

Universidad de Guadalajara

CUCEI



Materia: Sistemas Concurrentes y Distribuidos.

Maestra. Martha del Carmen Gutierrez.

Tarea 8 - Sistemas de Archivos.

Fecha: 31/10/24

Índice

Introducción	3
Aspectos de diseño de los sistemas de archivos	3
Tipos de sistemas de archivos de los diferentes sistemas operativos	3
Windows.	3
Linux.	4
MacOS.	5
Funcionamiento básico de los archivos	5
Características de los archivos	5
Aspectos de diseño de los sistemas de archivos	6
Casos prácticos de Sistemas de Archivos	7
Casos prácticos de Sistemas de Archivos	8
Ejemplificación real de un sistema de archivos distribuido.	8
Conclusión	10
Bibliografía:	10
Participantes: A -> Z	11

Introducción

Los sistemas de archivos son componentes esenciales que permiten a los usuarios y al sistema operativo almacenar, organizar y recuperar información de manera eficiente. Un sistema de archivos es un conjunto de estructuras y reglas que determinan cómo se almacenan y se acceden a los datos en un dispositivo de almacenamiento, como un disco duro, una unidad flash o una nube de almacenamiento.

Los sistemas de archivos desempeñan un papel crucial en la interacción entre los usuarios y la computadora, ya que permiten la creación, el acceso y la gestión de archivos y directorios. Estos sistemas establecen una jerarquía de organización que facilita la ubicación de datos y la administración de recursos de almacenamiento.

Es importante considerar de antemano que la definición de archivo en sistemas como UNIX y MS-DOS es una secuencia de bytes sin interpretación alguna. El significado y estructura queda a cargo de los programas que operen dichos archivos, pero no es de interés para el sistema operativo. Un archivo puede tener atributos, que son partes relativas a él, pero no son parte del archivo propiamente dicho, atributos como propietario, tamaño, fecha de creación, permiso de acceso.

Aspectos de diseño de los sistemas de archivos

Así como en los sistemas convencionales, en los sistemas distribuidos es fundamental que los archivos almacenen programas, datos y estos se encuentren disponibles cuando sea necesario. Comparten muchas características con sistemas centralizados, pero es importante revisar cuáles difieren de los sistemas centralizados.

En el caso de un sistema distribuido es importante distinguir entre un servicio de archivos y un servidor de archivos.

- **Servicio de archivos:** Se refiere al conjunto de utilerías que conforman el servicio que el sistema proporciona a los clientes. Describe las primitivas disponibles, los parámetros que se utilizan y las acciones que se llevan a cabo. Define para el usuario, los componentes del servicio con el que pueden contar, como una interfaz, pero no dice nada de su implantación.
- **Servidor de archivos:** Es un proceso que se ejecuta en una máquina y ayuda a implantar el servicio de archivos. Un sistema puede tener uno o varios servidores de archivos. Los clientes no conocen las características de los servidores de archivos o que éste es distribuido.

Un sistema distribuido de archivos tiene dos componentes distintos: el servicio de archivos y el servicio de directorios. El servicio de archivos se encarga de operaciones en archivos individuales, como lectura, escritura y adición. El servicio de directorios se encarga de crear, administrar directorios, añadir y eliminar archivos de los directorios.

El servicio de archivos proporciona primitivas para leer y escribir atributos relacionados a los archivos.

Tipos de sistemas de archivos de los diferentes sistemas operativos

Windows.

FAT: (File Allocation Table): Es uno de los sistemas de archivos más simples que existe desde la década de 1980, tiene varias versiones

- **FAT12 / FAT16:** FAT12 y FAT16 se aplicaron a los disquetes antiguos y hoy en día no se pueden encontrar con frecuencia. Los números en sus nombres indican la cantidad de bits utilizados para dirigirse a un solo clúster: 12 bits en FAT12, 16 bits en FAT16.
- **FAT32:** Todavía se usa mucho, principalmente debido a su amplia compatibilidad. Es que se puede acceder a él desde casi cualquier sistema operativo, incluidos macOS y Linux, lo que lo convierte en una buena opción para los dispositivos portátiles, como las tarjetas de memoria y memorias USB, pero con un límite de 4 GB por archivo.
- **exFAT:** (Extended File Allocation Table) Similar a FAT32 pero sin límite de tamaño, por lo que es útil en dispositivos de almacenamiento extraíble de gran capacidad.

NTFS: (New Technology File System) Es el sistema de archivos principal de Windows para sus discos duros y SSD. Soporta control de acceso, cifrado, compresión y recuperación ante fallos.

ReFS: ReFS ha sido diseñado para abordar ciertas deficiencias de NTFS, específicamente esas con respecto a la corrupción de datos. Tiene una tolerancia a fallas mucho más alta, diseñado para entornos de servidores y adecuado para grandes volúmenes de almacenamiento.

HPFS: (High Performance File System) fue creado por Microsoft en cooperación con IBM y lanzado al mercado con OS/2 1.20 en 1989 como un sistema de archivos para servidores que ofrece un rendimiento mucho mejor en comparación con FAT. Este sistema de archivos permite nombres de hasta 254 caracteres de doble byte y es más flexible con otras convenciones de nombres. Otro cambio que ofrece HPFS es que sustituye los clusters por sectores físicos de 512 bytes en las unidades de asignación para minimizar el espacio de disco desperdiciado.

Linux.

Ext: Se lanzó en 1992 como el primer formato diseñado específicamente para Linux.

- **Ext2:** Más eficiente debido a su estructura que se basa en el concepto de inodos. Dicho descriptor de índice (inodo) contiene los atributos de un objeto particular, como un archivo o un directorio, y apunta a las ubicaciones de sus datos.
- **Ext3:** Versión mejorada de Ext2 que soporta el registro por diario (journaling), el registro por diario está organizado como un archivo de registro (log) que almacena todos los cambios en el sistema de archivos y lo protege de la corrupción en caso de falla.
- **Ext4:** Cambió el método de asignación de datos de bloques individuales a extents. La idea detrás de esto es escribir la mayor parte de los datos del archivo en un área continua y luego anotar sólo la dirección de su primer bloque y la cantidad de bloques en la secuencia.

XFS: (Extended File System) Este tipo de sistema de ficheros está optimizado para almacenar archivos y volúmenes muy grandes en un solo host.

Btrfs: (B-Tree File System) Está ajustado para funcionar en una amplia gama de dispositivos, desde teléfonos inteligentes hasta servidores de gama alta.

F2FS: (Flash-Friendly File System) Ha sido diseñado específicamente para los dispositivos de almacenamiento basados en la memoria flash NAND y, por lo tanto, se utiliza más ampliamente en los teléfonos inteligentes modernos y medios de almacenamiento portátiles (extraíbles).

JFS: El sistema de archivos de diario (JFS) y el sistema de archivos de diario ampliado (JFS2) están incorporados en el sistema operativo base. Ambos tipos de sistema de archivos enlazan sus datos de archivo y directorio con la estructura utilizada por AIX Logical Volume Manager para el almacenamiento y la recuperación. La diferencia es que JFS2 está diseñado para acomodar un kernel de 64 bits y archivos más grandes.

MacOS.

APFS: (Apple File System) Este formato de nueva generación tiene como objetivo funcionar de manera eficiente con memorias flash y unidades de estado sólido. Incluye funciones como clonación de archivos y carpetas, cifrado completo y mejoras en el rendimiento.

HFS+: (Hierarchical File System Plus) Utilizado en versiones anteriores de macOS. sirvió como el sistema de archivos predeterminado para las computadoras Mac, así como para los productos iPod y Xserve, antes de la aparición de APFS

Funcionamiento básico de los archivos

Almacenamiento de Datos: Los archivos se utilizan para almacenar información de manera persistente en un sistema informático. Pueden contener diversos tipos de datos, como texto, imágenes, videos, programas, hojas de cálculo y más.

Nombres de Archivo: Los archivos se identifican mediante un nombre único. El nombre de archivo suele consistir en un nombre y una extensión separados por un punto (por ejemplo, "documento.txt"). La extensión indica el tipo de archivo y el programa que se utilizará para abrirlo.

Acceso: Los archivos se pueden leer, escribir y eliminar. Los permisos de acceso controlan quién puede hacer qué con un archivo. Por ejemplo, puedes tener permisos para leer un archivo pero no para editarlo.

Extensiones de Archivo: Las extensiones de archivo indican el formato del archivo y se utilizan para determinar qué programa debe abrirlo. Por ejemplo, los archivos con la extensión .docx suelen abrirse con Microsoft Word.

Operaciones Básicas: Las operaciones más comunes en archivos incluyen crear nuevos archivos, abrir archivos existentes, guardar cambios en archivos y eliminar archivos. Estas acciones suelen ser realizadas mediante aplicaciones de software.

Características de los archivos

Un archivo es un contenedor de información. La mayoría de los archivos que se utilizan contienen información (datos) en un formato determinado: un documento, una hoja de cálculo, un gráfico. El formato es la disposición de los datos dentro del archivo. El formato del archivo se conoce como tipo de datos.

Atributos de archivo : Los archivos en sistemas distribuidos poseen varios atributos que incluyen el tipo de archivo (como directorio, enlace simbólico, archivo especial, etc.), la longitud del archivo, el identificador del sistema que contiene el archivo, y la última vez que fue modificado. Los atributos pueden ser leídos o modificados según se necesite.

Operaciones de lectura y escritura : Las operaciones típicas de lectura y escritura permiten al cliente especificar la posición y la cantidad de bytes a manejar en el archivo. En el caso de sistemas como NFS, existen configuraciones para que los datos escritos se aseguren en un almacenamiento estable, lo cual es crucial para la durabilidad frente a fallas.

Gestión de enlaces simbólicos : Los sistemas como NFS utilizan operaciones especiales para manejar los enlaces simbólicos. Estas operaciones permiten a los clientes resolver nombres de ruta mediante el uso de enlaces, brindando flexibilidad en la organización y acceso a los archivos distribuidos.

Cuando el Gestor de archivos está en uno de sus modos de vista de iconos, puede identificar el tipo de datos de un archivo por el icono que se utiliza para representar el archivo. Cada tipo de datos tiene un icono diferente.

La mayoría de los programas de aplicación identifican un número limitado de tipos de datos. Por ejemplo, un editor de documentos probablemente no puede leer un archivo de hoja de cálculo. El escritorio le ayuda a reconocer diferentes tipos de archivos mediante una base de datos de tipos de datos. En la mayoría de los casos, cuando se hace doble clic en un archivo, el escritorio ejecuta automáticamente la aplicación que reconoce el tipo de datos de este archivo.

La longitud máxima permitida para un nombre de archivo varía de un sistema a otro. Algunos sistemas operativos no permiten nombres de archivo superiores a 14 caracteres. Si fuera necesario, consulte al administrador de su sistema.

Aspectos de diseño de los sistemas de archivos

- **Organización Jerárquica:** Los sistemas de archivos suelen organizarse de manera jerárquica en una estructura de árbol. Esto implica la existencia de directorios o carpetas que pueden contener archivos y subdirectorios. La jerarquía facilita la organización y la búsqueda de archivos.
- **Nombres de Archivos:** El sistema de archivos debe permitir nombres de archivos descriptivos y significativos para que los usuarios puedan identificar fácilmente el

contenido de un archivo. Además, es importante considerar restricciones en la longitud y los caracteres permitidos en los nombres de archivos.

- **Permisos de Acceso:** Los sistemas de archivos suelen tener mecanismos de control de acceso para determinar quién puede leer, escribir, ejecutar o eliminar archivos y directorios. Esto es esencial para garantizar la seguridad y la privacidad de los datos.
- **Tipos de Archivos:** Deben definirse y gestionarse los tipos de archivos que el sistema puede manejar. Esto incluye archivos de texto, imágenes, videos, ejecutables, archivos de configuración, etc. Cada tipo de archivo puede requerir un manejo específico.
- **Estructura de Datos:** El diseño debe considerar la estructura de datos utilizada para almacenar información sobre archivos y directorios. Esto incluye la tabla de asignación de archivos (File Allocation Table), nodos de índice, o cualquier otra estructura utilizada para rastrear la ubicación de los datos en el almacenamiento.
- **Recuperación de Errores:** Se deben implementar mecanismos de recuperación de errores para lidiar con posibles fallos de hardware o software que puedan afectar la integridad de los datos. Esto puede incluir la redundancia de datos y registros de transacciones.
- **Fragmentación:** El diseño debe abordar el problema de la fragmentación, que puede dividir los datos en fragmentos pequeños y dispersos en el almacenamiento. Se pueden utilizar técnicas como la fragmentación externa o interna para minimizar este problema.
- **Eficiencia en el Acceso:** El sistema de archivos debe ser eficiente en términos de acceso a los datos. Esto implica considerar la ubicación física de los datos en el almacenamiento para minimizar los tiempos de búsqueda y acceso.
- **Capacidad de Escalabilidad:** El sistema de archivos debe ser escalable para manejar un aumento en la cantidad de datos a lo largo del tiempo. Esto puede implicar la posibilidad de agregar unidades de almacenamiento adicionales o redistribuir datos de manera eficiente.
- **Compatibilidad:** En sistemas operativos multiplataforma, es importante considerar la compatibilidad entre sistemas de archivos. Los sistemas de archivos deben ser compatibles con diferentes sistemas operativos para facilitar el intercambio de datos.
- **Sistemas de Archivos Distribuidos:** En entornos distribuidos, el diseño debe abordar la replicación de datos, la consistencia y la tolerancia a fallos para garantizar la disponibilidad y la confiabilidad de los datos en múltiples ubicaciones.
- **Recuperación y Respaldo:** Debe existir un plan de recuperación y respaldo sólido para proteger los datos contra pérdidas accidentales o ataques cibernéticos.

Casos prácticos de Sistemas de Archivos

Un sistema de archivos distribuido, o DFS, es un esquema de almacenamiento y gestión de datos que permite a los usuarios o a las aplicaciones acceder a archivos de datos como PDF, documentos de Word, imágenes, archivos de vídeo, archivos de audio, etc., desde un almacenamiento compartido en cualquiera de los múltiples servidores en red. Sus datos

compartidos y almacenados en un clúster de servidores permiten a muchos usuarios compartir recursos de almacenamiento y archivos de datos en múltiples equipos.

Hay dos razones principales por las que una empresa usaría un sistema de archivos distribuidos (DFS):

Para almacenar datos de forma permanente en soportes de almacenamiento secundario. Para compartir información de forma fácil, eficiente y segura entre usuarios y aplicaciones. Como subsistema del sistema operativo del equipo, mediante sistemas de archivos distribuidos (DFS) se gestiona, organiza, almacena, protege, recupera y comparte los archivos de datos. Las aplicaciones o los usuarios pueden almacenar o acceder a los archivos de datos en el sistema como lo harían con un archivo local. Desde sus ordenadores o teléfonos inteligentes, los usuarios pueden ver todas las carpetas compartidas de DFS como una ruta única que se ramifica en una estructura arbolada a los archivos almacenados en varios servidores

El sistemas de archivos distribuidos (DFS) tiene dos componentes críticos:

Transparencia de la ubicación – esto significa que los usuarios verán un único espacio de nombres para todos los archivos de datos, independientemente del ordenador que utilicen para acceder o almacenar los archivos. Los usuarios no podrán saber dónde se almacenó el archivo por primera vez y podrán mover archivos dentro de las carpetas según sea necesario sin tener que cambiar el nombre de la ruta.

Redundancia – mediante una característica de replicación de archivos, DFS extiende copias de un archivo a través de los nodos del clúster, lo que significa que los datos permanecen altamente disponibles, incluso en caso de fallo del servidor.

Casos prácticos de Sistemas de Archivos

Ejemplificación real de un sistema de archivos distribuido.

Un ejemplo real de un sistema distribuido es Distributed File System (DFS) de Microsoft, específicamente su implementación más reciente conocida como DFS Replication (DFSR). DFSR es un sistema de archivos distribuido desarrollado por Microsoft que proporciona replicación eficiente de archivos en entornos empresariales distribuidos, ofreciendo tolerancia a fallos, ancho de banda optimizado y escalabilidad.

Arquitectura distribuida: DFSR está diseñado para entornos de red empresariales donde se requiere la replicación eficiente de archivos entre múltiples ubicaciones geográficas.

Replicación de archivos: DFSR permite la replicación bidireccional y multi-maestro de archivos entre servidores en diferentes ubicaciones. Esto significa que los archivos pueden ser actualizados tanto en el servidor central como en las réplicas locales, y los cambios se sincronizan automáticamente en todas partes.

Tolerancia a fallos: DFSR proporciona tolerancia a fallos al replicar archivos en múltiples servidores. Si un servidor falla, los usuarios pueden acceder a los archivos desde otro servidor replicado sin interrupción del servicio.

Ancho de banda eficiente: DFSR utiliza algoritmos de replicación eficientes para minimizar el uso del ancho de banda de la red. Solo se transfieren las diferencias entre los archivos (delta) en lugar de enviar todo el archivo cada vez que se produce una actualización.

Integración con Active Directory: DFSR se integra estrechamente con Active Directory, lo que facilita la administración y el control de acceso a los archivos a través de las políticas de grupo y las asignaciones de permisos.

Escalabilidad: DFSR es altamente escalable y puede manejar grandes cantidades de datos distribuidos en múltiples ubicaciones geográficas, lo que lo hace adecuado para empresas con operaciones globales.

Ejemplo: HDFS (Hadoop Distributed File System)

Caso Práctico: Análisis de Big Data en una Empresa de E-commerce

1. Contexto: Una empresa de e-commerce maneja grandes volúmenes de datos generados por usuarios, como historiales de compras, comportamientos de navegación y preferencias. Estos datos son esenciales para realizar análisis, generar recomendaciones personalizadas y optimizar la experiencia de usuario.

2. Necesidad: Para procesar y analizar estos datos masivos, la empresa requiere almacenar y gestionar petabytes de información de manera distribuida y confiable. Necesitan un sistema de archivos que permita acceder a estos datos desde múltiples servidores en paralelo.

3. Implementación: HDFS es implementado sobre un clúster de servidores, donde los datos se fragmentan en bloques y se distribuyen entre múltiples nodos. Cada archivo se divide en fragmentos de datos de tamaño fijo (normalmente 64 o 128 MB) y se almacena de manera redundante (típicamente con una réplica en tres nodos diferentes), asegurando así la disponibilidad en caso de fallos.

4. Ventajas: Con HDFS, la empresa puede:

- Ejecutar análisis de grandes volúmenes de datos en paralelo mediante herramientas como Apache Spark o MapReduce.
- Tolerar fallos de hardware gracias a la replicación de datos, lo que evita la pérdida de información.
- Escalar el sistema fácilmente añadiendo más nodos al clúster conforme aumentan las necesidades de almacenamiento y procesamiento.

Resultado

La empresa logra realizar análisis en tiempo casi real, permitiendo personalizar las recomendaciones y mejorar la experiencia de usuario. Además, optimizan el uso de los recursos computacionales, lo cual contribuye a una mejor eficiencia operativa y un ahorro en costos de infraestructura.

Este es solo uno de los ejemplos de cómo un sistema de archivos distribuido como HDFS se utiliza para resolver problemas complejos de almacenamiento y procesamiento de datos en empresas con grandes volúmenes de información

Conclusión

Los sistemas de archivos son fundamentales para la organización, almacenamiento y recuperación de datos, y su diseño impacta significativamente la eficiencia y confiabilidad del sistema operativo. Cada sistema operativo emplea distintos tipos de sistemas de archivos que se adaptan a sus necesidades específicas, desde los ampliamente compatibles como FAT y exFAT en Windows, hasta sistemas optimizados para dispositivos de estado sólido como APFS en macOS y sistemas robustos para servidores como Ext4 y Btrfs en Linux.

Los sistemas de archivos distribuidos, como DFS de Microsoft y HDFS en entornos de big data, representan avances que permiten a las organizaciones gestionar y acceder a enormes volúmenes de datos de manera eficiente, segura y tolerante a fallos. La arquitectura distribuida y las capacidades de replicación ofrecen una alta disponibilidad y escalabilidad, elementos clave para las aplicaciones modernas que exigen un acceso continuo y rápido a los datos. En conjunto, los sistemas de archivos facilitan la interacción entre el usuario y la máquina, asegurando que la información esté siempre disponible, segura y organizada.

Los sistemas de archivos son elementos esenciales en los sistemas operativos, permitiendo almacenar, organizar y recuperar información. Estos sistemas se diferencian según las necesidades y entornos de cada dispositivo, como Windows, Linux y macOS, que emplean diferentes estructuras y capacidades. Mientras que FAT32 y NTFS en Windows brindan compatibilidad y seguridad, sistemas como Ext4 y Btrfs en Linux ofrecen optimización para servidores. En entornos distribuidos, los sistemas de archivos como DFS y HDFS facilitan la administración y disponibilidad de grandes volúmenes de datos, esenciales para el análisis y gestión eficiente de la información. La arquitectura distribuida y la replicación de datos en estos sistemas aseguran alta disponibilidad y tolerancia a fallos, mejorando la eficiencia y escalabilidad en aplicaciones empresariales y de big data. En conclusión, el diseño de los sistemas de archivos impacta la interacción usuario-máquina, garantizando seguridad, disponibilidad y eficiencia en la gestión de la información.

Bibliografía:

- Tanebaum Andrew. (1995). *Sistemas Operativos Distribuidos*. España. Prentice-Hall Hisp.
- LLC SysDev Laboratories. (2023, abril 25). *Guía para los sistemas de archivos de Windows*. UFS Explorer; LLC SysDev Laboratories.
<https://www.ufsexplorer.com/es/articles/windows-file-systems/>
- LLC SysDev Laboratories. (2023, abril 25). *Guía para los sistemas de archivos de Linux*. UFS Explorer; LLC SysDev Laboratories.
<https://www.ufsexplorer.com/es/articles/linux-file-systems/>

- LLC SysDev Laboratories. (2023, abril 25). Guía para los sistemas de archivos de macOS. UFS Explorer; LLC SysDev Laboratories.
<https://www.ufsexplorer.com/es/articles/mac-os-file-systems/>
- Deland-Han. (2023, 26 diciembre). Introducción a los sistemas de archivos FAT, HPFS y NTFS - Windows Client. Microsoft Learn.
<https://learn.microsoft.com/es-es/troubleshoot/windows-client/backup-and-storage/fat-hpfs-and-ntfs-file-systems>
- AIX 7.3. (2023). <https://www.ibm.com/docs/es/aix/7.3?topic=types-jfs-jfs2>

Participantes: A -> Z

Aceves Moreno Ulises Emmanuel - 219748081
 Alvarez Garcia Allan Rick - 221350214
 Avila Calderon Jose Francisco - 220286903
 Barroso Monroy Maria Fernanda - 218128918
 Bravo Aguilera Steven - 214561056
 Bryan Eduardo Zarate Miramontes - 217444026
 Cisneros Casillas Luis Javier - 215424923
 Covarrubias Castillo Jose Aron - 215661135
 Espinosa Bramasco Rafael Antonio
 Garcia Garcia Hector Emmanuel - 217585177
 Garduño Arreola Eric David - 220791322
 Haro González Roberto - 216752436
 Hernandez Orozco Juan Pablo - 219294285
 Martínez Castillo Lisseth Abigail - 218292645
 Matuz Gómez Braulio Alanni - 219293637
 Menendez Gomez Jose Dario - 220286458
 Munguia Guizar Marlon Uriel - 220791039
 Muñoz Espinoza Martin - 216573604
 Muñoz Remus Miguel - 213417024
 Orozco Guzman Oscar Etienne - 219293823
 Padilla Perez Jorge Daray - 216584703
 Reyes Diaz Yoel - 216587265
 Rodriguez Saenz Óscar Fernando - 217570811
 Rodriguez Tarula Alan Michel - 219747964
 Veliz Luna Víctor Ismael - 216752665
 Veloz Alcaraz Axel Abraham - 217499726