

# Fundamentos de Arquivos e Armazenamento Secundário



---

Cristina D. A. Ciferri

Thiago A. S. Pardo

Leandro C. Cintra

M.C.F. de Oliveira

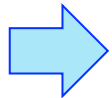
Moacir Ponti Jr.



# Armazenamento de Dados

---

- Armazenamento primário
  - memória primária (RAM)
  - *memória do computador*
- Armazenamento secundário
  - memória secundária
  - *disco, fita, CD-Rom, ....., os quais são acessados diretamente pelo computador*
- Armazenamento terciário
  - dados arquivados em *jukebox*
  - *disco, fita, CD-Rom, etc, os quais não são diretamente acessados pelo computador*





# Disco X Memória Primária

---

- Estimativas de tempo de acesso
  - HD: alguns milisegundos  $\sim 10\text{ms}$  ( $10^{-3}$ )
  - RAM: alguns nanosegundos  $\sim 10\text{ns} \dots 40\text{ns}$  ( $10^{-9}$ )

Ordem de grandeza da diferença entre os tempos de acesso:  $\sim 250.000$

HDs são 250.000 vezes mais lentos que memória RAM!



# Disco X Memória Primária

---

- Capacidade de Armazenamento
  - HD – muito alta, a um custo relativamente baixo
  - RAM – limitada pelo custo e espaço
- Tipo de Armazenamento
  - HD – não volátil
  - RAM – volátil



# Disco

---





# Organização

---

- **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

setor: menor porção endereçável do disco

# Organização

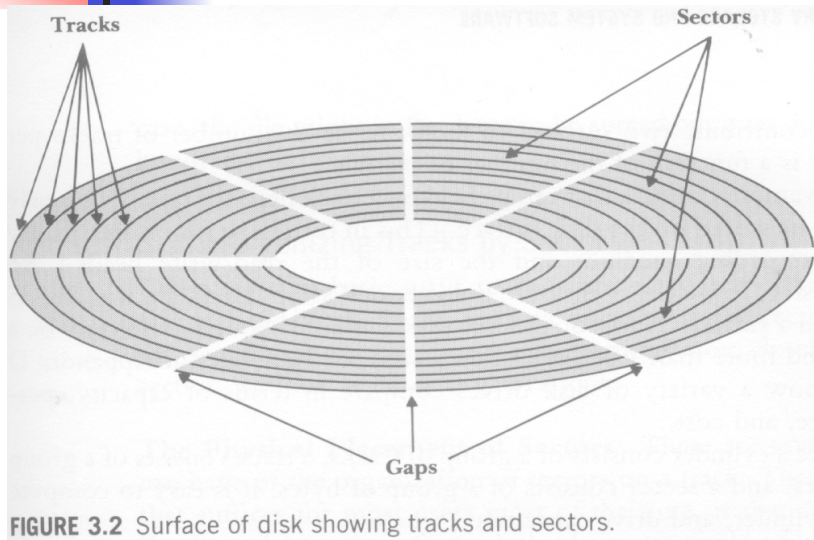


FIGURE 3.2 Surface of disk showing tracks and sectors.

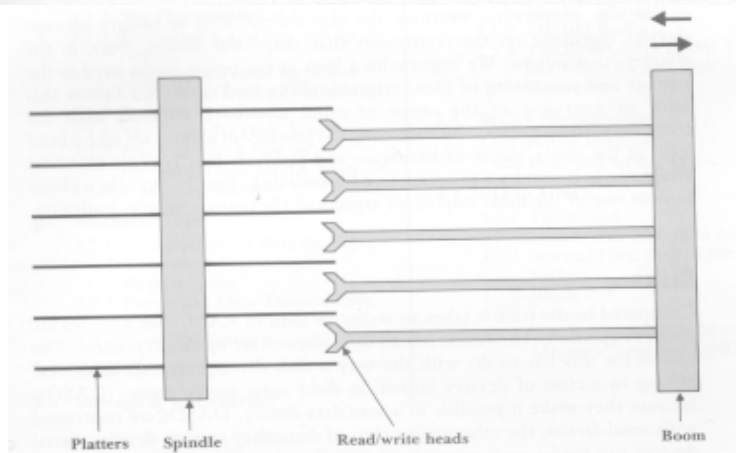
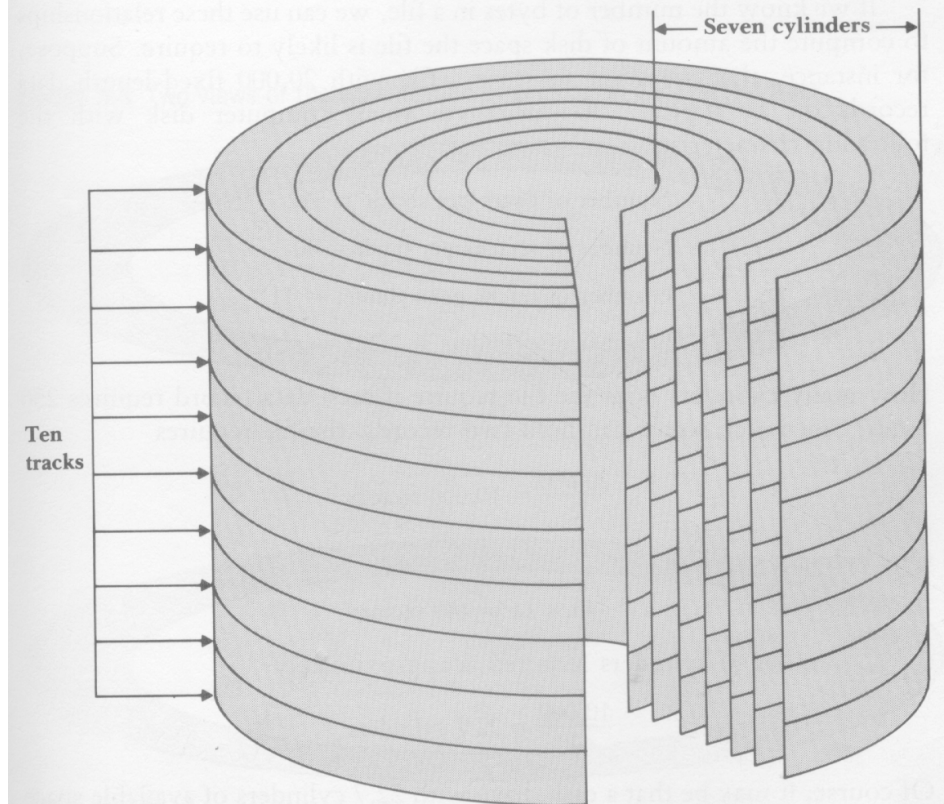


FIGURE 3.1 Schematic illustration of disk drive.

FIGURE 3.3 Schematic illustration of disk drive viewed as a set of seven cylinders.





# Capacidade do disco (nominal)

---

- Capacidade do **setor**
  - n<sup>o</sup> bytes (Ex. 512 bytes)
- Capacidade da **trilha**
  - n<sup>o</sup> de setores/trilha x capacidade do setor
- Capacidade do **cilindro**
  - n<sup>o</sup> de trilhas/cilindro x capacidade da trilha
- Capacidade do **disco**
  - n<sup>o</sup> de cilindros x capacidade do cilindro





# Seeking

---

- Movimento de posicionar a cabeça de L/E sobre a trilha/setor desejado
- O conteúdo de todo um **cilindro** pode ser lido com 1 único seeking
- É o movimento **mais lento** da operação leitura/escrita
- **Deve ser reduzido ao mínimo**



# Custo de Acesso a Disco

---

- Seek time (tempo de acesso)
  - tempo para posicionar a cabeça de leitora e gravação no cilindro correto
- Rotational delay (delay de rotação)
  - tempo para rotacionar o disco para que a cabeça de leitora e gravação seja posicionada no setor correto
- Transfer time (tempo de transferência)
  - tempo para transferir o dado para a memória primária



# Sistema de Arquivos

---

- Formatação física (Disco Físico)
  - a organização do disco em setores/trilhas/cilindros que já vem da fábrica
  - pode ser mudada por meio de partições
- Formatação lógica (Disco Lógico)
  - `instala' o sistema de arquivos no disco
  - subdivide o disco em regiões endereçáveis
  - introduz *overhead* relacionado ao espaço ocupado com informações para gerenciamento



# Sistema de Arquivos

---

- Faz parte do sistema operacional (S.O.)
- Fornece a infraestrutura básica para a manipulação de arquivos em memória secundária via software
- Oferece um conjunto de operações para a manipulação de arquivos

criar (create, open)

renomear (rename)

fechar (close)

escrever dados (write)

posicionar (seek)

destruir ou remover (delete)

abrir (open)

ler dados (read)

escrever dados no final (append)

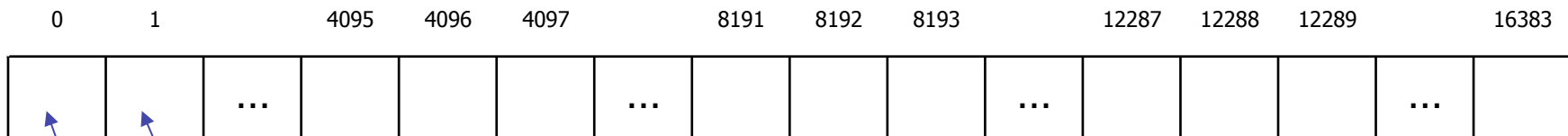
...



# Arquivo Físico

---

- Sequencia de bytes armazenados no disco



byte 1 = 2º byte  
byte 0 = 1º byte

➡ byte n = (n+1) byte

tamanho:  
16.384 bytes ou 16KB



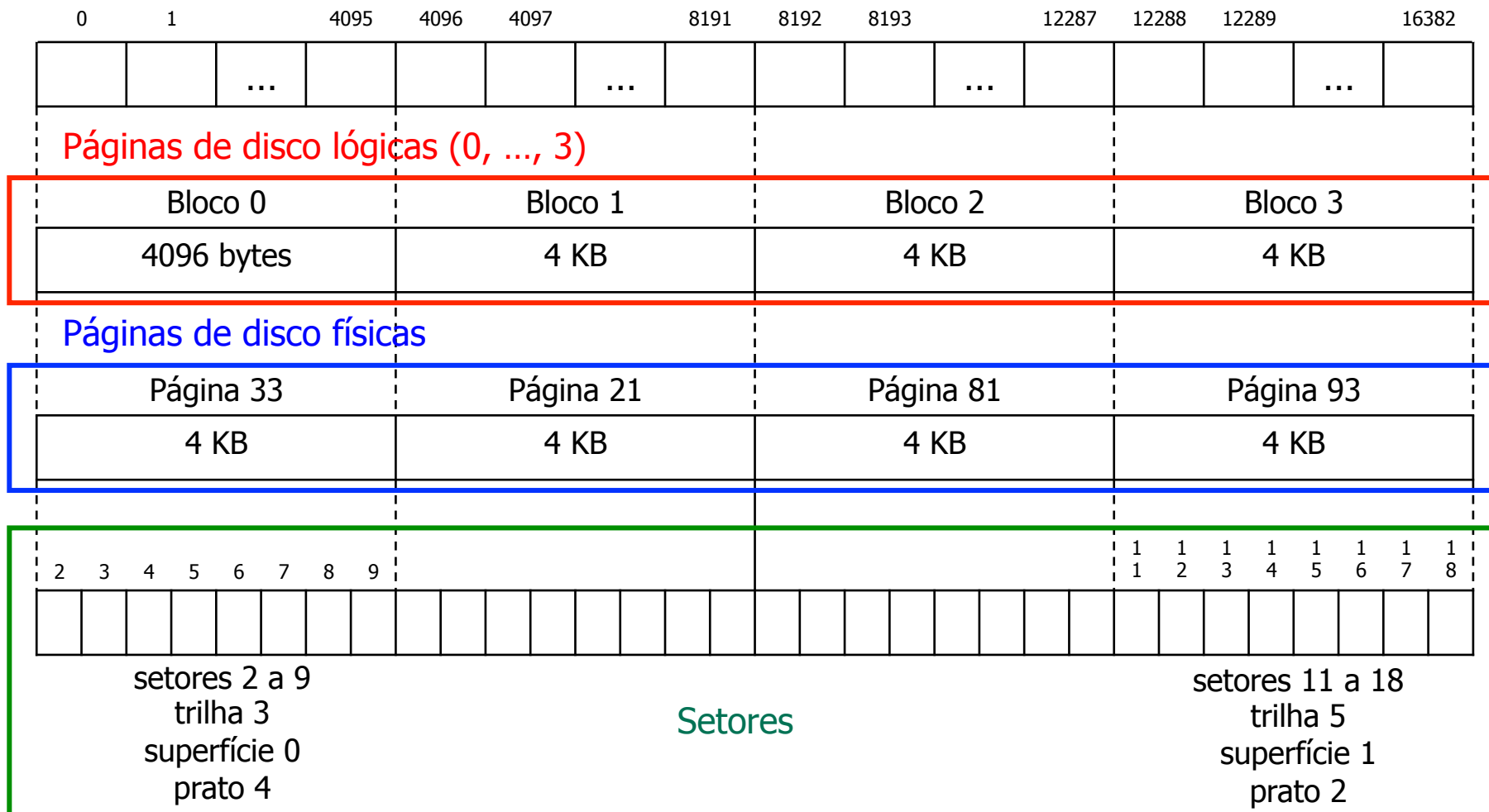
# Página de Disco

---

- Conjunto de setores logicamente contíguos no disco
- Um arquivo é visto pelo sistema de arquivos como um conjunto de páginas de disco
  - arquivos são alocados em uma ou mais páginas de disco

Também chamado de bloco de disco ou **cluster (livro)**

# Página de Disco





# Mapeamentos

---

- Páginas lógicas → páginas físicas
  - depende da técnica de alocação de espaço em disco (ex.: alocação contígua, alocação encadeada e alocação indexada)
- Páginas físicas → setores
  - feito por um programa especial chamado condutor de dispositivo (*device driver*)

Toda a gerência do espaço em disco é feita pelo sistema de arquivos com base nos conceitos de páginas de disco lógicas e físicas, e não no conceito de setores.



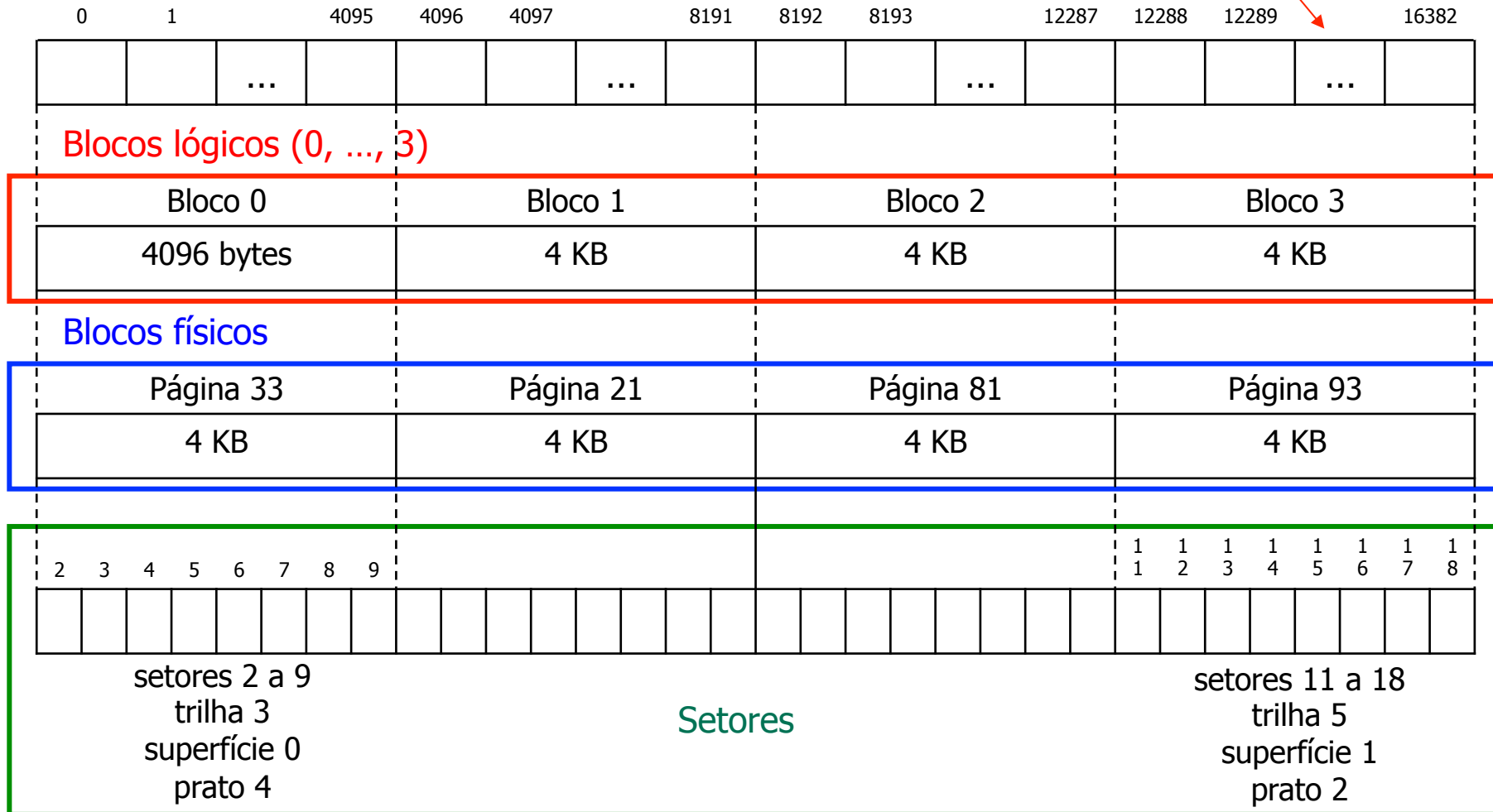


# Posição Corrente no Arquivo

---

- Abstração que permite a especificação de uma chamada do sistema para indicar o onde um arquivo deve ser lido ou escrito
- Características
  - a leitura e escrita acontecem a partir da posição corrente
  - a posição corrente é então avançada para imediatamente após o último byte lido ou escrito
  - é possível informar um endereço específico a ser lido, o qual faz com que a posição corrente seja a informada no endereço

READ (byte 12347)  
posição corrente (implícita)



Acessos: Bloco lógico 3, bloco físico 93, setor 11, trilha 5, superfície 1, prato 2



# Fragmentação Interna

---

- Perda de espaço útil decorrente da organização do arquivo em páginas de disco de tamanho fixo
- Exemplo
  - página de disco de 4K (4.096 bytes)
  - necessidade de se escrever 1 byte
  - *desperdício de 4.095 bytes nessa página de disco*



# Tamanho da Página de Disco

---

- Definido automaticamente pelo S.O. quando o disco é formatado
- Exemplo
  - (FAT Windows): sempre uma potência de 2
    - 2, 4, 8, 16 ou 32KB
  - Determinado pelo máximo que a FAT consegue manipular, e pelo tamanho do disco
    - FAT16: pode endereçar  $2^{16}$  clusters = 65.536 clusters
- Quanto maior a página de disco, maior a fragmentação interna!



# Arquivo

---

- Conteúdo
  - um registro de cabeçalho
  - registros de dados
- Descritor do arquivo
  - estrutura usada pelo sistema de arquivos para gerenciar cada arquivo existente
  - exemplos de conteúdo: nome do arquivo, tipo do arquivo, tamanho em bytes, proteção, data e hora do último acesso, data e hora da criação, identificação do proprietário, local onde os dados estão armazenados, ...



# TDAA e TAAP

---

- Para tornar mais rápido o acesso aos arquivos, o sistema de arquivos mantém na memória primária
  - uma tabela contendo os **descritores de arquivos abertos** por todos os processos (TDAA)
  - uma ou mais tabelas de **arquivos abertos por processo** (TAAP)



# TDAA e TAAP

---

- Cada entrada de **TDAA** armazena
  - uma cópia do descritor do arquivo mantido em disco, número de processos usando o arquivo
  - informações que não variam conforme o processo, como o tamanho do arquivo
- cada entrada de **TAAP** armazena
  - informações que variam conforme o processo, como posição corrente, modo de abertura do arquivo

# TDAA e TAAP

