Memória Secundária

6897/9895 – Organização e Recuperação de Dados

Profa. Valéria

UEM – CTC – DIN

Slides preparados com base no Cap. 3 do livro FOLK, M.J. & ZOELLICK, B. *File Structures*. 2nd Edition, Addison-Wesley Publishing Company, 1992.

News room > News releases >

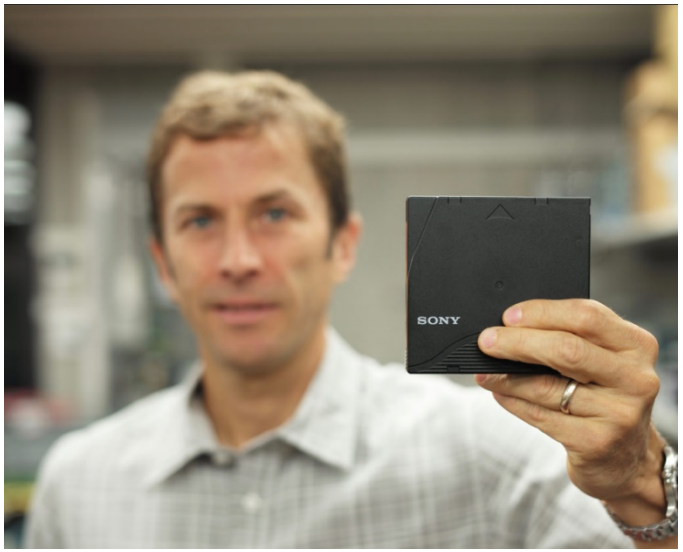
IBM sets new record for magnetic tape storage; makes tape competitive for cloud storage

Increased storage density demonstrates viability of scaling the tape roadmap for another decade

TSUKUBA, Japan - 02 Aug 2017: IBM (NYSE: IBM) Research scientists have achieved a new world record in tape storage – their fifth since 2006. The new record of 201 Gb/in^2 (gigabits per square inch) in areal density was achieved on a prototype sputtered magnetic tape developed by Sony Storage Media Solutions. The scientists presented the achievement today at the 28th Magnetic Recording Conference (TMRC 2017) here.

(...)

This new record areal recording density is more than 20 times the areal density used in current state of the art commercial tape drives such as the IBM TS1155 enterprise tape drive, and it enables the potential to record up to about 330 terabytes (TB) of uncompressed data* on a single tape cartridge that would fit in the palm of your hand. 330 terabytes of data are comparable to the text of 330 million books, which would fill a bookshelf that stretches slightly beyond the northeastern to the southwestern most tips of Ja



- Durabilidade → 20-30 anos

- Fonte:

  https://www-03.ibm.com/press/us/en/pressrelease/52904.wss

IBM's Tale of the Tape

More than 60 years of tape innovation

	2006	2010	2014	2015	2017
Aerial Density (bits per sq inch)	6.67 Billion	29.5 Billion	85.9 Billion	123 Billion	201 Billion
Cartridge Capacity (Terabytes)	8	35	154	220	330
# of Books Stored	8 Million	35 Million	154 Million	220 Million	330 Million
Track Width	1.5 µm	0.45 µm	0.177 µm	0.140 µm	103 nm
Linear Density (bits per inch)	400'000	518'000	600'000	680'000	818'000
Tape Material	Barium Ferrite	Barium Ferrite	Barium Ferrite	Barium Ferrite	Sputtered Media
Tape Thickness (micrometers)	6.1	5.9	4.3	4.3	4.7
Tape Length (meters)	890	917	1255	1255	1098

#5thtaperecord

IBM

COMPUTERS

Western Digital 14TB hard drive sets storage record

The enterprise-class Ultrastar Hs14 drive is the world's first single drive to launch with this massive capacity.

BY **JOSHUA GOLDMAN** / OCTOBER 4, 2017 7:19 AM PDT

Western Digital's Ultrastar Hs14 packs a massive 14TB capacity into a single drive -- a world's first -- and you can't have it. Not yet at least.

The new enterprise-class drive falls under the company's HGST brand and is built for cloud and data centers. According to Western Digital, it has 40 percent more capacity and more than twice the write performance of its predecessor.

- Fonte:
  https://www.cnet.com/news/western-digital-14tb-hard-drive-sets-storage-record/

Discos vs. RAM

- Por que os discos são tão lentos?
  - Assim como as fitas, eles são dispositivos magnéticos que envolvem partes mecânicas em oposição à memória RAM, que depende apenas de energia elétrica

- Como os discos funcionam e como são organizados?

Discos Magnéticos

- Pertencem à categoria **DASDs** (*Direct Access Storage Devices*)
  - Permitem o acesso direto ao dado de um endereço específico

- Em oposição aos dispositivos seriais, como as fitas magnéticas
  - Um dado não pode ser acessado antes que todos os anteriores sejam lidos

Discos Magnéticos

- **Composição**
  - Um ou mais pratos sobrepostos atravessados por um eixo
  - Pratos revestidos por uma camada magnética extremamente fina*
    - São aplicados campos magnéticos e as partículas reagem a esses campos
    - A cabeça de leitura e gravação é um eletroímã e sua polaridade pode ser alternada constantemente
    - Com o disco girando continuamente, variando a polaridade da cabeça de gravação, variamos também a direção dos polos positivos e negativos das moléculas da superfície magnética
    - De acordo com a direção dos polos, temos um bit 1 ou 0
  - Atualmente são utilizadas as duas superfícies de todos os pratos
  - A capacidade do disco está relacionada a densidade de armazenamento dos pratos e a quantidade de pratos:
    - Densidade → número de trilhas por superfície
    - Mais pratos → maior a capacidade → menor a espessura dos pratos → maior o custo do disco

* Fonte: https://www.gta.ufrj.br/grad/07_1/hd/func.html

Discos Magnéticos

- **Exemplos de discos magnéticos**
  - **Hard disks** (discos rígidos)
    - Alta capacidade
    - Baixo custo por bit
    - Lento em relação à RAM
  - **Floppy disks** (disquete)
    - Baixa capacidade
    - Baixo custo
    - Muito mais lento que o HD
  - **ZIP disks**
    - Capacidade de até 750 MB
    - Velocidade: Floppy < ZIP < HD

Curiosidade

- Primeiro disco rígido → IBM 350, construído em 1956
- 50 discos de 24 polegadas de diâmetro (~60 cm),
- Capacidade total de 4,36 MB
- 1,70m de altura e de comprimento, quase 1 tonelada
- Custava 35 mil dólares
- Chamado “unidade de disco”

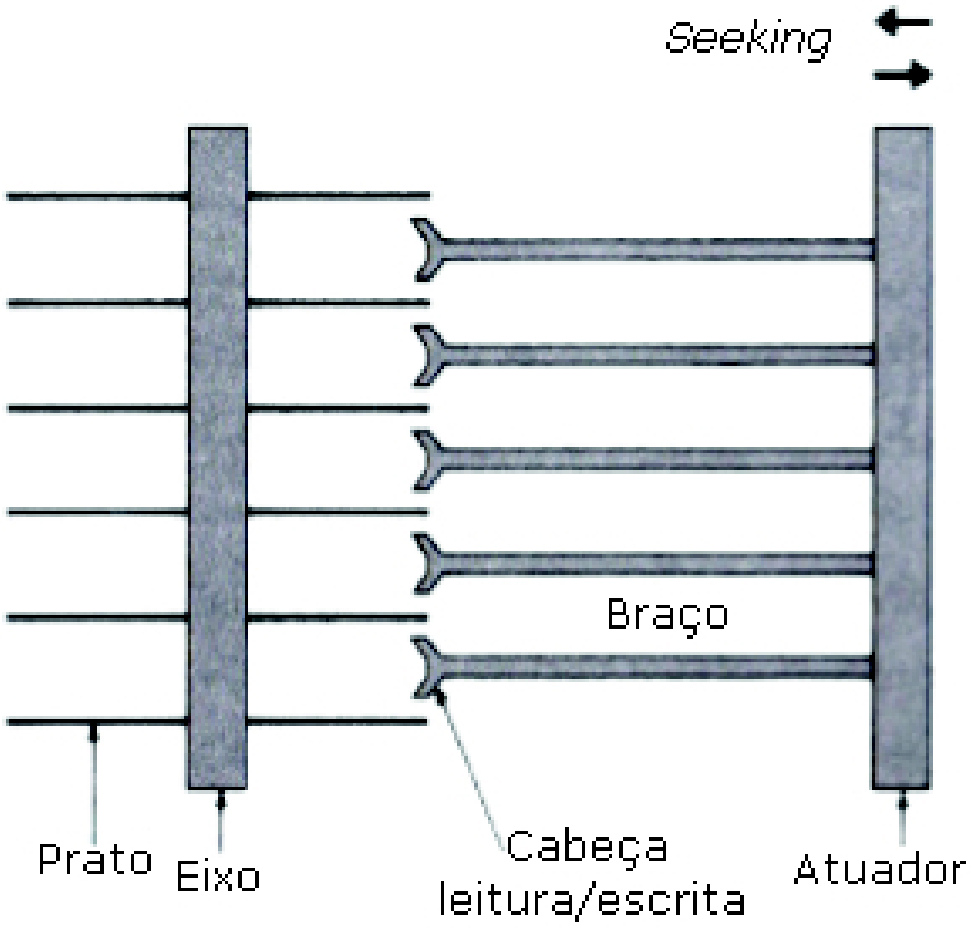
- Em 1973, IBM lançou o Winchester, que é considerado o pai dos HDs modernos

Discos Magnéticos

- **Funcionamento**
  - Os pratos giram em uma velocidade constante, que é medida em rotações por minuto (RPM)
  - O disco dispõe de **cabeças de leitura/escrita** posicionadas entre cada par de pratos, que realizam a leitura/escrita de dados nas superfícies
    - A distância entre a cabeça de leitura/escrita e um prato é extremamente pequena, mas existe!
    - Na verdade, o "colchão de ar" que se forma com a alta rotação afasta as cabeças de L/E da superfície
  - As cabeças de leitura/escrita são sustentadas por um braço e um atuador é responsável por posicioná-las na superfície do prato
    - Esse posicionamento é chamado de *seeking*

Discos Magnéticos

- **Esquema dos componentes de um disco**

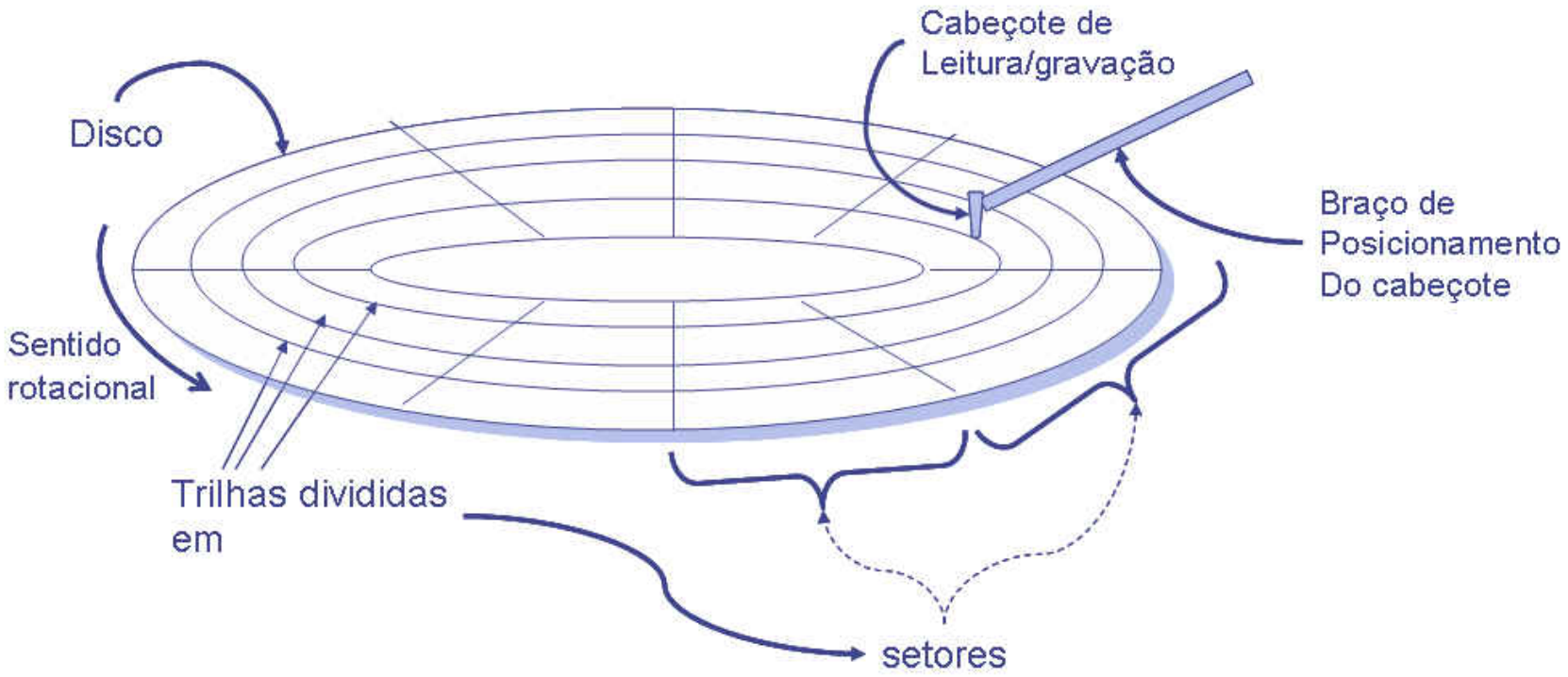


Discos Magnéticos

- **Organização dos discos**
  - Os dados são armazenados em trilhas sucessivas na superfície de um prato
    - A trilha mais externa é a de número zero
  - Cada trilha é dividida em setores
    - Um setor é a menor parte endereçável do disco (normalmente 512 bytes)
  - Quando é realizada uma chamada READ() por um byte específico
    - O S.O. solicita ao disco a busca pela superfície, a trilha e o setor corretos
    - Os dados de um setor inteiro são copiados para um *buffer*
    - A partir do *buffer*, se encontra o byte específico que foi requisitado
  - Normalmente os discos vêm setorizados de fábrica (formatação física)

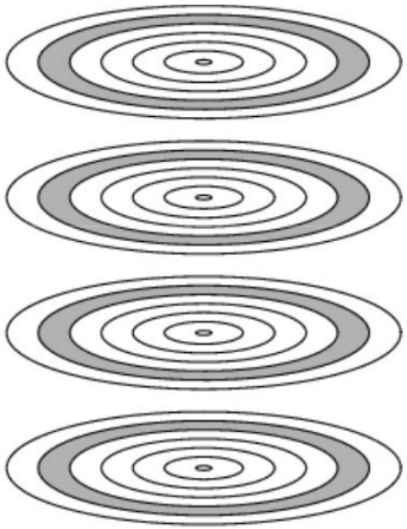
Discos Magnéticos

- **Organização de um prato**

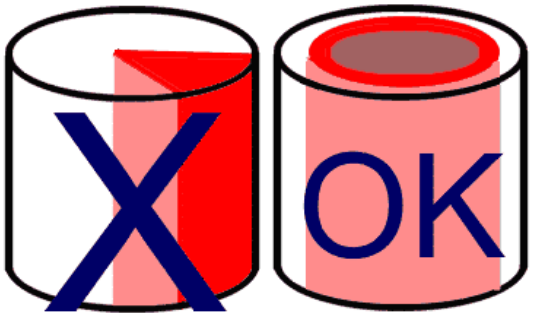


Discos Magnéticos

- **Organização dos discos**
  - Em um disco formado por diversos pratos, o conjunto de trilhas na mesma direção (sobrepostas em pratos diferentes) forma um cilindro



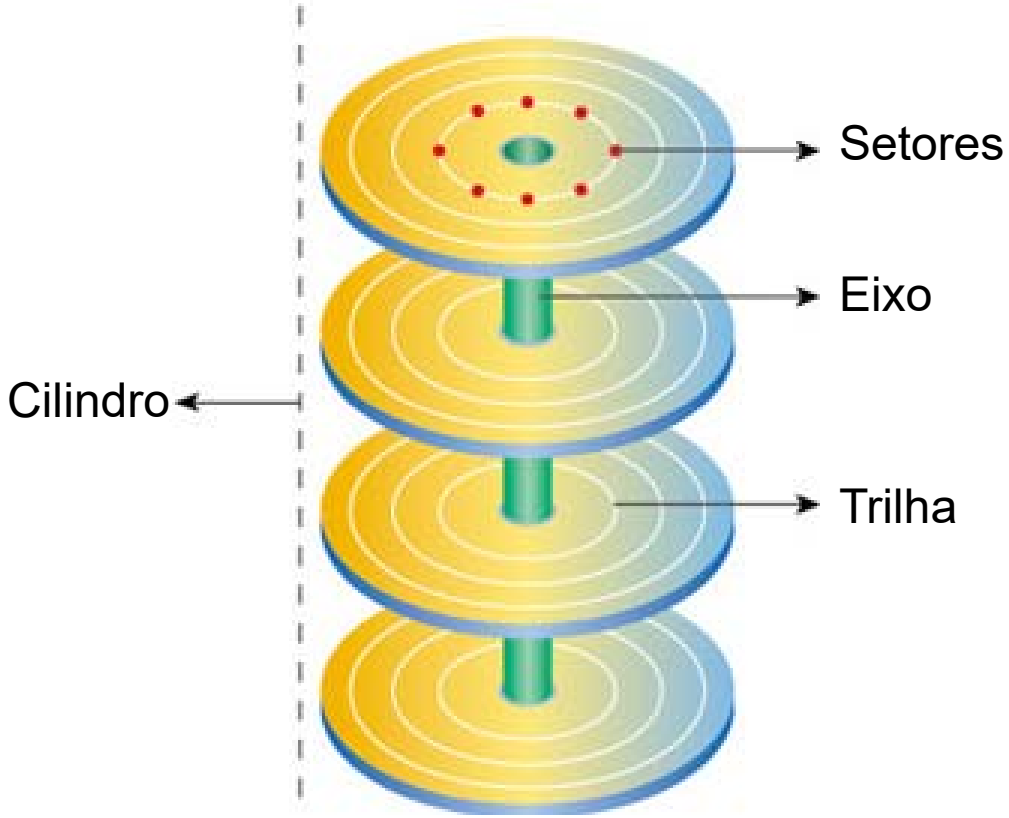
  - Toda a informação em um cilindro pode ser acessada sem *seeking* adicional
    - O *seeking* normalmente é a parte mais lenta em uma operação de leitura/escrita



Discos Magnéticos

- **Organização dos discos**

4 Discos
3 Trilhas/superfície
24 Trilhas
3 Cilindros
8 Setores/trilha
192 Setores



Discos Magnéticos

- **Estimativa de capacidade**
  - A quantidade de dados que pode ser armazenada em uma trilha depende do quão densamente os bits podem ser armazenados
    - A densidade depende da qualidade da mídia e do tamanho das cabeças de leitura/escrita
    - Exemplo:
      - Disco rígido 1TB SATA III Segate 7200RPM (± R$250,00)
      - 16.383 trilhas/superfície e 63 setores (de 4k) por trilha (256 KB/trilha)

Discos Magnéticos

- **Estimativa de capacidade**
- Sendo que:
  - Número de Cilindros = Número de Trilhas/Superfície
  - Número de Trilhas do Cilindro = 2 x Número de Pratos

- Temos que:
  - Capacidade da **Trilha** =

    Número bytes por Setor x Número de Setores por Trilha
  - Capacidade do **Cilindro** =

    Capacidade da Trilha x Número de Trilhas do Cilindro
  - Capacidade do **Disco** =

    Capacidade do Cilindro x Número de Cilindros

Discos Magnéticos

- **Estimativa de capacidade e espaço necessários**
- Exemplo
  - Propriedades do disco
    - 512 bytes/setor
    - 63 setores/trilha
    - 16 trilhas/cilindro
    - 4080 cilindros
  - Quantos **cilindros** são necessários para armazenar um arquivo de 50.000 registros de 256 bytes cada?

Discos Magnéticos

- **Estimativa de capacidade e espaço necessários**
- Exemplo
  - Propriedades do disco
    - 512 bytes/setor
    - 63 setores/trilha
    - 16 trilhas/cilindro
    - 4080 cilindros

Capacidades:
1 trilha = 63 setores de 512 bytes = 32.256 bytes
1 cilindro = 16 trilhas = 1.008 setores = 516.096 bytes

  - Quantos **cilindros** são necessários para armazenar um arquivo de 50.000 registros de 256 bytes cada?
    - Como cada setor tem 512 bytes, é possível armazenar 2 registros por setor, sendo necessário um total de 25.000 setores
    - Em 1 cilindro temos 1.008 setores (63 setores x 16 trilhas)
    - Então, para armazenar 25.000 setores são necessários 24,8 cilindros (25.000 setores/1.008 setores)

Discos Magnéticos

- **Clusters**
  - Organização lógica que visa aumentar o desempenho e mantida pelo **Gerenciador de Arquivos** (*File Manager),* presente no S.O.
    - Quando um arquivo é acessado por uma aplicação, é o gerenciador de arquivos quem associa o arquivo lógico às suas posições físicas
    - O arquivo é visto como uma série de *clusters*
  - Para fazer esse mapeamento, o gerenciador de arquivos utiliza uma tabela de alocação de arquivos (chamada de *File Allocation Table* (FAT) em alguns S.Os.)

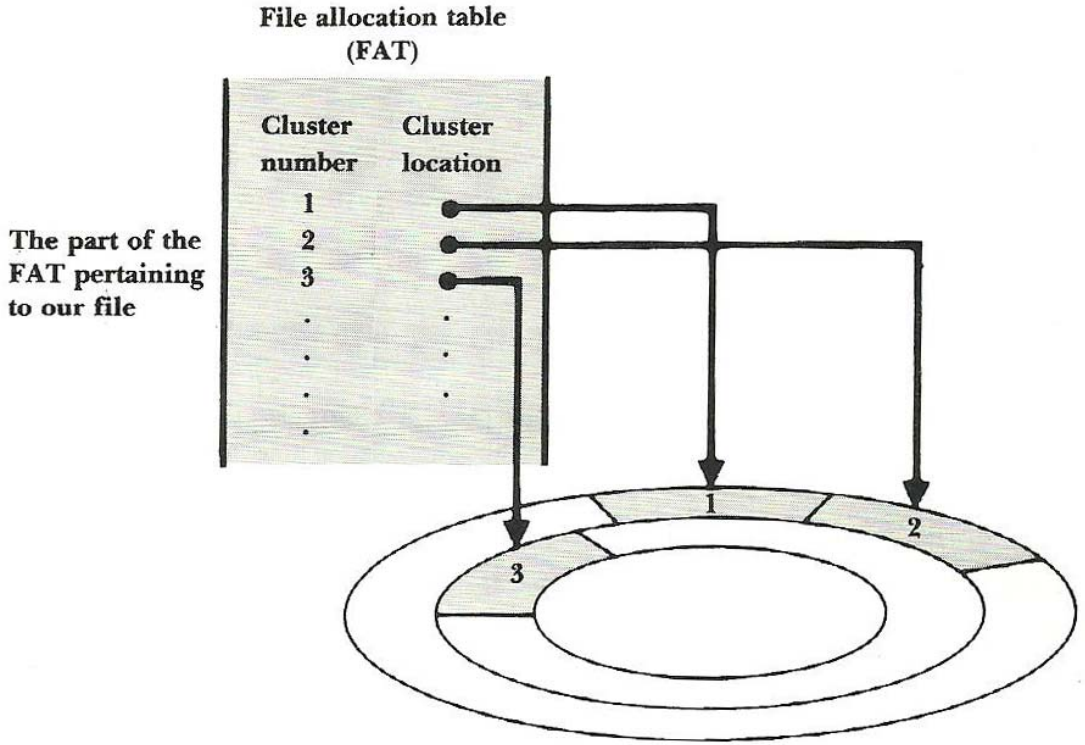
Discos Magnéticos

- ***Clusters***
  - Em vez da tabela de alocação endereçar setores, endereça *clusters*
    - Um *cluster* é um conjunto de um ou mais setores contíguos do disco
      - Todos os setores de um *cluster* podem ser lidos sem *seeks* adicionais e sem a necessidade de consultas adicionais à tabela de alocação
    - A tabela de alocação contém uma lista de todos os *clusters* de um arquivo, ordenados de acordo com a ordem lógica dos setores que eles contém

- O tamanho do setor é uma característica do disco
- O tamanho do cluster é uma característica do S.O.

Discos Magnéticos

- ***Clusters***
  - Cada *cluster* do disco é usado para um único arquivo, ou seja, em um mesmo *cluster* não haverá informações sobre mais de um arquivo



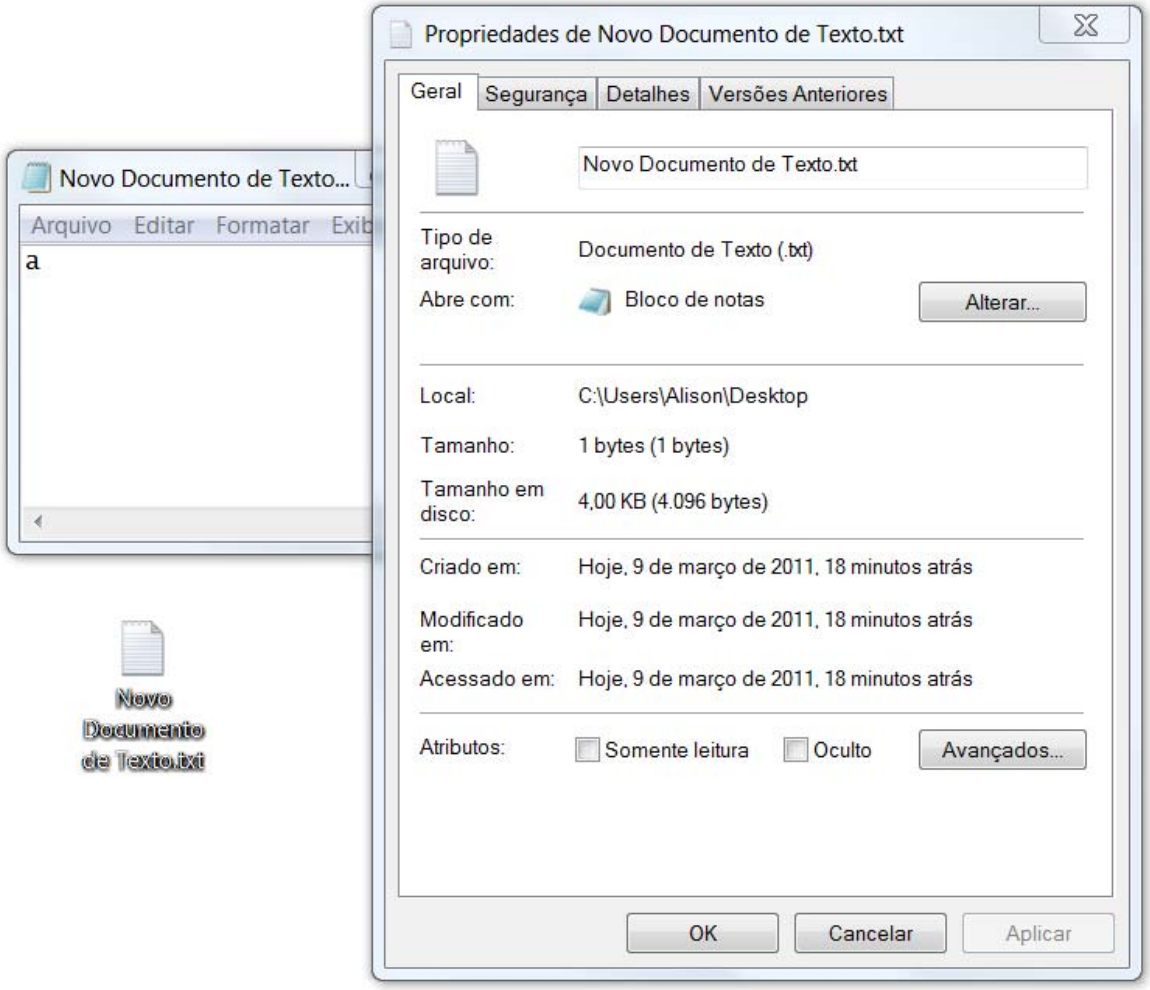
Discos Magnéticos

- **Comparação do tamanho de *cluster* entre partições**

Tamanho do volume	Tamanho do cluster de FAT16	Tamanho do cluster de FAT32	Tamanho do cluster de NTFS
7 MB–16 MB	2 KB	Não compatível	512 bytes
17 MB–32 MB	512 bytes	Não compatível	512 bytes
33 MB–64 MB	1 KB	512 bytes	512 bytes
65 MB–128 MB	2 KB	1 KB	512 bytes
129 MB–256 MB	4 KB	2 KB	512 bytes
257 MB–512 MB	8 KB	4 KB	512 bytes
513 MB–1,024 MB	16 KB	4 KB	1 KB
1,025 MB–2 GB	32 KB	4 KB	2 KB
2 GB–4 GB	64 KB	4 KB	4 KB
4 GB–8 GB	Não Compatível	4 KB	4 KB
8 GB–16 GB	Não compatível	8 KB	4 KB
16 GB–32 GB	Não compatível	16 KB	4 KB
32 GB–2 TB	Não compatível	Não compatível	4 KB

Discos Magnéticos

- ***Clusters***



Exemplo:
Arquivo de 1 byte
Ocupando 4 KB
(tamanho do *cluster*)

Discos Magnéticos

- ***Clusters***
  - *Clusters* maiores garantem a habilidade de se ler mais setores sem realização de *seek*
    - ⇧Aumento de desempenho quando o arquivo é processado sequencialmente
    - ⇩Pode ocorrer fragmentação interna ao *cluster*
- ***Extent***
  - É um conjunto de *clusters* fisicamente consecutivos, agrupando os setores/trilha/cilindro (se o arquivo for suficientemente grande), como um todo
    - Em um arquivo formado apenas por *clusters* consecutivos, haverá apenas **um** *seek* para sua leitura
    - Se o arquivo aumentar de tamanho, o *file manager* tenta manter um único *extent*, mas se não for possível, o arquivo é dividido em mais *extents*

Discos Magnéticos

- **Custos de acesso a disco**
  - Em termos de tempo, o custo de um acesso a disco é a soma de três tempos:
    - **Posicionamento** (*seek time*)
      - Tempo necessário para mover a cabeça leitura/escrita até a trilha correta – depende da distância percorrida
    - **Latência** (*rotational delay* ou *latency*)
      - Atraso necessário para a cabeça de leitura/escrita se posicionar no setor (ou bloco) correto
    - **Transferência** (*transfer time*)
      - Tempo necessário para um byte ser lido na superfície do disco e transferido para o *buffer* interno do controlador

Custos de Acesso ao Disco

- ***Seek time***
  - É impossível saber exatamente quantas trilhas serão percorridas durante as buscas
  - O que se faz é determinar o **tempo médio**, assumindo que as posições iniciais e finais para cada acesso são aleatórias
  - Por meio de estudos empíricos, foi descoberto que uma busca percorre, em média, 1/3 do total de trilhas, sendo que o tempo gasto para percorrer esse número de trilhas tem sido usado pelos fabricantes como indicador do **tempo médio de busca (*seek time*)**
    - Em 1991 o tempo médio divulgado pelos fabricantes era de 10 ms à 40 ms
    - Atualmente, temos algo entre 5 ms à 10 ms

Custos de Acesso ao Disco

- ***Seek time***
  - O que ocorre quando um arquivo está armazenado em cilindros consecutivos e é acessado sequencialmente?
    - O *seek time* é reduzido! (melhor situação)
  - O que ocorre quando dois arquivos, localizados em extremos opostos do disco (um no cilindro mais externo e outro no mais interno), são acessados alternadamente?
    - O *seek time* é alto! (pior situação)

- O tempo de *seek* tende a ser mais caro em sistemas multiusuários em que vários processos concorrem pelo uso do disco

Custos de Acesso ao Disco

- **Latência (*Rotational delay*)**
  - Tempo gasto para a cabeça de leitura/escrita sair de um posição aleatória e encontrar o setor desejado
    - Estima-se que esse tempo seja a metade do tempo de uma rotação (na prática, esse tempo costuma ser menor)
    - Também chamado de ***Average Latency* ou latência**
    - Em 1991, com discos de 3.600 RPM, a metade do tempo de uma rotação era de 8,3 ms
    - Atualmente os discos de 7.200 RPM têm latência média de 4,16 ms

Custos de Acesso ao Disco

- **Tempo de transferência (*Transfer time*)**
  - Uma vez que a cabeça de leitura/escrita está sob o setor, ele pode ser transferido
  - O tempo de transferência é dado por:
    - O tempo para transferir uma trilha inteira normalmente é o tempo de uma rotação
    - **(n° bytes transferidos/n° bytes na trilha) x tempo de rotação**
    - Exemplo:
      - *Transfer time* de 1 KB em um disco com 32 setores de 512 bytes por trilha (16.384b) e tempo de rotação de 8,2 ms:

        (1.024/16.384) x 8,2 = 0,51 ms
      - Neste exemplo o tempo está expresso em bytes/ms. Os fabricantes costumam utilizar MB/s

Custo de Acesso ao Disco

- O modo de acesso ao arquivo pode afetar drasticamente os custos de tempo de acesso
  - Dois modos de acesso:
    - Sequencial: o máximo do arquivo é processado em cada acesso
    - Aleatório: apenas um registro é acessado por vez

- **Exemplo**
  - Determinar o tempo necessário para ler um arquivo com 40.000 registros de 256 bytes cada
    - Vamos calcular o tempo para leitura do arquivo com acesso sequencial e aleatório e comparar os resultados

Exemplo

- Vamos considerar as seguintes especificações do disco:

Tempo médio de *seek*	13 ms
Tempo de latência	8,3 ms
Tempo de transferência	16,7 ms/trilha ou 1.229 bytes/ms
Bytes por setor	512
Setores por trilha	100
Trilhas por cilindro	12
Trilhas por superfície	1.748
Tamanho do *cluster*	10 setores (5.120 bytes)
Tamanho mínimo do *extent*	10 *clusters* (51.200 bytes) = 100 setores = 1 trilha

Exemplo

- Tamanho do arquivo → 40.000 registros de 256 bytes cada
  - Se cada *cluster* tem 5.120 bytes (10 setores), podemos armazenar 20 registros por *cluster* (5.120/256)
  - Será necessário um total de 2.000 *clusters* para armazenar todos os registros (40.000/20)
  - Como os *extents* possuem 10 clusters, serão necessários 200 *extents,* que correspondem à 200 trilhas
- Vamos assumir o pior caso, em que as 200 trilhas que armazenam o arquivo estão espalhadas aleatoriamente no disco
  - Situação extrema, mas que pode ocorrer em discos no limite da capacidade, especialmente com arquivos pequenos

Exemplo

- Custo com acesso sequencial
  - Para cada trilha, em que são lidos setores consecutivos, o processo de leitura envolve os seguintes custos:
    - Tempo médio de *seek*= 13 ms
    - Tempo de latência = 8,3 ms
    - Tempo de transferência de uma trilha = 16,7 ms

Para 200 trilhas =

(13 + 8,3 + 16,7) x 200 = 7.600 ms = **7,6 segundos**

Exemplo

- Custo com acesso aleatório
  - Para cada registro, a operação de leitura envolve os seguintes custos:
    - Tempo médio de *seek* = 13 ms
    - Tempo de latência = 8,3 ms
    - Tempo de transferência de um *cluster* = 1,67 ms

      (Tempo de 16,7 ms/trilha, como cada trilha = 10 clusters, então é gasto 16,7/10 ms por cluster)

Para 40.000 registros =

(13 + 8,3 + 1,67) x 40.000 = 918,8 s = **15,3 minutos**

Custo de Acesso

- A diferença entre o acesso sequencial e acesso aleatório é muito grande!
  - No exemplo anterior: 7,6 seg vs. 15,3 min
  - Essa diferença se deve a quantidade de *seeks* – 200 movimentos contra 40.000
- Por isso é aconselhável que o máximo de informação necessária seja lida em cada acesso
  - Evitando o frequente posicionamento da cabeça de leitura/escrita para cada registro
  - Esse custo pode ser minimizado com o uso de boas estruturas de dados

Discos Magnéticos

- **O disco é um gargalo**
  - Gargalo: quando um dispositivo mais lento afeta o desempenho de outros mais rápidos, se tornando um fator limitante do sistema
  - Discos são mais lentos que os processadores e as redes
    - A CPU espera tempos enormes para a transmissão para/do disco
    - A transmissão de dados na rede também pode ser mais rápida do que os discos
    - Quando isso ocorre, se diz que o processo é *disk-bound*
      - CPU e a rede têm que esperar pelos dados sendo transmitidos do/para disco