

Estruturas de Dados em Arquivos: Armazenamento Secundário

Prof. Paulo H. R. Gabriel

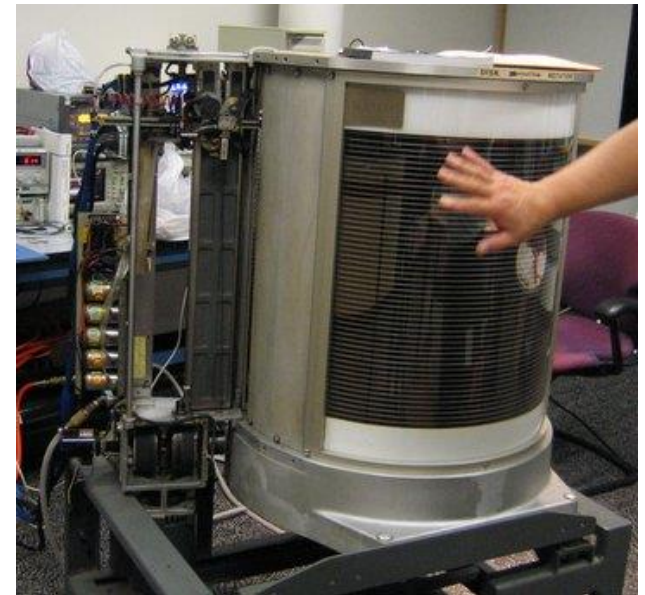
Créditos

- ▶ **Parte desse material foi baseado nas aulas de:**
 - ▶ Moacir Ponti Jr.
 - ▶ Thiago A. S. Pardo
 - ▶ Leandro C. Cintra
 - ▶ Maria C. F. de Oliveira



Breve Histórico: *Hard Disk*

- ▶ Com os discos:
 - ▶ Foram adicionados índices aos arquivos
 - ▶ Tornaram possível manter uma lista de chaves e ponteiros para acesso aleatório
 - ▶ Mais uma vez o crescimento dos arquivos de índice tornou difícil a sua manutenção
- ▶ Necessidade de *estruturas de dados mais eficientes*



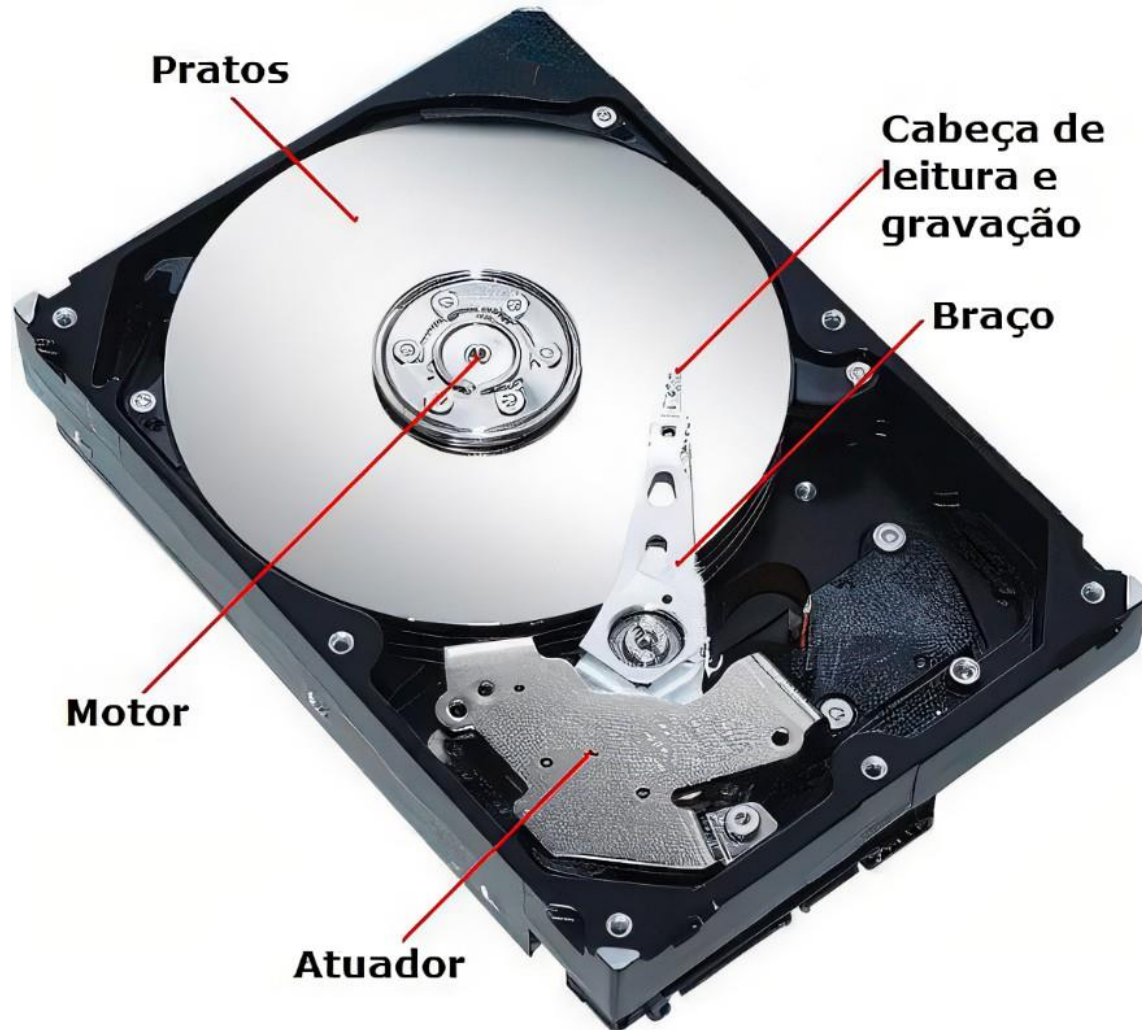
Fonte: Wikipedia (IBM 350)

HD: Componentes Importantes

- ▶ Discos (pode haver vários)
 - ▶ Substrato sólido (alumínio, vidro ou cerâmica)
 - ▶ Superfície magnética: filme magnético
- ▶ Cabeças de leitura/escrita: leem e escrevem nos discos enquanto eles giram (> 7000 RPM, em HDs comerciais)
 - ▶ Convertem entre sinais elétricos e pulsos magnéticos
- ▶ A informação lida é armazenada em um *buffer*, de onde é transferida para a memória



HD: Componentes Importantes



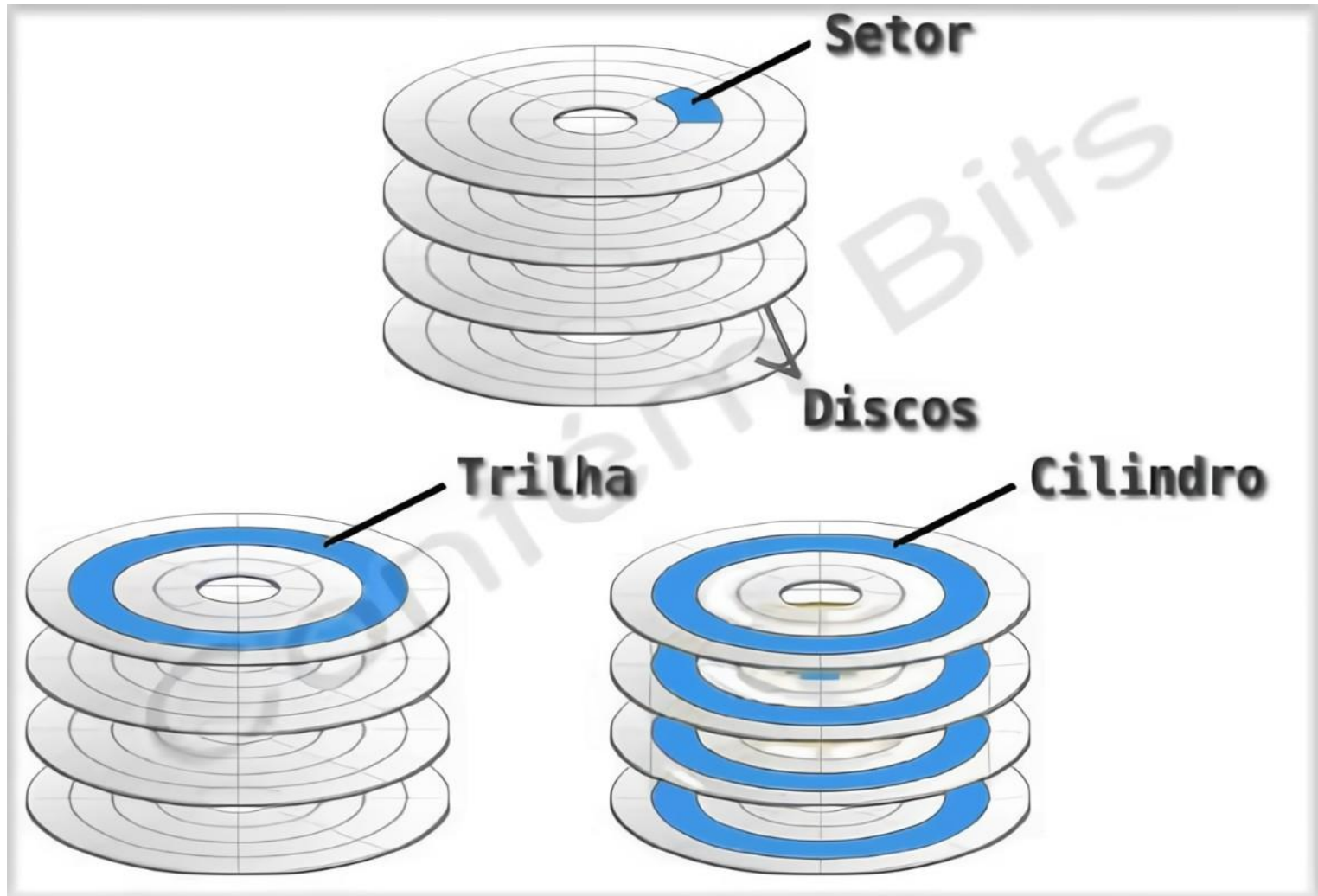
Fonte: [Blog do Disco Rígido \(memoriasecundaria.blogspot.com\)](http://memoriasecundaria.blogspot.com)

Organização da informação no disco

- ▶ **Disco:** conjunto de “pratos” empilhados
 - ▶ Dados são gravados nas superfícies desses pratos
- ▶ **Superfícies:** são organizadas em trilhas
- ▶ **Trilhas:** são organizadas em setores
- ▶ **Cilindro:** conjunto de trilhas na mesma posição



Organização da informação no disco



Capacidades de um disco

- ▶ Capacidade do setor
 - ▶ n° bytes (Ex. 512 bytes)
- ▶ Capacidade da trilha
 - ▶ n° de setores por trilha * capacidade do setor
- ▶ Capacidade do cilindro
 - ▶ n° de trilhas por cilindro * capacidade da trilha
- ▶ Capacidade do disco
 - ▶ n° de cilindros * capacidade do cilindro



Exemplo

- ▶ Você sabe o seguinte sobre seu HD
 - ▶ Número de bytes por setor: 512
 - ▶ Número de setores por trilha: 40
 - ▶ Número de trilhas por cilindro: 11
 - ▶ Número de cilindros: 1.331
- ▶ Há um conjunto de dados composto por 20.000 registros, sendo que cada registro tem 256 bytes
- ▶ Quantos cilindros são necessários para se armazenar esses 20.000 registros?



Exemplo

▶ Dados

- ▶ Número de bytes por setor: 512
- ▶ Número de setores por trilha: 40
- ▶ Número de trilhas por cilindro: 11
- ▶ Número de cilindros: 1.331
- ▶ Tamanho de cada um dos 20.000 registros: 256 bytes
- ▶ Cada setor, de 512 bytes, armazena dois registros (de 256 bytes cada)
- ▶ Portanto, são necessários 10.000 setores
- ▶ Um cilindro tem 11 trilhas, sendo que cada uma tem 40 setores
- ▶ Número de setores por cilindro: $11 * 40 = 440$ setores por cilindro
- ▶ Número de cilindros necessários: $10.000/440 = 22,7$ cilindros



Curiosidade

- ▶ O primeiro disco rígido foi lançado em 16/09/1957
- ▶ Era formado por 50 discos magnéticos contendo 50.000 setores
 - ▶ Cada disco suportava 100 caracteres alfanuméricos, totalizando uma capacidade de 5 *megabytes*
- ▶ Dimensões
 - ▶ 152,4 centímetros de comprimento
 - ▶ 172,72 centímetros de largura
 - ▶ 73,66 centímetros de altura



Endereços no disco

- ▶ Um **setor** é a menor porção endereçável do disco
- ▶ Ao ler 1 byte, o SO determina qual a superfície, trilha e setor em que se encontra esse byte
- ▶ O conteúdo do setor é carregado para uma memória especial (*buffer* de E/S) e o byte desejado é lido do *buffer* para a RAM
 - ▶ Se o setor necessário já está no *buffer*, o acesso ao disco torna-se desnecessário



Sistema de Arquivos

- ▶ A organização do disco em setores, trilhas e cilindros é uma formatação física (já vem da fábrica)
 - ▶ Pode ser alterada se o usuário quiser dividir o disco em partições
- ▶ É necessária uma formatação lógica, que “instala” o sistema de arquivos no disco
 - ▶ Subdivide o disco em regiões endereçáveis
- ▶ Sistema de arquivos: estruturas lógicas e sub-rotinas usadas para controlar acesso aos dados em disco



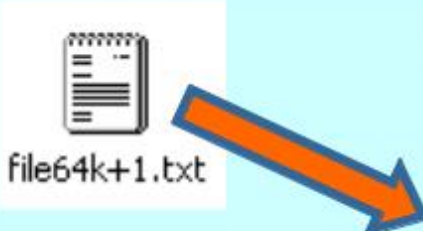
Sistema de Arquivos

O sistema de arquivos FAT

- ▶ Windows 9x e dispositivos de memória *flash*
- ▶ Não endereça setores, mas grupos de setores (*clusters*)
 - ▶ 1 cluster = 1 unidade de alocação
 - ▶ 1 cluster = n setores
- ▶ Um arquivo ocupa, no mínimo, 1 cluster
 - ▶ Unidade mínima de alocação
- ▶ Se um programa precisa acessar um dado, cabe ao sistema de arquivos do SO determinar em qual *cluster* ele está

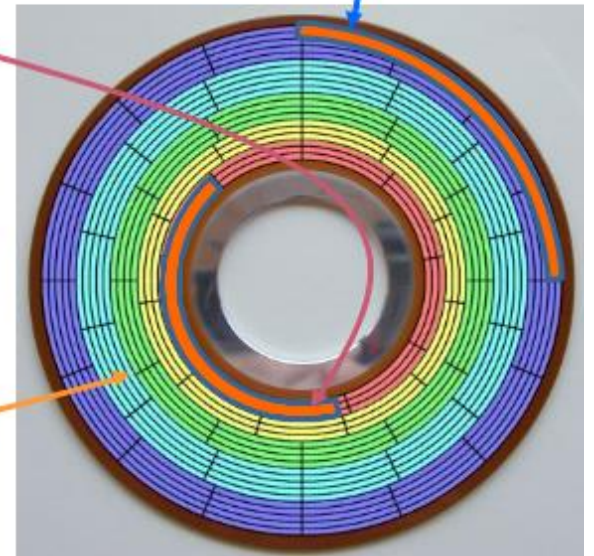


FAT: *File Allocation Table*



Cluster	Free / Next / Final	LBA
0	1	0
1	5	4
...		
10	14	40
11	free	44
12	free	48
13	free	52
14	final	56
...		
62	free	248
63	free	252

File Allocation Table



Tamanho do Cluster

- ▶ Definido automaticamente pelo SO quando o disco é formatado
- ▶ FAT (Windows): sempre uma potência de 2:
 - ▶ 2, 4, 8, 16, 32KB
- ▶ Determinado pelo máximo que a FAT consegue manipular e pelo tamanho do disco
 - ▶ FAT16: pode endereçar 2^{16} clusters = 65.536
- ▶ Quanto maior o cluster, maior a *fragmentação*!



Outros sistemas de arquivos

- ▶ **FAT32 (Windows 95 e posteriores)**
 - ▶ clusters de tamanho menor, endereça mais clusters, menos fragmentação
- ▶ **NTFS (New Technology File System)**
 - ▶ Sistemas OS/2 (IBM) e Windows NT
 - ▶ Lida melhor com arquivos maiores, a depender do tamanho do volume
 - ▶ Deu origem ao sistema atual do Windows (ReFS)
- ▶ **ext3 e 4 (UNIX)**
 - ▶ clusters entre 1 e 8KB



Custo de Acesso ao Disco

Combinação de três fatores:

- ▶ Tempo de busca (*seek*): tempo para posicionar o braço de acesso no cilindro correto
- ▶ Atraso (*delay*) de rotação: tempo para o disco rodar de forma que a cabeça de L/E esteja posicionada sobre o setor desejado
- ▶ Tempo de transferência: tempo p/ transferir os bytes

Tempo transferência = (n° de bytes transferidos/n° de bytes por trilha) * tempo de rotação

