

# INTRODUÇÃO AOS SISTEMAS DE ARQUIVO

Prof. Hélio Esperidião

# Definições de Arquivos

Um arquivo é basicamente um conjunto de dados armazenados em um dispositivo físico não-volátil, com um nome ou outra referência que permita sua localização posterior.

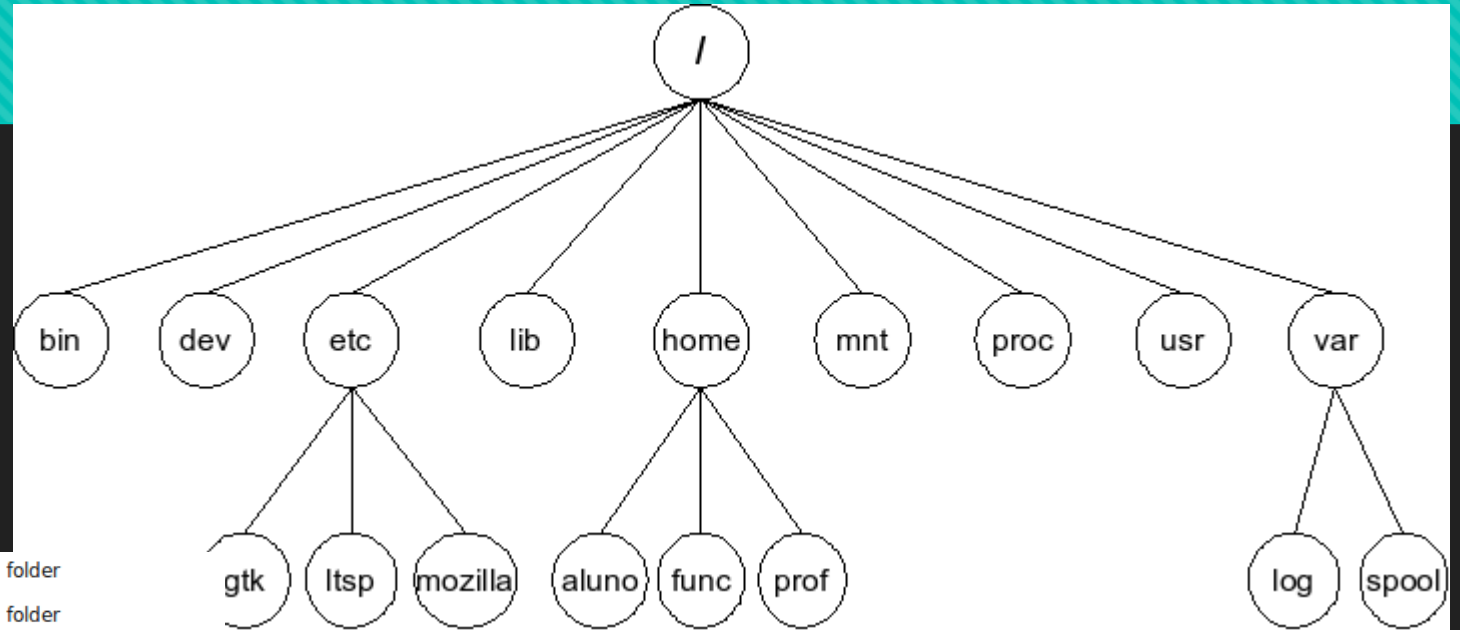
Do ponto de vista do usuário e das aplicações, o arquivo é a unidade básica de armazenamento de informação em um dispositivo não-volátil, pois para eles não há forma mais simples de armazenamento persistente de dados.

Arquivos são extremamente versáteis em conteúdo e capacidade: podem conter desde um texto ASCII com alguns bytes até uma sequência de vídeo com dezenas de gigabytes ou mais.

Como um dispositivo de armazenamento pode conter **milhões de arquivos**, estes são organizados em **estruturas hierárquicas** denominadas **diretórios**.

A organização física e lógica dos arquivos e diretórios dentro de um dispositivo é denominada sistema de arquivos.

# estruturas hierárquicas



[-] sites	3 items	folder
[-] all	4 items	folder
+ libraries	0 items	folder
[-] modules	3 items	folder
+ contrib	0 items	folder
[-] custom	1 item	folder
+ fs	0 items	folder
ABC README.txt	162 bytes	plain text document
[-] themes	3 items	folder
+ contrib	0 items	folder
+ custom	0 items	folder
ABC README.txt	161 bytes	plain text document
ABC README.txt	347 bytes	plain text document

# Sistemas de Arquivos

- Um sistema de arquivos pode ser visto como uma imensa estrutura de dados armazenada de forma persistente em um dispositivo físico.
- Existe um grande número de sistemas de arquivos, dentre os quais podem ser citados o **NTFS** (nos sistemas Windows), **Ext2/Ext3/Ext4** (Linux), **HPFS** (MacOS), **FFS** (Solaris) e **FAT** (usado em pendrives USB, máquinas fotográficas digitais e leitores MP3)

# Atributos de arquivos

- Cada arquivo é caracterizado por um conjunto de atributos, que podem variar de acordo com o sistema de arquivos utilizado. Os atributos mais usuais são:
  - **Nome:** string de caracteres que identifica o arquivo para o usuário, como “foto1.jpg”, “relatório.pdf”, “hello.c”, etc;
  - **Tipo:** indicação do formato dos dados contidos no arquivo, como áudio, vídeo, imagem, texto, etc. Muitos sistemas operacionais usam parte do nome do arquivo para identificar o tipo de seu conteúdo, na forma de uma extensão: “.doc”, “.jpg”, “.mp3”, etc.;

# Atributos de arquivos

- **Tamanho:** indicação do tamanho do conteúdo do arquivo, em bytes ou registros
- **Datas:** para fins de gerência, é importante manter as datas mais importantes relacionadas ao arquivo, como suas datas de criação, de último acesso e de última modificação do conteúdo;
- **Proprietário:** em sistemas multiusuários, cada arquivo tem um proprietário, que deve estar corretamente identificado;

# Atributos de arquivos

- **Permissões de acesso:** indicam que usuários têm acesso àquele arquivo e que formas de acesso são permitidas (leitura, escrita, remoção, etc.);
- **Localização:** indicação do dispositivo físico onde o arquivo se encontra e da posição do arquivo dentro do mesmo;



# Operações básicas sobre arquivos

- Criação (create)
- Escrita (write) e leitura (read)
- Reposicionamento em um ponto qualquer do arquivo (file seek)
- Remoção (delete)
- Abertura (open) e encerramento (close)
- Adicionalmente: renomeação (rename);

# Organização de Arquivos e acesso a dados

- **Partição:** Permite criar discos lógicos.
  - Dividir um disco físico em vários discos lógicos menores.
- **Blocos ou setores:** A menor unidade de armazenamento disponibiliza pelo disco.

# Alocação de Arquivos

## Alocação Contínua

- Este é o esquema mais simples de alocação de arquivos, onde cada arquivo é armazenado no disco como um bloco contíguo de dados. Neste esquema, em um disco com blocos de 1 KB, um pequeno arquivo de 20 KB seria armazenado em 20 blocos consecutivos.
- **Principais Vantagens:** simples implementação: controle de onde está cada arquivo no disco é feito por 1 único número (endereço em disco do 1º bloco).

# Alocação Contígua

- **Performance:** todo o bloco (arquivo) pode ser lido do disco de uma única vez. É necessário o tempo de somente um seek.
- **Problemas:** a estratégia só pode ser usada se o tamanho Máx. do arquivo for conhecido no momento de sua criação (devido a necessidade existente em saber o tamanho total do arquivo ou quantidade de blocos que ele ocupa).

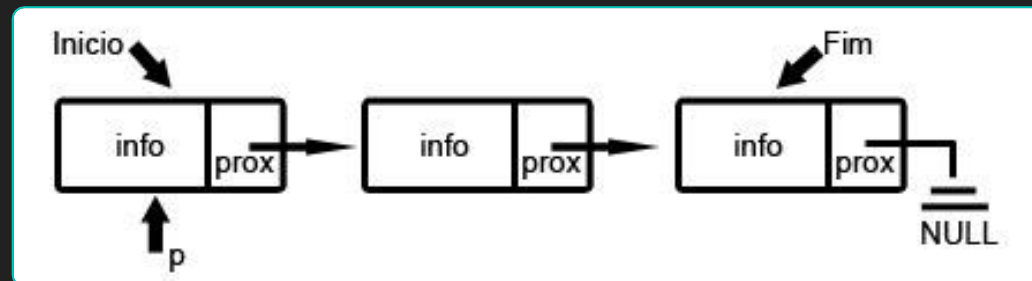
# Alocação Contígua

- **fragmentação do disco:** perde-se muito espaço útil com este esquema de alocação. Ao remover um arquivo a área ocupada pelo mesmo é liberada ocasionando lacunas por todo o disco. Necessidade de desfragmentação (custo alto).

# Alocação de Arquivos

## Alocação com Lista Encadeada

- Nesta estratégia de alocação, usamos uma lista encadeada para indicar os espaços ocupados em disco pelo arquivo. Assim, não é mais necessário que o arquivo seja armazenado em posições contíguas do disco. A primeira instrução de cada bloco é usada com um ponteiro para o próximo bloco e o restante do bloco é usado para armazenar as informações (dados) do arquivo.



# Alocação com Lista Encadeada

- **Principais Vantagens:** Qualquer bloco pode ser utilizado, permitindo que os arquivos cresçam indefinidamente enquanto houver espaço no disco.
- Só precisa armazenar o endereço do 1º bloco do arquivo (em cada bloco existirá um ponteiro para o próximo bloco do arquivo).
- **Problemas:** O acesso randômico é lento, pois existe a necessidade de percorrer a lista.