

## ARCHIVOS DE DATOS

### CONCEPTOS IMPORTANTES

#### **Archivo**

**Definición 1:** es un conjunto de registros homogéneos que contienen datos heterogéneos.

**Definición 2:** es un conjunto de registros homogéneos bajo una determinada organización (o bajo una estructura en particular.)

**Definición 3:** es un espacio de almacenamiento direccionable, nominable (con nombre), que contiene datos o información bajo una determinada organización o estructura.

**Definición 4:** es un conjunto de información localizada y almacenada como una unidad (posee nombre y dirección de ubicación).

**Definición 5:** es un conjunto de datos estructurados en una colección de entidades elementales o básicas denominadas registros o artículos (relacionados entre sí) que son de igual tipo y que constan, a su vez, de diferentes entidades de nivel más bajo denominadas campos.

#### **Registro**

**Definición 1:** es cada uno de los componentes del archivo que posee una cierta estructura idéntica para todos ellos con objeto de almacenar la información referente al tema general del archivo.

**Definición 2:** es un conjunto de campos relacionados que contienen elementos o datos elementales. Todos los registros tienen la misma estructura.

**Definición 3: Registro (Record)** es una colección de campos relacionados que pueden tratarse como una única unidad por un programa de aplicación. Por ejemplo: un registro de empleados va a contener campos como nombre, número de empleado, DNI, ciudad, etc.

Dependiendo del diseño, los registros pueden ser de longitud fija o de longitud variable.

- **Registros de longitud fija:** siempre tienen el mismo tamaño, esto ahorra tiempo de procesamiento, son más fáciles de manejar y cubren las necesidades de la mayoría de las aplicaciones. Sin embargo, no siempre hacen uso eficiente del espacio asignado, no todos los valores asignados a los campos ocupan el espacio reservado.
- **Registros de longitud variable:** el tamaño del registro puede variar debido a que los datos varían en longitud o debido a que el número de datos cambie de un caso a otro.

#### **Campo**

**Definición 1:** es una mínima parte constitutiva de los registros.

**Definición 2:** es la unidad mínima y elemental de información accesible en un registro.

**Definición 3:** es la unidad mínima de información de un registro, o también se dice que es un ítem o elemento de datos elementales, tal como un nombre, número de empleado, ciudad, DNI, etc.

**Definición 4:** (Field) es el elemento de datos básico. Un campo individual contiene un valor único. Está caracterizado por su longitud y por el tipo de datos. Dependiendo del diseño del archivo, los campos pueden ser de tamaño fijo o variable. Los datos contenidos en un campo se pueden dividir en subcampos; por ejemplo, el campo *fecha* se divide en los subcampos *día*, *mes*, y *año*.

#### **Clave**

**Definición 1:** una clave (key) o *indicativo* es un campo de datos que identifica al registro y lo diferencia de otros registros. Esta clave se utiliza para facilitar la extracción de información de un registro. Se espera identificar en forma unívoca cada registro almacenado en el archivo. Claves típicas son nombres o números de identificación. Condición necesaria y suficiente de la clave: unicidad y minimalidad, una clave por registro es la minimalidad y el dato de esa clave debe ser único, por ejemplo, DNI.

# BASES DE DATOS

---

## CLASIFICACION DE LOS ARCHIVOS DE ACUERDO AL CICLO DE VIDA

Los archivos se clasifican en persistentes y temporales.

- **PERSISTENTES:** se almacenan generalmente en memoria secundaria (discos rígidos, CD, etc.) y permanecen en ella desde su creación hasta su eliminación explícita.
- **TEMPORALES:** pueden residir en memoria secundaria o en memoria principal, y tienen un tiempo de vida determinado de acuerdo a la ejecución de uno o más procesos.

## CLASIFICACION DE LOS ARCHIVOS DE ACUERDO A LA INFORMACIÓN CONTENIDA

De acuerdo a la información contenida en los archivos, éstos se clasifican en:

- **Archivos de programa:** almacenan instrucciones para manipular los datos.
- **Archivos de datos:** almacenan datos en sentido estricto (en forma de registros).
  - **Archivo Maestro:** conjunto de registros acerca de un aspecto importante de las actividades de una organización. Pueden contener datos que describen el estado actual de eventos específicos o indicadores de la empresa. Son útiles sólo mientras se mantengan exactos y actualizados. Antes de que los archivos puedan utilizarse, deben ser mantenidos para reflejar, incluso, los eventos más recientes que afecten los datos en ellos. Esto se logra mediante el uso de Archivos de Transacciones (eventos que afectan a la organización y sobre los cuales se calculan datos).
  - **Archivo de Transacciones:** archivo temporal con dos propósitos: acumular datos acerca de los eventos al momento que ocurran y actualizar los archivos maestros para reflejar los resultados de las transacciones actuales. Los archivos maestros son permanentes y duran mientras exista el sistema. Sin embargo, los contenidos de los archivos cambian como resultado del procesamiento y actualización. Por otro lado, los archivos de transacciones son temporales, cuando no son necesarios se borran o se destruyen.
  - **Archivo de Reportes:** archivo temporal que se utiliza cuando en tiempo de impresión no está disponible para todos los reportes producidos, situación que surge con frecuencia en el procesamiento superpuesto. La computadora escribe el reporte a un archivo en disco, donde permanece hasta que pueda imprimirse. Este proceso se conoce como impresión por cola; es decir, la salida que no puede imprimirse cuando se produce, forma una cola en un archivo de reportes, en espera de respuesta a la solicitud de impresión. Los archivos de reportes se pueden utilizar con muchos otros dispositivos de salida.
  - **Archivo de Respaldo:** copia de un archivo maestro, o de transacciones, hecho para garantizar que se dispone de un duplicado si algo le ocurre al original.

## ORGANIZACION DE LOS ARCHIVOS

Se entiende por organización de los archivos a la forma en que se estructuran dentro del dispositivo de almacenamiento.

*Métodos de organización de archivos:*

El número de alternativas posibles en la organización de archivos es casi ilimitado. A continuación, se describen cuatro alternativas de diseño básico de archivo.

- *Archivo secuencial*
- *Archivo indexado*
- *Archivo secuencial indexado*
- *Archivos directo.*

# BASES DE DATOS

## ARCHIVO SECUENCIAL

**Descripción:** surgen de una organización llamada Apilo. Es la forma más simple de almacenar y recuperar registros en un archivo. Un fichero secuencial está organizado físicamente de tal forma que cada registro es adyacente al siguiente registro del mismo fichero; es decir, que los registros se almacenan uno tras otro ocupando espacios de memoria contiguos. No existen posiciones sin uso. No existe la posibilidad de acceder directamente a los registros y se accede a ellos en el orden en que fueron escritos en el fichero, del primero al último y de uno en uno.

Para localizar un registro es necesario examinar cada registro hasta encontrar el deseado.

	Nombre	Edad	Estatura	IQ
1	Arias	55	5'8"	95
2	Benitez	39	5'6"	75
3	Bona	36	5'7"	70
4	Bresan	25	5'6"	49
5	Calderón	27	5'11"	80
6	Ferrer	42	5'9"	178
7	García	61	5'6"	169
8	Portal	36	5'7"	83
9	Puch	31	5'6"	95
10	Yañez	59	5'5"	145
11	Klein	26	5'8"	47
12	Miller	27	5'2"	75

Archivo Secuencial

Una operación de lectura o escritura lee o escribe el registro y avanza al apuntador al siguiente registro.

- ▣ **Altas:** el nuevo registro se agrega a continuación del último dado de alta (altas al final).
- ▣ **Bajas:** no hay forma física de dar de baja un registro intermedio y realizar el corrimiento; en estos casos se debe hacer una baja lógica. La única forma de realizar la baja física es generar un archivo nuevo sin el registro marcado; el archivo original se borra.

Una baja lógica es marcar de alguna manera un registro dentro del archivo secuencial para no utilizarlo. Un ejemplo podría ser dar de baja un registro si el campo DNI es igual a 0.

D.N.I.	APELLIDO	NOMBRE	CIUDAD	...
23114589	Benítez	Claudia	Oro Verde	...
0	Miller	Carlos	Cerrito	...
27415777	Bona	Hernán	Viale	...
23066774	Bertolo	Pablo	Paraná	...
...	...	...	...	...

Las bajas lógicas las define el programador.

- ▣ **Modificaciones:** los sistemas más viejos no permitían realizar modificaciones. Actualmente, se permite reescribir un registro en un archivo secuencial.
- ▣ **Acceso:** los registros se acceden por su orden de aparición en el archivo. Para localizar un registro específico se utiliza una clave de búsqueda que es comparada con cada registro del archivo, comenzando desde el principio. Este proceso de lectura y comparación se repite hasta encontrar un valor coincidente, de lo contrario continua hasta alcanzar el final del archivo (EOF –End Of File).
  - Costo en fracaso =  $n + 1$  ( $n$  es la cantidad de registros).
  - Costo promedio =  $(n + 1) / 2$

**Nota:** si el arreglo está ordenado según frecuencia de uso y la búsqueda es secuencial, entonces tenemos un costo en fracaso =  $n$  y un costo en éxito =  $n / \ln n$

- ▣ **Ventajas y desventajas:** es el más óptimo en cuanto al aprovechamiento de memoria, ya que se ocupa espacio a medida que se necesita, no requiere espacio previo inicial.  
Si es necesario acceder a cada registro en el archivo para una aplicación particular, un archivo secuencial es un buen método de organización. Si, en promedio, alrededor de la mitad de los registros en el archivo se van a utilizar, la organización secuencial sigue siendo aceptable, ya que al menos uno de dos registros recuperados será utilizado y el tiempo necesario para examinar un registro adyacente es breve.  
Por otro lado, si el requerimiento es hallar un registro particular en un archivo muy grande, la organización secuencial del archivo se convierte en una desventaja. El programa debe comenzar al principio del archivo y leer cada registro hasta encontrar el correcto, una tarea que consume mucho tiempo, a menos que el registro deseado sea uno de los primeros en el archivo.  
En general, si se necesitan menos del 10% de los registros en un archivo durante una ejecución común de procesamiento, el archivo no debe establecerse como un archivo secuencial. Por otro lado, si se desea acceder más del 40% de los registros, el analista debe elegir la organización secuencial. Entre el 10% y el 40%, la decisión depende del tamaño del archivo, la frecuencia de uso, o si será actualizado con frecuencia o utilizado sólo para recuperación.  
Para las aplicaciones interactivas que incluyen peticiones o actualizaciones de registros individuales, los archivos secuenciales ofrecen un rendimiento pobre.  
Otra desventaja es el costoso mantenimiento de este tipo de organización.
- ▣ **Utilización de archivos secuenciales:** los archivos secuenciales son el tipo de archivo utilizado con mayor frecuencia en el procesamiento comercial de datos orientado al manejo por lotes.  
Cuando los datos se procesan cíclicamente, como en el caso de las aplicaciones mensuales de facturación o nómina, la efectividad de este enfoque es difícil de lograr mediante otros métodos. Sin embargo, los datos conservados en archivos secuenciales son difíciles de combinar con otros datos para proporcionar información con algún propósito específico, y es necesario el manejo cronológico del acceso al archivo, si la información solicitada debe ser actual.  
La organización secuencial es conveniente para el procesamiento exhaustivo de todos los registros de un fichero (por ejemplo, cálculo de nóminas, listados completos de empleados, actualización de parámetros, recuento de registros, cálculo de medias, varianzas, etc.).  
Además, el uso de los archivos secuenciales es idóneo para aplicaciones fuera de línea (*off line*).  
Ejemplo de organización secuencial: cassette.

## ORGANIZACIÓN INDEXADA

**CARACTERÍSTICAS GENERALES:** Representa una especie de balance entre la organización de archivos secuenciales y archivos directos relativos (directos), ya que el método usa un examen secuencial del índice, seguido del acceso directo a la dirección física apropiada en el área de datos.

El principio fundamental de las organizaciones y métodos de acceso indexados es el de asociar a la clave de un registro su dirección relativa dentro del fichero mediante una 'tabla de contenido' del fichero, llamado índice. El índice contiene las claves de todos los registros asociadas a la dirección relativa de cada uno de ellos. La tabla de índices puede almacenarse en el mismo fichero o en un fichero separado.

Un archivo con organización indexada consta de un área de índice y un área de datos.

El *área de datos* es el área de almacenamiento que contiene a los registros y está direccionada por el área de índice.

El *área de índice* incluye una clave de registro y la dirección de almacenamiento asociada; dicha dirección es el lugar físico donde se encuentran los datos o información del registro en particular.

El área de índice puede considerarse como una tabla pero en realidad generalmente se implementa en un Árbol B, en el cual sus ramas son información de cómo se van generando los índices y los nodos terminales tienen las direcciones físicas del registro asociado a la clave. Esto se debe a que la estructura de árbol es más rápida que una tabla para la administración de claves. El área de índices es de constante actualización ya que debe respetar un orden en cuanto a la ubicación de las claves (generalmente ascendente). Un archivo puede tener más de un índice y éste puede tener uno o varios campos claves.

El fichero indexado puede contener las claves de todos los registros o solamente un subconjunto de ellas. Se llama densidad del índice al cociente resultante de dividir el número de claves contenidas en el índice entre el número de registros del fichero. Si el índice contiene las claves de todos los registros del fichero, decimos que el índice es denso, y si la densidad del índice es menor que la unidad, decimos que el índice es no denso.

Si el índice es no denso, los registros se almacenan ordenadamente en bloques conteniendo el índice, la clave mayor contenida en el bloque y la dirección del bloque para cada bloque del fichero.

Los índices ocupan espacio adicional pero proporcionan un método rápido de localización de registros.

Es más rápido que el archivo secuencial pero más lento que el direccionamiento directo.

Flexibilidad para procesar el archivo en cualquier forma dependiendo del requerimiento en el momento.

Es conveniente utilizar un archivo indexado cuando se producen muchas variaciones de volumen, ya que no es necesario dimensionarlo previo a su creación.

Cuando se crea un archivo como indexado, los registros deben estar ordenados con respecto a su clave y el administrador de archivos crea tanto un área de datos como el área de índices.

Para hallar un registro cuando se desconoce la dirección de almacenamiento, es necesario examinar todos los registros; sin embargo, la búsqueda será más fácil si se usa un índice, ya que toma menos tiempo buscar en un índice que en un archivo completo de datos.

El procedimiento a seguir para localizar un registro específicos:

1. Leer el índice.
2. Ubicar la clave.
3. Tomar la dirección
4. Buscar el dato en el área de datos.

## ARCHIVO INDEXADO PURO

- ▣ **Descripción:** consta siempre de dos áreas: un área de índices y un área de datos. El área de datos puede estar estructurada en espacios contiguos o bien cada registro estar distribuido de forma individual. Lógicamente es indistinto, ya que de todas formas se conserva un nexo entre los registros cuando están separados. El área de índice almacena una clave y una dirección asociada a esta clave por cada registro en el área de datos.

Una clave es un conjunto de uno o más campos pertenecientes al registro del archivo que identifica unívocamente al registro (claves únicas). Para que un conjunto de una o más claves pueda ser clave, debe cumplir con la unicidad y la minimalidad.

- **Unicidad:** que una clave pertenezca a un registro, que no existan duplicaciones y que no tenga valores nulos. Dicho de otra manera, no hay dos claves con el mismo valor.
- **Minimalidad:** si la clave es compuesta (más de un campo), cualquier campo que se quite de la clave hace perder el criterio de unicidad.

El área de claves es un área de búsqueda organizada generalmente con un Árbol B, por la clave propiamente dicha.

**Área de Índice**

Clave de Registro	Dirección
AB	1021
AC	1021
AD	1022
BC	1018

**Área de Datos**

1021	24613787	Alvarez	Andrea
1021	24592208	Torres	Diego
1022	84507114	Alvarez	José
1018	24300116	Silva	Mariana

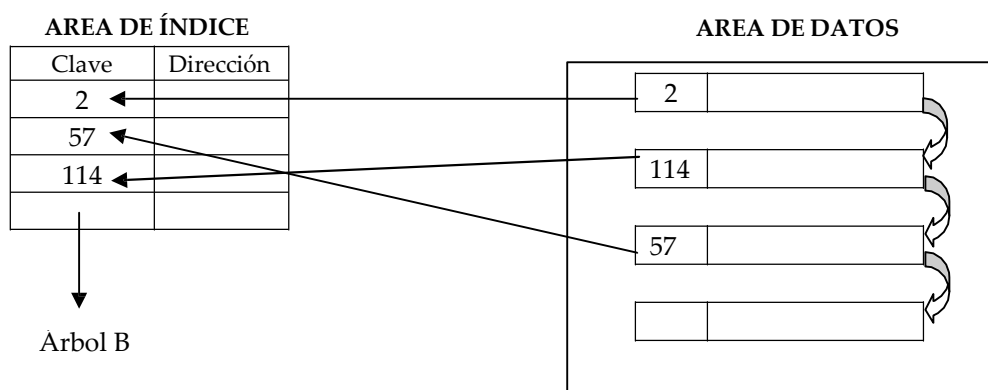
# BASES DE DATOS

- ▣ **Altas:** físicamente, el administrador de archivos ubica la nueva clave en el área de índice, consulta y reserva una dirección en el área de datos, que asocia en el área de índice, luego graba la información del registro en el área de datos reservada, y por último establece el encadenamiento del nuevo registro con los existentes.
- ▣ **Bajas:** permite las bajas físicas (área de índice) y lógicas (área de datos). Una baja física puede realizarse de dos formas:
  - se busca la clave y se obtiene la dirección de los datos. En el área de índice se realiza una baja física, y en el área de datos se marca el registro. No se pierde el registro pero tampoco se puede utilizar, dado que se pierde la dirección para localizarlo. Para evitar desperdiciar espacio de memoria y mantener actualizado el archivo, es necesario depurarlo. Haciendo uso del encadenamiento, se verá qué registros están dados de baja lógicamente y luego se convertirá en memoria disponible ese sector y se deben modificar los enlaces correspondientes para la reorganización del archivo (primero se da de baja en el AD, luego en el AI).
  - Primero se ubica el registro en el área de datos, se libera el lugar que ocupaba y recién luego se elimina físicamente la clave del área de índice.

Cuando el archivo no está en uso, el área de índice posee dos lugares de trabajo, en el disco (medio magnético, lugar físico) y en la memoria principal (área de trabajo por excelencia).

- ▣ **Modificaciones:** se puede regrabar directamente sobre el registro. Se ubica la clave en el área de índice, se localiza el registro en el área de datos, el que se carga en memoria para ser modificado y posteriormente grabado. Si se cambia la clave, se debe reorganizar el área de índice.
- ▣ **Acceso:** la única forma de acceder a un registro de un archivo indexado es a través de una clave. Dada una clave, se la busca en el área de índice; si se la encuentra, se busca la información en el área de datos.
- ▣ **Utilización de archivos indexados:** se utilizan en sistemas *on line*; por ejemplo, dentro de la facultad, en el sistema de alumnado. Se usan principalmente en áreas en las que la oportunidad temporal de la información sea muy crítica. Se encuentran ejemplos de esto en los sistemas de reservación de las aerolíneas, en los trabajos bancarios, en los sistemas militares de datos y en otras aplicaciones de tipo inventario. En estos casos, es raro que los datos se procesen serialmente, excepto en las revisiones de existencias ocasionales, tal vez exclusivamente anuales.

Otras instancias en las que resultan convenientes los archivos indexados se presentan cuando los datos son sumamente variables y dinámicos.



## RECONSTRUCCION DE UN ARCHIVO INDEXADO

Es factible la pérdida de información por cortes de energía si se está trabajando en Memoria Principal, o por rupturas de pistas si la información está alojada en disco.

Si se pierde el área de índice completa, no existe forma de acceder, es por ello que todo administrador de archivos prevé una regeneración en el área de índices.

Para reconstruir el archivo se utiliza una herramienta llamada regenerador que va directamente al área de datos.

Si es contigua y la clave es parte del registro, arma el área de índice con la clave y la dirección; si en el registro no está la clave nunca podría generarse el área de índice. Si el espacio no es contiguo, es fundamental tener la dirección del primer registro hasta llegar al último; es importante guardar la clave en el registro de datos para poder reparar el índice; una vez realizado esto se habrá generado el área de índice.

Es fundamental tener la dirección del primer registro, por eso todo administrador de archivos lo hace (en DOS se guarda en la FAT, en FOX → REINDEX, en COBOL → RECOVERY).

♦ Nota: si se pierde la dirección del primer registro, se podría recuperar manualmente trabajando con las direcciones reales (DEBUG).

## • ARCHIVO SECUENCIAL INDEXADO (ISAM: Indexed Sequential Access Method)

▣ **Descripción:** Los ficheros secuenciales indexados se organizan por bloques de registros donde los índices apuntan a cada bloque. Un fichero ISAM consta de tres zonas lógicas:

- ☒ **Área primaria de datos**, donde se almacenan los registros.
- ☒ **Área de desbordamiento**, también llamada de desborde, de *overflow* o de derrame, donde se almacenan los registros cuando se llena el área primaria de datos.
- ☒ **Área de índice**, donde se almacenan los índices.

El área primaria está organizada por bloques, es decir, almacena bloques individuales de registros (en orden secuencial). Dentro de cada bloque hay una cantidad fija de registros, ordenados por clave. Esa cantidad (n) de registros la determina el administrador de archivos. Los registros contenidos en los bloques están dispuestos en forma consecutiva. El acceso a cada bloque es directo, y en el bloque se busca el registro (clave) secuencialmente.

El área de índice sólo almacena las direcciones físicas correspondientes al comienzo de cada bloque y su clave asociada. Es necesario sólo un índice para cada bloque.

El área de saturación se utiliza para almacenar aquellos registros que en las actualizaciones no han tenido cabida en el área primaria. Posee una organización secuencial.

El área de desborde aparece cuando los archivos han sido dimensionados; sirve para cubrir todas las posibilidades de fallas en el dimensionamiento.

Generalmente, la dimensión del área de overflow es un 15% o 20% del área de datos.

Si el área de desborde crece mucho, quiere decir que el dimensionamiento elegido no fue el correcto.

Si permanece estable, implica que el dimensionamiento es adecuado.

### AREA DE INDICE

Clave del registro	Dirección de comienzo del bloque
1115	1345
1315	1349
1429	1346
1725	1350



# BASES DE DATOS

## AREA DE DATOS

1345	1010	1011	1013	1014	1017	1019	1110	1113	1115					
1346	1316	1317*	1321	1323	1324	1410	1414	1415	1417	1418	1419	1427	1428	1429
1349	1117	1120	1121	1210	1211	1212	1215	1217	1218	1221	1310	1313	1315	
1350	1510	1521	1522	1617	1619	1620	1721	1724	1725					
O- F-	*1318													

- Altas:** si no existe el archivo, se crea un bloque (definiendo su tamaño y su porcentaje de ocupación) y se graban los datos con la clave en el primer lugar del bloque; por último, se actualiza el área de índice con la clave y la dirección del bloque. Si existe el archivo, se debe buscar en qué bloque cabría la clave a dar de alta considerando que si el por ciento de ocupación es mayor al previsto se abre un nuevo bloque y se inserta el registro, esto se realiza así para asegurar que siempre exista lugar. Luego se debe indicar en el índice que esa clave está en el nuevo bloque, lo que implica incrementar el área de índice cuando se dan de alta nuevos bloques.

Al querer adicionar registros y no contar con el espacio suficiente para reacomodar los registros dentro del bloque y mantener el orden secuencial con respecto a su clave, se crea un área de saturación. De este modo, físicamente el archivo principal ya no es secuencial; sin embargo, desde el punto de vista del programador, el archivo es lógicamente secuencial. Los registros en las pistas están encadenados a los registros en las pistas de saturación, para que el archivo sea accesible como si estuviera en orden físico secuencial.

Dentro de cada bloque se reserva un GAP de protección (30/40%); una vez cubierto el resto del bloque, un alta de clave superior a la última clave genera la apertura de un nuevo bloque, quedando el GAP de protección para los nuevos registros que tengan el valor entre la mínima del bloque y la mínima del siguiente bloque.

Una vez que el bloque se llena, todos los nuevos registros correspondientes a ese bloque se graban en el área de desborde.

En síntesis, para dar de alta un nuevo registro, se siguen los siguientes pasos:

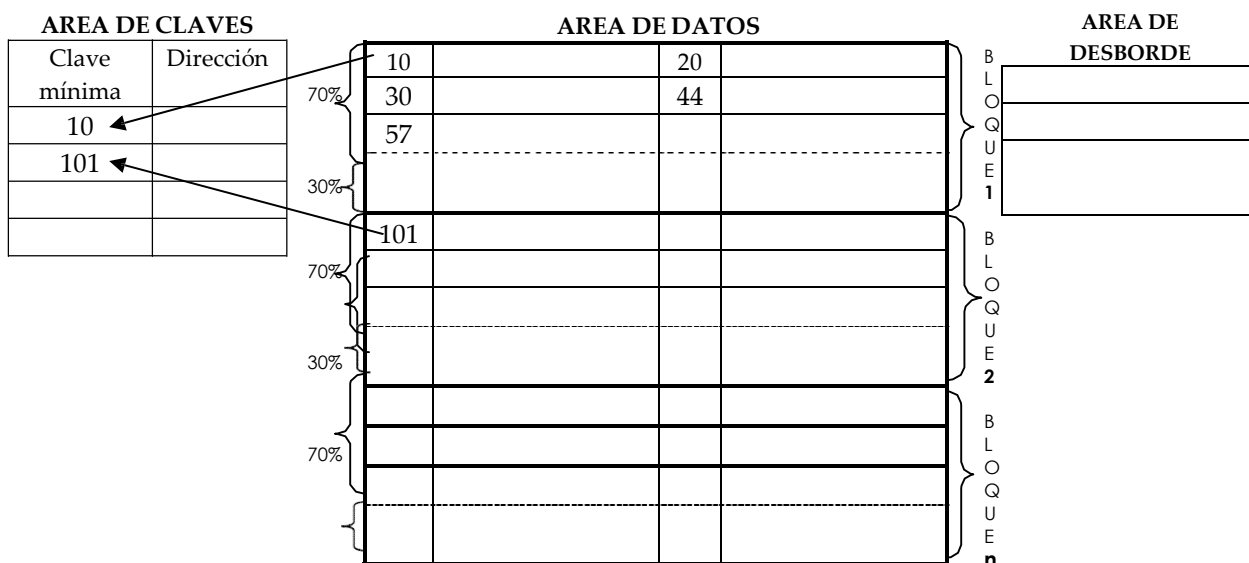
- ☒ Se recorre hasta encontrar una clave mayor, sabiendo así que el nuevo registro se ubicará en el bloque anterior.
- ☒ Luego de tener la dirección del bloque, se graba el registro al final del mismo.
- ☒ Si se tiene lugar en el bloque de datos, se da de alta.
- ☒ Si no hay lugar dentro del bloque, se lo graba en el área de desborde, actualizando los links.
- Bajas:** se puede marcar lógicamente el registro dentro del bloque, o bien producir un corrimiento. En el caso que el registro borrado sea el mismo del índice (clave), deberemos modificar el área de índices. Se utilizan los bloques en memoria, se recorren en memoria. Si el registro es único, desaparece el bloque y su referencia en el área de índices, es decir, que sólo es posible dar de baja físicamente si es el registro del bloque.
- Modificaciones:** se puede regrabar directamente sobre el registro. Para localizar un registro, en primer lugar el programa accesa al índice, empezando con el registro del principio, compara la clave en el índice con la clave de búsqueda, hasta ubicar la clave correspondiente al bloque en el que se encuentra el registro a modificar. El programa le dice al sistema que comience a leer registros en la dirección asociada a la clave y compara la clave de cada registro con la clave de búsqueda. Cuando éstas coinciden, se cargan los datos en memoria para ser modificados y posteriormente grabados. Si se cambia la clave, se debe reorganizar el área de índice

- Este tipo de organización de ficheros es muy utilizada porque mantiene un buen equilibrio entre el tamaño de la tabla de índices y el tiempo de acceso a los registros.

El área de saturación puede crecer excesivamente, lo que supone una disminución en la rapidez en su tratamiento y una mayor dificultad en su procesamiento.

- Este tipo de archivo es adecuado para el acceso en línea u orientado a terminales. Esto no es posible con los archivos de tipo secuencial.

También se utilizan comúnmente para manejar consultas, con la restricción de que la consulta deberá especificar el atributo llave. Son casos comunes de este empleo las consultas de facturación basadas en números de cuenta.



## ARCHIVOS DIRECTOS

- ❑ **Descripción:** Medio u organización que permite procesar o acceder en forma rápida a los registros haciendo referencia directamente a su posición relativa en el soporte de almacenamiento, sin necesidad de pasar por la información anterior. Son el único tipo de archivo donde su acceso es directo. Se debe establecer la dimensión (longitud y cantidad de registros) previo a su utilización. Esto nos obliga a hacer una estimación sin saber la cantidad de registros que se van a utilizar. Para dimensionarlo, hay que tener en cuenta:

- Cantidad de registros a almacenar
- Longitud de cada registro.

Cada registro tiene asociado un **Número Relativo de Registro** (numeración de los registros), que no es la dirección física que tiene realmente en el disco, es el Sistema Operativo quien se encarga de relacionar el N° Relativo de Registro con la dirección física.

Todo registro posee un campo de identificación (clave). Es conveniente al guardar la información que este campo esté en relación con el número relativo de registro, a fin de poder acceder por medio de él. No es necesario su orden. Como todo tipo de archivo dimensionable, posee un área de overflow o de desborde, que se define en el mismo momento de definir al área de datos.

- ❑ **Altas:** para dar de alta a un registro, es necesario posicionarse según el número relativo de registro y luego grabar la información, siempre haciendo referencia al número de registro. Como se accede directamente a los N registros, entonces es posible dejar espacios entre uno y otro. Si se accede por simulación de acceso secuencial, se da de alta como en archivos secuenciales.
- ❑ **Bajas:** las bajas son lógicas, pero es posible crear un archivo de menor dimensión e igual tipo, luego copiar los registros que no están marcados como dados de baja en éste, para finalmente borrar el archivo de origen; o se puede usar el campo marcado lógicamente para determinar cuando es posible sobrescribir los datos de un registro (que lógicamente está borrado).
- ❑ **Modificaciones:** Para realizar modificaciones se debe ubicar el registro, leerlo y una vez disponibles los datos en memoria hacer los cambios pertinentes para luego reposicionar el registro y grabarlo.
- ❑ **Acceso:** ante la petición de un registro determinado, el método calcula la dirección del bloque físico que lo contiene y accede a él directamente. El acceso es rápido y tiene costo 1. También es posible acceder al registro mediante el contenido de su campo clave si éste corresponde exactamente a un N° Relativo de Registro (NRR). Sin embargo, ya que no es usual que las claves de registro cumplan este requerimiento, a menudo se utiliza el *Hashing* para proporcionar un medio de acceso directo. Con Hashing, se obtiene un NRR aplicando una función matemática de transformación de la clave que divide, extrae o dobla la clave del registro. Es necesario que el número relativo de registro esté en un rango determinado con relación a la dimensión del archivo ( $1 - N$ ) y que los valores devueltos no se repitan, de lo contrario no se puede asegurar que se refiere a un registro en especial. Además, los valores obtenidos deben distribuirse de manera uniforme en todo el espacio asignado en vez de acumularse todos juntos; esto garantiza una recuperación más rápida de los registros y hace un mejor uso del espacio. Si no se puede transformar una clave porque no se tiene una función para transformarla, se simula una lectura secuencial, que consiste en acceder al primer registro del archivo, leer los datos y determinar, mediante una comparación, si es el buscado; si no lo es, se incrementa en uno la posición y se compara nuevamente; así sucesivamente hasta llegar al registro deseado. Este procedimiento de búsqueda es muy lento (incluso más que el secuencial).
- ❑ **Ventajas y desventajas:** el método de acceso directo es rápido, ya que se evitan las operaciones intermedias de archivo. Si se puede asociar una clave de registro con un NNR, el archivo directo es 100% eficiente. Si no se puede hacer la coincidencia, el archivo prácticamente no sirve. Entre no servir y el ideal, existe el Hashing. Por ejemplo, teniendo 100 proveedores (con número de proveedores contiguos), se asocia Número de proveedor 1 con NNR 1, y así sucesivamente. En cambio, por ejemplo con DNI, no resulta eficiente.

## BASES DE DATOS

- ▣ **Utilización de archivos directos:** se usan cuando la información no tiene gran variabilidad, tanto los números de registros como en el contenido de cada registro. En aplicaciones en que los tamaños de registro son pequeños y fijos, donde el acceso rápido es esencial y donde el acceso a los datos siempre se efectúa en forma simple, la organización de archivo directo es muy adecuada. En este caso, acceso simple significa emplear una sola llave para recuperar, y no acceso en serie. A modo de ejemplo, los archivos directos se utilizan con frecuencia en directorios, tablas de precios, programación de tiempos, listas de nombres, etc.

Los archivos directos también desempeñan un importante papel como componentes de organizaciones más complejas de archivos.

### ANÁLISIS ENTRE COSTO Y BENEFICIO

	Archivo Secuencial	Archivo Indexado	Archivo Secuencial Indexado	Archivo Directo
<b>Uso de la memoria</b>	Óptimo.	Utiliza más memoria que un archivo secuencial, pero es justificado ya que tiene un acceso bastante rápido.	Bueno.	Regular, ya que se debe dimensionar el archivo antes de crearlo, con lo que puedo llegar a malgastar memoria si sobran registros, o puede llegar a ser poca la cantidad de registros reservados.
<b>Eficiencia en búsquedas y consultas</b>	Es ineficiente, ya que se debe recorrer uno a uno todos los registros hasta encontrar el que se desea. El costo de las búsquedas y consultas es alto; para disminuirlo se sugiere ordenar los registros (en forma ascendente o descendente) en función de algún campo de los registros.	Son bastante eficientes, más rápido que el secuencial y que el secuencial indexado, pero más lento que el directo.	Buena, es más rápida que la secuencial, aunque más lenta que la directa.	Bastante eficiente, ya que posee costo de acceso 1.