

Gestión de Archivos

César Pedraza Bonilla

Universidad Nacional

capedrazab@unal.edu.co

18 de mayo de 2023

- 1 Generalidades
- 2 Organización y acceso de archivos
- 3 Directorios.

- Es un software que provee a los usuarios y aplicaciones, servicios relativos al uso de archivos.
- Requisitos:
 - Garantizar la integridad de la información almacenada.
 - Optimizar el rendimiento del sistema de administración de disco.
 - Ofrecer soporte para E/S.
 - Minimizar la posibilidad de pérdida de información.
 - Ofrecer una forma estándar de acceso a los recursos.
 - Ofrecer soporte para múltiples usuarios.

Archivos

Un archivo es un conjunto de datos con propiedades tales como:

- Existencia a largo plazo.
- Intercambiable entre procesos.
- Estructura. Dependiendo del archivo se dispone de una estructura interna.

El sistema de archivos provee ciertas funciones sobre los archivos:

- Creación.
- Borrado.
- Apertura.
- Cierre.
- Lectura.
- Escritura.

Términos comunes en archivos:

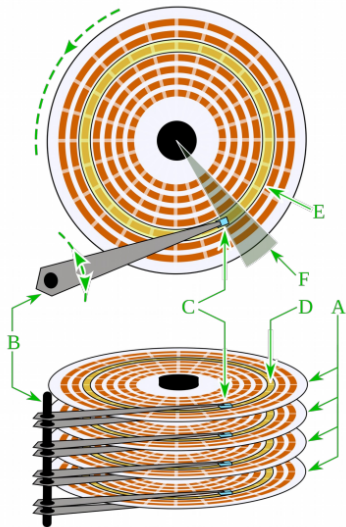
- Campo. Es el elemento básico de datos. Un campo contiene un valor simple, tal como un número, cadena o estructura.
- Registro. Es una colección de campos relacionados, que pueden ser tratados como una unidad en ciertas aplicaciones.
- Archivo. Es una colección de registros similares.
- Base de datos. Es una colección de datos relacionados.

Archivos

- Bits (b): Es la unidad más básica de información digital y puede representar un valor de 0 o 1.
- Bytes (B): Un byte está compuesto por 8 bits y es la unidad básica de almacenamiento en la computadora. Puede representar un carácter o un número.
- Kilobyte (KB): Equivale a 1,024 bytes (2^{10} bytes).
- Megabyte (MB): Equivale a 1,048,576 bytes (2^{20} bytes) o aproximadamente un millón de bytes.
- Gigabyte (GB): Equivale a 1,073,741,824 bytes (2^{30} bytes) o aproximadamente mil millones de bytes.
- Terabyte (TB): Equivale a 1,099,511,627,776 bytes (2^{40} bytes) o aproximadamente un billón de bytes.

Archivos

- Petabyte (PB): Equivale a 1,125,899,906,842,624 bytes (2^{50} bytes) o aproximadamente mil billones de bytes.
- Exabyte (EB): Equivale a 1,152,921,504,606,846,976 bytes (2^{60} bytes) o aproximadamente un trillón de bytes.
- Zettabyte (ZB): Equivale a 1,180,591,620,717,411,303,424 bytes (2^{70} bytes) o aproximadamente mil trillones de bytes.
- Yottabyte (YB): Equivale a 1,208,925,819,614,629,174,706,176 bytes (2^{80} bytes) o aproximadamente un cuatrillón de bytes.



- Disco
- Brazo
- Cabeza
- Cilindro
- Pista
- Sector

Sistemas de archivos en Linux

Un sistema de archivos es una colección organizada de archivos regulares y directorios. Se crea utilizando el comando `mkfs`.

Linux ofrece soporte para una amplia variedad de sistemas de archivos, que incluyen:

- Varios sistemas de archivos nativos de UNIX, como Minix, System V y BSD.
- Sistemas de archivos FAT, FAT32 y NTFS de Microsoft.
- Sistema de archivos ISO 9660 para CD-ROM.
- Sistema de archivos HFS de Apple Macintosh.
- El tradicional sistema de archivos `ext2`.
- Diversos sistemas de archivos de red, como NFS, SMB, NCP y Coda.
- Varios sistemas de archivos con registro, como `ext3`, `ext4`, Reiserfs, JFS, XFS y Btrfs.

Sistema de archivos

La unidad básica para asignar espacio en un sistema de archivos es un bloque lógico, que consiste en múltiples bloques físicos contiguos en el dispositivo de disco donde reside el sistema de archivos. Por ejemplo, el tamaño del bloque lógico en ext2 es de 1024, 2048 o 4096 bytes.

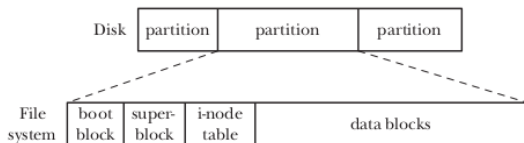


Figura: Sistema de archivos

Un sistema de archivos consta de las siguientes partes:

- Bloque de arranque: Es el primer bloque en un sistema de archivos y contiene información para arrancar el sistema operativo.
- Superbloque: Es un solo bloque que sigue al bloque de arranque y contiene información de parámetros del sistema de archivos, como el tamaño de la tabla de i-nodos y el tamaño del sistema de archivos.
- Tabla de i-nodos: Cada archivo o directorio en el sistema de archivos tiene una entrada única en la tabla de i-nodos, que registra información sobre el archivo.
- Bloques de datos: La mayoría del espacio en un sistema de archivos se utiliza para los bloques de datos que forman los archivos y directorios en el sistema.

Arquitectura de un sistema de archivos

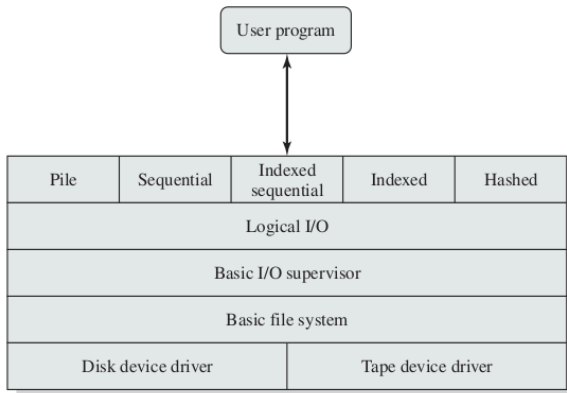


Figura: Arquitectura general de un sistema archivos.¹

¹Tomado de Stallings, Operating Systems.

Arquitectura de un sistema de archivos

- Drivers. Capa de software que se comunica con el disco.
- Basic file system, physical I/O. Manipula bloques de datos.
- Basic IO supervisor. Inicia y termina las sesiones de IO. Selecciona el dispositivo sobre el que se realiza las operaciones.
- Logical IO. Permite a los usuarios y aplicaciones acceder a registros. Convierte los datos de usuario o aplicación a conjuntos de bloques.
- Access method. Provee una forma estándar a los usuarios para acceder a disco.

Sistema de archivos

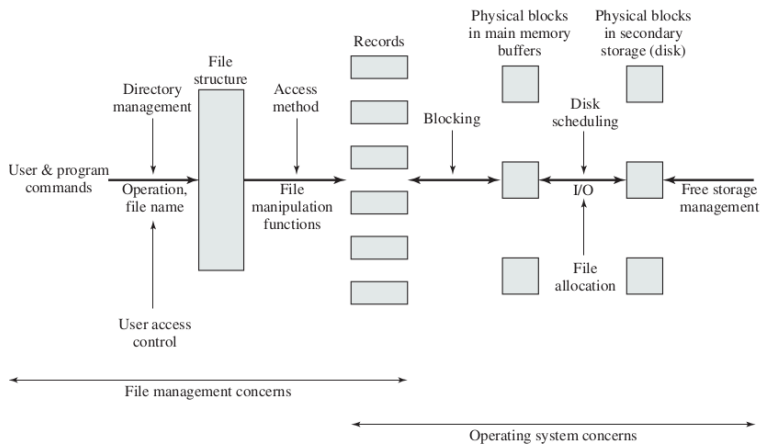


Figura: Conversión de datos en un sistema de archivos.²

²Tomado de Stallings, Operating Systems.

i-node

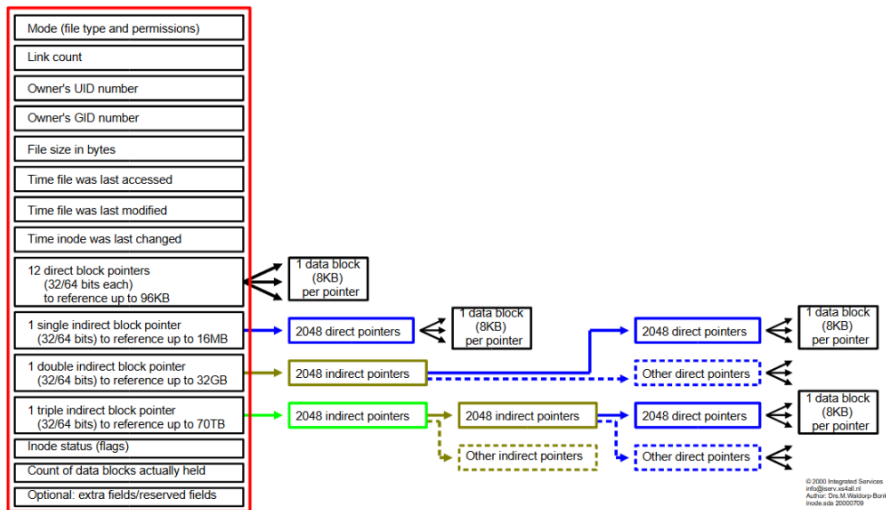


Figura: inode.³

La tabla de i-nodos de un sistema de archivos contiene un i-nodo (índice de nodo) para cada archivo en el sistema. Los i-nodos se identifican numéricamente por su ubicación secuencial en la tabla de i-nodos. El número de i-nodo (o simplemente número de i-nodo) de un archivo es el primer campo mostrado por el comando `ls -li`. La información mantenida en un i-nodo incluye lo siguiente:

- Tipo de archivo (por ejemplo, archivo regular, directorio, enlace simbólico, dispositivo de caracteres).
- Propietario (también conocido como ID de usuario o UID) del archivo.
- Grupo (también conocido como ID de grupo o GID) del archivo.
- Permisos de acceso para tres categorías de usuario: propietario, grupo y otros.

La información mantenida en un i-nodo incluye lo siguiente (continuación):

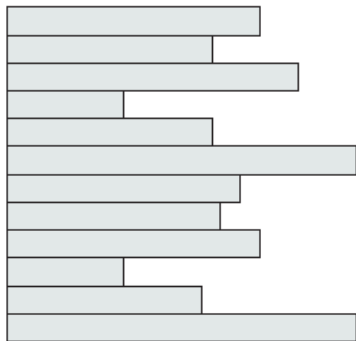
- Tres marcas de tiempo: última vez que se accedió al archivo (mostrada por `ls -lu`), última vez que se modificó el archivo (tiempo predeterminado mostrado por `ls -l`), y última vez que se cambió el estado (último cambio en la información del i-nodo, mostrado por `ls -lc`).
- Número de enlaces duros al archivo.
- Tamaño del archivo en bytes.
- Número de bloques asignados al archivo, medido en unidades de bloques de 512 bytes.
- Punteros a los bloques de datos del archivo.

Existen varias formas de organizar de forma lógica los archivos. Entre las más importantes se encuentran:

- La pila.
- El archivo secuencial.
- El archivo secuencial indexado.
- El archivo indexado.
- El archivo directo o *hash*.

Organización pila

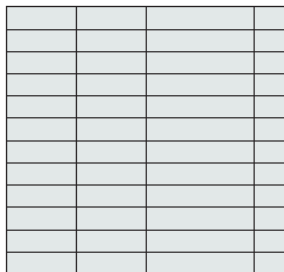
Los registros son almacenados conforme van llegando. Si se desea buscar un campo, es necesario recorrer el archivo hasta encontrarlo.



Variable-length records
Variable set of fields
Chronological order

Archivo secuencial

Se emplea un formato fijo para los registros. Los tamaños son fijos.



The diagram illustrates a sequential file structure as a grid of 12 rows and 4 columns. Each row represents a record, and each column represents a field within that record. The grid is empty, representing the uniform structure of fixed-length records in a sequential file.

Fixed-length records
Fixed set of fields in fixed order
Sequential order based on key field

Figura: Archivo organizado de forma secuencial.⁵

⁵Tomado de Stallings, Operating Systems.

Archivo secuencial indexado

Se almacenan registros de forma secuencial pero se añaden índices para acceder más rápidamente a cualquier registro, basándose en el índice que se encuentra al inicio del archivo.

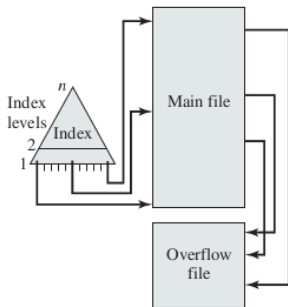
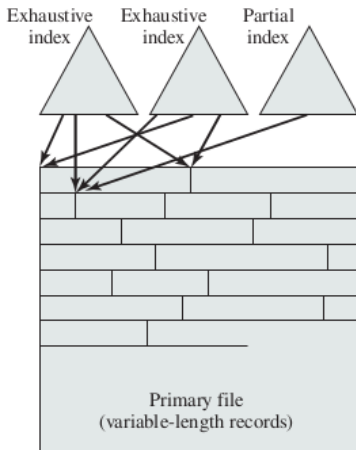


Figura: Archivo organizado de forma secuencial indexada.⁶

⁶Tomado de Stallings, Operating Systems.

Archivo indexado

Se manejan varios tipos de índices. Uno exhaustivo que permite localizar una entrada para cada registro del archivo. El índice parcial contiene las entradas para los campos de interés dentro de los registros.



Los arboles-B son una estructura de datos de árbol utilizada en algunos sistemas de archivos como HFS y Ext4, para organizar los bloques de los archivos.