utilización que se hace del archivo

se utilizan los siguientes parámetros, que se describen a continuación: volumen, actividad, volatilidad y crecimiento.

- Volumen. Es el espacio, en caracteres, que ocupa el archivo en el soporte. Puede estimarse el volumen, antes de crear el archivo, multiplicando el número previsto de registros por su longitud media. Si el volumen es un problema crítico, interesa usar organizaciones que aprovechen el espacio.
- Actividad. Caracteriza la importancia de los procesos de consulta y modificación del archivo. Se especifica con los siguientes parámetros:
 - Tasa de consulta o modificación. Es el porcentaje de registros consultados o modificados en cada tratamiento del archivo, respecto al total de registros en el archivo.
 - Frecuencia de consulta o modificación. Es el número de veces que se accede al archivo para consultar o modificar en un período de tiempo fijo (mes o año, por ejemplo). En algunos casos se determina la frecuencia de consulta de cada registro del archivo o de cada campo, para optimizar la distribución de los registros en el archivo o de los campos en el registro.
- Volatilidad. Es una indicación del peso de los procesos de inserción y borrado en el tratamiento del archivo. Se especifica con los parámetros tasa de renovación y frecuencia de renovación, con los mismos significados anteriores.
- Crecimiento. Caracteriza la variación de volumen del archivo. Se mide mediante la tasa de Crecimiento, que es el porcentaje de registros en que aumenta el archivo en cada tratamiento.

7 Conclusiones.

Los datos que se encuentran en memoria masiva suelen organizarse en archivos. Un fichero es un conjunto de información sobre un mismo tema, tratada como una unidad de almacenamiento y organizada de forma estructurada para la búsqueda de un dato individual. Un fichero está compuesto de registros homogéneos que contienen información sobre el tema.

El sistema operativo utiliza como unidad de transferencia de datos al disco el bloque. Un bloque es el espacio de disco mínimo que puede reservarse para almacenar un fichero. Para minimizar así la frecuencia de acceso al disco, el sistema operativo mantiene una caché en memoria para alojar los bloques del disco de más reciente uso.

El número de registros lógicos que puede contener un registro físico se denomina factor de bloqueo. Este valor es importante porque influye en el rendimiento del sistema.

Podemos clasificar los archivos como archivos permanentes y archivos temporales. Los archivos permanentes se pueden clasificar en archivos maestros, constantes, o históricos. Los archivos temporales se pueden clasificar en intermedios, de maniobras, o de resultados.

La organización de un fichero determina el modo de estructurar los registros pertenecientes a un fichero en almacenamiento secundario.

Las organizaciones consecutivas sitúan los registros físicamente uno a continuación de otro. Esta es la única organización de archivos susceptible de ser

gestionada en un dispositivo no direccionable. La secuencia en que aparecen los registros en el archivo puede estar determinada por el valor de algún campo o ser simplemente temporal. Si el archivo se encuentra sobre un soporte direccionable es posible almacenar junto a cada registro un puntero con la dirección del registro siguiente.

Las organizaciones direccionadas almacenan los registros según una relación establecida entre el valor de la clave de direccionamiento y la dirección física del registro, siendo posible realizar un acceso directo a un registro si se conoce la clave de direccionamiento. Podemos distinguir entre organizaciones direccionadas directas y dispersas.

Sobre las organizaciones consecutivas y directas es posible superponer estructuras que mejoran la eficiencia en el acceso a los registros que cumplan ciertas condiciones de búsqueda; se trata de las organizaciones indexadas. Una índice se define sobre uno o más campos de un fichero, de forma que cada uno de los valores de la clave de indexación se asocia a la dirección del registro al que le corresponde dicho valor.