

Marco Teórico: Sistemas de archivos distribuidos

Simón Gómez, Sara María Castrillon y Manuela Tolosa

Universidad EAFIT

ST0263 TÓPICOS ESPECIALES EN TELEMÁTICA

Profesor: *Edwin Nelson Montoya Munera*

10 de septiembre de 2023

Medellín, Antioquia

Sistema de Archivos

El sistema de archivos es la forma más elemental que tiene un computador para controlar la forma que se almacenan, buscan y recuperan archivos, es una parte esencial de cada sistema operativo. El sistema de archivos es por lo general, jerárquico, donde esta jerarquía está definida por las rutas donde la parte más a la izquierda tiene mayor jerarquía.

Entre sus funciones están:

1. Asignar un espacio de memoria persistente para cada archivo.
2. Gestionar el acceso a datos protegidos.
3. Administrar el espacio libre.
4. Estructurar la información en el hardware de almacenamiento (disco duro, ssd...).

Sistema de archivos distribuidos

Un sistema de archivos distribuido (DFS) es un esquema de almacenamiento y gestión de datos que permite a los usuarios o 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. Los 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 DFS: para almacenar datos de forma permanente en soportes de almacenamiento secundario y para compartir información de forma fácil, eficiente y segura entre usuarios y aplicaciones. Como subsistema del sistema operativo del equipo, mediante 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.

El DFS tiene dos componentes críticos: la transparencia de la ubicación y la redundancia. La transparencia de la ubicación 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. La redundancia se logra mediante una característica de replicación de archivos, donde 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.

Sistema de Archivos de red

El Network File System (NFS) es un protocolo de nivel de aplicación que se utiliza para sistemas de archivos distribuidos en un entorno de red de computadoras de área local. Fue desarrollado originalmente por Sun Microsystems en 1984. Este permite que distintos sistemas conectados a una misma red accedan a ficheros remotos como si se tratara de locales.

El sistema NFS está dividido al menos en dos partes principales: un servidor y un cliente. Los clientes acceden de forma remota a los datos que se encuentran almacenados en el servidor. Las estaciones de trabajo locales utilizan menos espacio de disco debido a que los datos se encuentran centralizados en un único lugar pero pueden ser accedidos y modificados por varios usuarios.

Los usuarios no necesitan disponer de un directorio “home” en cada una de las máquinas de la organización. Los directorios “home” pueden crearse en el servidor de NFS para posteriormente poder acceder a ellos desde cualquier máquina a través de la infraestructura de red. Todas las operaciones sobre ficheros son atómicas. Esto significa que la operación solo retorna cuando el servidor ha completado todo el trabajo asociado para esa operación.

En el caso de Windows Server, NFS proporciona una solución de uso compartido de archivos para empresas que tienen entornos heterogéneos que incluyen equipos Windows y equipos que

no son de Windows. Con el protocolo NFS se pueden transferir archivos entre equipos que ejecutan Windows y otros sistemas operativos que no son Windows, como Linux o UNIX.

En resumen, NFS es una herramienta poderosa y flexible que permite compartir archivos y recursos entre diferentes sistemas operativos, mejorando la eficiencia y la colaboración en entornos de red.

Almacenamiento de objetos

El almacenamiento de objetos, también conocido como almacenamiento basado en objetos, es una arquitectura de almacenamiento de datos que maneja los datos como objetos, en lugar de otras arquitecturas de almacenamiento como los sistemas de archivos, que manejan los datos como una jerarquía de archivos, y el almacenamiento de bloques, que maneja los datos como bloques dentro de sectores y pistas⁴. Normalmente, cada objeto incluye el propio dato, una cantidad variable de metadatos y un identificador global único.

Algunos sistemas de archivos distribuidos utilizan una arquitectura basada en objetos, donde los metadatos del archivo están almacenados en servidores de metadatos y los datos del archivo están almacenados en servidores de almacenamiento de objetos. El almacenamiento distribuido es un sistema de almacenamiento definido por software que permite el acceso a los datos en cualquier momento, desde cualquier lugar y solo a aquellas personas que queramos que accedan.

El almacenamiento de objetos se utiliza para propósitos como almacenar fotos en Facebook, canciones en Spotify o archivos en servicios de colaboración en línea, como Dropbox⁴. El lugar de almacenamiento real (local o remoto) es irrelevante. A diferencia del almacenamiento clásico en archivos, en el almacenamiento de objeto los metadatos se pueden elegir libremente. Por lo tanto, se pueden introducir metadatos específicos para cada archivo. Esto permite integrar información de la aplicación correspondiente o datos de usuario.

Para acceder a los datos (y comunicarse con el sistema de almacenamiento) se utiliza el protocolo de transferencia de hipertexto (HTTP) y una API REST. Esto simplifica en gran medida la conexión entre el cliente y el almacenamiento de objeto, ya que el protocolo HTTP es compatible con todos los ordenadores. La comunicación con el sistema de almacenamiento solo requiere unos pocos comandos: PUT para crear un objeto, GET para leer un objeto, DELETE para borrar un objeto y LIST para crear una lista de todos los objetos.

Cuales son los problemas que puede generar usar los sistemas de archivos distribuidos

Los sistemas de archivos distribuidos (DFS) son una forma de almacenamiento y gestión de datos que permite a los usuarios o aplicaciones acceder a archivos de datos desde un almacenamiento compartido en múltiples servidores en red. Aunque los DFS tienen muchas ventajas, también pueden presentar algunos desafíos y problemas. Algunos de los problemas que pueden surgir al usar sistemas de archivos distribuidos incluyen:

- **Inconsistencia:** El uso de caché puede introducir problemas de inconsistencia. Una posible solución es la escritura a través de caché, donde cuando se modifica una entrada de la caché, el nuevo valor se envía también de inmediato al servidor. Sin embargo, esto puede aumentar la latencia y reducir el rendimiento.
- **Complejidad:** Los sistemas distribuidos pueden ser más complejos y menos intuitivos que otros tipos de sistemas informáticos debido a problemas interrelacionados como fallas independientes y falta de determinación. Además, las fallas en los sistemas distribuidos pueden ocurrir de muchas formas y no siempre es posible saber si algo ha dejado de funcionar.

- Latencia: La latencia puede ser un problema en los sistemas distribuidos, ya que los datos deben transmitirse a través de la red entre diferentes servidores. Esto puede afectar el rendimiento y la velocidad de acceso a los datos.
- Gestión de errores: La gestión de errores en sistemas distribuidos puede ser más difícil debido a la complejidad y la variedad de formas en que pueden ocurrir fallas. Es importante tener un plan sólido para manejar errores y recuperarse rápidamente en caso de fallos del sistema.

Referencias

- Sandberg, R. Goldberg, D. Kleiman, S. Walsh D. Lyon, B. (June 1985). «Design and Implementation of the Sun Network File System» (en inglés). Proceedings of the Summer 1985 Usenix Conference.
- ¿Qué es el almacenamiento distribuido? (s/f). Nutanix. Recuperado el 6 de septiembre de 2023, de <https://www.nutanix.com/es/info/distributed-storage>
- Object storage: la solución de almacenamiento moderna. (2023, marzo 1). IONOS Digital Guide; IONOS. <https://www.ionos.es/digitalguide/servidores/know-how/object-storage/>
- (S/f). Amazon.com. Recuperado el 6 de septiembre de 2023, de <https://aws.amazon.com/es/builders-library/challenges-with-distributed-systems/>
- *Funciones del sistema de archivos distribuido (DFS)*. (s/f). Microsoft.com. Recuperado el 6 de septiembre de 2023, de <https://learn.microsoft.com/es-es/windows/win32/dfs/distributed-file-system-dfs-functions>
- (S/f). Usal.es. Recuperado el 6 de septiembre de 2023, de <http://vis.usal.es/rodrigo/documentos/sisdis/teoria/6-archivos.pdf>