

## **TEMA 4**

### **SISTEMAS DE ARCHIVOS**

**4.1 INTRODUCCION:** Todas las aplicaciones computarizadas necesitan almacenar y recuperar la información. Las condiciones para almacenar información a largo plazo son:

- almacenar gran cantidad de información.
- La información debe sobrevivir a la terminación del proceso que la usa.
- Posibilitar que varios procesos accedan a la información.

➔ La solución es el almacenamiento de la información en discos y otros medios externos en unidades llamadas **archivos**. Los archivos deben ser **persistentes** (no le afecta la creación o terminación de un proceso).

#### **4.2 FUNCIONES DEL SISTEMA DE ARCHIVO**

El archivo es una colección de datos con nombre. Se manipulan como una unidad por operaciones como (abrir, cerrar, crear, destruir, copiar); los elementos de datos individuales dentro del archivo se manipulan por operaciones como (leer, escribir, insertar, borrar).

**4.4 ARCHIVOS:** Un archivo es un conjunto de registros relacionados”. Los nombres de los archivos deben ser únicos dentro de un directorio. Algunos sistemas se distinguen con mayúscula y minúscula otro no, el nombre tiene dos partes separados por un punto. *nombre.extensión* (que indica algo del archivo)

La estructura de un archivo puede ser secuencia de bytes, secuencia de registros y árbol (con un campo clave).

La estructuras de archivos pueden ser: secuencia de bytes, secuencia de registros y árbol (con un campo clave).

**Los tipos de archivos son:**

Archivos regulares con información de los usuarios

Directorios: archivos de sistema para mantener una estructura del sistema de archivos.

Archivos especiales de caracteres: (E/S) y archivos especiales de bloques (modelan discos)

**Los tipos de acceso son:** \*\*Secuencialmente (de principio a fin) \*\*Aleatoriamente (en cualquier orden)

Con amor floppy

**4.3 EL SISTEMA DE ARCHIVOS.** Es la parte del S.O. que permite compartir controladamente la información de los archivos, con varios tipos de acceso (lectura, escritura, ejecución).

El sistema de archivos debe proporcionar respaldo y recuperación para prevenir la pérdida accidental o la destrucción de información. Se puede referenciar a un archivo con un nombre, brindando independencia de dispositivo.

El sistema de archivo debe brindar una interface favorable al usuario, suministrando una visión lógica de los datos y las funciones que se van a ejecutar.

El sistema de archivos es un componente importante de un S.O. y suele contener:

- ♣ Métodos de acceso: a los datos almacenados en los archivos.
- ♣ Administración de archivos: para que sean almacenados, compartidos, etc.
- ♣ Administración del almacenamiento auxiliar: para asignar espacio a los archivos
- ♣ Integridad de archivo: se relaciona más con la administración del espacio de almacenamiento secundario (discos); se lo organiza con una raíz que indica que comienza el disco, directorios y archivos.
- ♣ En sistemas jerárquicos el nombre de un archivo es su trayectoria del directorio raíz al archivo.

**4.5 DIRECTORIOS:** Los utiliza el S.O. para llevar un registro de los archivos, en muchos sistemas son también archivos. La organización de directorios es muy variada y puede ser:

- ♣ Directorio único: un directorio raíz y todos los archivos de los usuarios.
- ♣ Directorio por usuario: un directorio por cada usuario, a parte el raíz.
- ♣ Árbol de directorios: cada usuario tiene muchos directorios.

El nombre de las rutas de acceso son: ➔ absoluta: trayectoria raíz => archivo

➔ Relativo: se utiliza junto con el concepto de directorio de trabajo o directorio activo.

Las operaciones con directorios: las llamados al sistema permitidos por el manejo de directorios son ➔ crear, eliminar (debe estar vacío), abrir, cerrar, leer, ligar, desligar, etc.

#### **4.6 IMPLANTACION DEL SISTEMA DE ARCHIVOS Y SU RELACION CON LA SIGNACION Y LIBERACION DE ESPACIO:**

Se tiene en mente la forma de almacenamiento de archivos y directorios, la administración de espacio en disco, y la forma de hacerlo bien eficiente y confiable. También hay que tener en cuenta el problema de la fragmentación creciente del espacio en disco, qué trae problemas de performance porque hace que los archivos se dispersen a través de los bloques separados. Para aliviar esto se utiliza la técnica de recogida de residuos periódicamente a una reorganización de archivos.

**Implantación de archivos:** el aspecto clave es el registro de los bloques asociados a cada archivo. Los métodos usados son la “asignación contigua o adyacente del almacenamiento secundario, es fácil porque solo necesita el número de bloque del inicio para localizar un archivo y rinde bien en la E/S; el problema es que debe conocer el tamaño máximo del archivo al crearlo y produce fragmentación de discos. En la asignación “NO contigua” el almacenamiento es más dinámico. Los archivos están dispuestos en todo el disco y produce largas búsquedas; el mantenimiento implica sobrecarga. Uno de los esquemas que utiliza este método es La asignación por “Bloques” o por “Nodos-i” (nodos índice) el cual se asocia a cada archivo pequeño tabla o “nodo índice” que contiene los atributos y direcciones en disco de los bloques del archivo, va del disco a la memoria primaria al abrirlo.

**Implantación de directorios:** para abrir un archivo el S.O. utiliza información del directorio que contiene la dirección del bloque en el disco. La función del sistema de directorios es asociar el nombre del archivo con la información necesaria para localizar los datos.

**Archivos compartidos:** conviene que estos archivos compartidos aparezcan en distintos directorios de distintos usuarios. La conexión con un directorio y un archivo de otro directorio el cual comparten se llama “ENLACE”. El enlace se produce haciendo que el sistema cree un nuevo archivo de tipo “LINK”. El archivo LINK ingresa al directorio del usuario que accede a un archivo de otro directorio y el usuario; solo contiene el nombre de la ruta de acceso del archivo al cual se enlaza, esto se llama “ENLACE SIMBOLICO”. En búsqueda genérica se suele encontrar el mismo archivo por distintas rutas.

**Administración del espacio en disco:** existen dos estrategias para almacenar un archivo. → asignar “N” bytes consecutivos de espacio en el disco. → Dividir el archivo en bloques no necesariamente adyacentes (mas usadas).

**El tamaño del Bloque:** (unidad de asignación) tienen:

- \* Sector: es **chico** y con archivo constará de muchos bloques.
- \* Pista: de ½ K. 1 K. o 2K.
- \* Cilindro: es **grande** y desperdicia espacio en disco.

**DISK CUOTAS:** Para evitar que los usuarios se apropien de un espacio excesivo en disco, El S.O. multiusuario proporciona un mecanismo que establece cuotas en el disco. Un administrador del sistema asigne a cada usuario una proporción máxima de archivos y bloques y el S.O. garantice que el usuario no exceda su cuota.

**Confiability del sistema de archivos:** se debe proteger la información que está en el sistema de archivo efectuando resguardos; así evitando las consecuencias de la pérdida de los sistemas de archivo. Las pérdidas se suelen deber a problemas de hardware, software, hechos externos, etc.

Si hay bloques defectuosos se solucionan por hardware (con una lista de bloques defectuosos) o por software (con un archivo de bloques defectuosos).

#### **Respallos:**

\* **Copias de seguridad o de back-up:** se respalda los archivos con frecuencia que consiste en realizar copias completas del contenido del disco (flexible o rígido).

\* Otra estrategia es: **El vaciado por incrementos o respaldo incremental:** consiste en obtener una copia de respaldo periódicamente llamada copia total y luego obtener una copia diaria solo de aquellos archivos modificados desde la última copia total.

**Desempeños de los sistemas de archivos:** el acceso al disco es más lento que el acceso a la memoria, los tiempos se miden en Milisegundos para los discos, y nanosegundos para la memoria. Se debe reducir el número de accesos a disco, mediante la técnica del:

- \* **Bloque de Caché (buffer caché):** se utiliza el término de « ocultamiento » del francés “cacher = ocultar”. Un caché es una colección de bloques que pertenecen al disco pero se mantienen en memoria para mejor rendimiento.

✱ Otra técnica para mejorar el rendimiento de un sistema de archivo es reducir la cantidad de movimientos del brazo del disco (mecanismos de acceso): se debe colocar los bloques que probablemente tengan acceso secuencial próximos entre si preferentemente en el mismo cilindro.

**4.7 DESCRIPTOR DE ARCHIVOS o bloque de control de archivos:** Es un “bloque de control” que contiene información que el sistema necesita para administrar un archivo”. Es una estructura dependiente del sistema; y contiene: nombre del archivo, localización del archivo en el almacenamiento secundario, organización del archivo (métodos de acceso); tipo y fechas (creación y destrucción). Los descriptores de archivos suelen mantenerse en el almacenamiento secundario y se pasan al almacenamiento primario al abrir el archivo.

El descriptor de archivo es controlado por el sistema de archivos.

**4.8 SEGURIDAD:** Los sistemas de archivos contienen información valiosa para el usuario que deben proteger.

La seguridad es la garantía de que los archivos no sean leídos o modificados por personal no autorizado; incluyendo aspectos técnicos, de administración, legales y políticos.

El S.O. utiliza mecanismos de protección para resguardar la información de la computadora.

**Dos grandes problemas de la seguridad son:**

1. **Pérdida de datos:** ➔ *Hechos:* (incendios, inundaciones, guerra, ...) ➔ *errores de hardware o software* (falla en el disco, CPU, ...) ➔ *errores humanos:* (pérdida de un disco)
2. **Intrusos:** ➔ *pasivos* (solo lee los archivos no autorizados a leer) ➔ *Activos* (hace cambios no autorizados a los datos)
3. **Ataque del caballo de Troya** modificar un programa normal para que haga cosas adversas además de su función usual.
4. **Gusano de Internet:** consta de un programa arrancador y un gusano. Un gusano es un programa completo que hace daño.
5. **Virus:** son fragmentos de programa que se añaden a programas existentes con la intención de infectar (dañar) a otros. Se propaga a través de copias ilegítimas de programas. El virus se ejecuta o intenta reproducirse cuando el programa que la aloja se ejecuta. El virus es más fácil de evitar que curar.

**Algunas soluciones para diseñar sistemas mas seguros son:**

➔ **Añadir contraseñas:** se debe prever intentos de penetración consistentes en combinaciones de nombres y contraseñas. Deben almacenarse cifradas (encriptadas).

➔ **Identificación física del usuario:** (huellas digitales): forma o longitud de los dedos, huellas digitales, vocales, firmas.

➔ **Medidas preventivas:** (registrar los accesos fallidos) tender trampa a los intrusos.

**4.9 MECANISMOS DE PROTECCION:** El sistema los utiliza para resguardar información. Muchos objetos del sistema necesitan protección (CPU, memoria, disco, impresora, archivos).

➔ **Dominio de protección:** cada objeto se referencia por un nombre y un subconjunto de operaciones que se pueden realizar sobre él. Un dominio es un conjunto de parejas (objeto, derechos)

➔ Un **derecho** es el permiso para realizar algunas de las operaciones.

➔ **Listas de control de acceso:** asocia a cada objeto una lista ordenada con todos los dominios que pueden tener acceso al objeto, y la forma de dicho acceso (lectura, escritura, ejecución).

**4.10 RESPALDOS Y RECUPERACION:** La destrucción de la información, ya sea accidental o intencional (falla de hardware y software, fenómenos meteorológicos, fallas de energía, robos) debe tenerse en cuenta por el S.O. en general pero por el sistema de archivo en particular. Las técnicas utilizadas son ➔ **Respaldos periódicos:** hacer copia de los archivos. ➔ **Pasar todas las transacciones a un archivo:** copiándolo en otro. ➔ **Respaldo incremental:** se resguarda solo aquellos archivos modificados.