

## Unidad 2: SISTEMAS ARCHIVOS

Objetivo específico: Concepto de archivo son los métodos y estructuras de datos que un sistema operativo utiliza para seguir la pista de los archivos de un disco o partición.

### 2.1. Concepto de archivo

(File System). Un sistema de archivos es un método para el almacenamiento y organización de archivos de computadora y los datos que estos contienen, para hacer más fácil la tarea encontrarlos y accederlos. Los sistemas de archivos son usados en dispositivos de almacenamiento como discos duros y CD-ROM e involucran el mantenimiento de la localización física de los archivos.

Más formalmente, un sistema de archivos es un conjunto de tipo de datos abstractos que son implementados para el almacenamiento, la organización jerárquica, la manipulación, el acceso, el direccionamiento y la recuperación de datos. Los sistemas de archivos comparten mucho en común con la tecnología de las bases de datos.

Son los métodos y estructuras de datos que un sistema operativo utiliza para seguir la pista de los archivos de un disco o partición; es decir, es la manera en la que se organizan los archivos en el disco. El término también es utilizado para referirse a una partición o disco que se está utilizando para almacenamiento, o el tipo del sistema de archivos que utiliza.

La diferencia entre un disco o partición y el sistema de archivos que contiene es importante. Unos pocos programas (incluyendo, razonablemente, aquellos que crean sistemas de archivos) trabajan directamente en los sectores crudos del disco o partición; si hay un archivo de sistema existente allí será destruido o corrompido severamente. La mayoría de programas trabajan sobre un sistema de archivos, y por lo tanto no utilizarán una partición que no contenga uno (o que contenga uno del tipo equivocado).

Antes de que una partición o disco sea utilizada como un sistema de archivos, necesita ser iniciada, y las estructura de datos necesitan escribirse al disco. Este proceso se denomina *construir un sistema de archivos*.

La mayoría de los sistemas de archivos UNIX tienen una estructura general parecida, aunque los detalles exactos pueden variar un poco. Los conceptos centrales son *superbloque*, *nodo-i*, *bloque de datos*, *bloque de directorio*, y *bloque de indirección*. El superbloque tiene información del sistema de archivos en conjunto, como su tamaño (la información precisa aquí depende del sistema de archivos). Un nodo-i tiene toda la información de un archivo, salvo su nombre. El nombre se almacena en el directorio, junto con el número de nodo-i. Una entrada de directorio consiste en un nombre de archivo y el número de nodo-i que representa al archivo. El nodo-i contiene los números de varios bloques de datos, que se utilizan para almacenar los datos en el archivo. Sólo hay espacio para unos pocos números de bloques de datos en el nodo-i; en cualquier caso, si se necesitan más, más espacio para punteros a los bloques de datos son colocados de forma dinámica.

### 2.2. Estructura del sistema de archivos

El sistema de archivos o ficheros es el componente del sistema operativo encargado de administrar y facilitar el uso de las memorias periféricas, ya sean secundarias o

terciarias. Sus principales funciones son la asignación de espacio a los archivos, la administración del espacio libre, y la administración del acceso a los datos resguardados. Estructuran la información guardada en una unidad de almacenamiento (normalmente un disco duro de una computadora), que luego será representada ya sea textual o gráficamente utilizando un gestor de archivos. La mayoría de los sistemas operativos manejan su propio sistema de archivos.

Lo habitual es utilizar dispositivos de almacenamiento de datos que permiten el acceso a los datos como una cadena de bloques de un mismo tamaño, a veces llamados sectores, usualmente de 512 bytes de longitud (También denominados clústers). El software del sistema de archivos es responsable de la organización de estos sectores en archivos y directorios y mantiene un registro de qué sectores pertenecen a qué archivos y cuáles no han sido utilizados. En la práctica, un sistema de archivos también puede ser utilizado para acceder a datos generados dinámicamente, como los recibidos a través de una conexión de red (sin la intervención de un dispositivo de almacenamiento).

La estructura de un sistema de archivos de un sistema operativo es el nivel más básico de organización. Casi todas las formas en que un sistema operativo interactúa con sus usuarios, aplicaciones y modelos de seguridad dependen de la manera en que almacena y organiza los archivos en los dispositivos de almacenamiento. El proporcionar una estructura de sistema de archivos común asegura que los usuarios y programas pueden acceder y escribir a los archivos.

Los sistemas de archivos dividen los archivos en dos categorías lógicas:

- archivos compartibles vs. no compartibles
- archivos variables vs. estáticos

Los archivos *compartibles* son aquellos a los que se puede acceder desde varios hosts; mientras que los archivos *no compartibles* sólo están disponibles localmente. Los archivos *variables*, tales como documentos, pueden cambiar en cualquier momento; los archivos *estáticos*, tales como binarios, no cambian sin una actuación por parte del administrador de sistemas.

La razón para visualizar a los archivos de esta manera es para ayudar a correlacionar la función del archivo con los permisos otorgados a los directorios que los sostienen. El modo en que el sistema operativo y sus usuarios interactúan con un archivo dado determina el directorio en el cual estos archivos están ubicados, si ese directorio está montado como de sólo lectura o sólo escritura y el nivel de acceso que cada usuario tiene a ese archivo. El nivel superior de esta organización es crucial. El acceso a los directorios inferiores puede estar restringido o se pueden manifestar problemas de seguridad si el nivel superior es dejado sin organizar o no sigue ninguna estructura rígida.

Rutas y nombre de archivos

La estructura de directorios suele ser jerárquica, ramificada o "en árbol", aunque en algún caso podría ser plana. En algunos sistemas de archivos los nombres de archivos son estructurados, con sintaxis especiales para extensiones de archivos y números de versión. En otros, los nombres de archivos son simplemente cadenas de texto y los metadatos de cada archivo son alojados separadamente.

En los sistemas de archivos jerárquicos, usualmente, se declara la ubicación precisa de un archivo con una cadena de texto llamada "ruta" —o path en inglés—. La nomenclatura para rutas varía ligeramente de sistema en sistema, pero mantienen por lo general una misma estructura. Una ruta viene dada por una sucesión de nombres de directorios y subdirectorios, ordenados jerárquicamente de izquierda a derecha y separados por algún carácter especial que suele ser una diagonal ('/') o diagonal invertida ('\') y puede terminar en el nombre de un archivo presente en la última rama de directorios especificada.

Ejemplo de 'ruta' en un sistema Unix

Así, por ejemplo, en un sistema tipo Unix como GNU/Linux, la ruta para la canción llamada "La canción.ogg" del usuario "Alvaro" sería algo como:

*/home/Alvaro/Mi música/La canción.ogg*

dónde:

'/' representa el directorio raíz donde está montado todo el sistema de archivos.

'home/Alvaro/Mi música/' es la ruta del archivo.

'La canción.ogg' es el nombre del archivo, que se establece como único.

Ejemplo de 'ruta' en un sistema Windows

Un ejemplo análogo en un sistema de archivos de Windows (específicamente en Windows 8) se vería como:

*C:\Users\Alvaro\Music\canción.mp3*

dónde:

'C:' es la unidad de almacenamiento en la que se encuentra el archivo.

'\Users\Alvaro\Music\' es la ruta del archivo.

'canción' es el nombre del archivo.

'mp3' es la extensión del archivo, este elemento, parte del nombre, es especialmente relevante en los sistemas Microsoft Windows, ya que sirve para identificar qué tipo de archivo es y la aplicación que está asociada con el archivo en cuestión, es decir, con qué programa se puede editar o reproducir el archivo. Para la mayoría de los sistemas operativos modernos la extensión del archivo es un complemento burocrático solo útil para la observación del usuario, ya que los entornos de administración de archivos y aplicaciones varias, analizan la información contenida en el principio del interior del archivo (MIME headers) para determinar su función o asociación, la cual normalmente está catalogada en la tabla MIME Content-Type en el sistema. El sistema Windows permite ocultar la extensión de los archivos si el usuario lo desea, de no hacerlo la extensión aparece en los nombres de todos los archivos. En el sistema operativo Windows XP, si el usuario cambia la extensión de un archivo, éste puede quedar inutilizable si la nueva extensión lo asocia a un programa que no tenga la capacidad de editar o reproducir ese tipo de archivo. Algunos usuarios aun así habilitan la

visualización de las extensiones en los sistemas Windows como medida de precaución para evitar virus que utilicen íconos o nombres parecidos a los archivos personales del usuario, ya que la extensión permite identificar a los ficheros .EXE, los ejecutables en Windows.

#### Tipo de sistemas de archivos

- Sistemas de archivos de disco

Un sistema de archivo de disco está diseñado para el almacenamiento de archivos en una unidad de disco, que puede estar conectada directa o indirectamente a la computadora.

- Sistemas de archivos de red

Dentro de esta clasificación encontramos dos tipos de sistemas de archivos: los sistemas de archivos distribuidos (no proporcionan E/S en paralelo) y los sistemas de archivos paralelos (proporcionan una E/S de datos en paralelo).

- Sistemas de archivos de propósito especial

(Special purpose file system). Aquellos tipos de sistemas de archivos que no son ni sistemas de archivos de disco, ni sistemas de archivos de red. Ejemplos: acme (Plan 9), archfs, cdafs, cfs, devfs, udev, ftpfs, Infs, nntpfs, plumber (Plan 9), prodfs, ROMFS, swap, sysfs, TMPFS, wikifs, LUFFS, etc.

#### Integridad de un sistema de archivo

La corrección y validación de un sistema de archivos puede ser comprobada utilizando el comando **fsck**, para reparar cualquier problema menor que encuentre, y alertar al usuario si hay errores irreparables. El código implementado en los sistemas de archivos puede estudiarse de forma muy efectiva y normalmente son causados por fallos de alimentación, hardware defectuoso, o errores de operación; por ejemplo, no apagar el sistema adecuadamente.

#### Sistema de archivo y sistema operativo

En realidad, la elección de un sistema de archivos depende en primer lugar del sistema operativo que esté usando. Generalmente, cuanto más reciente sea el sistema operativo, mayor será el número de archivos que admita. Por esto, se necesita contar con FAT16 en DOS y en las primeras versiones de Windows 95.

#### Coexistencia de varios sistemas de archivo

Cuando existen varios sistemas operativos, se debe elegir un sistema de archivos para cada uno, teniendo en cuenta que es posible que deba acceder a los datos de un sistema operativo desde otro. Una solución sería la de usar particiones FAT para los sistemas, con particiones no mayores a 2 GB.

#### Sistemas de archivo virtuales

Los sistemas de archivos virtuales son sistemas de archivos basados en memoria que proporcionan acceso a información de núcleo especial y utilidades. La mayoría de los sistemas de archivos virtuales no utilizan espacio en disco de sistemas de archivos. Además, algunos sistemas de archivos virtuales, como el sistema de archivos temporal (TMPFS), utiliza el espacio de intercambio en un disco.

#### Sistemas de archivo temporal

El sistema de archivos temporal (TMPFS) utiliza memoria local para lecturas y escrituras de sistemas de archivos. Con TMPFS pueden mejorar el rendimiento del sistema al evitar el costo de leer y escribir archivos temporales en un disco local o en la red.

#### Ejemplo de sistema de archivo a utilizar

Actualmente, el más popular sistema de archivos es ext3, debido a que es un sistema de archivos con bitácora. Reiserfs es otra elección popular porque también posee bitácora. Dependiendo de la sobrecarga del listado de estructuras, velocidad, fiabilidad (percibible), compatibilidad, y otras varias razones, puede ser aconsejable utilizar otro sistema de archivos.