# Gerência de Sistema de Arquivos

Armazenar e recuperar dados da memória secundária é um processo que ocorre em todos programas. Um processo deve ser capaz de ler e gravar dados em unidades de armazenamento e compartilhar esses dados com outros processos.

A **gerência** **de** **sistema** **de** **arquivos** é a **parte do SO responsável pelas operações de arquivos,** pela **segurança** **de** **acesso** e pelo **compartilhamento** **de arquivos** na rede, entre os processos e entre os usuários.

## Arquivo

É um mecanismo de abstração, uma coleção nomeada de dados que pode conter um ou mais registros físicos. Um **registro físico ou bloco físico** é a unidade de informação que é lida e escrita no disco (como um conjunto de linhas em .txt).

Um arquivo pode ser composto de registros com tamanho fixo(mesmo tamanho) ou pode ser composto de registros com tamanho variado, nos dois casos **o total do tamanho dos registros equivale ao tamanho do arquivo.**

### Nomeação de arquivos

Uma das partes importantes do mecanismo que abstrai arquivos é o nome, é através dele que processos terão acesso ao mesmo.

As **regras para nomear um arquivo variam de acordo com o SO**, mas todo SO atual permite cadeias de no mínimo 8 caracteres e a maioria suporta 256 caracteres e alguns são case-sensitive (como o Linux).

#### Extensões de arquivos

Uma **extensão** é um sufixo que define o formato e a função de um arquivo, no Windows todo tipo de arquivo tem uma extensão.

**Cada extensão pode pedir um software diferente.** As extensões podem ser **divididas** **em** **áudio** (mp3, wav, etc), **vídeo** (mp4, mpeg, etc), **imagem** (gif, png, jpeg, etc), **compactadores** (zip, rar, 7z, etc) e **documentos** (txt, doc, xls, etc).

No **Linux** é possível ter mais de uma extensão e não há limite de caracteres na extensão, já no **Windows** só é possível criar extensões com até 4 caracteres e somente a última extensão é considerada.

### Organização de Arquivos

Trata de como os dados/instruções estão armazenados internamente, a estrutura pode variar dependendo da função e do tipo do arquivo e quem a determina é o programa que criou o arquivo. As organizações **mais comuns são:**

* **Sequencial**: Os registros são ordenados de maneira sequencial e o tamanho dos arquivos é fixo
* **Não sequencial:** consiste em armazenar os dados em uma sequência não estruturada de bytes e sem uma ordem lógica.
* **Indexado**: Os registros estão desordenados e são usados índices para encontrá-los o tamanho pode ser fixo ou variável
* **Relativo**: Os registros são ordenados de maneira relativa aos outros e o tamanho é variável

#### Métodos de Acesso aos dados nos Arquivos

Dependendo de como o arquivo é organizado ele pode ser acessado das seguintes maneiras:

* **Acesso sequencial:** O acesso é feito na ordem de gravação, é necessário **percorrer todos antecessores** de um registro para acessá-lo e a **inserção** é feita no **final**
* **Acesso direto:** A leitura/gravação é feita diretamente em uma posição conhecida**.** A **posição do registro é necessária para qualquer operação e o tamanho deve ser fixo** ou não é possível calcular os índices. Depois de percorrer uma organização sequencial é possível fazer acesso direto com base nas posições percorridas.
* **Acesso Indexado ou Acesso por Chave:** Tem como base o acesso direto, em arquivos com organização indexada é possível **encontrar um arquivo dado o valor da chave de seu ponteiro.**

### Operações em arquivos

Arquivos podem ser manipulados através de operações como: abrir (preparar o arquivo para ser referido), fechar (impedir que um arquivo seja referido), criar, destruir, copiar, renomear e listar (imprimir conteúdo), etc

Os dados nos arquivos também podem ser manipulados por operações realizadas entre a memória principal e a memória secundária como: ler (copiar o conteúdo para memória), escrever (copiar dados da memória para o arquivo), atualizar (modificar) e apagar.

Além disso é possível classificar um arquivo de acordo com seus atributos como tamanho, localização, permissões, tipo, volatilidade (frequência de alterações) e atividade (frequência de uso).

### Tipos de arquivos

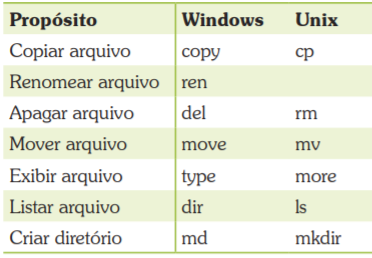
**Do ponto de vista de SO um arquivo é só um conjunto de registros**, mas **para** **o** **usuário** um arquivo pode ser de **3 tipos:**

* **Arquivo executável:** É composto por **instruções** **em** **linguagem** **de** **máquina** que serão **executadas** **pelo** **SO**.
* **Arquivo de dados:** É composto por **dados** **estruturados**, um número determinado de **campos**, **posição** **inicial**, **posição** **final**, etc e pode representar um .txt por exemplo
* **Arquivo multimídia:** Contém imagem, áudio, som, vídeo, etc

## **Sistema de arquivos**

É um sistema responsável por organizar e gerenciar o acesso aos dados de um disco, sendo responsável por garantir que os arquivos salvos fiquem disponíveis, sejam compartilhados, fiquem em segurança e que não sejam corrompidos.

O sistema de arquivos pode ser acessado via CLI ou por uma interface gráfica.



### **Funções** do sistema de arquivos

* Gerenciamento de espaço dos dispositivos de armazenamento (especialmente HD)
* Compartilhamento de arquivos de maneira segura e controlada
* Abstração de características do dispositivo (especialmente de localização de arquivos, deve ser possível referenciar um arquivo por seu nome, sem conhecer seu endereço e disco).
* Segurança por meio de backups e quando necessário criptografia e decriptamento

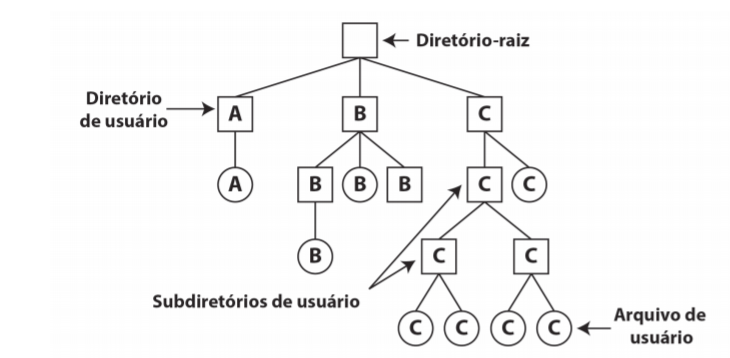
### Diretórios

Para organizar e controlar os arquivos os sistemas de arquivos usam diretórios (pastas) que costumam ser arquivos que contem a localização física e o nome dos arquivos organizados lá dentro pelo usuário/sistema.

O sistema de diretórios mais simples é o **sistema de arquivo de nível único ou diretório raiz**, que guarda todos arquivos num único diretório, de modo que dois arquivos não podem ter o mesmo nome nunca.

O sistema de diretório mais apropriado é o **sistema de arquivo estruturado hierarquicamente**, neste modelo, uma raiz(diretório raiz), onde começa o armazenamento, aponta para um diretório de cada usuário que contem os arquivos daquele usuário.

Nos diretórios de usuário existem outros diretórios com arquivos/diretórios pertencentes aquele usuário que guardam a localização física do arquivo no dispositivo de armazenamento, de modo que um nome de arquivo só é exclusivo naquele diretório e o sistema é uma árvore.



A **maioria dos SO implementam o sistema hierárquico**, mas o nome do diretório raiz e o delimitador de diretórios podem variar de SO para SO. O Windows nomeia o diretório raiz como uma letra seguida de dois pontos (C:) e separa os diretórios usando \ enquanto sistemas Unix usam / como diretório raiz e usam / para separar diretórios.

#### Caminhos de arquivos absolutos e relativos

Um **caminho de arquivo absoluto** é um caminho que começa no diretório raiz e é único (C:\Users\PC\Documents).

Um **caminho relativo** é um caminho que parte de um diretório de trabalho (workspace) ou do diretório atual até chegar a um arquivo (.\Documents). A maioria dos sistemas que usa diretórios hierárquicos se refere ao **diretório atual como “.”** e ao **diretório pai como “..”**, possibilitando navegar pela árvore a partir de um caminho relativo.

#### **Link de ligação**

É um arquivo que aponta para outro arquivo em um diretório diferente, normalmente são criados para facilitar a navegação. No **Unix** é conhecido como **link simbólico,** no **Windows** é conhecido como **atalho** e no **MAC** é conhecido como **apelido(alias).**

Um **link pode ser estrito ou não estrito**, ele é **estrito** quando aponta para a localização física de um arquivo e **não é estrito** quando aponta para endereço com a localização do arquivo.

#### Desfragmentação

A desfragmentação é usada para remover fragmentos de arquivos deletados em um disco e melhorar seu desempenho. Durante a operação **um arquivo pode mudar de localização, fazendo com que seja necessário atualizá-lo no sistema e invalidando um link estrito**, criando a necessidade de atualizar o ponteiro.

### Metadados de arquivos

A maioria dos sistemas armazena dados sobre os arquivos como localização de blocos livres próximos ao arquivo, seu horário de modificação, seu horário de criação, etc. Essas informações são chamadas de metadados, **protegem a integridade dos arquivos e só podem ser modificadas pelo sistema.**

### **Superbloco**

Quando um dispositivo é formatado é criada uma lista de blocos livres, a data e hora de modificação, a localização do diretório raiz e informações sobre falhas, esse conjunto de informações é chamado de **superbloco**, quando há problemas com o superbloco **podem ocorrer falhas ao acessar arquivos, então, para impedir perdas o sistema cria e distribui cópias do superbloco.**

## Montagem

O sistema de arquivos de um SO vem montado, mas as vezes é necessário adicionar outros sistemas de arquivos, como um segundo HD, e por essa razão é possível montar vários sistemas de arquivos que são combinados em um único espaço de nome.

Os primeiros pontos de montagem(NTFS) do Windows sempre iniciavam da raiz, sendo que C: continha o SO, o segundo disco era o D: e assim por diante.

No **Unix e a nas versões mais recentes do NTFS** (5.0 +) os **pontos de montagem podem ser iniciados em qualquer lugar do sistema de arquivos**, embora seja comum que Unix use /mnt/ como raiz.

**Tabelas de montagem** são tabelas com informações sobre nomes, caminhos dos pontos de montagem e o disco de cada arquivo montado e são **usadas pelo sistema de arquivos para gerenciar diretórios montados.**