Banco de Dados 2
Armazenamento - Níveis de RAID

Prof. Silvio R. R. Sanches

UTFPR
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
Cornélio Procópio

Introdução

- O SGBD oferece uma visão de alto nível dos dados

- No entanto:
  - Os dados precisam ser armazenados como bits em um ou mais dispositivos de armazenamento

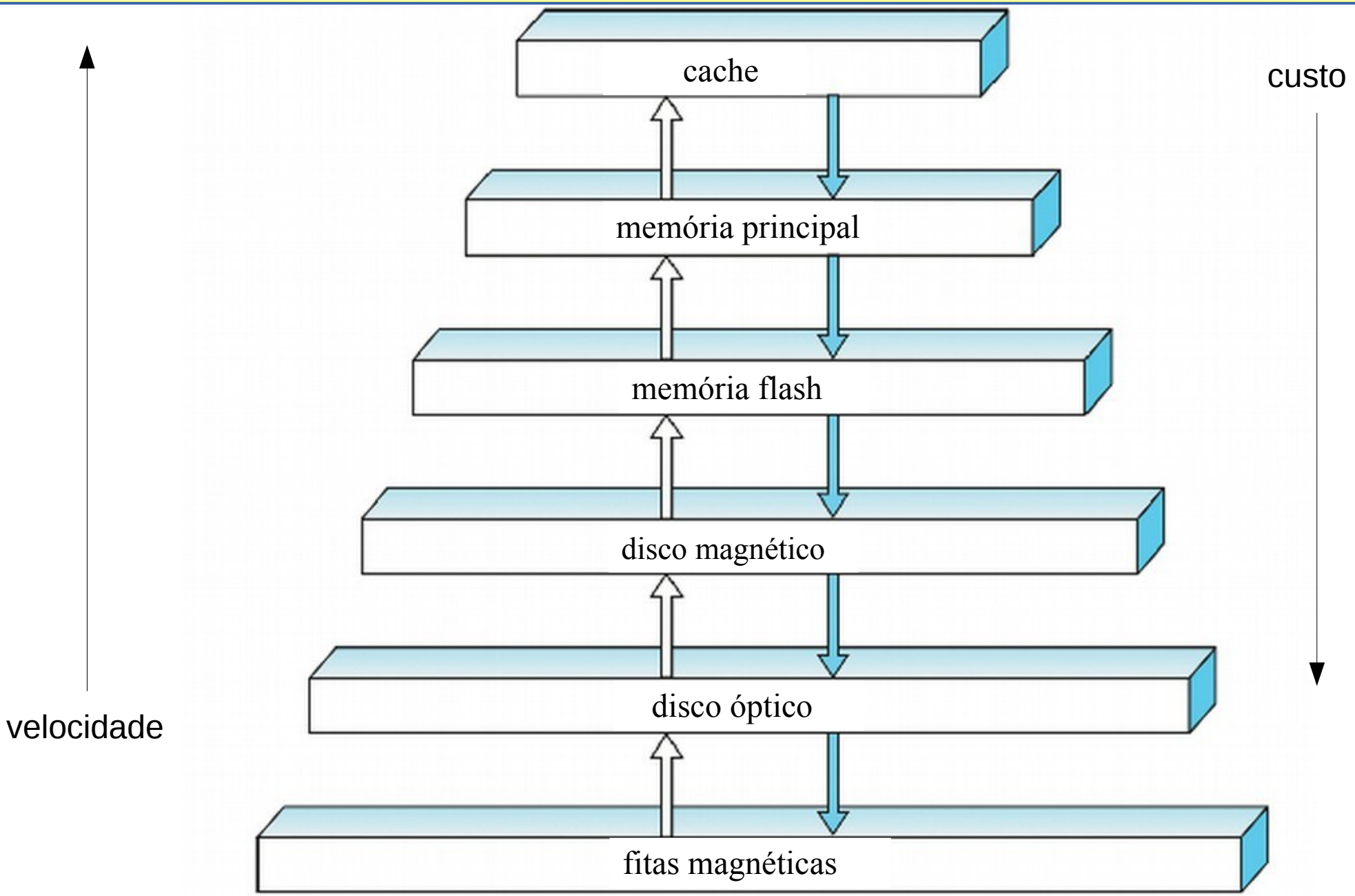
Meios de Armazenamento Físico

- Pode-se diferenciar meios de armazenamento:
  - **Armazenamento volátil:**
    - Conteúdo é perdido quando a energia é desligada.
  - **Armazenamento não volátil:**
    - Conteúdo persiste mesmo quando a energia é desligada.
    - Inclui armazenamento secundário e terciário

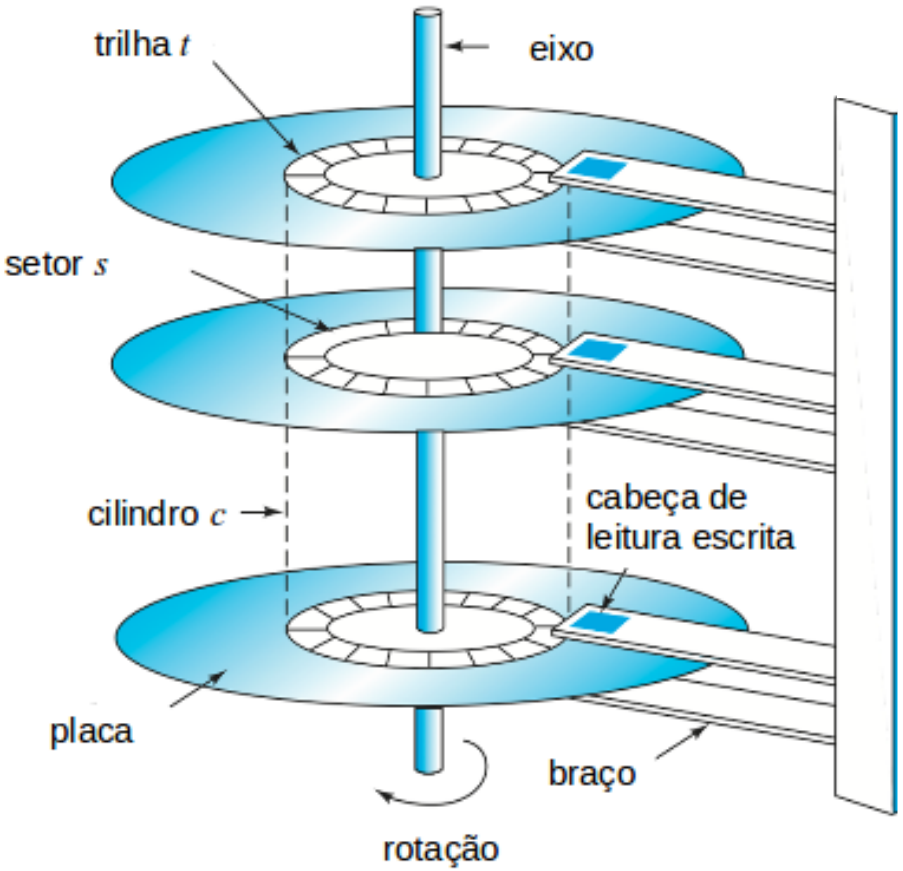
Meios de Armazenamento Físico

- Cache
- Memória Principal
- Memória Flash
- Disco Magnético
- Armazenamento Óptico
- Armazenamento em Fita

Hierarquia: Velocidade e Custo



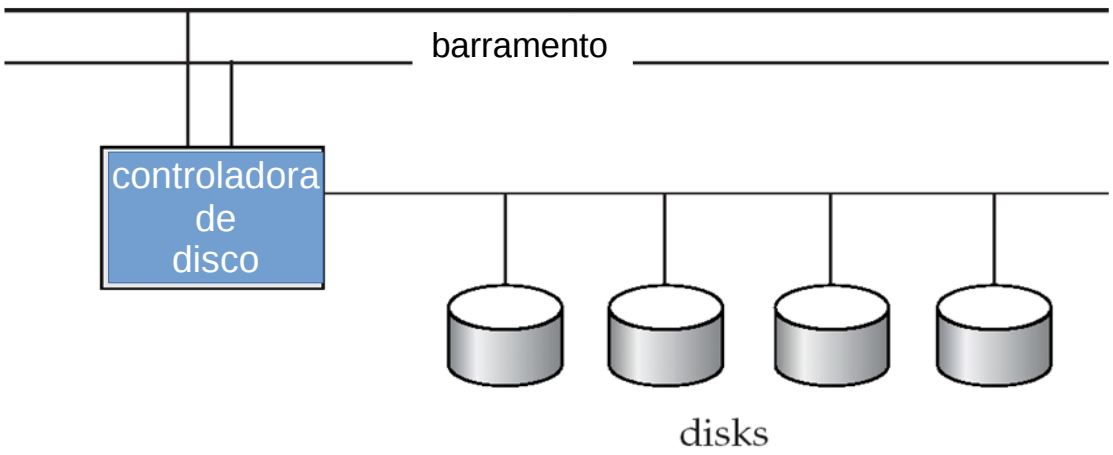
Disco Magnético



- Cabeça de leitura-escrita
  - Posicionadas muito próximas à superfície da placa
  - Lê ou grava informações magneticamente codificadas.
- Superfície da placa é dividida logicamente em trilhas circulares
- Para ler/escrever um setor:
  - Braço do disco movimenta para posicionar cabeça na trilha certa
  - Placa gira continuamente; os dados são lidos/escritos quando o setor passa sob a cabeça de leitura/escrita

Disco Magnético

- Controladora de Disco:
  - Interface entre o sistema de computador e o hardware real da unidade de disco
    - Inicia ações para mover o braço do disco para a trilha correta
    - Colocam ***checksums*** nos setores gravados
    - Remapeia setores defeituosos



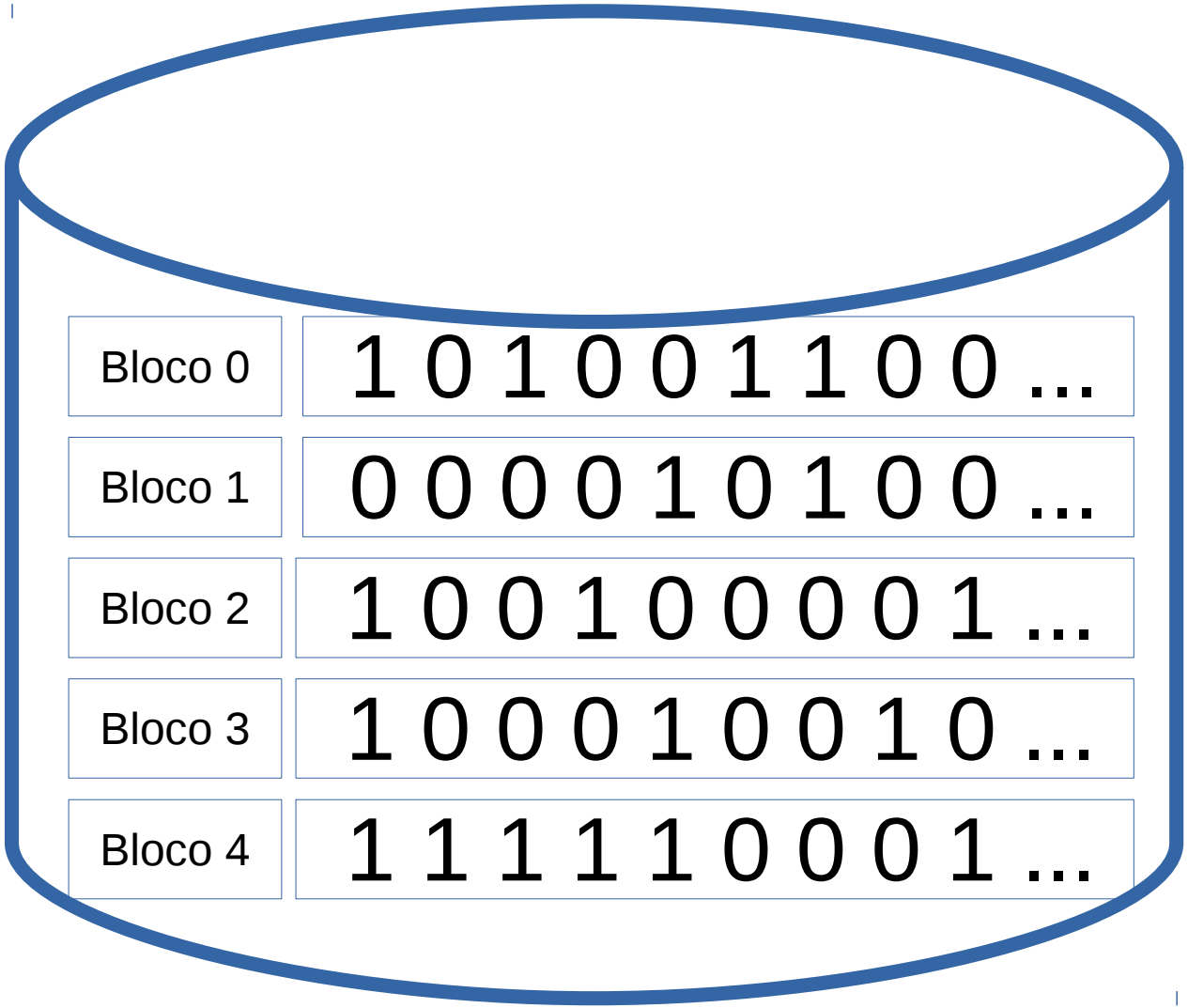
- Vários discos ligados a um sistema de computador através de uma controladora
- Interfaces de disco comuns
  - ATA, SATA, SCSI, SAS

Níveis de RAID

- Espelhar todos os discos oferece alta confiabilidade, mas é caro
- Espalhar os dados por vários discos oferece altas taxas de transferência de dados, mas não melhora a confiabilidade
- Diversos esquemas alternativos buscam melhorar a redundância com menor custo
- Esses esquemas possuem diferentes opções de custo-desempenho
  - São classificadas como **níveis de RAID (Redundant Array of Independent Disks)**
- No RAID, cada arquivo pode ser dividido em nível de **bloco** ou em nível de **byte**
  - O RAID nível 2, não mais utilizado na prática, trabalha no nível de bit

RAID

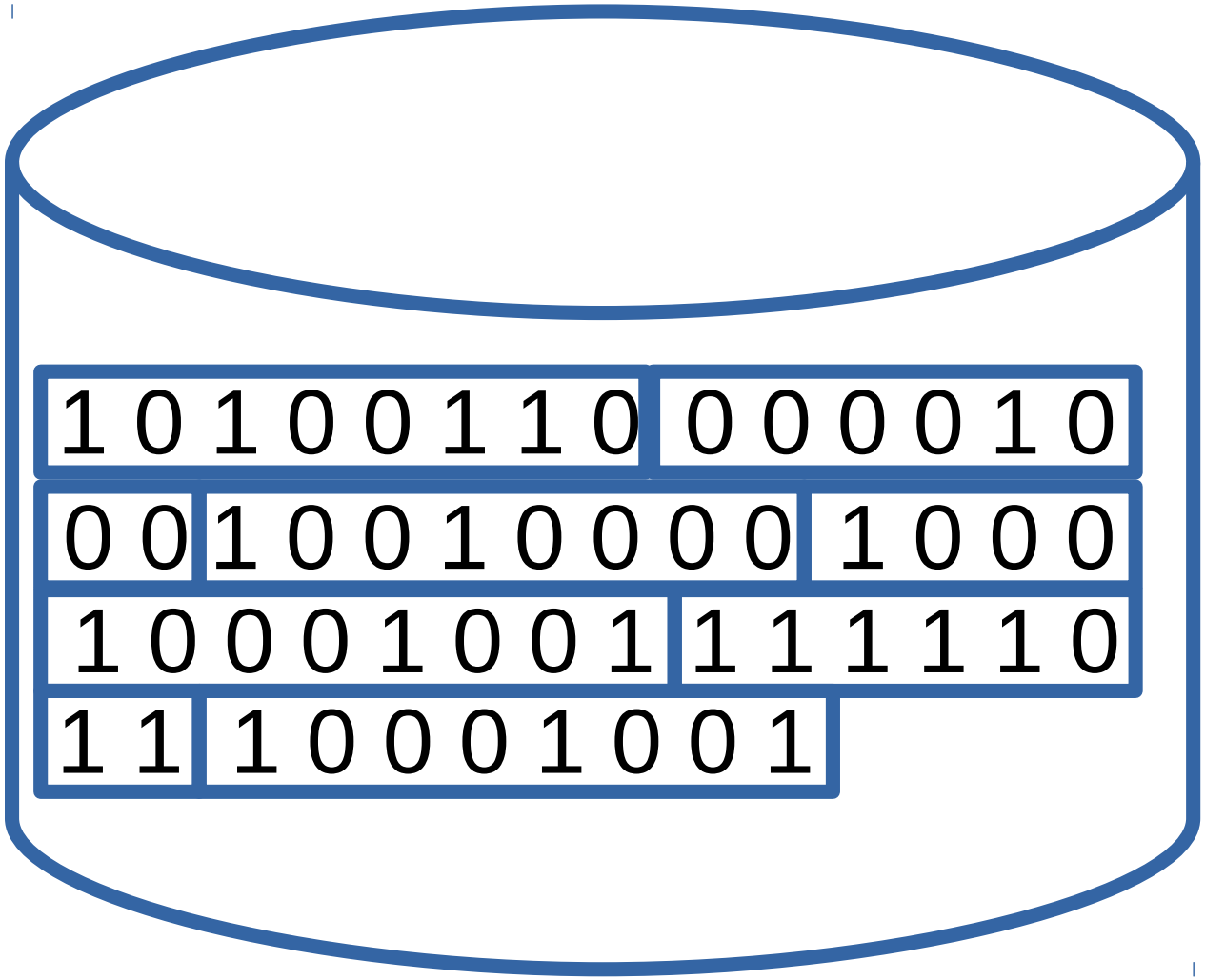
Arquivo visualizado em nível de **bloco** em um único disco



- Nível de bloco:
  - Cada arquivo é dividido em partes de um **bloco** de tamanho.
  - Cada bloco possui tipicamente de 4 a 8 kilobytes

RAID

Arquivo visualizado em nível de **byte** em um único disco



- Nível de byte:
  - Cada arquivo é dividido em partes de um **byte** de tamanho.

RAID

- Nível de bloco: cada arquivo é dividido em partes de um bloco de tamanho

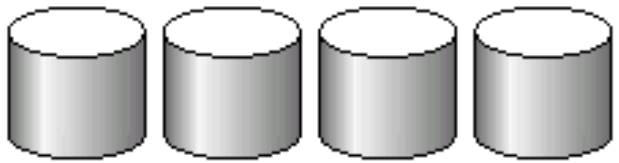
Disco 1	Disco 2	Disco 3	Disco 4
Arquivo 1, bloco 1	Arquivo 1, bloco 2	Arquivo 1, bloco 3	Arquivo 2, bloco 1
Arquivo 2, bloco 2	Arquivo 3, bloco 1	Arquivo 3, bloco 2	...

- Nível de byte: cada arquivo é dividido em partes de um byte de tamanho

Disco 1	Disco 2	Disco 3	Disco 4
Arquivo 1, byte 1	Arquivo 1, byte 2	Arquivo 1, byte 3	Arquivo 1, byte 4
Arquivo 1, byte 5	Arquivo 1, byte 6	Arquivo 1, byte 7	Arquivo 2, byte 1
Arquivo 2, byte 2	Arquivo 3, byte 1	...	...

RAID nível 0

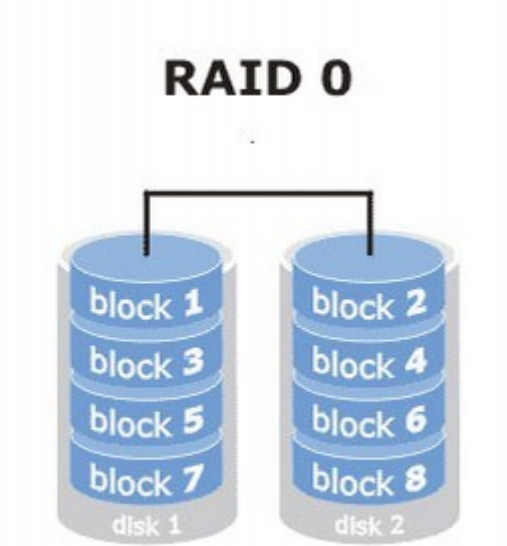
- Oferece aos *arrays* de disco o espalhamento no nível de blocos
- Não existe qualquer redundância
- Exemplo:
  - Array de tamanho 4:



RAID 0: espalhamento não redundante

RAID nível 0

- Melhora o desempenho usando vários discos:
  - Exemplo:
    - Gravar 1GB de dados:
    - 500MB fica armazenado em um disco e os outros 500MB, em outro disco.
    - Quando os dados precisam ser lidos, é obtido um pedaço de cada disco



- Vantagem:
  - É mais rápido do que realizar a leitura de apenas um disco.

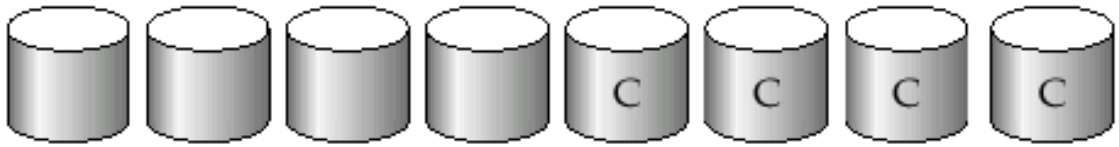
- Desvantagem:
  - Se um dos discos falhar, todos os dados são perdidos.

http://www.baboo.com.br/hardware

RAID nível 1

- Espelhamento de disco com espalhamento de bloco

- Exemplo:
  - Organização espelhada que mantém quatro discos de dados



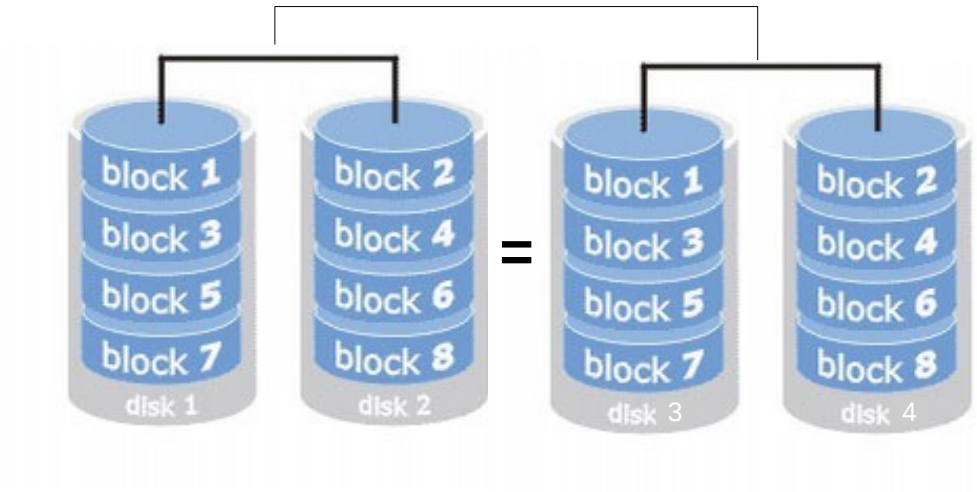
RAID 1: discos espelhados

C indica uma segunda cópia de dados

RAID nível 1

- Os discos são programados para serem espelhados:
  - Quando o computador grava 100MB de dados, ele também armazenará os 100MB no outro conjunto de discos.
  - Se um dos discos falhar, o outro possui uma cópia atualizada de todo seu conteúdo.

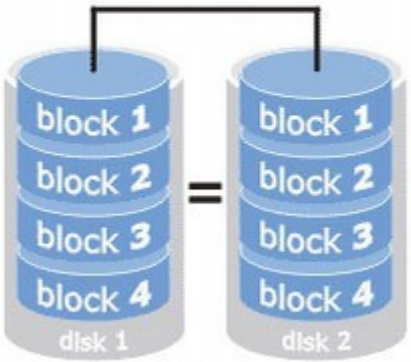
RAID 1



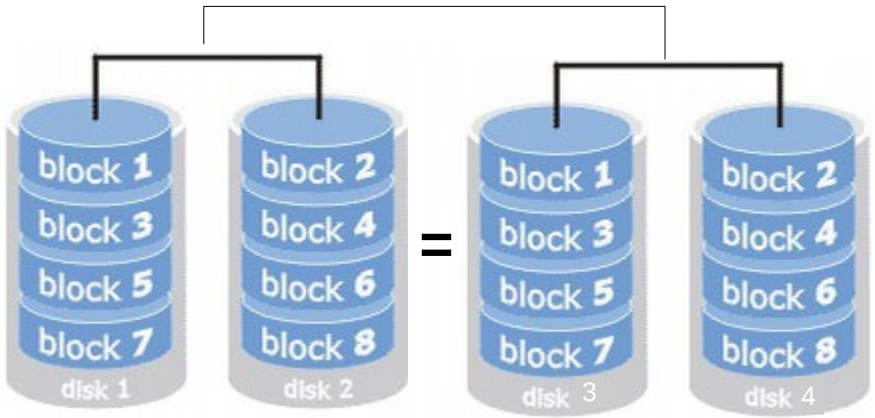
RAID nível 1

- Alguns fornecedores usam:
  - O termo RAID 1+0 para se referir ao espelhamento com espalhamento
  - O termo RAID 1 para se referir ao espelhamento sem espalhamento

RAID 1



RAID 1 + 0



Exercícios

- ...

RAID nível 2

- Conhecido como organização por código de correção de erro no estilo da memória
  - Emprega bits de paridade

- **Paridade Par:**
  - O bit anexado serve para tornar o número total de 1's par

01001 ⇨ 001001
10110 ⇨ 110110

- **Paridade Impar:**
  - O bit anexado serve para tornar o número total de 1's impar

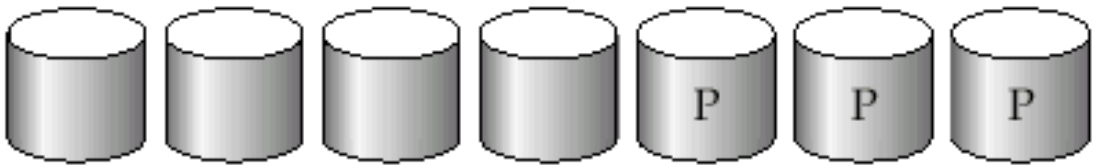
01001 ⇨ 101001
10110 ⇨ 010110

RAID nível 2

- Todos os erros de 1 bit serão detectados pelo sistema de memória
- Esquema de correção de erro que armazenam dois ou mais bits extras podem reconstruir os dados (se um único bit foi danificado)

- Exemplo em banco de dados:
  - O primeiro bit de cada byte poderia ser armazenado no disco 0
  - O segundo no disco 1, e assim por diante, até que o oitavo bit seja armazenado no disco 7
  - Os bits de de correção de erro seriam armazenados em outros discos

Dados rotulados com P armazenam os bits de correção de erro

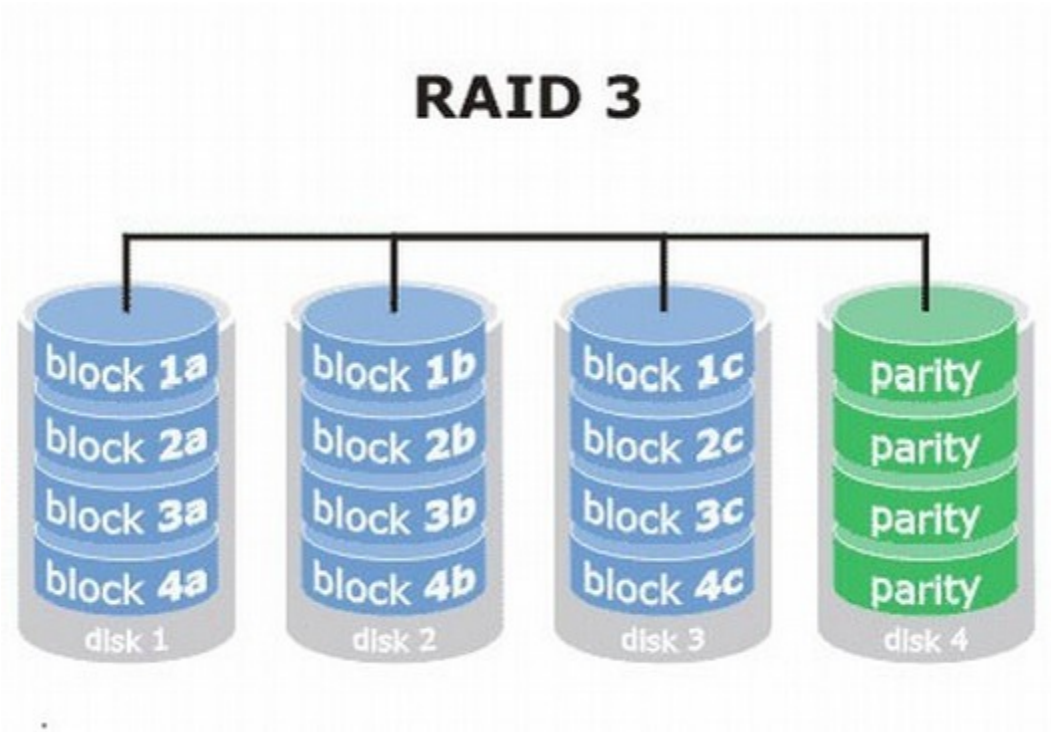


RAID nível 3

- Organização com paridade intercalada por bit
  - Explora o fato de que as controladoras de disco podem detectar se um setor foi lido corretamente
  - Se um setor for danificado, o sistema sabe exatamente qual é esse setor:
  - Um único bit de paridade pode ser usado para correção de erro, além da detecção

RAID nível 3

- Separa os arquivos em bytes, não em blocos
- Um disco é utilizado para paridade



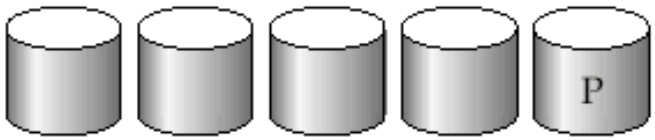
RAID nível 3

- O RAID nível 3 é:
  - Tão bom quanto o nível 2
  - Menos dispendioso no número de discos extras
    - Gasta apenas um disco extra
  - O nível 2 não é usado na prática
  - Apesar de conter leitura e gravação rápida, os discos giram em sincronia para obter os dados. Leitura aleatória de dados dentro do disco também sofre com desempenho.

- Benefícios em relação ao nível 1
  - Só precisa de 1 disco de paridade para vários regulares
    - RAID nível 1 precisa de um espelho para cada disco
  - Leitura e escrita mais rápida

RAID nível 4

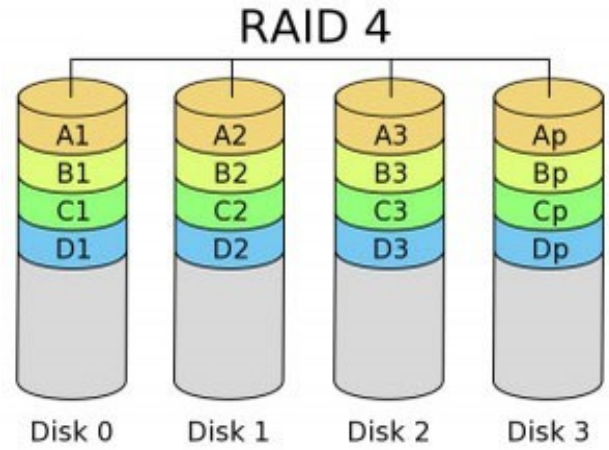
- Organização de paridade intercalada por bloco
- Utiliza o espalhamento no nível de bloco
- Mantém um bloco de paridade em um disco separado, para os blocos correspondentes de N outros discos
- Se um disco falhar o bloco de paridade pode ser usado com os blocos correspondentes a partir de outros discos, para restaurar os blocos de disco que falhou



RAID 4: paridade intercalada por bloco

RAID nível 4

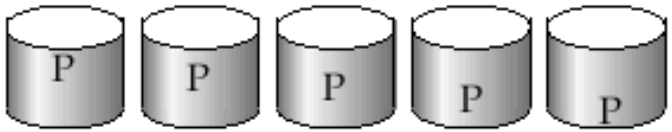
- Uma leitura de bloco acessa apenas um disco, permitindo que outras solicitações sejam processadas pelos outros discos



- Portanto:
  - A taxa de transferência de dados para cada acesso é mais lenta
    - Não paraleliza uma única operação de acesso a bloco
  - Por outro lado, vários acessos de leitura podem prosseguir em paralelo, levando a taxa de E/S geral é mais alta

RAID nível 5

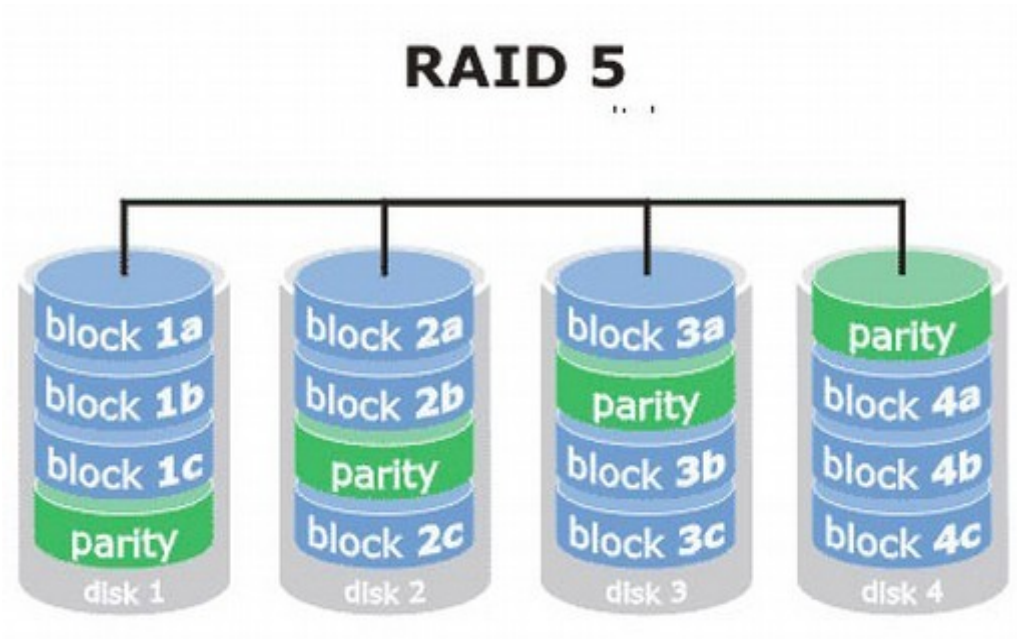
- Paridade distribuída intercalada por bloco
- Melhora o nível 4:
  - Particiona dados e paridade entre todos os $N+1$ discos, em vez de armazenar os dados em N discos e a paridade em um disco
- Todos os discos podem participar satisfazendo solicitações de leitura:
  - No nível 4, o disco de paridade não pode participar



RAID nível 5

- Aumenta o número total de solicitações que podem ser atendidas em determinada quantidade de tempo

- Para cada conjunto de N blocos lógicos, um dos discos armazena a paridade, e os outros N discos armazenam os blocos



Exemplo

- Com um array de cinco discos, o bloco de paridade, rotulado como Pk para os blocos lógicos $4k$, $4k$+1, $4k$+2, $4k$+3, é armazenado no disco k mod 5

- Os blocos correspondentes dos outros quatro discos armazenam os quatro blocos de dados de $4k$ a $4k$+3

A tabela a seguir indica como os 20 primeiros blocos, 0 a 19, e seus blocos de paridade, são distribuídos

Disco 0	Disco 1	Disco 2	Disco 3	Disco 4
P0	0	1	2	3
4	P1	5	6	7
8	9	P2	10	11
12	13	14	P3	15
16	17	18	19	P4

Bloco de Paridade 3
3 mod 5 = 3
P3 no disco 3

Bloco de Paridade 4
4 mod 5 = 4
P4 no disco 4

Blocos lógicos
$4k$ = 4*3 = 12
$4k$+1 = 4*3+1 = 13
$4k$+2 = 4*3+2 = 14
$4k$+3 = 4*3+3 = 15

Blocos lógicos
$4k$ = 4*4 = 16
$4k$+1 = 4*4+1 = 17
$4k$+2 = 4*4+2 = 18
$4k$+3 = 4*4+3 = 19

Bloco de paridade armazenado no disco *k* mod 5

RAID nível 5

- Vantagens:
  - O nível 5 oferece melhor desempenho de leitura-escrita com o mesmo custo que o nível 4
  - O nível 4 não é usado na prática

Por quê não se pode armazenar a paridade para blocos no mesmo disco?

Um bloco de paridade não pode armazenar a paridade para blocos no mesmo disco, pois uma falha no disco resultaria em perda de dados e também de paridade (não seria recuperável)

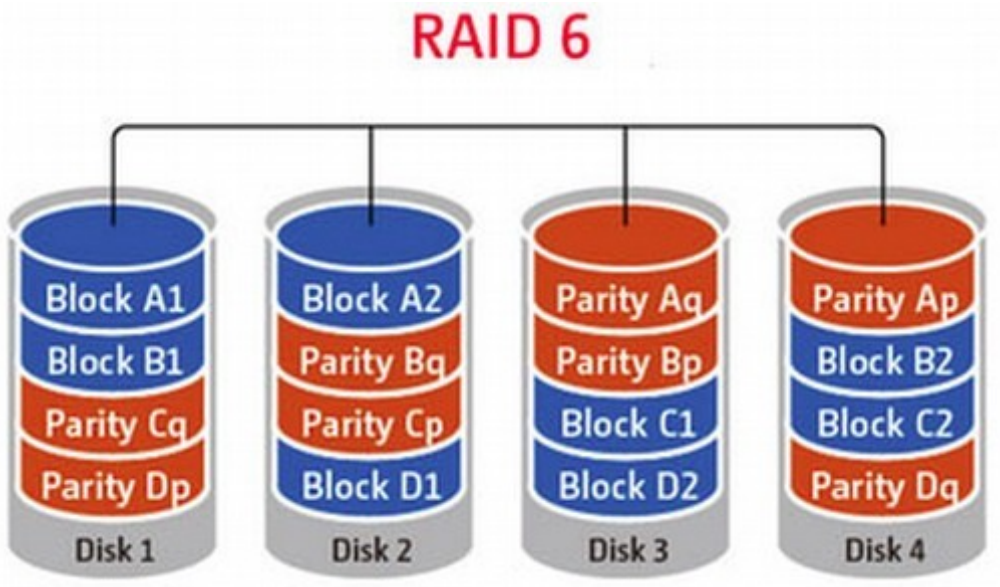
RAID nível 6

- Esquema $P + Q$
  - Muito semelhante ao RAID 5, mas armazena informações redundantes extras para proteger contra múltiplas falhas de disco
  - Usa códigos de correção de erro, como os códigos Reed-Solomon
  - Exemplo:
    - 2 bits de dados redundantes são armazenados para cada 4 bits de dados (no nível 5, é armazenado um bit de paridade)
    - O sistema pode tolerar falhas em dois discos

RAID 6: Esquema $P + Q$

RAID nível 6

- Se dois discos falharem:
  - Com RAID 5, você não terá seus dados armazenados
  - Com RAID 6 poderá ter seus arquivos salvos.



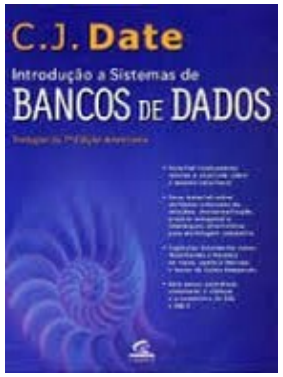
Escolha do nível de RAID

- RAID 0 é usado nas aplicações de alto desempenho em que a segurança dos dados não é crítica
- RAID 2 e 4 não são usados, pois foram substituídos pelos níveis 3 e 5
- Nível 3 é inferior ao nível 5
  - Espalhamento de blocos oferece boas taxas de transferência de dados para grandes transferências
  - Para transferências pequenas, o benefício de transferências paralelas é pequeno, pois o tempo de acesso ao disco domina
- RAID 6 é raramente usado porque os níveis 1 e 5 oferecem a segurança adequada para a maioria das aplicações

Exercícios

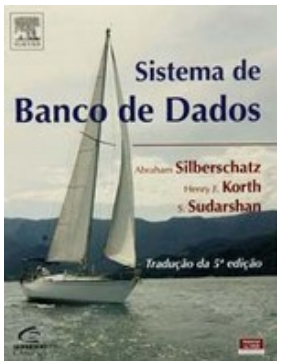
- ...

Bibliografia Básica



- DATE, C. J. Introdução a sistemas de bancos de dados. Rio de Janeiro, RJ: Campus, 2000.

- ELMASRI, Ramez; NAVATHE, Shamkant B. Sistemas de banco de dados. 4. ed. São Paulo: Pearson Addison-Wesley, 2005.



- SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. Sistema de banco de dados. Rio de Janeiro, RJ: Elsevier, 2006.