

ARCHIVOS DE DATOS

Cátedra de Bases de Datos - Teoría

- **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.

Definiciones - Archivos

- **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.

Definiciones - Registros

- **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.

Definiciones - Campo

- 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.

Definiciones - Clave

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.

Clasificación según el ciclo de vida

- Archivos de programa: almacenan instrucciones para manipular los datos.
- Archivos de datos: almacenan datos en sentido estricto (en forma de registros).
 - Archivo Maestro
 - Archivo de Transacciones.
 - Archivo de Reportes.
 - Archivo de Respaldo.

Clasificación según la información

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.

Organización de los archivos

Descripción: surgen de una organización llamada Apilo. Es la forma más simple de almacenar y recuperar registros en un archivo.

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.

Archivo secuencial

	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

Archivo secuencial

Altas: El nuevo registro se agrega al final del archivo.
(append)

Bajas: No existe la baja física. Se deben marcar los registros con un valor testigo.

Modificaciones: Originalmente, los registros no se podrían modificar. Actualmente existe tecnología para **reescribir** en un registro.

Archivo secuencial

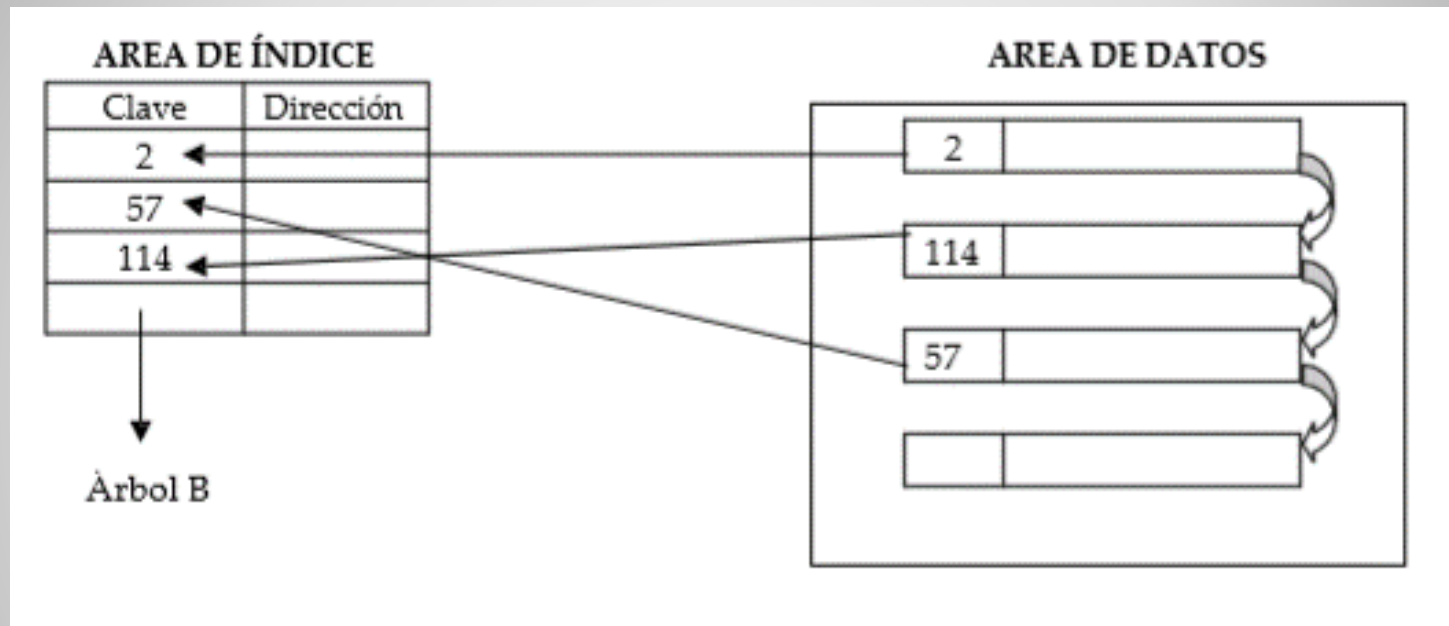
Representa un balance entre la organización de archivos secuenciales y archivos directos relativos (directos),

El principio fundamental de la 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 **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.

Archivo indexado



Archivo indexado

Altas: Grabar el registro en el área de datos, guardando la dirección de memoria. Insertar la clave en el área de índices con la clave y la dirección. El área de índices se mantiene ordenada. Nota: versiones más avanzadas guardan un puntero entre los registros del área de datos para una posterior reconstrucción del área de índices.

Bajas: Existen bajas físicas. Simplemente se elimina el registro del área de índices y se cambia el puntero (o se marca el registro como eliminado) del área de datos.

Modificaciones: Buscar en el área de índices el registro, obtener la dirección de memoria en el área de datos, ubicar el registro en el área de datos y modificarlo.

Archivo indexado

Descripción: Se organizan por bloques de registros donde los índices apuntan a cada bloque.

Consta de tres zonas lógicas:

- **Área primaria de datos**, se divide en **bloques** donde se almacenan los registros, con un tamaño previamente definido fijo y no se puede redimensionar.
- **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 que consiste en la primera clave del área de datos y la dirección de memoria de la misma.

Archivo secuencial indexado

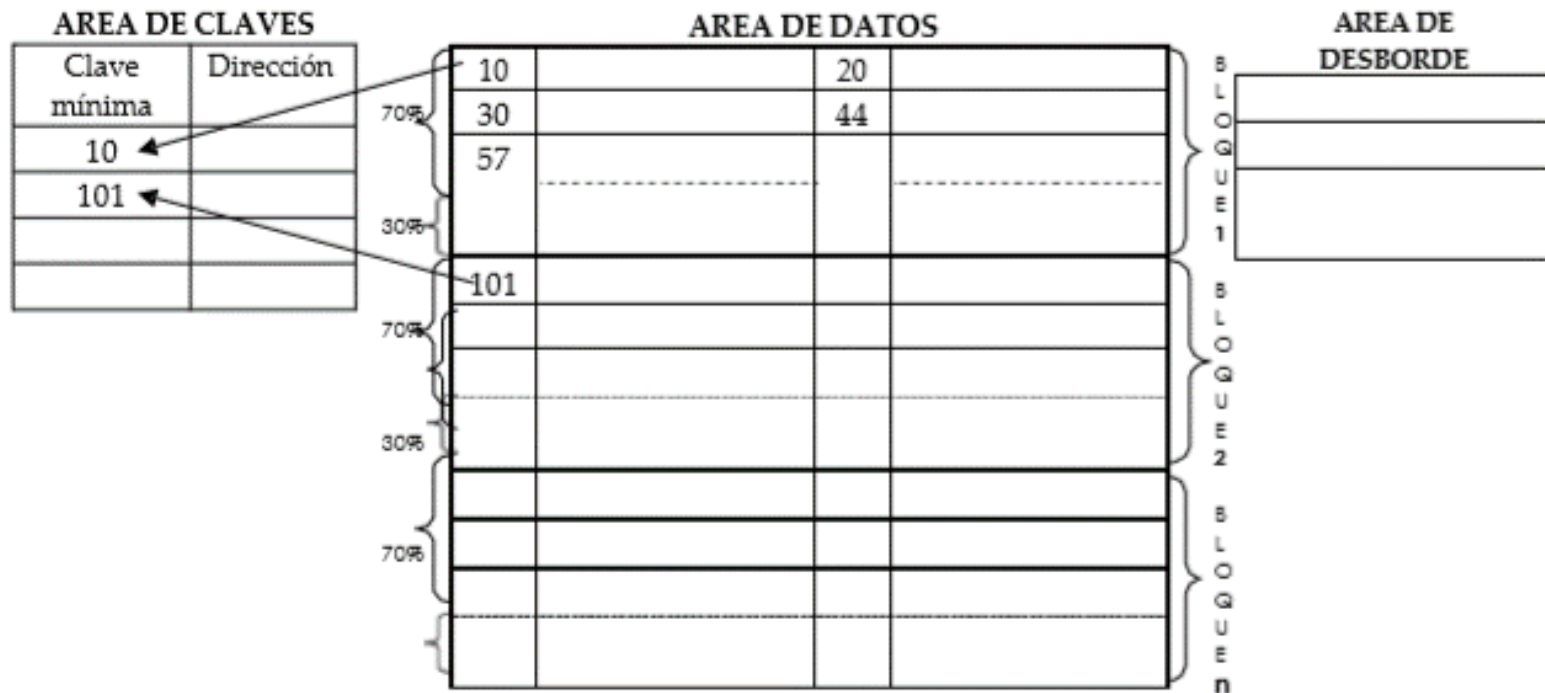
El área primaria está organizada por bloques.

Dentro de cada bloque hay una cantidad fija de registros, **ordenados por clave**. Los registros contenidos en los bloques están dispuestos en forma consecutiva. El acceso a cada bloque es directo a través del área de índice, 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. Si al momento de reordenar un bloque del área primaria, cambia la clave del primer registro, se debe entonces modificar el área de índices.

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

Archivo secuencial indexado



Archivo secuencial indexado

Descripción: 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. SEEK.

Son el único tipo de archivo donde su acceso es directo.

Se debe establecer la dimensión (filesize es el tamaño del archivo que está establecido en base a la longitud del registro por cantidad de registros) previo a su utilización. Esto nos obliga hacer una estimación sin saber la cantidad de registros que se van a utilizar.

Archivo directo

ANÁLISIS ENTRE COSTO Y BENEFICIO

	Archivo Secuencial	Archivo Indexado	Archivo Secuencial Indexado	Archivo Directo
Uso de la memoria	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.
Eficiencia en búsquedas y consultas	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.

Análisis costo-beneficio