

Gerenciamento de arquivo

Gustavo de Oliveira e Keven Gustavo

Sumário

- Sistema de arquivos.
- Arquivos.
- Diretórios.
- Implementação do sistema de arquivos.
- Proteção e segurança.

Gerenciamento de arquivos

A parte do sistema operacional responsável por essa gerência é chamada de **Sistema de arquivos**:

- A parte mais visível do sistema operacional, pois é uma atividade frequentemente realizada pelos usuários.
- Armazenar e recuperar os dados de forma persistente.
- Persistência - Capacidade de guardar a informação após a finalização e criação de diversos processos.
- Deve ocorrer de maneira **uniforme** independente dos diferentes dispositivos de armazenamento.

Arquivos

Tópicos

- Definição.
- Extensões.
- Estrutura.
- Tipos.
- Atributos.
- Acesso.
- Operações.

Arquivos - Definição

- Um Arquivo é um conjunto de registros definidos pelo sistema e podem ser armazenados em diferentes dispositivos físicos.
- É constituído de informações logicamente relacionadas, podendo representar programas ou dados.
- Esse armazenamento torna possível a manipulação, leitura ou restauração desses dados.
- Representados por um nome e uma extensão (Ex.: aula.docx).

Arquivos - Extensões

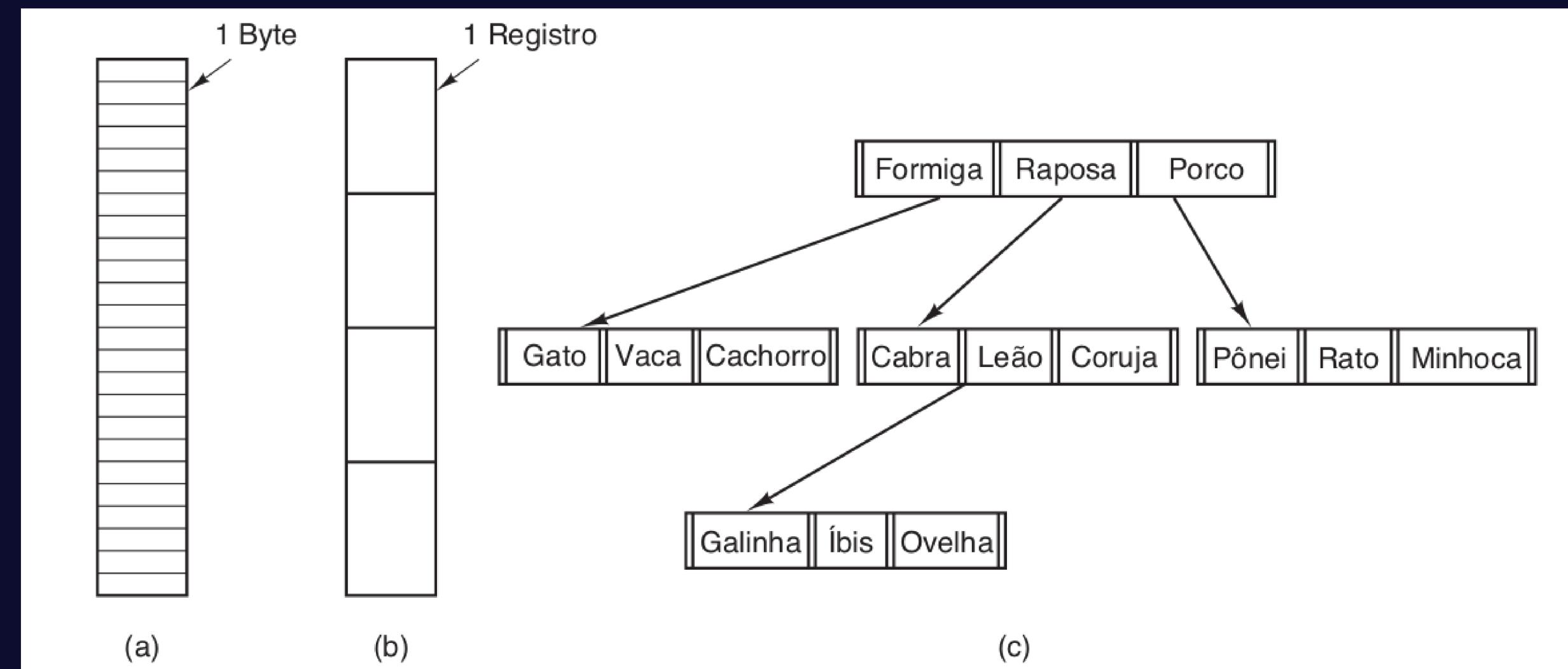
- A extensão define o conteúdo que tal arquivo apresenta.
- Ela não é obrigatória, exemplo em sistemas UNIX e WINDOWS.
- Geralmente associada a algum aplicativo.

Extensão	Significado
.bak	Cópia de segurança
.c	Código-fonte de programa em C
.gif	Imagem no formato Graphical Interchange Format
.hlp	Arquivo de ajuda
.html	Documento em HTML
.jpg	Imagem codificada segundo padrões JPEG
.mp3	Música codificada no formato MPEG (camada 3)
.mpg	Filme codificado no padrão MPEG
.o	Arquivo objeto (gerado por compilador, ainda não ligado)
.pdf	Arquivo no formato PDF (Portable Document File)
.ps	Arquivo PostScript
.tex	Entrada para o programa de formatação TEX
.txt	Arquivo de texto
.zip	Arquivo compactado

Arquivos - Estrutura

A estrutura do arquivo consiste na forma com que os seus dados estão internamente armazenados, possuindo três estruturas principais:

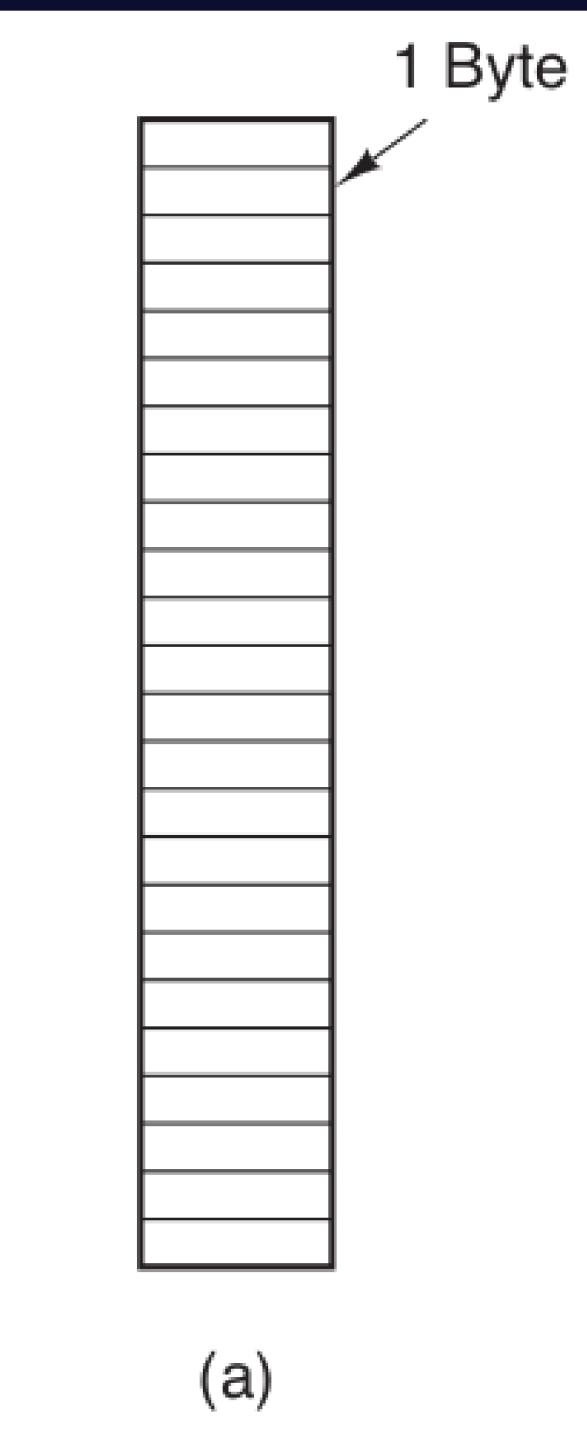
- (a) Sequencia não estruturada de bytes.
- (b) Sequencia de registros de tamanho fixo.
- (c) Arvores de registros(tamanho variado).



Arquivos - Estrutura

Sequencia não estruturada de bytes(a):

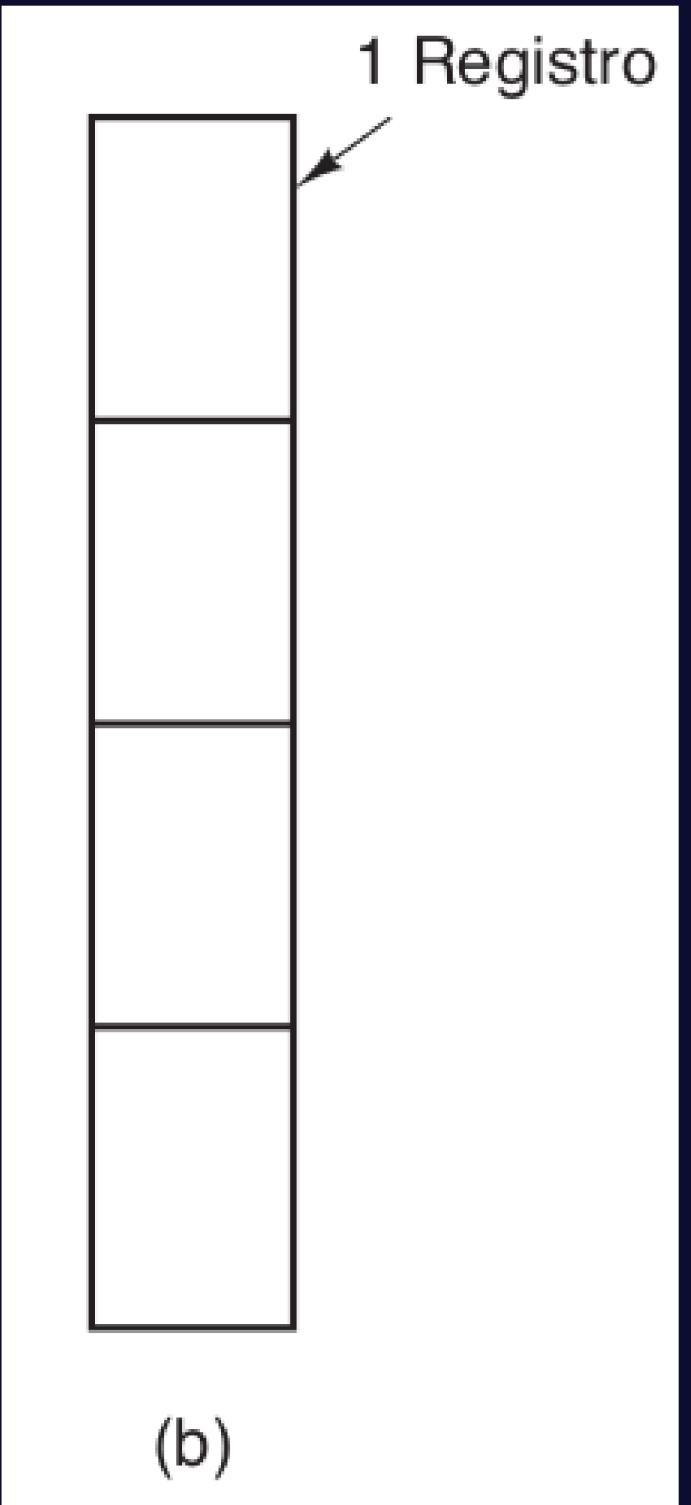
- São apenas conjuntos de bytes para o SO, ou seja, ele não se importa com o conteúdo do arquivo.
- A aplicação que define o significado do arquivo.
- Vantagem: Flexibilidade, os usuários colocam o que quiserem e nomeiam como quiserem.
- Utilizado em UNIX e Windows.



Arquivos - Estrutura

Sequencia de registros de tamanho fixo(b):

- Cada qual com sua estrutura interna.
- Ex.: 80 a 132 caracteres.
- Leitura/Escrita são realizadas em registros.
- 80 caracteres de um cartão perfurado, ou do monitor.
- 132 dos caracteres de uma impressora.
- SO's mais antigos, como o mainframes.
- Sistemas atuais não utilizam.

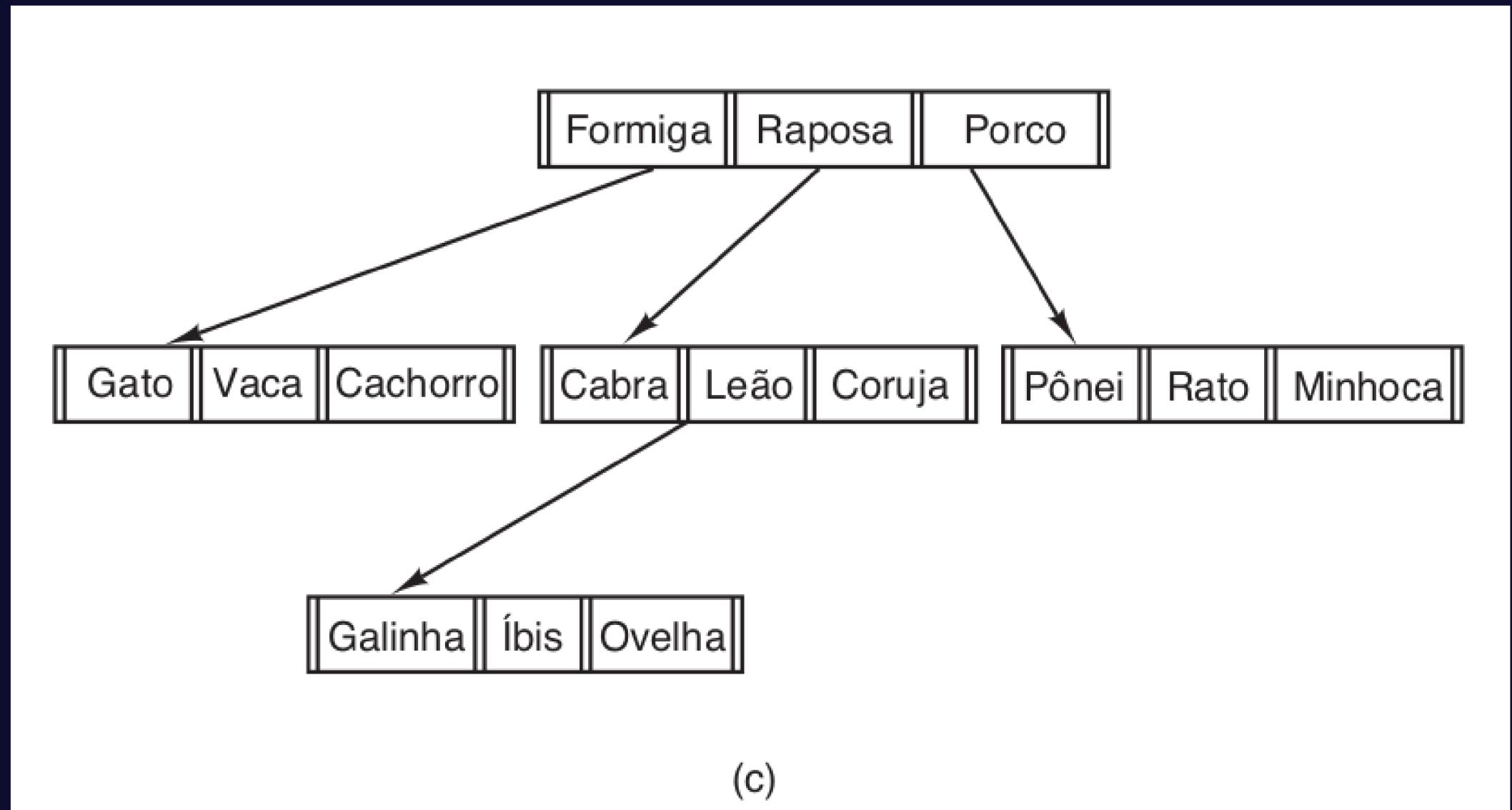


Fonte: Tanenbaum 3º edição

Arquivos - Estrutura

Arvores de registros(tamanho variado)(c):

- Cada parte do arquivo: com um campo chave em uma posição fixa.
- O arquivo consiste em árvore.
- A operação: obter o registro com uma certa chave.
- SO decide onde colocar novos registros, não o usuários.
- Usado em mainframes atuais.



Arquivos - Tipos

Os principais tipos de arquivo são:

- Arquivos Regulares;
- Diretórios;
- Arquivos de Metadados;
- Arquivos especiais de caractere;
- Arquivos especiais de bloco;

Arquivos - Tipos

Arquivos Regulares:

- São aqueles que contêm informação do usuário, o tipo mais comum de arquivo.
- Armazenam os dados que são criados, lidos, modificados e gerenciados pelos usuários e aplicativos.
- São categorizados em dois tipos, arquivos ASCII ou arquivos Binários.
- ASCII: consiste em linhas de texto, facilmente exibidos ou impressos como são, podem ser editados por qualquer editor de texto, e são mais portáveis e acessíveis.
- Binário: todo arquivo não ASCII, estrutura interna conhecida apenas pelas aplicações que os usam.
- Exemplo: Um arquivo do Microsoft Word “trabalho.docx”.

Arquivos - Tipos

Diretórios:

- São arquivos de sistema usados para estruturar o sistema de arquivos
- Não contêm dados, mas são utilizados como contêineres para agrupar um conjunto de arquivos que possuem alguma relação.
- No sistema, os diretórios são arquivos especiais que possuem uma lista de entrada de arquivo e seus caminhos.
- Exemplo: a pasta de 'Downloads', onde por padrão recebe todos os arquivos baixados pelo navegador.

Arquivos - Tipos

Arquivos de Metadados:

- São arquivos que não contêm dados reais, mas informações descritivas sobre outros arquivos ou objetos.
- Utilizados para fornecer informações sobre os arquivos normais, como: informações de autoria, data de criação, tamanho, tipo de arquivo, permissões de acesso, entre outros.
- Os arquivos de metadados são frequentemente usados para indexar, organizar e pesquisar outros arquivos, são essenciais para sistemas de gerenciamento de arquivos, bancos de dados e motores de busca.
- Exemplo: No sistema de arquivos NTFS do Windows, os arquivos "\$MFT" e "\$LogFile", contêm metadados essenciais sobre todos os arquivos e diretórios no sistema.

Arquivos - Tipos

Arquivos Especiais de Caractere:

- São arquivos que são usados para a comunicação com dispositivos de caractere, como impressoras ou dispositivos de comunicação.
- Eles permitem que os aplicativos interajam com dispositivos de caractere como se estivessem lidando com arquivos normais, enviando e recebendo dados por meio desses arquivos especiais
- Exemplo: o dispositivo de caractere “/dev/ttyS0” no Linux, é usado para se comunicar com uma porta serial.

Arquivos - Tipos

Arquivos Especiais de Bloco:

- São arquivos que são usados para a comunicação com dispositivos de bloco, como discos rígidos ou unidades de CD/DVD
- Eles permitem a comunicação eficaz com dispositivos de armazenamento, tratando-os como arquivos regulares, mas fornecendo acesso de baixo nível a setores de dados.
- Exemplo: o dispositivo de bloco “/dev/sda” no Linux, representa um disco rígido.

Arquivos - Acesso

Os arquivos podem ser acessados de duas formas, sendo a primeira delas o **Acesso Sequencial**:

- Utilizado em sistemas operacionais antigos.
- Nesse processo é feito a leitura em ordem de byte a byte (registro a registro).
- Não podia pular bytes ou lê-los fora da ordem.
- É possível voltar ao inicio do arquivo e ler quantas vezes fossem necessárias.
- O acesso sequencial era conveniente quando o meio de armazenamento era uma fita magnética, em vez de disco.

Arquivos - Acesso

A outra forma de acesso é o **Acesso Aleatório**:

- Utilizado por sistemas operacionais modernos.
- Os bytes ou registros dos arquivos podem ser lidos em qualquer ordem, por meio de uma chave.
- Dois meios são usados para saber onde a leitura deve começar:
- No primeiro, cada operação `read` indica a posição no arquivo na qual a leitura deve começar, a partir do qual sera sequencial.
- No segundo, uma operação especial, **Seek**, é fornecida para marcar a posição corrente, após essa operação, o arquivo procura a marcação que então vai ser lido de forma sequencial a partir da posição corrente marcada.

Arquivos - Atributos

Todos os arquivos possuem atributos, que além do seu nome e o seu conteúdo, são informações associadas ao arquivo pelo sistema operacional.

- Exemplo: a data e hora em que o arquivo foi criado e o tamanho do arquivo.
- Esses atributos também são chamados de metadados.
- Cada Sistema Operacional possui sua propria lista de atributos.

Arquivos - Atributos

Atributo	Significado
Proteção	Quem tem acesso ao arquivo e de que modo
Senha	Necessidade de senha para acesso ao arquivo
Criador	ID do criador do arquivo
Proprietário	Proprietário atual
Flag de somente leitura	0 para leitura/escrita; 1 para somente leitura
Flag de oculto	0 para normal; 1 para não exibir o arquivo
Flag de sistema	0 para arquivos normais; 1 para arquivos de sistema
Flag de arquivamento	0 para arquivos com backup; 1 para arquivos sem backup
Flag de ASCII/binário	0 para arquivos ASCII; 1 para arquivos binários
Flag de acesso aleatório	0 para acesso somente sequencial; 1 para acesso aleatório
Flag de temporário	0 para normal; 1 para apagar o arquivo ao sair do processo
Flag de travamento	0 para destravados; diferente de 0 para travados
Tamanho do registro	Número de bytes em um registro
Posição da chave	Posição da chave em cada registro
Tamanho da chave	Número de bytes na chave
Momento de criação	Data e hora de criação do arquivo
Momento do último acesso	Data e hora do último acesso do arquivo
Momento da última alteração	Data e hora da última modificação do arquivo
Tamanho atual	Número de bytes no arquivo
Tamanho máximo	Número máximo de bytes no arquivo

Arquivos - Operações

Operações comuns relacionadas ao sistema de arquivos:

- **Create:** O arquivo é criado sem dados;
- **Delete:** O arquivo é removido do disco;
- **Open:** Permite que o sistema operacional busque os atributos e a lista de endereços de disco na memoria principal;
- **Close:** Libera o espaço ocupado pelo Open, força que o ultimo bloco de dados seja escrito no disco;
- **Read:** Lê do arquivo para um buffer;
- **Write:** Escreve os dados no arquivo;
- **Append:** Escreve dados no final do arquivo;
- **Seek:** Para acesso aleatório - procurar dados;
- **Get attributes:** Obtém os atributos de um arquivo;
- **Set attributes:** Modifica algum atributo do arquivo;
- **Rename:** Muda o nome do arquivo;

Diretórios

Tópicos

- Definição.
- Organização.
- Caminho.
- Operações.

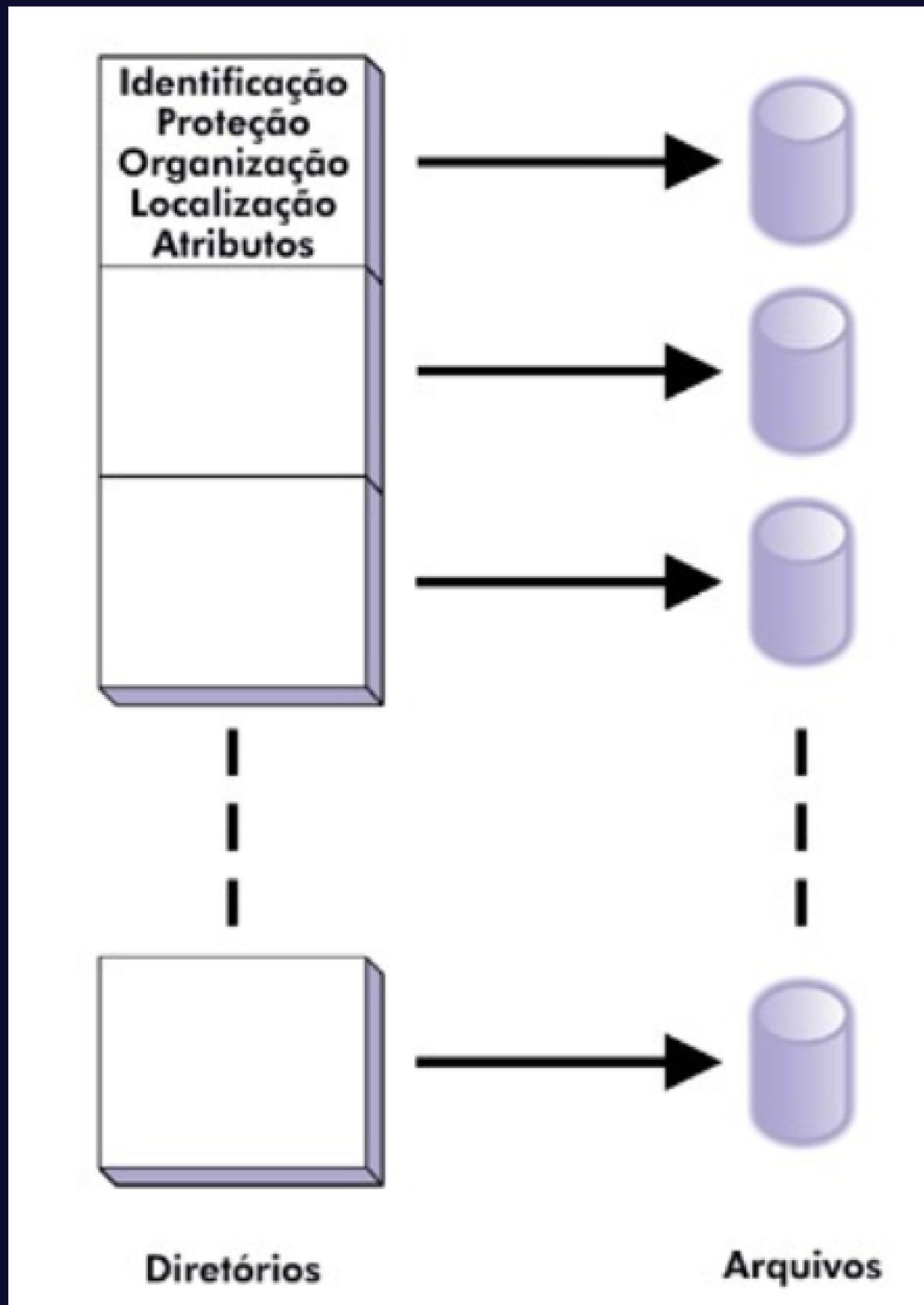
Diretórios - Definição

- É uma estrutura de organização, usada pelo sistema de arquivos, que serve para agrupar e organizar arquivos.
- São em si arquivos.
- Facilitam a localização, acesso e gerenciamento de arquivos relacionados, ajudando os usuários a manter seus dados organizados de maneira eficiente
- Cada diretório pode conter vários arquivos, bem como outros subdiretórios, criando uma árvore de diretórios.

Diretórios - Organização

A estrutura mais simples de um sistema de diretório é chamada de **Diretório em nível único**:

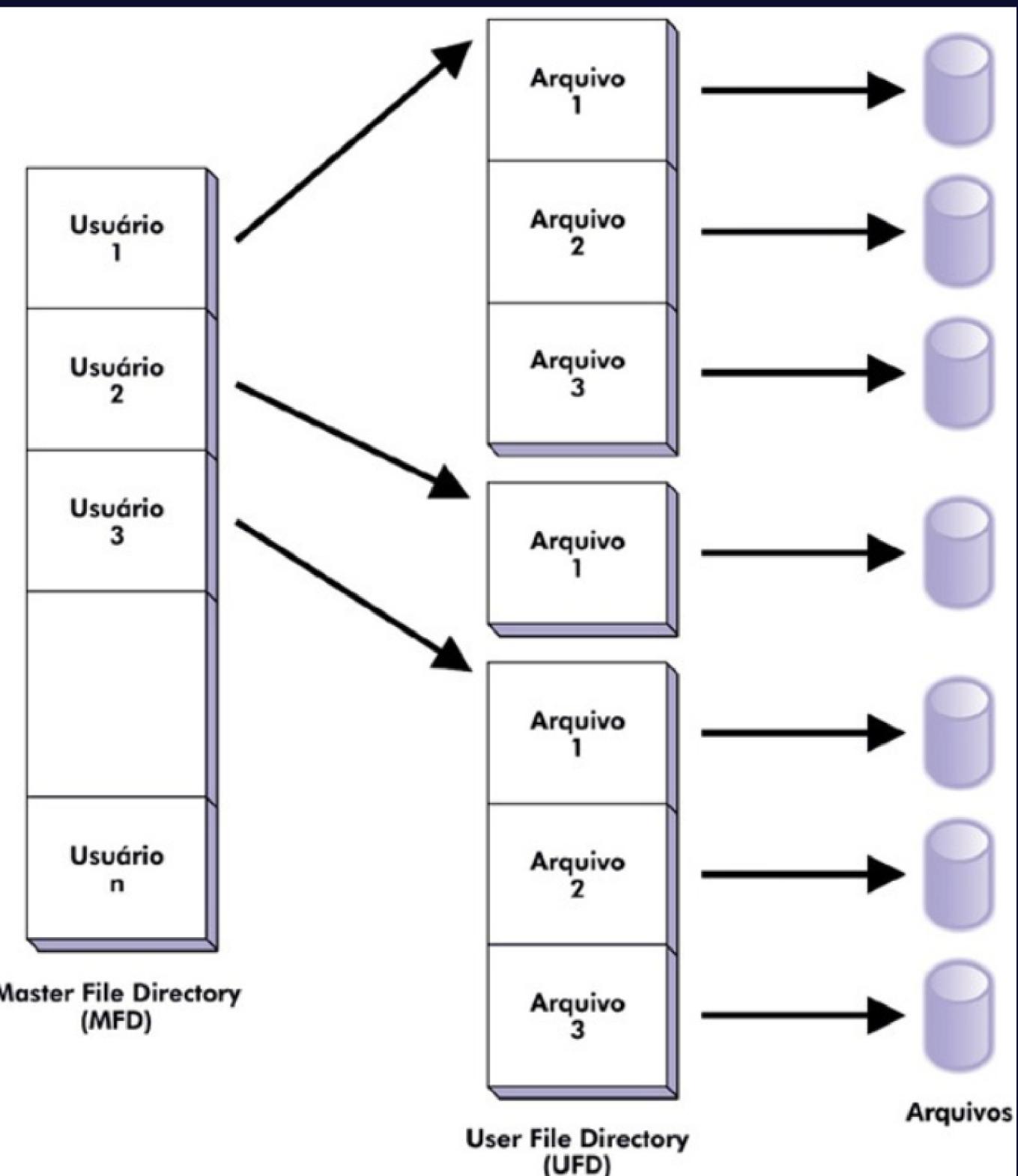
- Implementação mais simples.
- Um único diretório contendo todos os arquivos, sem possuir subdivisões adicionais em subdiretórios.
- É bastante limitado, já que não permite que usuários criem arquivos com o mesmo nome, o que ocasionaria um conflito no acesso aos arquivos.



Diretórios - Organização

A partir da estrutura de nível único foi desenvolvida outra estrutura, conhecida como **Diretório com dois níveis**:

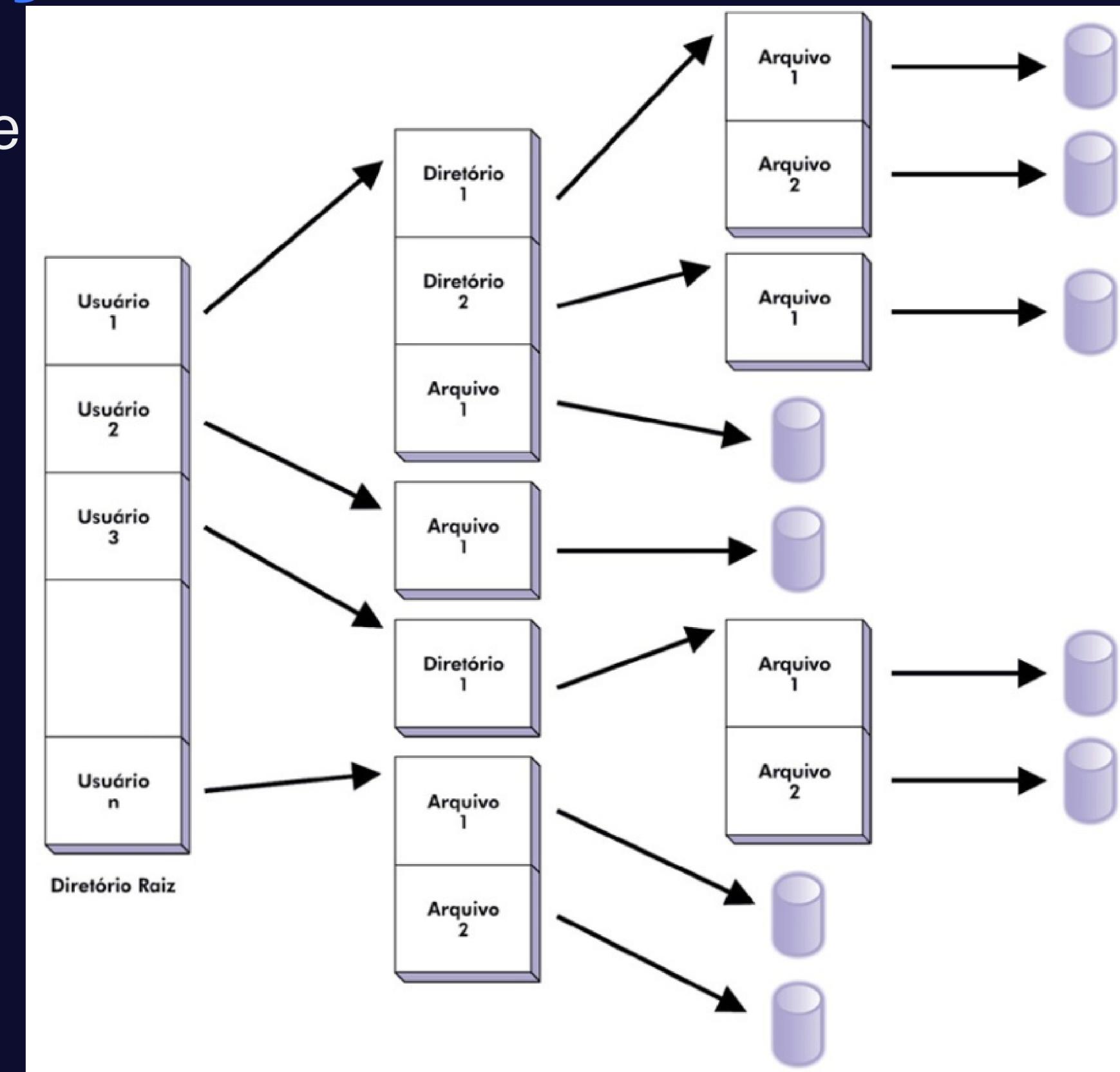
- Possui um diretório para cada usuário(UFD), possibilitando a criação de arquivos com qualquer nome, sem a preocupação de conflito com os demais arquivos do disco.
- O diretório 'MFD' é usado para controlar os diretórios individuais e localizar arquivos na estrutura.



Diretórios - Organização

O ultimo tipo de organização é a estrutura de **Diretórios em Árvore**:

- Adotada pela maioria dos sistemas.
- Extensão do modelo de dois níveis para um de múltiplo níveis.
- Cada usuário pode criar diversos níveis de diretórios, chamados de subdiretórios.
- A localização eficiente de arquivos por meio de subdiretórios é facilitada pela categorização logica e estruturada dos dados.



Diretórios - Caminho

Nomes de caminhos:

- É uma representação textual que descreve a localização de um arquivo ou diretório dentro de uma hierarquia de diretórios em sistemas de arquivos.
- Um caminho é construído especificando cada diretório pelo qual você deve passar para chegar ao arquivo desejado, começando a partir do diretório raiz.
- **Nome de Caminho Absoluto:** único, usando o nome do caminho completo do diretório raiz do sistema de arquivos até o arquivo desejado.
- O seu uso evita conflitos de nome e simplificam a manutenção.
- **Nome de Caminho Relativo:** considerando o diretório atual, usa o caminho relativo até o arquivo que deseja acessar.
- Exemplo: “C:\Program Files (x86)\Adobe\Acrobat Reader DC”.

Diretórios - Operações

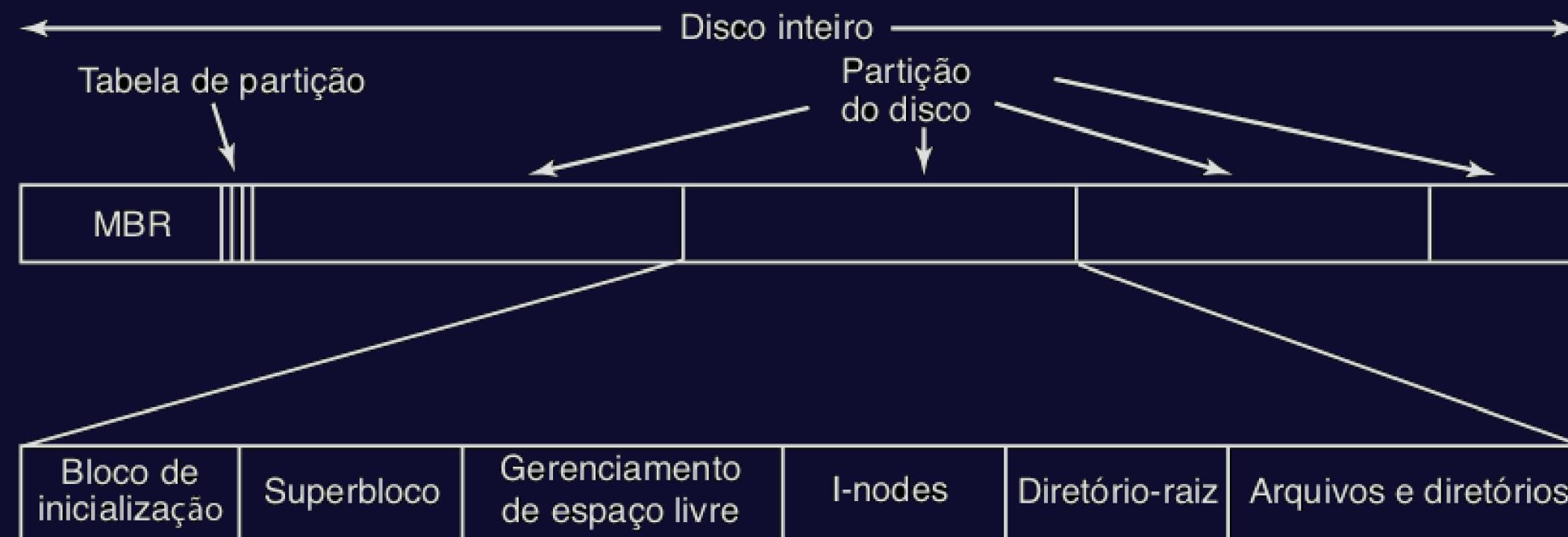
- **Create:** Criar um novo diretório.
- **Delete:** Remover um diretório.
- **Opendir:** Listar todos os arquivos em um diretório.
- **Closedir:** Fechar um diretório.
- **Readdir:** Ler as entradas de um diretório aberto.
- **Rename:** Renomear diretórios.
- **Link** (Criar um Vínculo para um Arquivo): A operação de atalho permite que um arquivo existente seja acessado a partir de diferentes diretórios.
- **Unlink** (Remover uma Entrada de Diretório): Quando uma entrada de diretório é removida.

Esquema do Sistema de Arquivos

Características importantes dos sistemas de arquivo:

- Os discos podem ser divididos em uma ou mais partições.
- Cada partição pode conter seu próprio sistema de arquivos.
- O setor 0 do disco sempre vai conter o MBR (Master Reboot Record).
- O BIOS lê e executa o MBR que localiza a partição ativa, lê o primeiro bloco, e executa-o.

Esquema do Sistema de Arquivos



TANENBAUM, 3º Ed. 2010

Implementação de arquivos

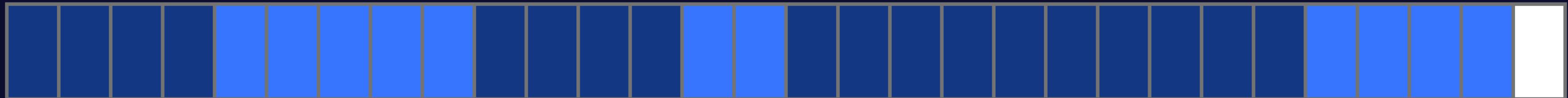
Métodos:

- Alocação contígua.
- Alocação encadeada
- Alocação encadeada por tabela na memória.
- I-nodes.

Alocação Contígua

Os blocos são dispostos de forma sequencial no disco.

Arquivo A
(4 Blocos)



Arquivo C
(4 Blocos)

Arquivo E
(10 Blocos)

Arquivo B
(5 Blocos)

Arquivo D
(2 Blocos)

Arquivo F
(4 Blocos)

Alocação contígua

Vantagens

- É mais simples de implementar.
- Tem um alto desempenho de leitura e escrita, pois requer menos operações de busca no disco.

Desvantagens

- Com o tempo de uso causa a fragmentação externa do disco.
- A alocação contígua torna complexa a realocação de um arquivo ao crescer, sendo um processo demorado.

Alocação Contígua

Arquivo A
(4 Blocos)

Arquivo C
(4 Blocos)

Arquivo E
(10 Blocos)



Arquivo B
(5 Blocos)

Arquivo D
(2 Blocos)

Arquivo F
(4 Blocos)

...

Alocação Contígua

Arquivo A
(4 Blocos)

Arquivo C
(4 Blocos)

Arquivo E
(10 Blocos)



Arquivo B
(5 Blocos)

Arquivo D
(2 Blocos)

Arquivo F
(4 Blocos)

Arquivo A
(4 Blocos)

Arquivo C
(4 Blocos)

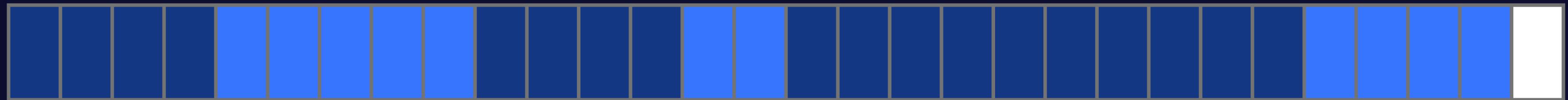
Arquivo E
(10 Blocos)



Arquivo F
(4 Blocos)

Alocação Contígua

Arquivo A
(4 Blocos)



Arquivo C
(4 Blocos)

Arquivo E
(10 Blocos)

...

Arquivo B
(5 Blocos)

Arquivo D
(2 Blocos)

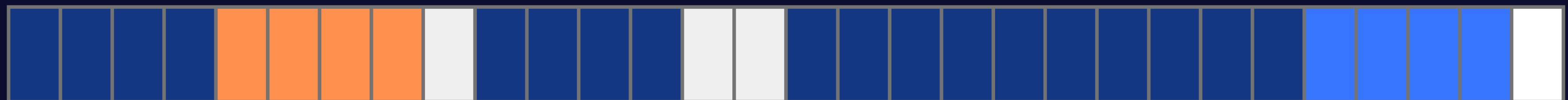
Arquivo F
(4 Blocos)

Arquivo A
(4 Blocos)

Arquivo C
(4 Blocos)

Arquivo E
(10 Blocos)

...



Arquivo G
(4 Blocos)

Arquivo F
(4 Blocos)

Alocação contígua

As 3 principais estratégias de alocação contígua:

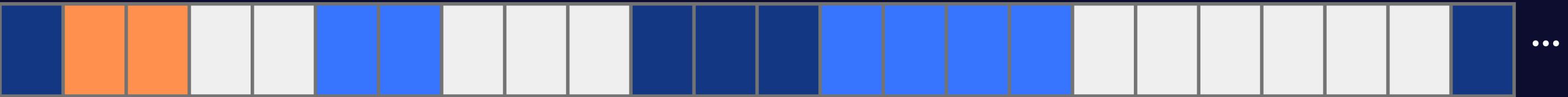
- First-fit: o primeiro segmento livre com tamanho suficiente para alocar o arquivo é selecionado. A busca na lista é sequencial, sendo interrompida tão logo se encontre um segmento adequado.
- Best-fit: seleciona o menor segmento livre disponível com tamanho suficiente para armazenar o arquivo. A busca em toda a lista se faz necessária para a seleção do segmento, a não ser que a lista esteja ordenada por tamanho.
- Worst-fit: o maior segmento é alocado e a busca por toda a lista se faz necessária, a menos que exista uma ordenação por tamanho.

Situação inicial



...

First-fit:



...

Best-fit



...

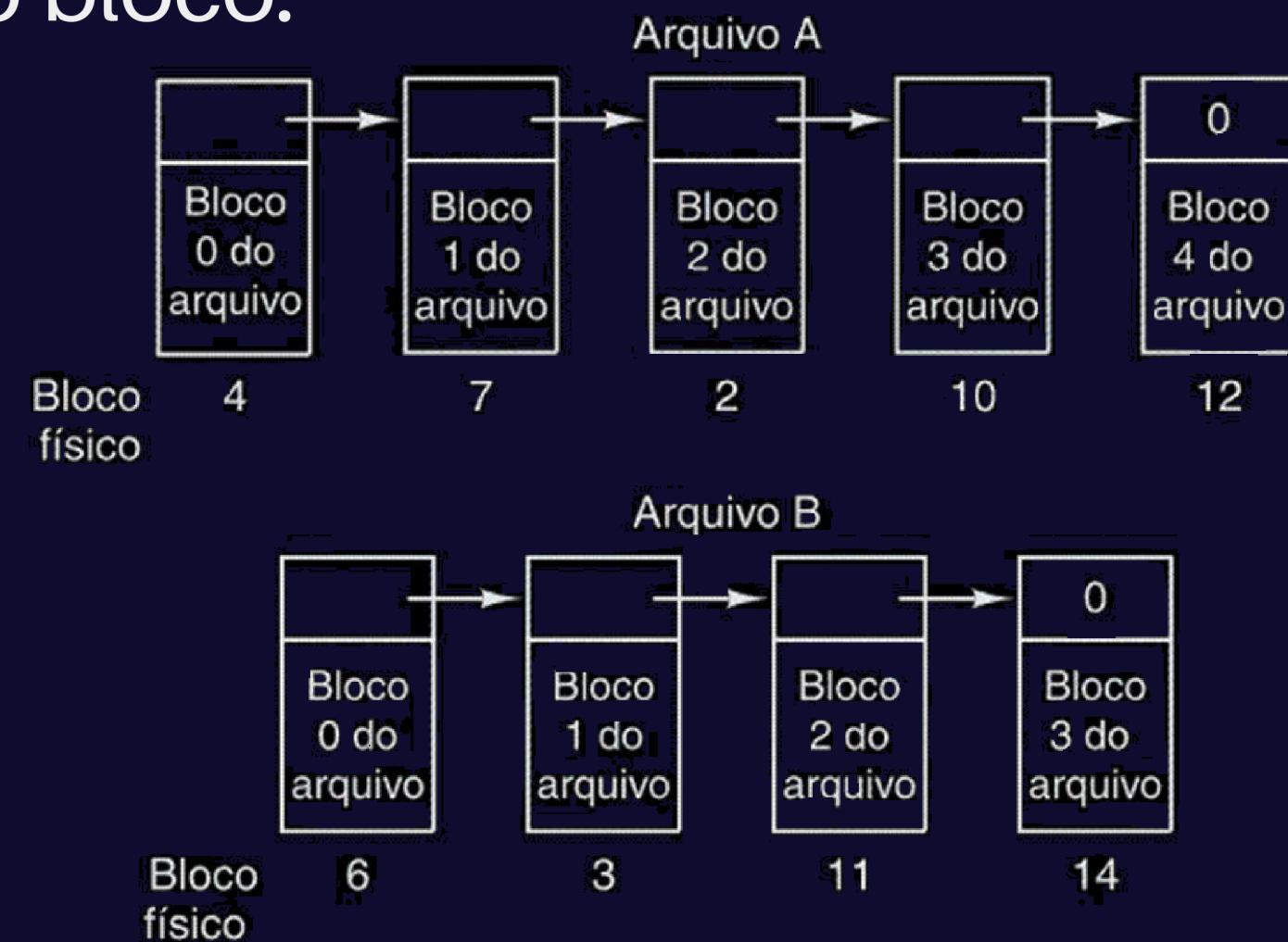
Worst-fit:



...

Alocação Encadeada

Na alocação encadeada os blocos de disco ficam dispersos no disco e a primeira palavra dos blocos servem como ponteiro para o próximo bloco.



Alocação encadeada

Vantagens

- Todos os blocos do disco podem ser utilizados.
- O sistema operacional só precisa manter o endereço do primeiro bloco.

Desvantagens

- O acesso aleatório é lento.
- A quantidade de dados dos blocos não são mais potências de 2.

Alocação Encadeada por tabela na memória

Nesse tipo de alocação diferentemente da alocação os ponteiros que indicam onde estão os próximos blocos ficam na memória, ao invés do disco.

A tabela onde os ponteiros ficam se chama de tabela FAT, que significa File Allocation Table.

Alocação Encadeada por tabela na memória

0	
1	11
2	14
3	
4	7
5	1
6	
7	13
8	
9	5
10	
11	-1
12	
13	2
14	-1
15	

← O arquivo A começa aqui.

← O arquivo B começa aqui.

← Bloco sem uso.

Alocação encadeada por tabela na memória

Vantagens

- O acesso é bem mais rápido se comparar com a alocação encadeada.
- Blocos voltam a ser de tamanhos de potencia de 2.
- Não sofre com as desvantagens das alocações anteriores.

Desvantagens

- A tabela inteira precisa estar na memória o tempo inteiro.
- Se o disco for grande muita memória será consumida.

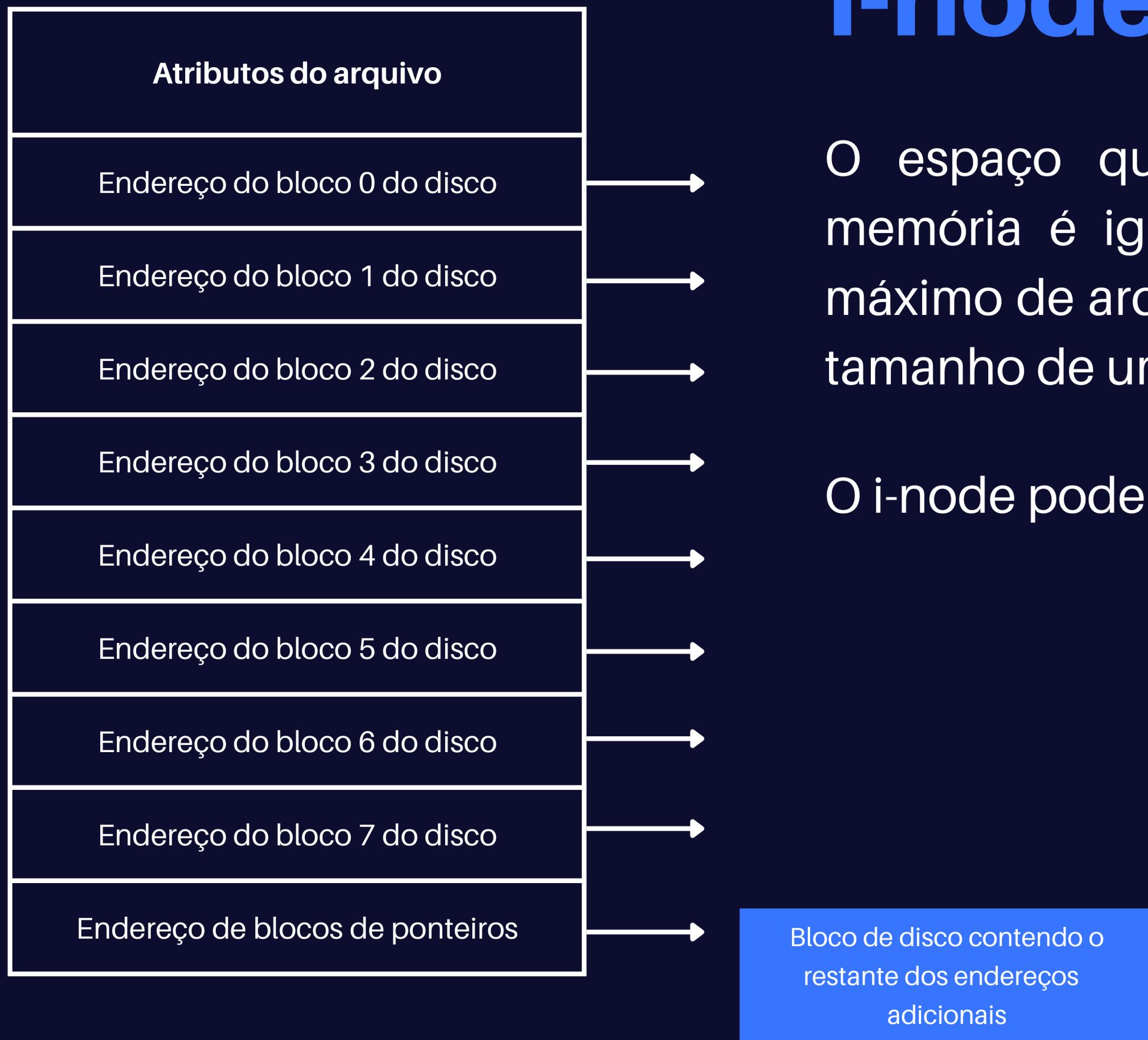
I-nodes

Nesse método, cada arquivo é associado a uma estrutura de dados chamada de i-nodes.

Os i-nodes mantém atributos dos arquivos e os endereços dos blocos de discos que contém os arquivos.

O i-node fica no disco e só vai para a memória, quando o arquivo está aberto.

I-nodes

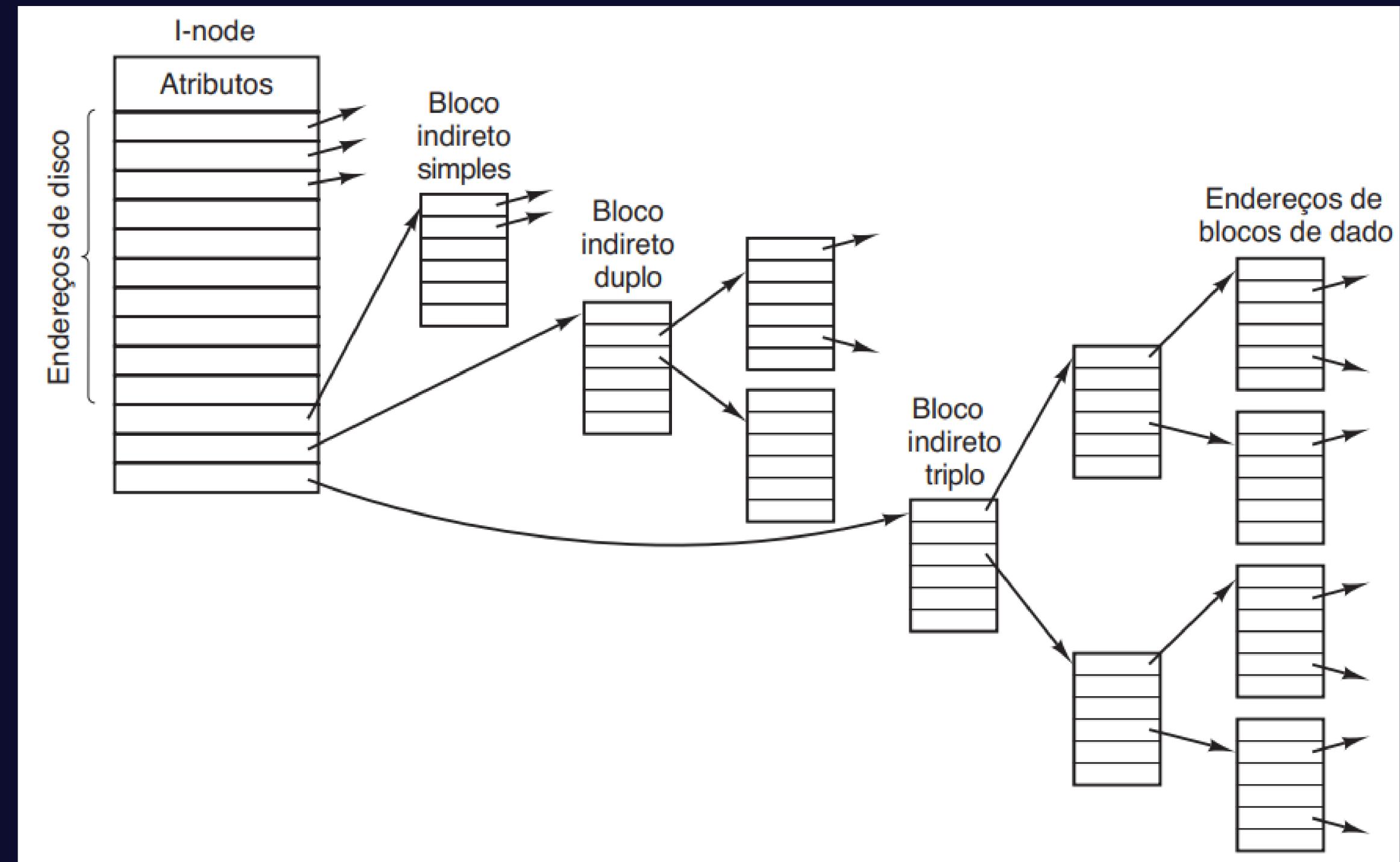


O espaço que é antecipadamente reservado na memória é igual a $k * n$, onde "k" é um número máximo de arquivos abertos simultaneamente e "n" o tamanho de uma estrutura de um i-node.

O i-node pode conter até X apontadores para blocos .

I-nodes

Exemplo de um i-node do Unix



TANENBAUM, 4º Ed. 2016

Proteção de arquivo

Tópicos

- Tipos de acesso de leitura.
- Senhas de acesso.
- Grupos de usuários.
- Lista de controle de acesso.

Proteção de arquivo

Um sistema de arquivo precisa possuir mecanismos próprios para proteger o acesso às informações gravadas em disco, além habilitar o compartilhamento de arquivos entre usuários.

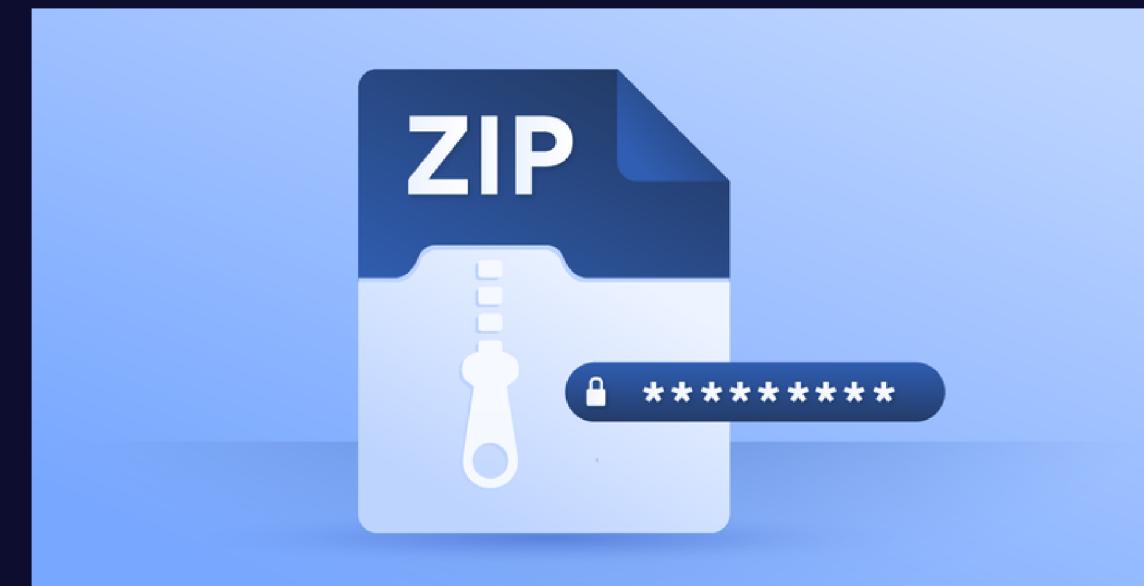
Em geral, o tipo de acesso a arquivos é implementado mediante a concessão ou não dos diferentes acessos que podem ser realizados. Os tipos de acesso:

- Leitura.
- Gravação.
- Execução.
- Eliminação.

Senhas de acesso

O controle de acesso se resume ao usuário ter conhecimento da senha e, consequentemente, ter a liberação do acesso ao arquivo concedida pelo sistema.

Se cada arquivo possui uma senha, ou o acesso é liberado totalmente ou o acesso não é liberado nada. Com isso não dá para se assegurar com quais operações o usuário vai realizar.



Google imagens

Grupos de usuários

Este tipo de proteção tem como princípio a associação de cada usuário do sistema a um grupo.

Os grupos de usuários são organizados logicamente com o objetivo de compartilhar arquivos e diretórios. Os usuários que desejam compartilhar arquivos entre si devem pertencer a um mesmo grupo.

Grupos de usuários

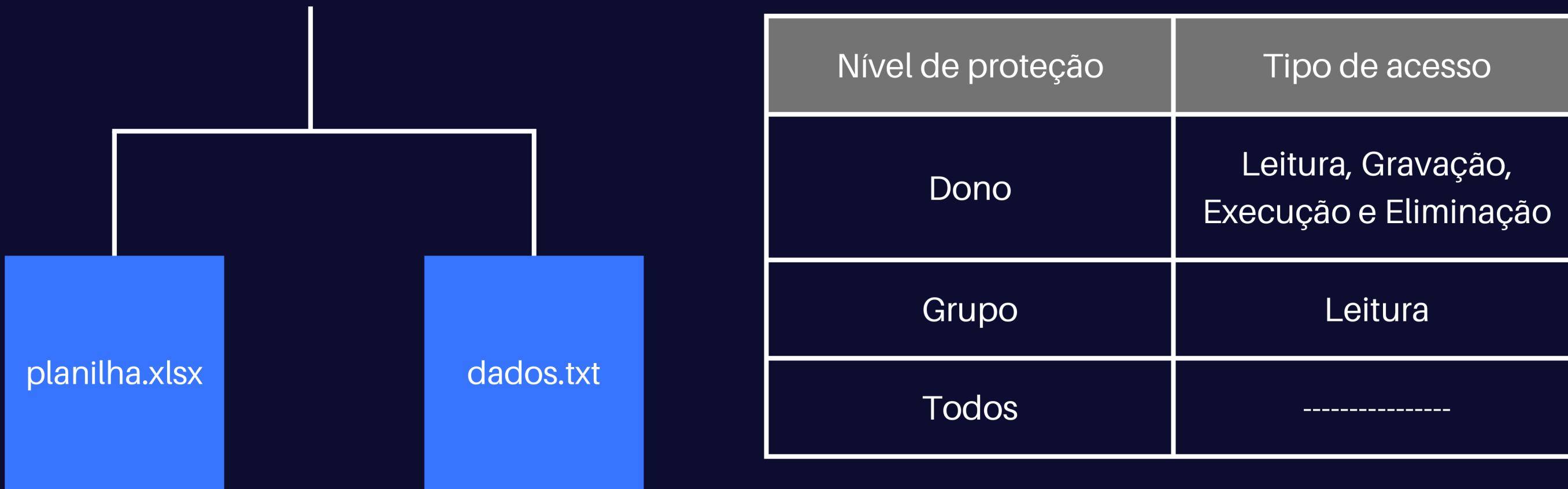
Nesse mecanismo são implementados três níveis de proteção:

- Owner (Dono).
- Group (Grupo).
- All (Todos).

E durante a criação do arquivo o usuário deve especificar se o arquivo deve ser acessado apenas pelo dono, pelo grupo que ele pertence ou por todos os usuários.

Devem ser especificados quais tipos de acesso devem ser concedidos (leitura, gravação, execução e eliminação) aos três níveis de proteção.

Grupos de usuários

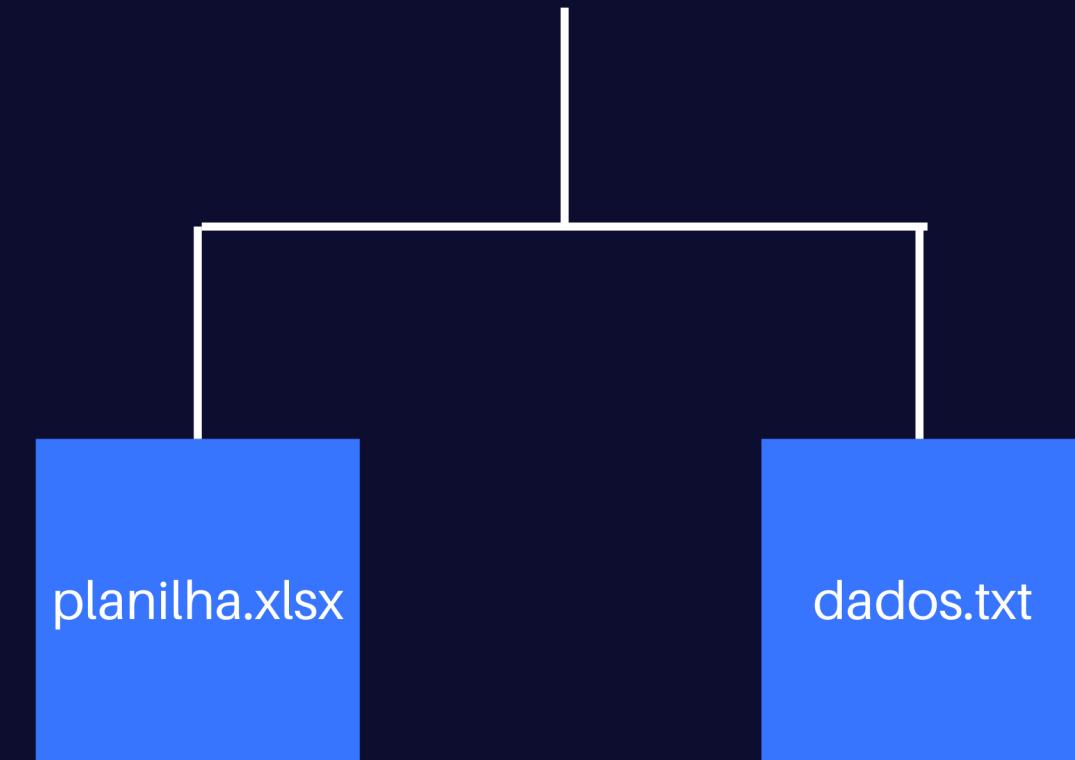


Lista de controle de acesso

Consiste em uma lista associada a cada arquivo, onde são especificados quais usuários e tipos de acesso permitidos. Quando um usuário tenta acessar um arquivo o sistema verifica se lista de controle autoriza a operação realizada por ele.

Lista de controle de acesso

Usuário	Tipo de acesso
João	Leitura e Gravação
Maria	Execução
Pedro	Leitura



Usuário	Tipo de acesso
João	Gravação
Maria	Eliminação
Pedro	Leitura e Gravação

Referências

- Tanenbaum, Andrew S. Sistemas Operacionais - Projeto E Implementação - 3^a Ed. BOOKMAN, 2010.
- Tanenbaum, Andrew S. Bos, Hebert. Sistemas Operacionais Modernos - 4^a Ed. São Paulo: Pearson Education do Brasil, 2016.
- Machado, Maia, Fundamentos De Sistemas Operacionais, 1a Edição, LTC, 2011.
- <https://www.youtube.com/watch?v=W9W9fRdzWjQ&t=430s>

Agradeço!

Obrigado pela atenção!

Reconhecimentos e Direitos Autorais

@autor: Gustavo de Oliveira Rego Moraes e Keven Gustavo dos Santos Gomes

@ contato: gustavo.moraes@discente.ufma.br e keven.gustavo@discente.ufma.br

@data última versão: 09/12/2023

@versão: 1.0

@outros repositórios: <https://github.com/gustvo-olive> e <https://github.com/KevenGustavo>

@Agradecimentos: Universidade Federal do Maranhão (UFMA), Professor Doutor Thales Levi Azevedo Valente, e colegas de curso.

@Copyright/License

Este material é resultado de um trabalho acadêmico para a disciplina **SISTEMAS OPERACIONAIS**, sobre a orientação do professor Dr. THALES LEVI AZEVEDO VALENTE, semestre letivo 2023.2, curso Engenharia da Computação, na Universidade Federal do Maranhão (UFMA). Todo o material sob esta licença é software livre: pode ser usado para fins acadêmicos e comerciais sem nenhum custo. Não há papelada, nem royalties, nem restrições de "copyleft" do tipo GNU. Ele é licenciado sob os termos da licença MIT reproduzida abaixo e, portanto, é compatível com GPL e também se qualifica como software de código aberto. É de domínio público. Os detalhes legais estão abaixo. O espírito desta licença é que você é livre para usar este material para qualquer finalidade, sem nenhum custo. O único requisito é que, se você usá-lo, nos dê crédito.

Copyright © 2023 Educational Material

Este material está licenciado sob a Licença MIT. É permitido o uso, cópia, modificação, e distribuição deste material para qualquer fim, desde que acompanhado deste aviso de direitos autorais.

O MATERIAL É FORNECIDO "COMO ESTÁ", SEM GARANTIA DE QUALQUER TIPO, EXPRESSA OU IMPLÍCITA, INCLUINDO, MAS NÃO SE LIMITANDO ÀS GARANTIAS DE COMERCIALIZAÇÃO, ADEQUAÇÃO A UM DETERMINADO FIM E NÃO VIOLAÇÃO. EM HIPÓTESE ALGUMA OS AUTORES OU DETENTORES DE DIREITOS AUTORAIS SERÃO RESPONSÁVEIS POR QUALQUER RECLAMAÇÃO, DANOS OU OUTRA RESPONSABILIDADE, SEJA EM UMA AÇÃO DE CONTRATO, ATO ILÍCITO OU DE OUTRA FORMA, DECORRENTE DE, OU EM CONEXÃO COM O MATERIAL OU O USO OU OUTRAS NEGOCIAÇÕES NO MATERIAL.

Para mais informações sobre a Licença MIT: <https://opensource.org/licenses/MIT>.