

TEMA 11

SISTEMAS DISTRIBUIDOS DE ARCHIVOS

11.1 INTRODUCCION A LOS SISTEMAS DISTRIBUIDOS DE ARCHIVOS: En un sistema distribuido se distinguen entre los conceptos de *servicio de archivos* y el *servidor de archivos*.

- El **servicio de archivos**: especifica los servicios que el sistema de archivos ofrece a sus clientes sin decir nada de su implantación.
- El **despachador (servidor) de archivos**: Es un proceso que se ejecuta en una maquina y ayuda a implantar el servicio de archivos. Puede haber uno o varios en un sistema.

11.2 DISEÑO DE LOS SISTEMAS DISTRIBUIDOS DE ARCHIVOS: Los componentes de un sistema distribuido de archivos son:

- El verdadero servicio de archivos: realiza operaciones en los archivos como lectura, escritura.
- El servicio de directorios: crea y maneja directorios, añade y elimina archivos de los directorios.

11.3 LA INTERFAZ DEL SERVICIO DE ARCHIVOS: La protección en los sistemas distribuidos usa las mismas técnicas de los sistemas uniprosesador.

Posibilidades: permiso de cada usuario para acceder a un objeto.

Listas para control de acceso: cada archivo tiene una lista de usuarios que pueden acceder al archivo.

Los servicios de archivos se pueden clasificar en:

- **Modelo carga / descarga:** Las principales operaciones son la lectura y escritura en un archivo. Los archivos se pueden almacenar en memoria o en un disco local.
- **Modelo de acceso remoto:** El sistema de archivos se ejecuta con todas las funciones en los servidores y no en los clientes.

11.4 LA INTERFAZ DEL SERVIDOR DE DIRECTORIOS: Proporciona operaciones para crear y eliminar directorios, nombrar y cambiar el nombre de archivos y mover archivos de un directorio a otro. Se utiliza un **sistema jerárquico de archivos**, representado por un árbol de directorios.

En sistemas que utilizan varios servidores de archivos mediante el montaje remoto, los clientes tienen una visión distinta del sistema de archivo, el sistema no se comporta como un único sistema de tiempo compartido.

La **transparencia con respecto a la posición** significa que el nombre de la ruta de acceso no sugiere la posición del archivo.

Resumiendo, los Métodos usuales para nombrar los archivos y directorios en un sistema distribuido son:

- Nombre maquina + ruta de acceso.
- Montaje de sistemas de archivos remotos en la jerarquía local de archivos.
- Un único espacio de nombres que tenga la misma apariencia en todas las maquinas.

11.5 SEMANTICA DE LOS ARCHIVOS COMPARTIDOS: se necesita definir una semántica de lectura-escritura cuando se comparten archivos.

En sistemas monoprocesador que permiten a los procesos compartir archivos (ej.: unix) la semántica es:

- Si una lectura sigue a una escritura, la lectura debe regresar el valor recién escrito.
- Si dos escrituras se realizan en serie y luego se ejecuta una lectura, el valor que se debe regresar es el almacenado en la última escritura.

En un sistema distribuido esta semántica se logra fácil si:

- Solo existe un servidor de archivos.
- Los clientes no ocultan los archivos.

Problemas que se presentan:

- Retrasos en la red (un read llega antes que un write, se obtiene un valor viejo)
- Desempeño pobre de un sistema distribuido, donde todas las solicitudes pasan a un único servidor.

Otra solución para el uso de archivo compartidos en un sistema distribuido es usar las “TRANSACCIONES ATOMICAS”, se garantiza que todas las llamadas contenida en la transacción se lleven a cabo en orden y no habrá interferencia de otras transacciones concurrentes.

11.6 IMPLANTACION DE UN SISTEMA DISTRIBUIDO DE ARCHIVOS :

- **Uso de archivos**
- **La estructura del sistema**

- El ocultamiento
- La duplicación o replica
- El control de la concurrencia

11.7 USO DE ARCHIVOS. Antes de implantar un sistema de archivos se deberá analizar los “*patrones de uso*”. Las principales propiedades observadas son: la mayoría de los archivos son pequeños, la lectura es mas común que la escritura, los accesos son secuenciales, los archivos son de corta vida, es poco usual compartir archivos.

11.8 ESTRUCTURA DEL SISTEMA: en algunos sistemas no existe distinción entre el software de un cliente y el de un servidor, todas las maquinas ejecutan lo mismo.

En otros sistemas el servidor de archivos y el de directorios son solo programas del usuario (se ejecuta el software de cliente y servidor en la misma máquina).

También los clientes y servidores pueden ser máquinas distintas de hardware o software.

Formas de estructurar el servicio a directorios:

- Combinar el servicio a directorios y archivos en un único servidor que administre todas las llamadas.
- Separar el servicio a directorios y archivos utilizando un servidor de directorios y un servidor de archivos.

Otro aspecto estructural es si los servidores de archivos, directorios o de otro tipo deben contener la *información de estado de los clientes*.

Si los servidores no contienen los estados, el servidor lleva a cabo la solicitud y no guarda información en relación a los clientes y las solicitudes. La longitud del mensaje es mayor porque la solicitud incluye información.

Si los servidores incluyen información de estado de los clientes entre las solicitudes. Al abrir un archivo el servidor guarda información que relaciones a los clientes con los archivos abiertos. Al abrir un archivo el cliente recibe un descriptor de archivo que existe para identificar el archivo.

11.9 OCULTAMIENTO: En un sistema cliente-servidor, cada uno con su memoria principal y un disco, Existen cuatro lugares donde se pueden almacenar los archivos o partes de ellos:

- El disco del servidor.
- La memoria principal del servidor.
- El disco del cliente (si existe).
- La memoria principal del cliente.

Cuando el cliente desea leer un archivo se transfiere del disco del servidor a la memoria principal del servidor, y luego recién a la memoria principal del cliente a través de la red.

Esto trae problemas de desempeño y se lo mejora “OCULTANDO” (conservando): los archivos de uso más recientes en la memoria principal del servidor. El cliente lee el archivo desde el cache del servidor y elimina la transferencia de archivo.

11.10 REPLICA: en los sistemas distribuidos de archivo se proporciona la réplica de archivos como un servicio:

- Hay varias copias de archivos
- Cada réplica esta en un servidor de archivo independiente
- La réplica se realiza para:
 - Aumentar la confiabilidad al disponer de respaldos independientes de cada archivo
 - Permite el acceso a archivos aun cuando falle un servidor de archivos.
 - Reparten la carga de trabajo entre varios servidores.

Un sistema de transporte de la réplica si la misma se administra sin intersección del usuario.

11.11 CONCLUSIONES IMPORTANTES RESPECTO DE LA IMPLANTACION DE UN SISTEMA DISTRIBUIDO DE ARCHIVOS: Los principios del diseño de un sistema distribuido de archivos son:

- Las estaciones de trabajo tienen ciclos que hay que utilizar: (si hay que elegir para hacer algo entre una estación de trabajo o un servidor). Se elige una estacion de trabajo porque sus ciclos de CPU son menos costosos que los ciclos de un servidor.
- Utilizar el cache el máximo posible porque ahorra tiempo de cómputo y ancho de banda de la red.
- Explotar las propiedades de uso (eficiencia y sencillos)
- Minimizar el conocimiento y modificación a lo largo del sistema.
- Crear lotes de trabajo mientras sea posible para el mejor desempeño.

11.12 TENDENCIA EN LOS SISTEMAS DISTRIBUIDOS DE ARCHIVOS: Los cambios en el hardware tendrán un efecto importante en los sistemas distribuidos de archivo futuro.

También el cambio de expectativa del usuario.

11.13 CONSIDERACIONES RESPECTO DEL HARDWARE: el abaratamiento de la memoria principal permitirá disponer de servidores con memorias cada vez mayores. Se podrá alojar directamente en memoria el sistema de archivo logrando mayor sencillez y mejor desempeño. También se podrá obtener respaldos continuos ante el corte de energía eléctrica.

La disponibilidad de redes de fibra óptica de alta velocidad permitirá una simplificación del software.

La posible construcción de interfaces de red que resuelvan por hardware problema difícil de soportar por software.

11.14 ESCALABILIDAD: la tendencia de los sistemas distribuidos es hacer sistemas cada vez más grandes. Algunos sistemas distribuidos que operan bien para cientos de máquinas.

Los algoritmos centralizados no se escalan bien porque el servidor puede convertirse en un cuello de botella.

11.15 REDES EN UN AREA AMPLIA. Los sistemas distribuidos se asocian con redes de área local (LAN), pero cada vez será mayor la necesidad de conectar entre si cubriendo grandes áreas (nacionales, regionales, continentales, etc.).

Los sistemas de archivo deberán soportar estas necesidades teniendo en cuenta la heterogeneidad de los equipos, formatos y códigos. Deberá atenderse a los cambios de tendencia de los requerimientos de las aplicaciones.

El ancho de banda de la red en los sistemas distribuido puede ser insuficiente para el desempeño esperado.

11.16 USUARIOS MOVILES: son usuarios de equipos móviles. Están gran parte del tiempo desconectados del sistema de archivos de su organización.

Lo deseable sería un sistema distribuido totalmente transparente para su uso simultáneo por millones de usuarios móviles que frecuentemente se desconectan.

11.17 TOLERANCIA DE FALLOS. La difusión de los sistemas distribuidos incrementa la demanda de sistemas que esencialmente nunca fallen.

Los sistemas tolerantes a fallos requerirán cada vez más redundancia en: Hardware, comunicaciones, software, datos, etc.

La replica de archivos sería un requisito; También que los sistemas funcionen aun con la carencia de parte de los datos.

Los tiempos de fallo aceptables por los usuarios serán cada vez menores.