

Sistemas de arquivos distribuídos

EC0036 - Sistemas Paralelos e
Distribuídos



Sistemas de arquivos distribuídos

- Daniel Nogueira 20938
- Felipe Castro Simões 21525

Sumário

1. Introdução
2. Sistemas de arquivos.
3. Sistemas de arquivos distribuídos.
4. Requisitos de um SAD.
5. Arquitetura de um SAD.
6. Estudo de caso: NFS.
7. Conclusão.
8. Referências bibliográficas.

Introdução

“Um sistema de arquivos distribuídos permite aos programas armazenarem e acessarem arquivos remotos exatamente como se fossem locais, possibilitando que os usuários acessem arquivos a partir de qualquer computador em uma rede.” (COULOURIS, et. al, p. 284).

Sistemas de arquivos

- Foram originalmente desenvolvidos como um recurso do S.O que fornece uma interface de programação conveniente para armazenamento em disco.
- São responsáveis pela organização, armazenamento, recuperação, atribuição de nomes, compartilhamento e proteção de arquivos.
- Projetados para armazenar e gerenciar um grande número de arquivos, com recursos para criação atribuição de nomes e exclusão de arquivos.

Sistemas de arquivos (2)

- Arquivos contêm dados e atributos.

Tamanho do Arquivo
Horário de Criação
Horário de Acesso (Leitura)
Horário de Modificação (Escrita)
Horário de Alteração de Atributo
Contagem de Referência
Proprietário
Tipo de Arquivo
Lista de Controle de Acesso

Sistemas de arquivos (3)

- Diretório é um arquivo de tipo especial;
- Fornece um mapeamento dos nomes textuais para identificadores internos;
- Podem incluir nomes de outros diretórios.

Sistemas de arquivos distribuídos

- Objetivo: permitir que os programas armazenem e acessem arquivos remotos exatamente como se fossem locais.
- Permitem que vários processos compartilhem dados por longos períodos, de modo seguro e confiável.
- O desempenho e segurança no acesso aos arquivos armazenados em um servidor devem ser compatíveis aos arquivos armazenados em discos locais.

Requisitos de um sistema de arquivos distribuídos

- Transparência;
- Atualização concorrente de arquivos;
- Replicação de arquivos;
- Heterogeneidade;
- Tolerância a falha;
- Consistência;
- Segurança;
- Eficiência.

Transparência (1)

“A transparência é definida como sendo a ocultação, para um usuário final ou para um programador de aplicativos, da separação dos componentes em um sistema distribuído de modo que o sistema seja percebido como um todo, em vez de uma coleção de componentes independentes.” (COULOURIS et al., 2008, p. 34)

- Acesso;
- Localização;
- Mobilidade;
- Desempenho;
- Mudança de escala.

Transparência (2)

- Transparência de acesso: Programas clientes não devem ter conhecimento da distribuição de arquivos.
- Transparência de localização: Programas clientes devem ver um espaço de nomes de arquivos uniforme, portanto os arquivos podem ser deslocados de um servidor para outro sem alteração de seus nomes de caminho.

Transparência (3)

- Transparência de mobilidade: Nem os programas clientes nem as tabelas de administração de sistema nos nós clientes precisam ser alterados quando os arquivos são movidos.
- Transparência de desempenho: Programas clientes devem continuar a funcionar satisfatoriamente, mesmo com a oscilação de carga sobre o serviço.
- Transparência de escala: Serviço deve ser extensível para acomodar mudanças no sistema distribuído.

Atualização concorrentes de arquivos

- Alterações feitas em um arquivo por um cliente não devem interferir nas operações de outros clientes, mesmo que esses clientes estejam manipulando o mesmo arquivo.
- Exclusão mutua.
- Serviços de arquivos atuais seguem padrões UNIX.
 - Travamentos em nível de arquivo ou registro.

Replicação de arquivos

- Um arquivo pode ser representado por várias cópias de seu conteúdo em diferentes locais.
- Permite servidores compartilharem a carga de fornecimento de arquivos;
- Melhora a tolerância a falhas.

Heterogeneidade

- As interfaces de serviço devem ser definidas de modo que o software cliente e servidor possa ser implementado para diferentes sistemas operacionais e computadores.

Tolerância a falha

- Por ser parte essencial nos sistemas distribuídos, é essencial que o serviço de arquivo distribuídos continue a funcionar diante de falhas de clientes e servidores.

Consistência

- Só deve haver uma versão disponível de cada arquivo no SAD, ou seja, se algum cliente ler um arquivo X em um momento, e outro ler o mesmo arquivo X em outro momento ou lugar, se X não foi modificado, ambos os clientes devem visualizar exatamente o mesmo conteúdo.

Segurança

- Mecanismos de controle de acesso baseados no uso de listas de controle de acesso.
- O SAD deve ter preocupação com permissão, autenticação, privacidade e integridade.

Eficiência

- Um serviço de arquivos distribuídos deve oferecer recursos que tenham pelo menos o mesmo poder e generalidade daqueles encontrados nos sistemas de arquivos convencionais e deve obter um nível de desempenho comparável.

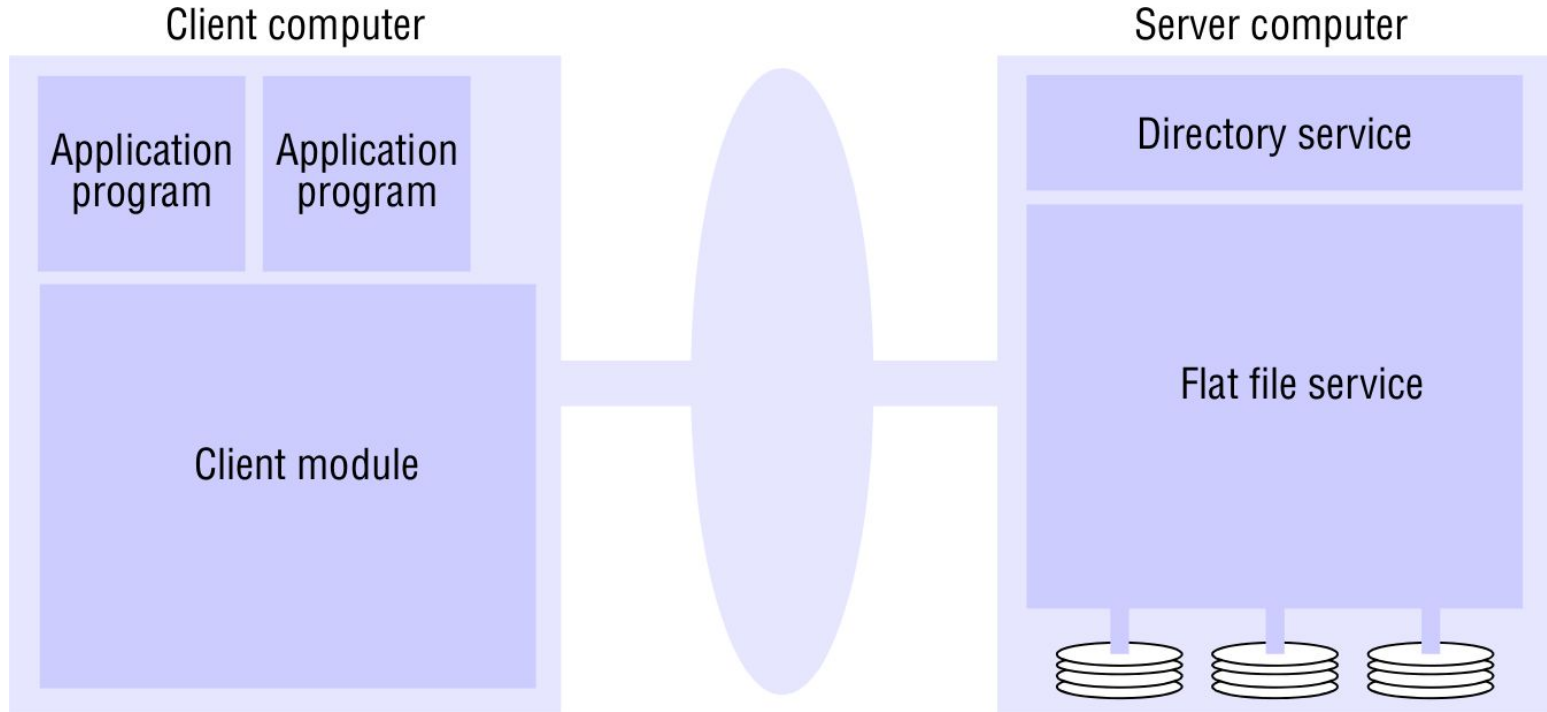
Arquitetura do SAD

- Modelo abstrato de arquitetura que serve para o Network File System (NFS) e Andrew File System (AFS).
- Divisão de responsabilidades entre três módulos.
 - Cliente;
 - Serviço de arquivos planos;
 - Serviço de diretórios.

Arquitetura do SAD (2)

- Design aberto
 - Diferentes módulos cliente podem ser utilizados para implementar diferentes interfaces.
 - Simulação de operações de arquivos de diferentes S.O.
 - Otimização de performance para diferentes configurações de hardware de clientes e servidores.

Arquitetura do SAD (3)



Serviço de arquivos planos

- Se preocupa com a implementação de operações sobre o conteúdo dos arquivos.
- Utiliza UFIDs (Identificadores Únicos de arquivos) de forma que cada arquivo tenha um identificador único no SD.
 - Os UFIDs são utilizados para referenciar arquivos em todas as requisições.

Serviço de diretórios

- Fornece um mapeamento entre nomes textuais de arquivos e seus UFIDs.
- Oferecem as funções necessárias para:
 - Gerar diretórios;
 - Adicionar novos nomes de arquivos a eles e obter suas UFIDs.
- Funciona como um cliente para o serviço de arquivos plano.

Módulo cliente

- Executa em qualquer computador cliente, integrando e estendendo as operações do serviço de arquivos plano e do serviço de diretório.
- Mantém informações sobre a localização de rede dos servidores de arquivos e diretórios.
- Mantém cache de blocos de arquivos recentemente utilizados.

Sun Network File System (1)

- Desenvolvido pela Sun Microsystems em 1984.
- Permite que um usuário, utilizando um computador cliente acesse arquivos em uma rede de forma similar a como o armazenamento local é acessado.
- Módulo servidor no kernel de cada computador que atua como servidor.
- Desenvolvido para UNIX, porém atualmente é independente do S.O.

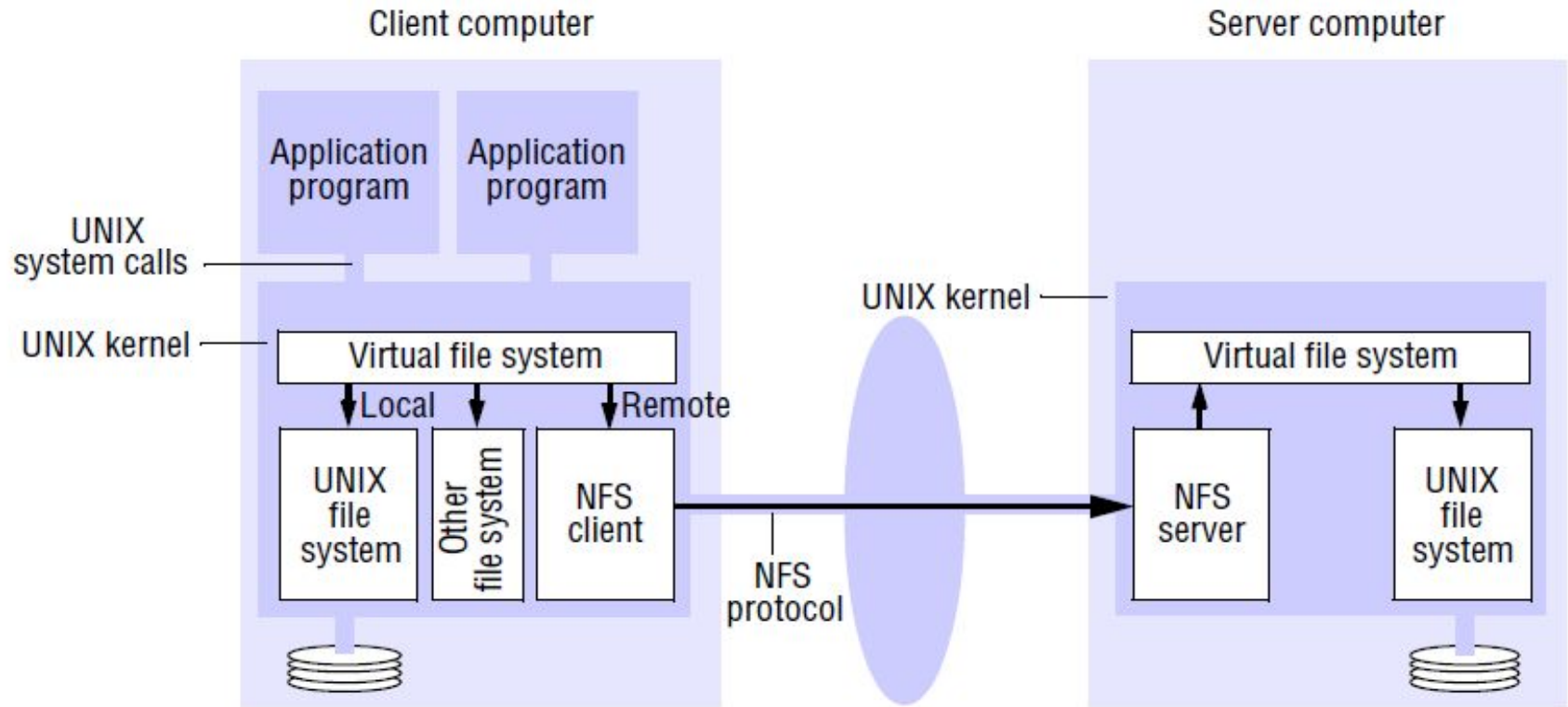
Sun Network File System (2)

- Por que utilizar o NFS?
 - Os dados acessados por todos os usuários podem ser mantidos numa máquina central.
- Exemplo:
 - Manter todas as contas de usuários em uma única máquina, e ter todas as máquinas da rede montando os diretórios pessoais a partir daquela máquina.
 - Dados que consomem muito espaço em disco podem ser mantidos em uma única máquina.

Sun Network File System (3)

- Sistema de arquivos virtual (VFS)
 - Mecanismo pelo qual o NFS prove transparência de acesso.
 - Operações de arquivo para sistemas locais ou remotos sem distinção.
 - Distinção entre os identificadores de arquivos utilizados pelo NFS e os utilizados localmente.

Sun Network File System (4)



Sun Network File System (5)

- Uso de cache.
 - Módulo cliente armazena em cache os resultados das operações para reduzir o número de requisições feitas aos servidores.
 - Implica na possibilidade de existirem diversas versões de arquivos.
 - Os cliente são responsáveis por fazerem uma consulta sequencial no servidor, verificando se a cache que eles possuem são atuais.
- Read-ahead.
- Delay-write.
 - Sync.

Conclusão

- Sistemas de arquivo distribuídos devem ser vistos pelos clientes como um sistema de arquivo local;
- A transparência é muito importante para seu bom funcionamento;
- É altamente necessário um bom controle de concorrência no acesso;
- Cache é importante.

Bibliografia

[1] COULOURIS, G., DOLLIMORE, J., KINDBERG, T. **Sistemas distribuídos conceitos e projetos**. Bookman. Quarta edição. 2005

[2] Sistemas distribuídos. Disponível em: <<http://pt.slideshare.net/computacaodepressao/sd08-si-sistemas-de-arquivos-distribudos>>. Acesso em 11/05/2016

[3] Prof. Dr. Francisco Isidro Massetto. Sistemas de arquivos distribuidos. Disponível em: <<http://professor.ufabc.edu.br/~francisco.massetto/sd/07-Aula7-SistemasArquivos.pdf>>. Acesso em 11/05/2016



Obrigado!

