

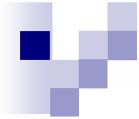
IFSULDEMINAS, campus Muzambinho
Curso de Ciência da Computação



Manipulação de Dados e Dispositivos de Armazenamento

Prof. Ricardo José Martins
ricardo.martins@muz.ifsuldeminas.edu.br

Curso de Bacharelado em Ciência da Computação
AED III – Algoritmo e Estruturas de Dados III



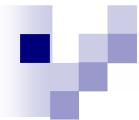
Armazenamento de Dados

Existe uma diversidade de dispositivos de armazenamento de dados num sistema computacional. Em geral, estes dispositivos são classificados da seguinte forma:

Dispositivos de Armazenamento Primário (exemplos: memória RAM e cache): estes dispositivos são manipulados diretamente pela CPU; em geral, são voláteis e possuem uma alta taxa de transferência de dados, mas capacidade limitada devido ao alto custo.

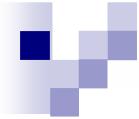
Dispositivo de Armazenamento secundário (exemplos: HD, DVD, fitas): estes dispositivos caracterizam-se por não serem voláteis e possuírem, em sua maioria, uma maior capacidade de armazenamento a um custo menor; possuem o problema de serem muito mais lentos que os dispositivos primários.

Outras metodologias de classificação podem ser definidas segundo alguns critérios, tais como: **velocidade, o custo por unidade de dados; confiabilidade, entre outros.**

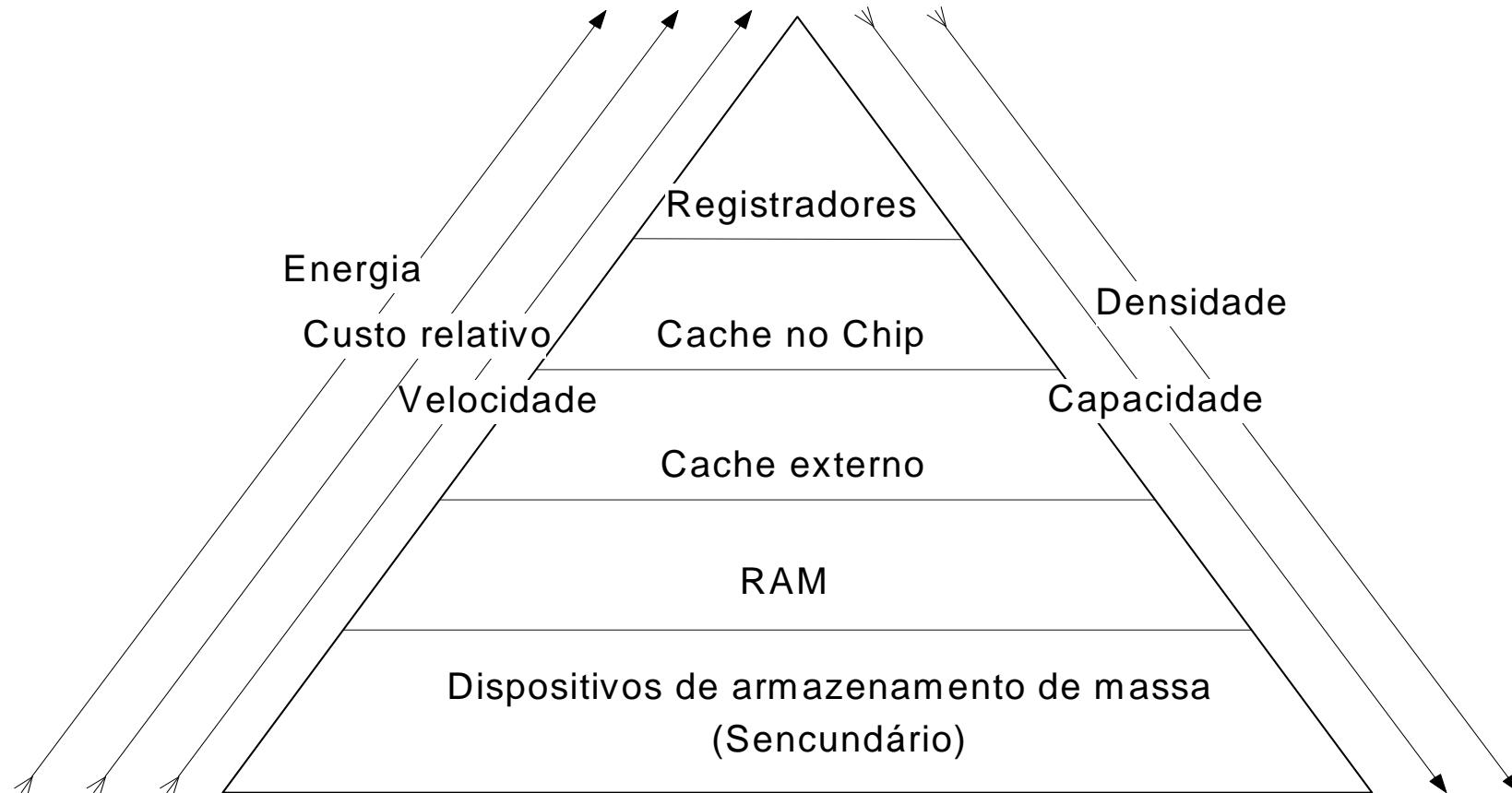


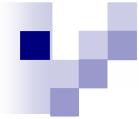
Exemplos de Dispositivos de Armazenamento de Dados

- **Registradores** – pequena porção de memória interna à CPU para armazenamento de resultados intermediários.
- **Cache** – memória de pequeno porte extremamente rápida se comparada à memória principal. Geralmente apresenta-se organizada em dois níveis, cache interna e externa.
- **Memória principal** - meio de armazenamento volátil no qual os dados e instruções estão disponíveis para uso pelo processador.
- **Armazenagem em disco** – É o principal meio de armazenamento de dados persistentes, ou seja, não voláteis. Para serem manipulados pela CPU é necessária à transferência destes para a memória principal, normalmente este processo é feito por um hardware específico de controle do dispositivo. Caracterizam-se pela alta capacidade de armazenamento e pelo baixo custo.
- **Armazenagem em fita** - Normalmente usadas para armazenamento *off-line* (cópias de segurança). É bem barato, mas seu acesso é muito mais lento e seqüencial.



Hierarquia de Memória

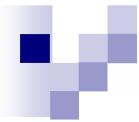




Características da Memória Secundária

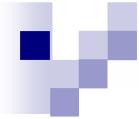
Em relação aos dispositivos de armazenagem secundária, pode-se considerar na sua análise ou aquisição as seguintes características:

- **Capacidade** - é a quantidade total de dados que pode ser armazenada no dispositivo (KB, MB, GB, TB, dentre outros).
- **Portabilidade** - as fitas magnéticas, CDR, CDRW, DVD, *Zip drive* e alguns outros mecanismos são removíveis, permitindo o armazenagem *off-line* de arquivos e o transporte destes dados, melhorando a confiabilidade das cópias de segurança.
- **Custo relativo** - é definido como o preço (custo) por unidade de armazenagem. Geralmente, o valor de um dispositivo aumenta proporcionalmente com a velocidade e forma de acesso.
- **Tamanho do registro** - representa o tamanho de um conjunto de dados contínuos que podem ser endereçados pelo mecanismo.
- **Método de acesso** - permite acesso direto e acesso seqüencial ou apenas seqüencial.
- **Velocidade de transferência de acesso** - é a velocidade na qual os dados podem ser transferidos entre a memória interna e o dispositivo.



Características da Memória Secundária

- **Tempo de busca (seek time) ou posicionamento** – para os dispositivos de disco (HD, CD-ROM, dentre outros), a cada operação de leitura/gravação antecede uma procura, que fisicamente consiste no tempo que leva para posicionar a cabeça de leitura/gravação sobre a trilha (cilindro) que contém os dados (registro) a serem recuperados.
- **Tempo de latência** – após o mecanismo posicionar a cabeça de leitura/gravação sobre a trilha desejada, este leva um tempo adicional de espera da rotação do cilindro, antes que o início do setor ou registro a ser acessado fique posicionado sob a cabeça de leitura/gravação; fita magnética é o tempo para alcançar a velocidade de operação a partir da posição “parada”.
- **Tempo de transferência ou transmissão** – é o tempo gasto para a que os dados do disco, cujo início se encontra sob a cabeça de leitura/gravação, seja apresentado a interface do disco e se torne disponível para manipulação. O tempo de transferência = (bytes transferidos / bytes por trilha) X tempo de rotação
- **Densidade da área de armazenagem** – é a capacidade relativa de armazenagem (número de bits, por exemplo) do dispositivo. É definido como o número de bits ou bytes que podem ser armazenados por unidade de área.



Dispositivos Existentes

Dispositivos de Acesso Seqüencial

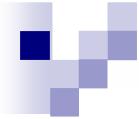
Exemplo típico: Fita

- Extremamente lentos.
- Utilizados para armazenar dados que serão acessados todos de uma só vez e não serão alterados posteriormente.
- Uso típico: Backup
- Gerência de arquivos não se ocupa destes dispositivos.

Dispositivos de Acesso Randômico (Aleatório)

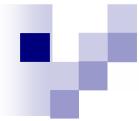
Exemplo típico: Disco

- Relativamente rápidos.
- Utilizados para dados que sofrem freqüentes pesquisas, dados muito alterados ou não definitivos.
- Uso típico: área de trabalho ou banco de dados.
- Base para as técnicas que serão vistas.



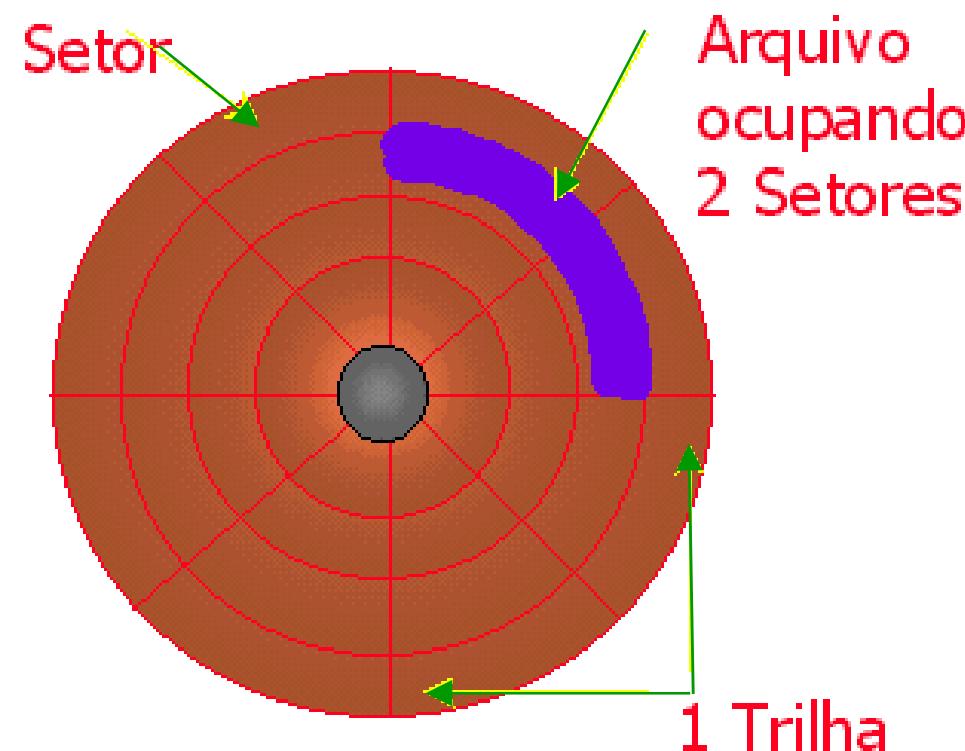
Estrutura dos Discos

- Exemplo típico para uma memória secundária de acesso aleatório.
- Compreender a forma de armazenamento é importante.
- Modelo de estruturação do disco serve para todos os outros meios secundários.
- Dados são organizados em superfícies, trilhas e setores, blocos.
- Um arquivo pode ser imaginado como sendo constituído por uma seqüência de dados no disco.
- Acesso é feito através do posicionamento de um cabeçotes de **r/w** em qualquer ponto.
- Uma unidade de alocação do disco (um bloco ou um setor) possui um endereço físico no disco.

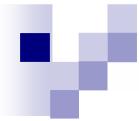


Estrutura dos Discos

**Modelo
Genérico**



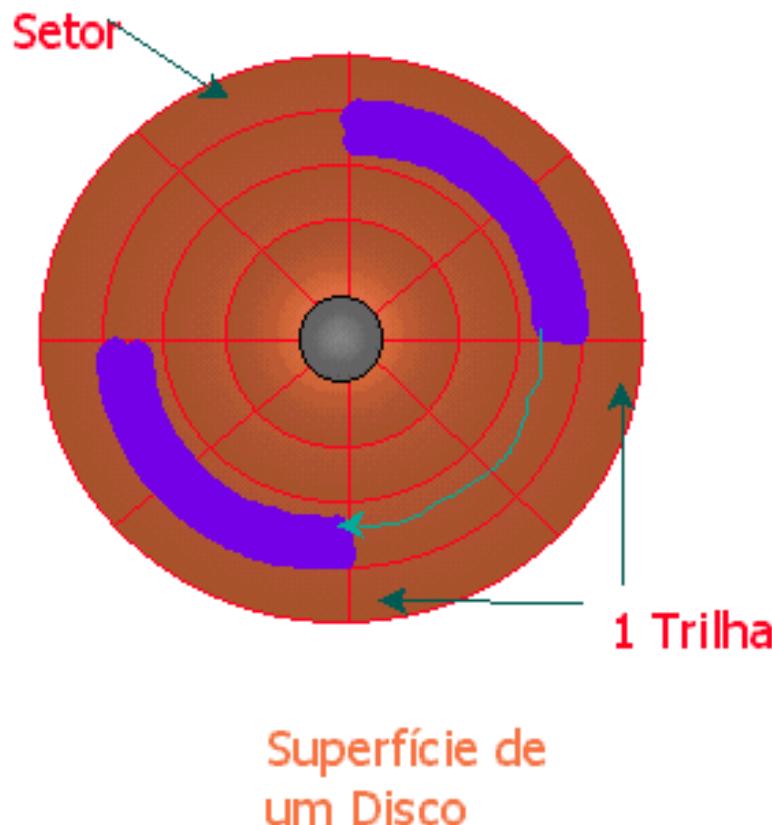
**Superfície de
um Disco**

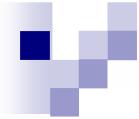


Estrutura dos Discos

Modelo mais realista

O disco pode apresentar fragmentação: i.e., os dados não estão exatamente em espaços contíguos, mas em unidades de alocação longe umas das outras, encadeadas como uma lista.

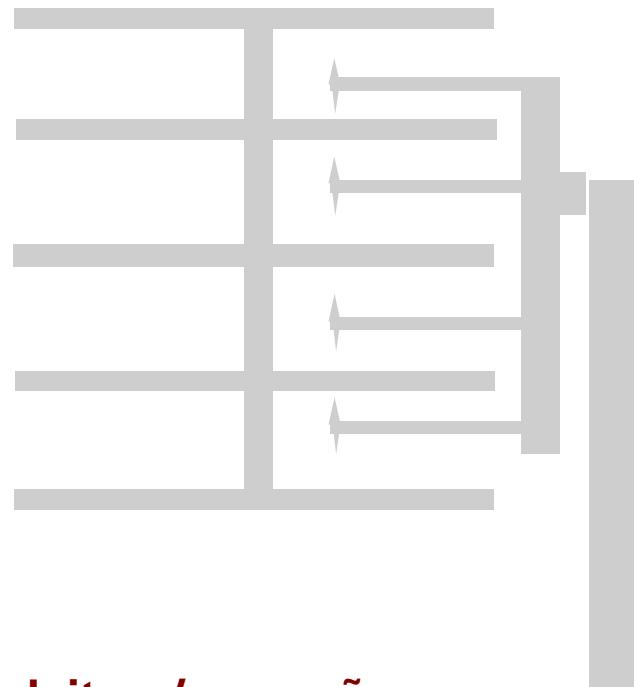




Estrutura dos Discos

■ Exemplo: Disco Rígido

Pratos



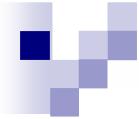
Eixo (*Spindle*)

Braços de controle de acesso

Cilindro
atual de
leitura

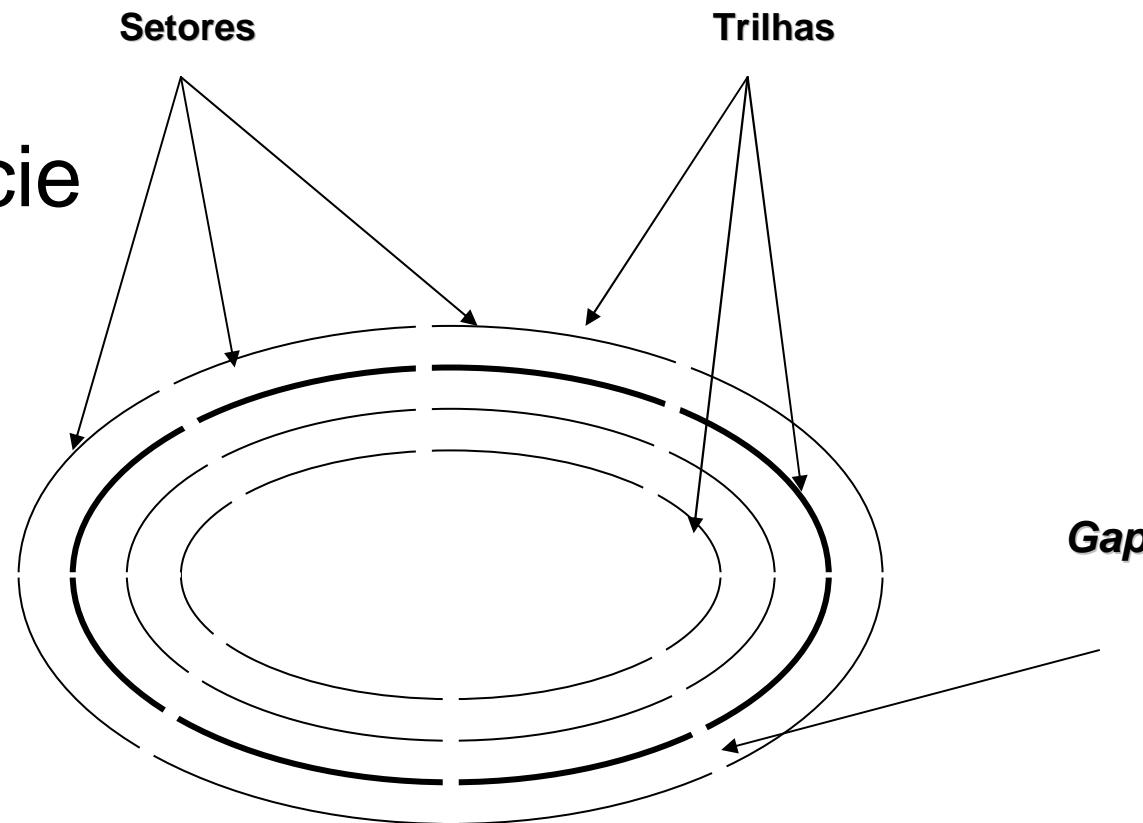
Cabeças de leitura/gravação

deslocamento

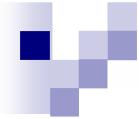


Estrutura dos Discos

■ Superfície



Organização de um prato (trilhas e setores)



Estrutura dos Discos

- **Capacidade de Armazenamento**

capac. da trilha = número de setores por trilha X número bytes por setor

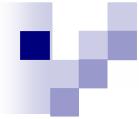
capac. do cilindro = número de trilhas por cilindro X capac. da trilha

capacidade do HD = número de cilindros X capacidade do cilindro

- **Tempo de Acesso**

O **tempo de acesso** (*access time*) é dado pelo tempo de posicionamento adicionado do tempo de latência, e o tempo de transferencia.

- Ler um arquivo **seqüencialmente** é mais rápido, pois o acesso seqüencial minimiza o **tempo de busca** e de **latência**.
- Para uma **leitura aleatória**, quase todos os acessos terão tempo de busca, tempo de latência e tempo de transferência.

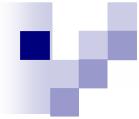


Exercício Dirigido

Considere uma unidade de disco magnético com as seguintes características :

- Setor : 512 bytes
- Taxa de transferencia : 20 Mbytes/seg
- Tempo de seek médio : 8 ms
- Tempo latência médio : 4 ms
- Capacidade por trilha : 85 Kbytes
- Capacidade total : 9100 Mbytes
- No. de pratos/superfícies : 10/20

- a. Qual é o numero de cilindros do disco ?
- b. Quantos registros de 128 bytes podem ser armazenados em 1 cilindro do disco ?
- c. Calcule o tempo gasto para a leitura seqüencial de 50.000 setores.
- d. Calcule o tempo gasto para a leitura aleatória de 50.000 setores.



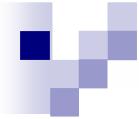
Soluções para as Questões do Exercício Dirigido

a. Qual é o numero de cilindros do disco ?

$$9100 \text{ MB} / 20 \text{ superfícies} = 455 \text{ MB} / \text{superfície}$$

$$455 * 1024 * 1024 \text{ Bytes} / (85 * 1024 \text{ Bytes} / \text{trilha}) = 5.481 \text{ trilhas/superfície.}$$

O número de cilindros no disco é igual ao número de trilhas por superfície. Portanto, há 5.481 cilindros no disco.

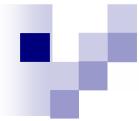


Soluções para as Questões do Exercício Dirigido

- b. Quantos registros de 128 bytes podem ser armazenados em 1 cilindro do disco ?

Em um cilindro do disco cabem $(85 * 1024 \text{ bytes / trilha}) * (20 \text{ trilhas / cilindro}) = 1.740.800 \text{ bytes}$.

Portanto, cabem $1.740.800 \text{ bytes} / (128 \text{ bytes / registros}) = 13.600 \text{ registros em um cilindro.}$



Soluções para as Questões do Exercício Dirigido

c. Calcule o tempo gasto para a leitura seqüencial de 50.000 setores.

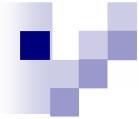
A leitura seqüencial envolve um tempo de **seek**, um tempo de latência e o tempo gasto com a transferência.

Tempo de seek: 8 ms

Tempo de latência: 4ms

Tempo de transferência: $50.000 \text{ setores} * 512 \text{ bytes / setor} = 25.600.000 \text{ bytes} / (20 * 1024 * 1024 \text{ bytes / seg}) = 1,22 \text{ seg}$

Tempo total: $0,008 + 0,004 + 1,22 = 1,23 \text{ seg}$

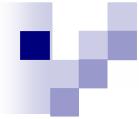


Soluções para as Questões do Exercício Dirigido

d. Calcule o tempo gasto para a leitura aleatória de 50.000 setores.

A leitura de cada setor requer um tempo de **seek**, um tempo de latência e o tempo de transferência. Dessa maneira, serão necessários 50.000 **seeks** e 50.000 latências no total, além do tempo de transferência.

Tempo total: $50.000 * (0,008 + 0,004 \text{ seg}) + 1,22 \text{ seg}$ (tempo de transferência, calculado na questão anterior) = 601,22 seg



Soluções para as Questões do Exercício Dirigido

e. Calcule o tempo gasto para a leitura seqüencial de 50.000 setores.

A leitura seqüencial envolve um tempo de **seek**, um tempo de latência e o tempo gasto com a transferência.

Tempo de seek: 8 ms

Tempo de latência: 4ms

Tempo de transferência: $50.000 \text{ setores} * 512 \text{ bytes / setor} = 25.600.000 \text{ bytes} / (20 * 1024 * 1024 \text{ bytes / seg}) = 1,22 \text{ seg}$

Tempo total: $0,008 + 0,004 + 1,22 = 1,23 \text{ seg}$