**Archivos**

Los **archivos** son estructuras que tienen las siguientes características:

* Deben ser capaces de contener grandes cantidades de información.
* La información contenida debe permanecer no importando los procesos que la generen.
* Distintos procesos deben poder acceder al archivo concurrentemente.

Cada archivo tiene atributos o metadatos:

* Nombre simbólico (dado por el usuario)
* Fecha y hora de creación
* Fecha y hora de actualización
* Bits de protección (lectura, escritura)
* Propietario
* Contraseña de acceso
* Numero de bytes por registro
* Capacidad máxima
* Capacidad actualmente ocupada
* Información sobre su ubicación
* Información que se quiere almacenar

**Directorio**

Es un archivo, constituido por una tabla con una fila por cada archivo o un subdirectorio que está integrado en el directorio, con metadatos sobre el mismo. La característica de un directorio es:

* Permiten al usuario organizar sus archivos.
* Lo constituyen un grupo de archivos y/o subdirectorios
* Al igual que un archivo tiene un nombre simbólico, propietario, bloques que ocupa.
* Enlaces a posiciones físicas.

Tienen una estructura arbórea incluyendo en sus nodos a los subdirectorios y en sus hojas a los archivos.

**Ruta**

**Ruta** o también conocida como path de un archivo es la lista de los directorios atravesados hasta llegar al directorio o archivo deseado. Existen dos tipos: path absoluto, que es indicar la dirección empezando por la raíz y el path relativo que es indicar la dirección desde el directorio actual.

**Tabla descriptora de archivos**

La **tabla descriptora de archivos,** esta contiene una lista de todos los archivos abiertos en un momento dado, cada vez que se abre un archivo, a esta tabla se añade una fila con su nombre, metadatos, tamaño, enlace a su primer bloque, etc. Cuando el archivo se cierra se elimina la información de la tabla.

**Gestión de archivos**

La **gestión de archivos** por el sistema operativo, a partir de los conceptos de archivo y directorio logra que el usuario no tenga que conocer los detalles físicos del periférico. Es decir, al usuario no se le muestran los problemas físicos de almacenamiento.

**Funciones del SO para la gestión de archivos**

Tabla

Descripción generada automáticamente

La gestión de archivos se puede ver desde dos puntos de vista: lógico y físico.

* **Visión lógica:**

Dentro de la visión lógica se tiene el sistema de gestión de E/S y gestión de archivos, esto ultimo se encarga de ubicar en bloques, dar acceso a varios usuarios, bloquear el uso de los archivos, así como de administrar el espacio libre. En la visión lógica un archivo es una sucesión de registros lógicos.

Los **registros lógicos,** estructurado en campos**,** mientras que un campo simplemente son conjuntos de caracteres

* **Visión física:**

Desde un punto de vista físico, los archivos son una sucesión de registros físicos y estos se almacenan en el dispositivo de memoria masiva en los llamados **clústeres** o **bloques** que son conjuntos de bytes que se transfieren en una operación de E/S.

El **tamaño del bloque** es una elección difícil, ya que un bloque pequeño representa poca eficiencia debido a las interrupciones que se generan al momento de transmisión de cada bloque, además de que las tablas de enlace sean mayores por lo cual la gestión de esta se torna más difícil. Por otra parte, los bloques grandes tienen la desventaja de la fragmentación interna, ya que se desaprovecha el ultimo “contenedor” o bloque, estadísticamente, en cuanto mas grande sea el tamaño del clúster mayor será este desperdicio. Este tamaño de bloque lo define el SO y por el dispositivo físico

**Asignación de espacio en disco**

El sistema de archivo en disco es la organización del almacenamiento de archivos de forma que estos puedan ser leídos o almacenados directa o indirectamente.   
Existen diferentes sistemas de archivos en disco, por ejemplo, el SO Windows reconoce: FAT, FAT32, NTFS, mientras que Linux reconoce:  XFS, JFS, UFS, ISO9660, FAT, FAT32.   
Se hacen grupos contiguos de bloques físicos (clústeres) para conseguir una mayor eficiencia.

**Clústeres de un archivo**

Son un grupo contiguo de bloques físicos, estos pueden almacenarse de múltiples formas:

* **Lista de enlaces:**

Cada disco dispone de una tabla en la que la posición de cada elemento se corresponde únicamente con cada clúster y contiene un apuntador al lugar donde se encuentra el siguiente bloque de archivo.

Cuando se are un archivo, el sistema de archivos carga sobre la memoria principal la lista de enlaces (apuntadores) así como el enlace al primer elemento

**Desventaja:** Si hay demasiados clústeres, la lista de apuntadores ocupa en demasía la capacidad de la memoria principal. Además de que los archivos nuevos deben encajar en los espacios libres dejados por los archivos eliminados, de esta forma, los clusters quedan muy separados unos de otros (lo conoce como un archivo fragmentado).

**Ventaja:** Se pueden recuperar archivos de una manera sencilla, ya que, al borrar un archivo en este tipo de almacenamiento no es necesario borrar físicamente los datos del archivo.

* **I-nodos o fichero de índices**

Cada archivo tiene asociado un nodo de información (i-nodo) que es una pequeña tabla de tamaño fijo, en esta se guardan los atributos del archivo y 13 direcciones; las primeras 10 direcciones indican las posiciones de los primeros clústeres donde se encuentran los datos de archivos, mientras que las ultimas 3 indican direcciones de los bloques del archivo de forma indirecta, ya que estos bloques contienen índices, con simple, doble y triple indirección respectivamente.

Cada bloque de índices puede llegar a tener muchísimas direcciones a otros bloques de índices o a los clústeres de un archivo.

* **Árboles B+**

Es un tipo de estructura que representa una colección de datos ordenados de manera que permite el acceso ya sea de lectura, inserción o borrado muy eficiente.

Los enlaces a los clústeres se guardan en las hojas que direccionan a los clústeres (que se encuentran en el nivel más bajo), los nodos internos contienen claves y apuntadores, mientras que los nodos hoja están enlazados entre sí, para permitir una búsqueda secuencial.

Un sistema de archivo de disco NTFS organiza los accesos siguiendo una estructura de este tipo (Árbol B+), en la que se pude notar que los tiempos de acceso muy bajos. Las unidades de almacenamiento que se consideran en este tipo de sistema de archivos son:

* **sector** (unidad de almacenamiento más pequeña),
* **clúster** (uno o más sectores contiguos, en la misma pista) y
* **volumen** (porción lógica de un disco compuesto de uno o más clústeres).