

# Cloud Computing

---

Gledson Scotti

# Sistemas de Arquivos Distribuídos



# Sistemas de arquivos distribuídos

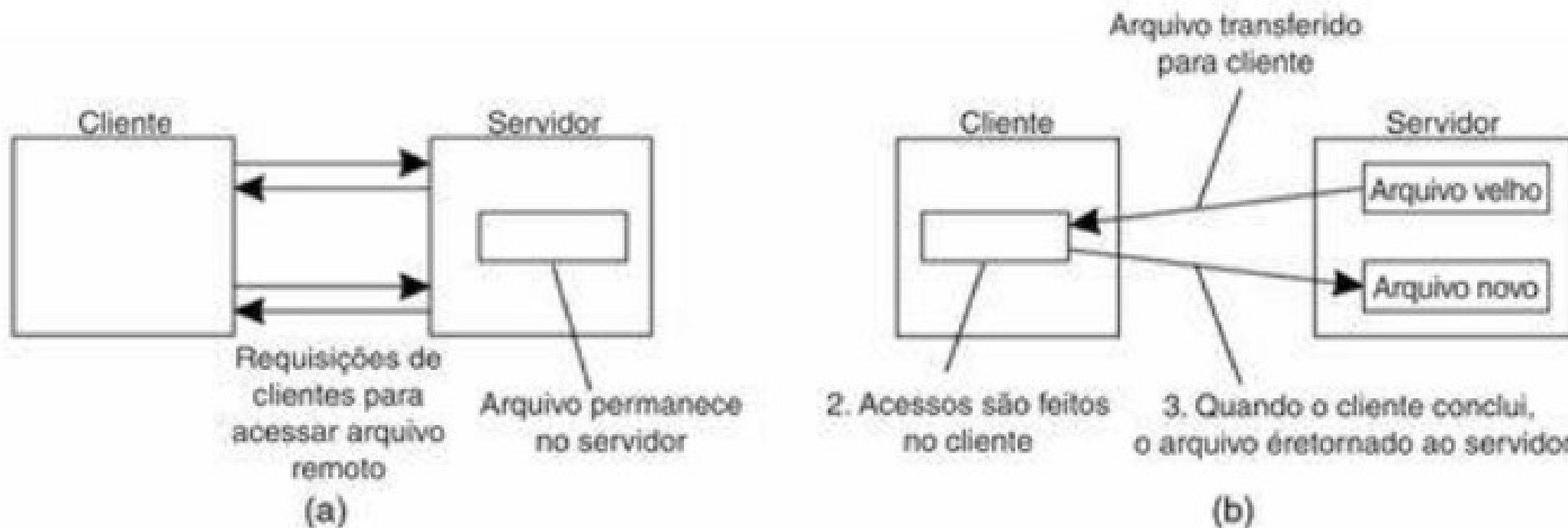
---

- Os SADs permitem o compartilhamento de informações através de recursos de hardware e software. Quando bem projetado, um sistema de arquivo dá acesso a arquivos armazenado em um servidor com desempenho e confiabilidade semelhantes aos arquivos armazenados em discos locais.



# Sistemas de arquivos distribuídos

Organizados no modelo da arquitetura cliente/servidor:



**Figura 11.1** (a) Modelo de acesso remoto. (b) Modelo de carga/atualização.

## Características Importantes:

- Diferentes clientes remotos podendo acessar arquivos em servidores de arquivos;
- Esquema de compartilhamento bem estruturado;
- Clientes dispersos;
- Ponto de vista centralizado.



# Sistemas de arquivos distribuídos

Por que SAD ?

- Devido ao problema do compartilhamento de recursos;
- Compartilhamento de espaço em disco
  - cada maquina não tem que armazenar arquivos que irá acessar;
- Compartilhamento da administração
  - administração e backup fáceis de fazer.
- Acesso de arquivos particulares de diferentes computadores.

O quê um SAD provê ?

- Provê acesso a dados armazenados nos servidores de arquivo através das Interfaces de sistemas de arquivo.
- Interfaces são partes do sistema de arquivo com as funções de:
  - Abrir arquivo, checar estado, fechar;
  - Ler ou escrever dados em um arquivo;
  - Bloquear um arquivo ou partes dele;
  - Listar arquivos em um diretório;
  - Apagar, renomear um arquivo ou diretório, etc.

O que o SAD deve suportar:

- Compartilhamento de informações remotas
- Permitir que um arquivo seja acessado transparentemente em qualquer nó, independente da sua localização.
- Mobilidade do Usuário
- Usuário pode trabalhar em qualquer nó em diferentes momentos (Flexibilidade)

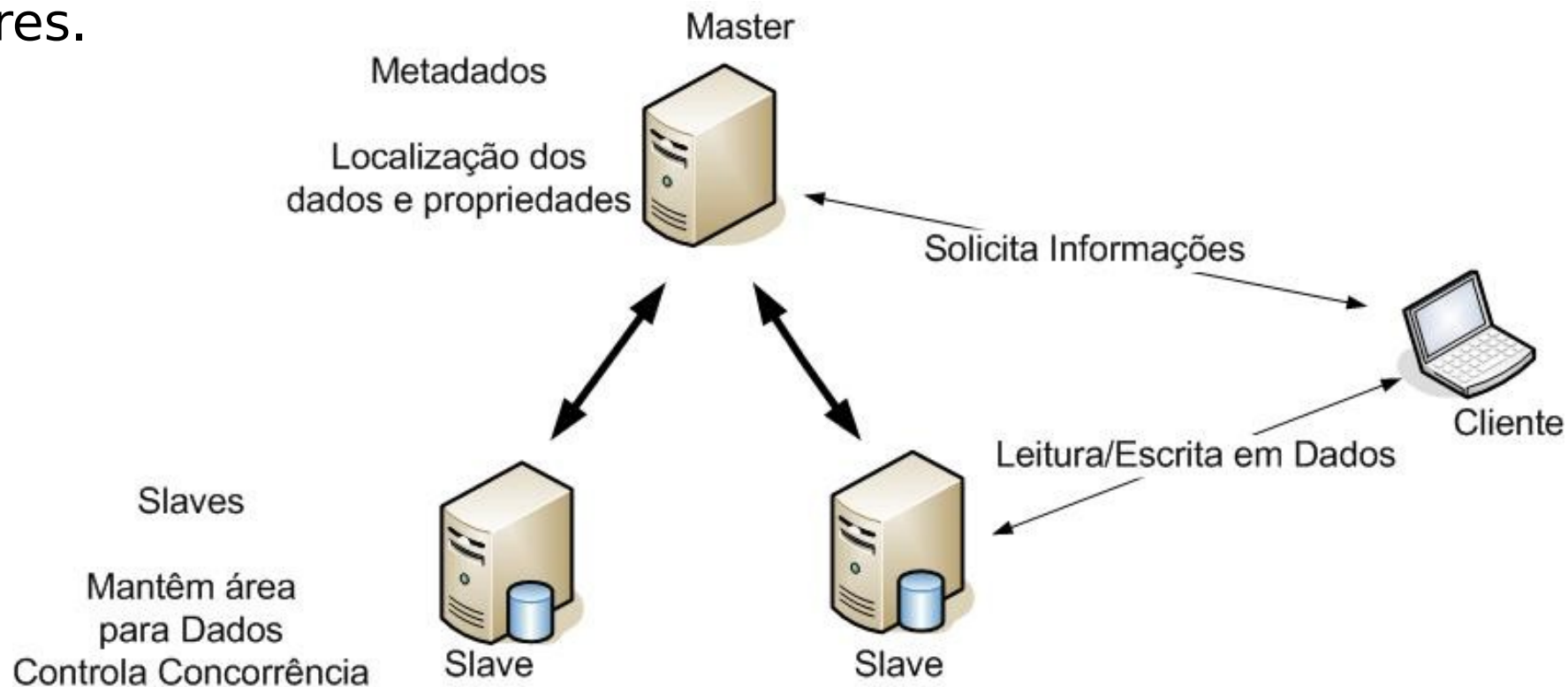




# Sistemas de arquivos distribuídos

Formas de armazenamento:

- O sistema de arquivo inteiro é armazenado em um único servidor.
- Arquivos são armazenados nos vários rígidoss de diferentes computadores.





## **Serviços Ofertados por um SAD:**

- Serviço de Nomes Distribuído:
  - O serviço de nomes cuida de indicar a localização de um determinado arquivo dado o seu nome ou caminho.
- Serviço de Arquivos Distribuído:
  - Responsável por fornecer operações sobre os arquivos que compõe o sistema.
  - Os arquivos podem ser armazenados de diferentes formas, dependendo do seu tipo e uso.



## **Serviços Ofertados por um SAD:**

- Serviço de Diretórios Distribuído:

- Responsável por manter a organização dos arquivos armazenados no sistema.

- Ele fornece uma interface para que os usuários possam arranjar seus arquivos num formato hierárquico, que é estruturado em diretórios e subdiretórios.



## **Características procuradas em um SAD:**

- Tolerância a Falhas: Se um servidor cair ou ficar fora do ar ou da rede, o sistema de arquivos não pode perder informações e nem ficar indisponível total ou parcialmente;
- Acesso Concorrente: Vários usuários podem acessar vários arquivos, ou os mesmos arquivos, sem sofrer danos, perda de performance ou quaisquer outras restrições;
- Replicação de Arquivos: Com esta funcionalidade, a confiança e a eficiência do serviço de arquivos é aumentada significativamente;



## **Características de um SAD:**

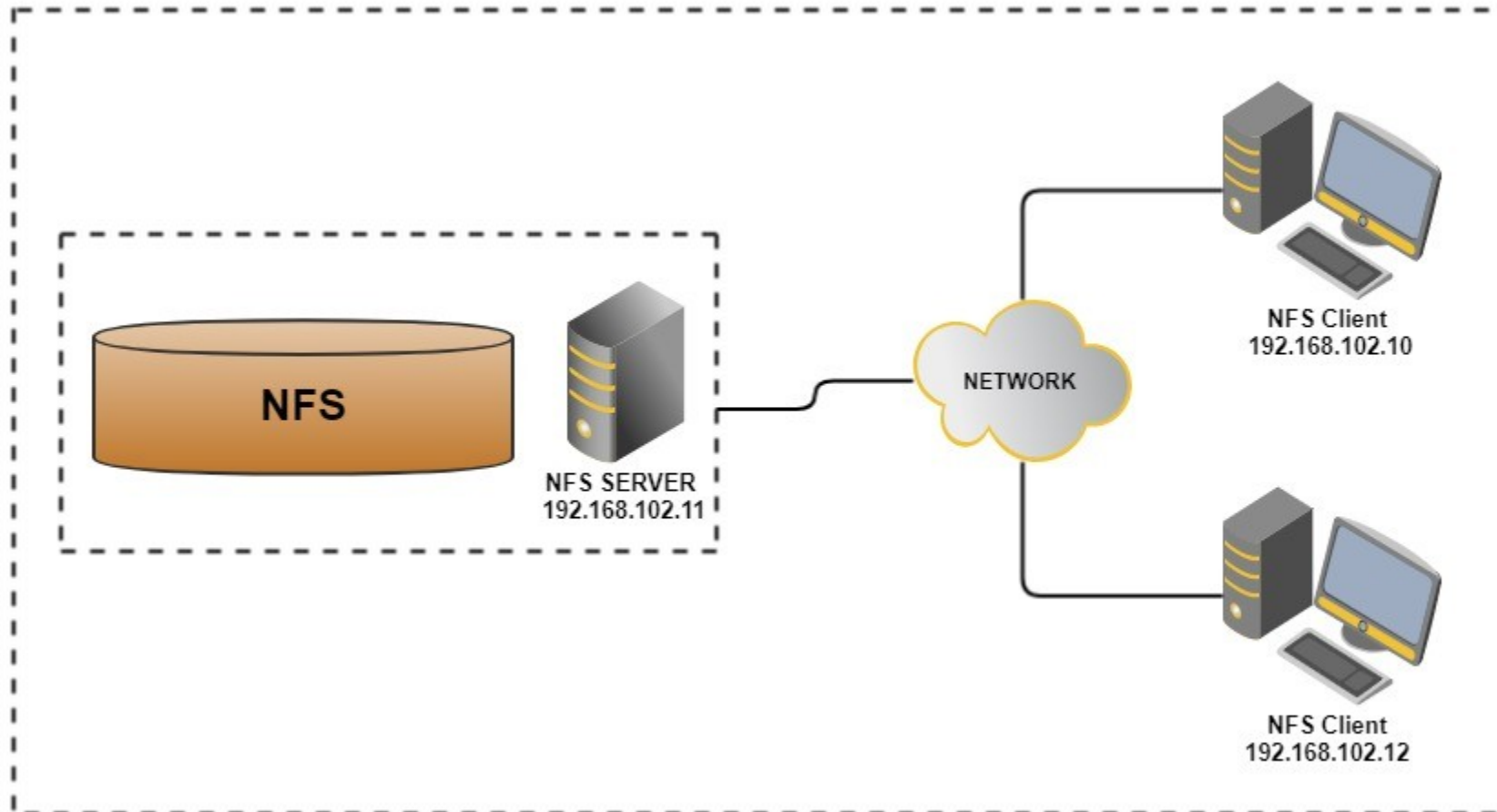
- Transparência: Usuários acessam pastas locais;
- Concorrência: Usuários concorrendo pelo mesmo arquivo;
- Replicação: Arquivos são replicados;
- Heterogeneidade: Diferentes recursos. Hardware, S.O., Recursos de Rede;
- Segurança: Controle de acesso / autenticação de clientes;
- Eficiência: Escalabilidade e usuários concorrentes. Latência;



## **Características de um SAD:**

- Tolerância a falhas: Queda de servidores ou comunicação;
- Consistência: Todas as cópias dos arquivos como se fossem uma apenas;
- Escalabilidade: Prever o crescimento de nós e usuários no sistema;
- Integridade dos dados: Arquivo é compartilhado por vários usuários.

## Arquitetura exemplo de SAD:



## **NFS é um exemplo canônico de SAD:**

- Arquitetura simples Cliente-Servidor;
- Muito popular no Unix;
- Ideia Fundamental: Cada servidor de Arquivos fornece uma visão padronizada de seu sistema local de arquivos;
- Fornece ao cliente acesso transparente a um sistema de arquivos que é gerenciado por um servidor remoto;
- É utilizada um sistema de Arquivo Virtual (VFS);
- Todas as requisições ao File System do servidor são realizadas via RPC (Remote Procedure Call);





# Sistemas de arquivos distribuídos

## NFS

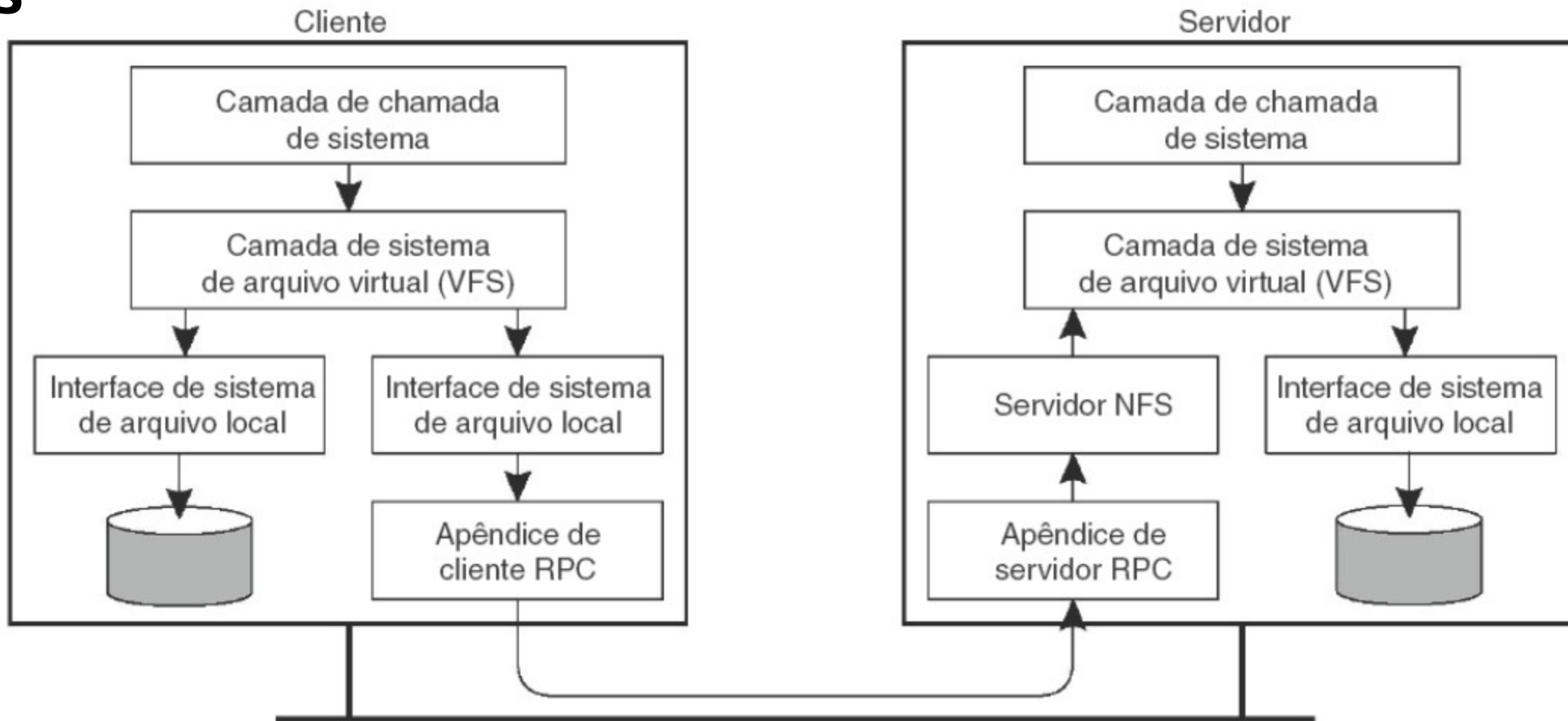


Figura 11.2 Arquitetura NFS básica para sistemas Unix.



## **SAD baseados em Cluster**

- Clusters normalmente são usados para computação paralela;
- Algo semelhante pode ser feito em SAD como file stripping (arquivos em tiras) através dos servidores;
  - Funciona bem se os dados tiverem uma estrutura bem regular;
  - Para aplicações de uso geral e dados irregulares e com diferentes estruturas esse modelo não será muito efetivo;
- Outra abordagem possível é armazenar arquivos inteiros em servidores, sem particionar o arquivo.



# Sistemas de arquivos distribuídos

## Modelo de armazenamento paralelo de arquivos em SAD:

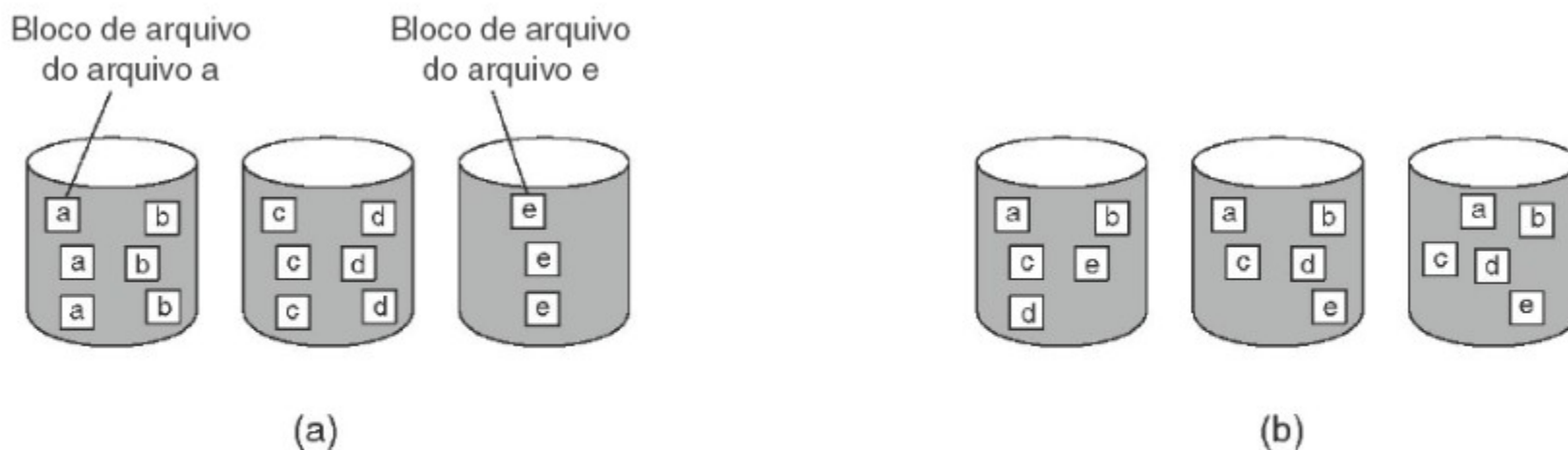


Figura 11.3 Diferença entre (a) distribuir arquivos inteiros por vários servidores e (b) desmembrar arquivos em tiras para acesso paralelo.



## Exemplos de SAD em Cluster:

- NFS
- **GFS**(Google FileSystem)
- **GlusterFS**
- PVFS (parallel virtual file system)
- AFS (Andrew FileSystem)
- DFS
- CODA (Constant Data Availability)
- SUN Network Filesystem
- **Hadoop** (HDFS)
- **Ceph**
- XteemFS
- Lustre