QXD-005 - Arquitetura de Computadores

Arquitetura RAID

Prof. Pedro Botelho

Nas Aulas Passadas…

- Visão de Alto Nível do Computador
- Memória Cache
- Memória Interna
- Memória Externa
- Questão: Como organizar várias unidades de armazenamento?

Nesta Aula…

- Arquitetura RAID

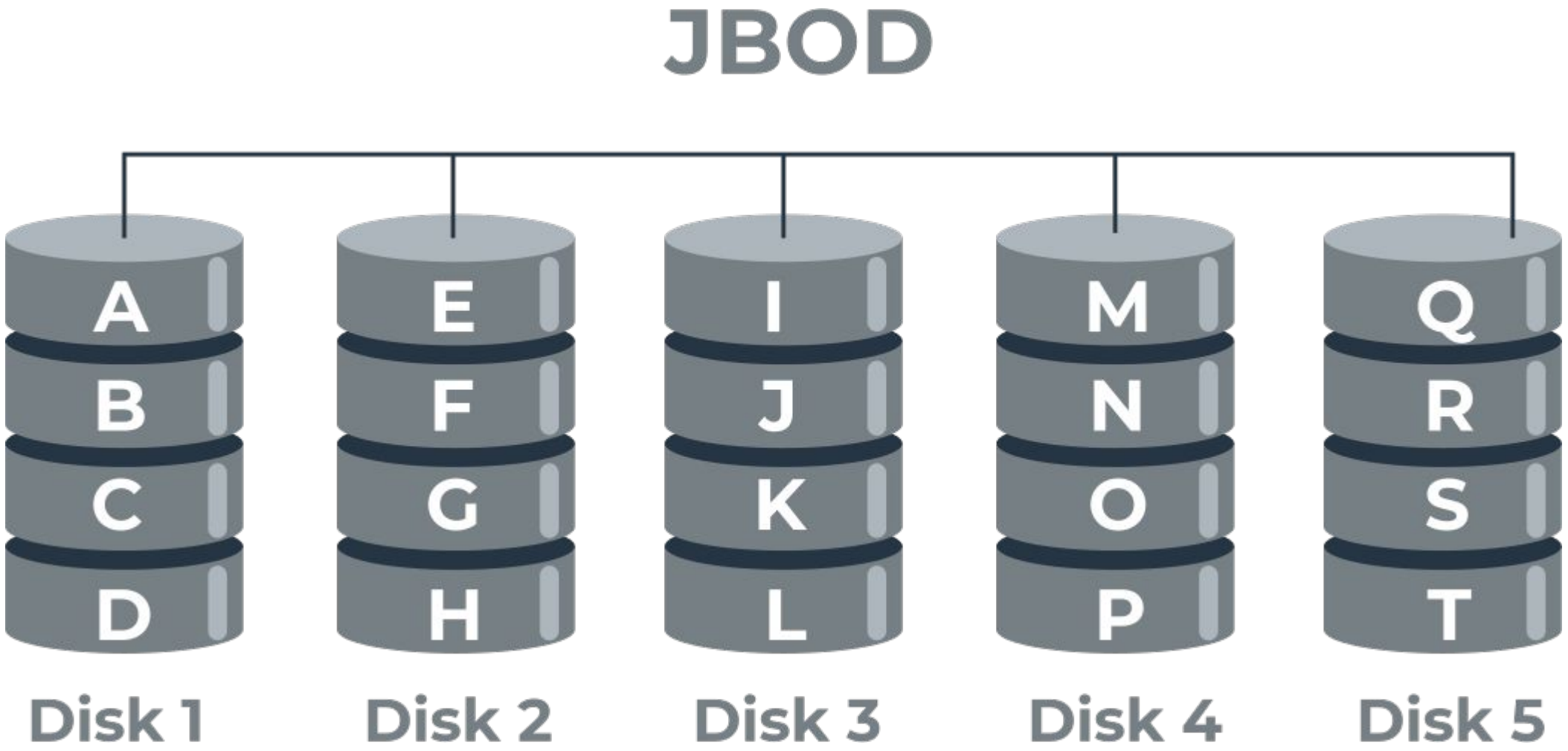
Arquitetura RAID

Arquitetura RAID

Como manter vários “discos”?

- Suponha um computador com 4 HDDs (ou SSDs) de 120GB...
  - Como gerenciá-los?
- Sistema Operacional reconhece as unidades como **“discos” individuais**
  - Por exemplo: Disco C, D, E no Windows
- Porém pode combinar todos os discos em uma única **“unidade lógica”**
- Configuração **JBOD** (*Just a Bunch of Disks*):
  - Une as 4 unidades físicas em uma lógica de 480GB
  - Muito **simples**, porém não oferece **confiabilidade** e **desempenho**
  - E se um dos discos falhar?
- Solução: **Arquitetura RAID**

Exemplo: Configuração JBOD



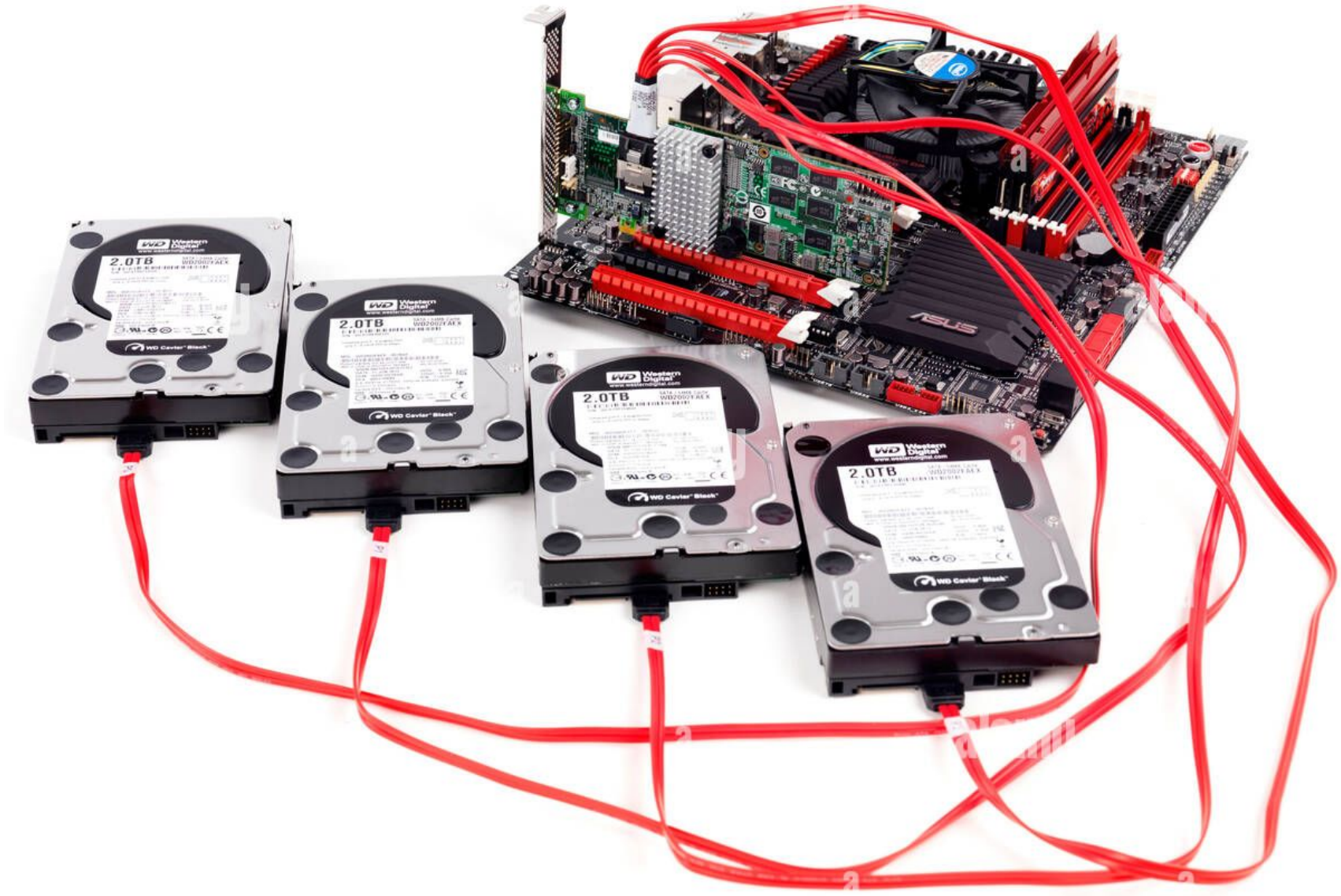
Introdução à Arquitetura RAID

- *Redundant Array of Independent/Inexpensive Discs*
- Conjunto de discos físicos vistos como uma única unidade lógica pelo SO
- As duas palavras-chave são:
  - **Redundante**: Dados redundantes em vários discos fornecem tolerância a falhas
  - **Array**: Vários discos acessados em paralelo com maior taxa de transferência do que um único disco.
- Vários níveis: Vantagens e desvantagens
- Métricas importantes:
  - **Capacidade**: Espaço dos discos disponível para dados
  - **Velocidade de Acesso:** Leitura e/ou escrita (desempenho)
  - **Segurança**: Redundância
  - **Custo**: Quantidade de discos

Gerenciamento do RAID

- Gerenciamento a nível de Software e Hardware
- **Hardware**: Placa com Controladora RAID dedicada
  - Possui chips DRAM e vários slots para discos: Pode suportar vários níveis RAID
- **Software**: Sistema Operacional ou um programa específico e.g. **mdadm** no Linux
  - Utiliza a memória DRAM e os slots da placa mãe: **Linux** suporta 0, 1, 4, 5, 6, 1+0 e **Windows** 0, 1 e 5
- Controladora de Hardware é bem mais **eficiente** que de Software, porém mais **cara**

Exemplo de Controladora RAID PCI Express

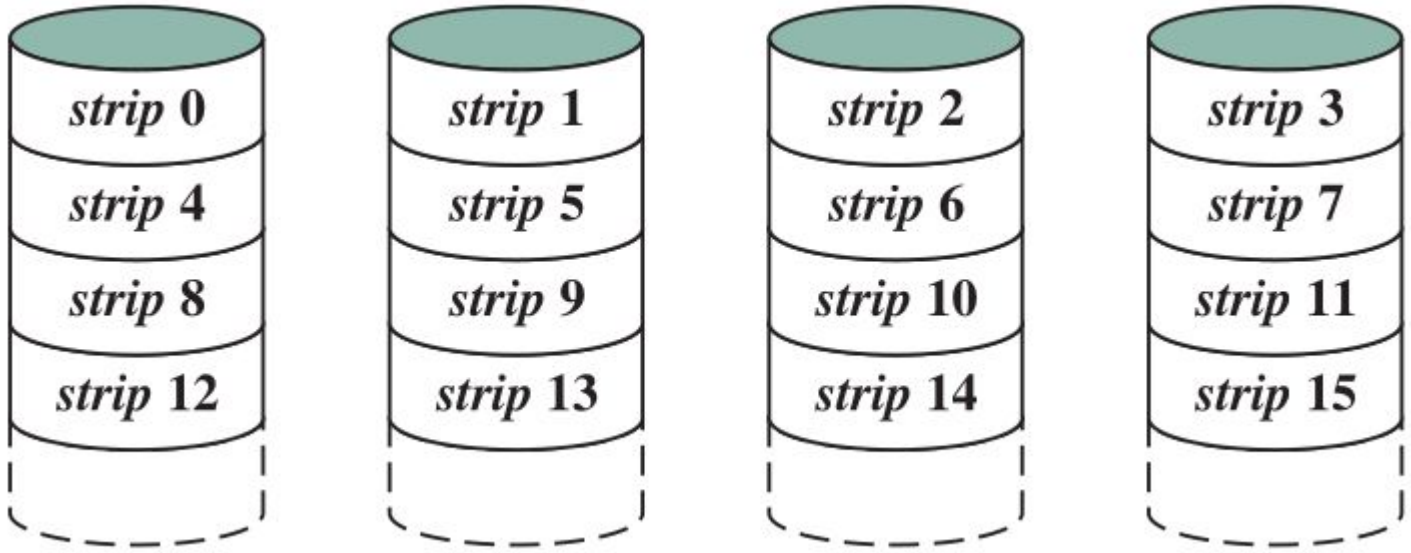


Arquitetura RAID

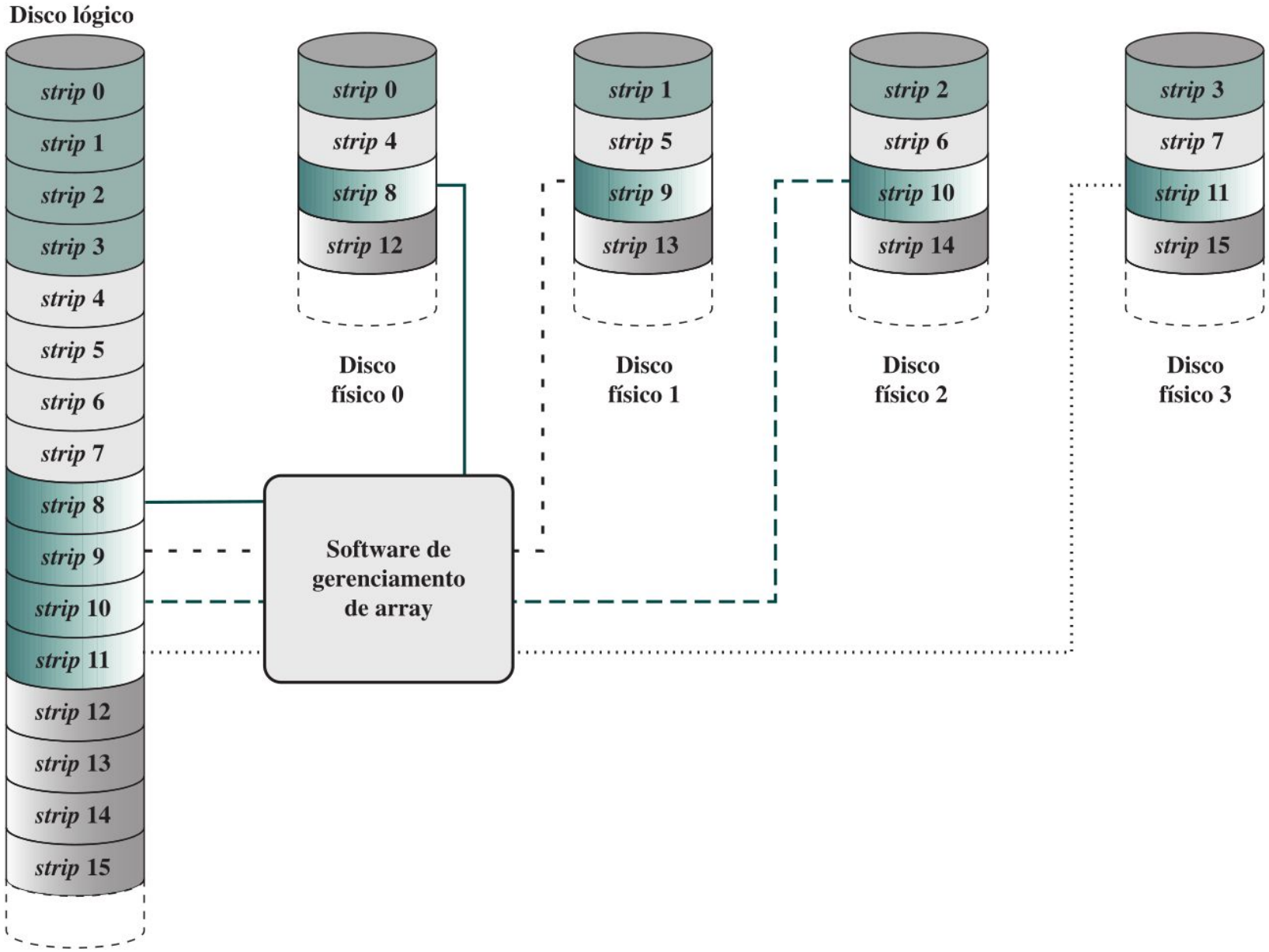
Níveis RAID

RAID 0 (Não redundante)

- Dados distribuídos em todos os discos: ***Stripping*** com *Round Robin*
- Aumento da velocidade
  - Múltiplas solicitações de dados provavelmente não estão no mesmo disco
  - Os discos buscam em **paralelo**
  - Um conjunto de dados provavelmente será **distribuído em vários discos**
- **Problema**: Sem redundância

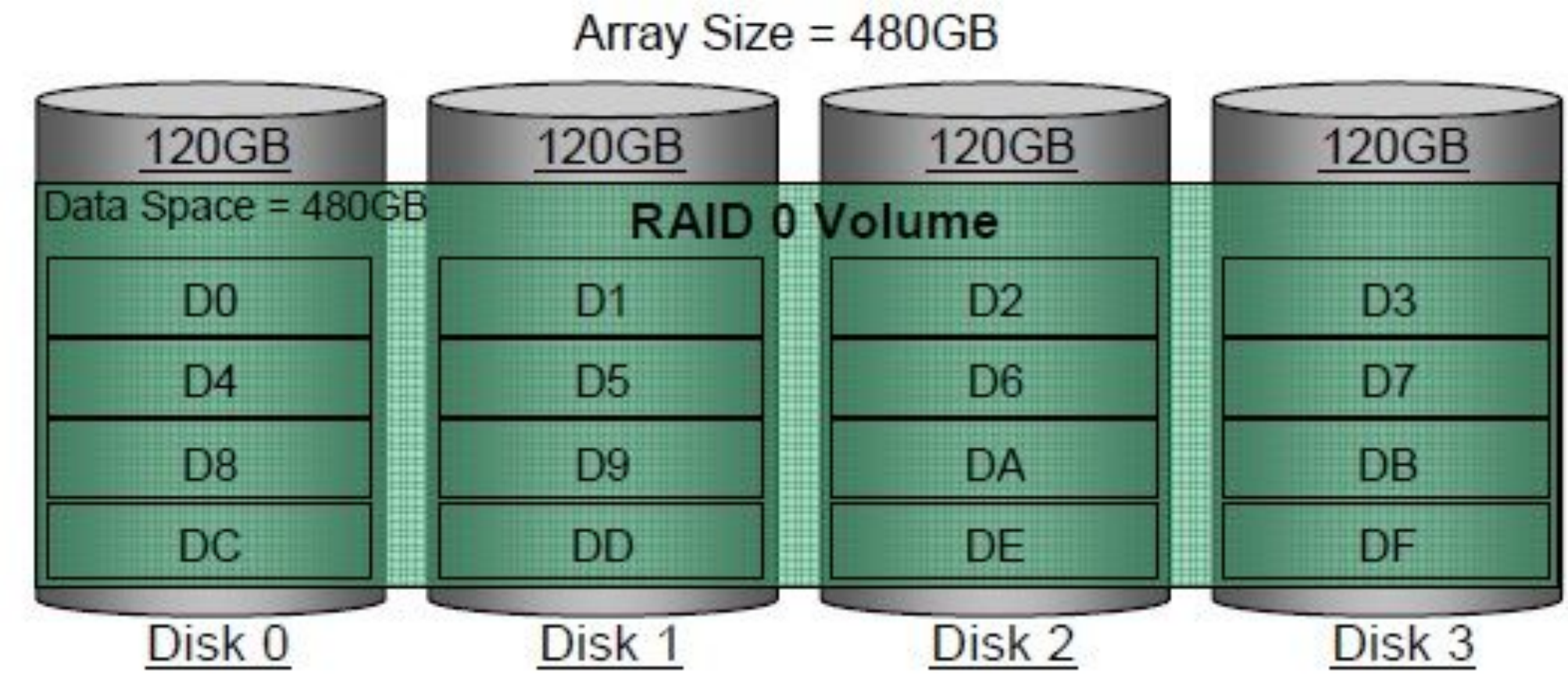


Mapeamento de Dados em um *Array* usando RAID 0



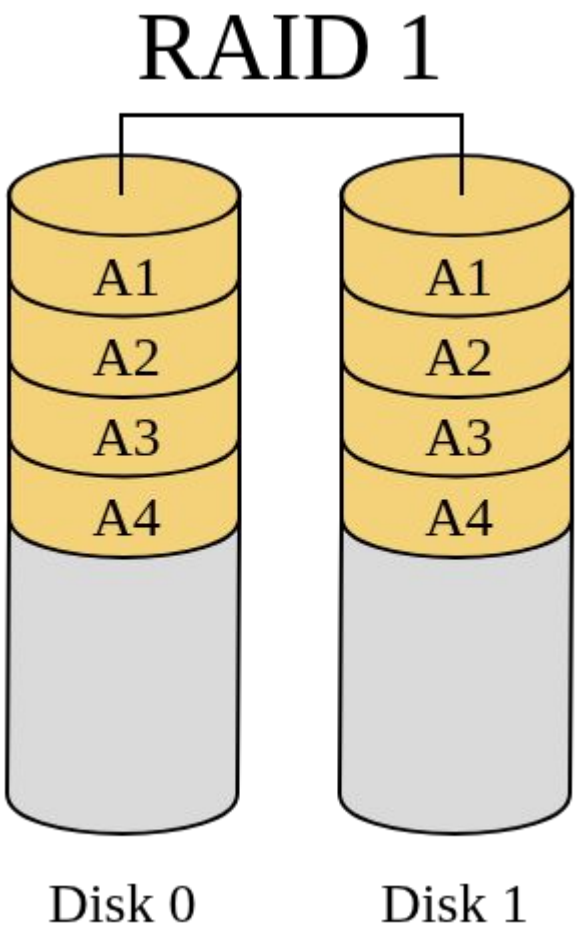
Exemplo: Discos usando RAID 0

- 4 discos de 120GB em RAID 0 →Capacidade de 480GB



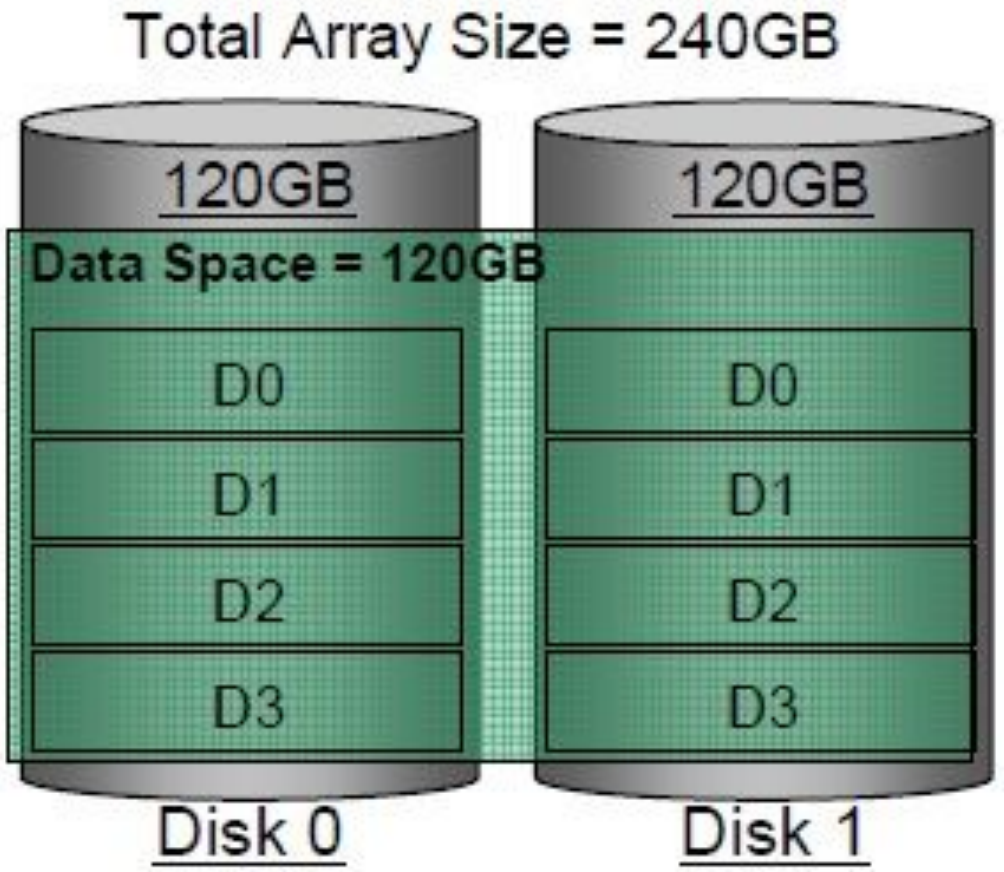
RAID 1 (Espelhado)

- Redundância alcançada pela **duplicação** de todos os dados
- Solicitação de leitura pode ser atendida por qualquer um dos dois discos
  - Desempenho ditado pelo mais rápido
- Uma solicitação de gravação requer que ambos os discos sejam atualizados
  - Desempenho ditado pelo mais lento
- Vantagem: Recuperação simples
  - Se o driver falhar, os dados estarão disponíveis no segundo
- Geralmente considerado junto com stripping: RAID 1+0



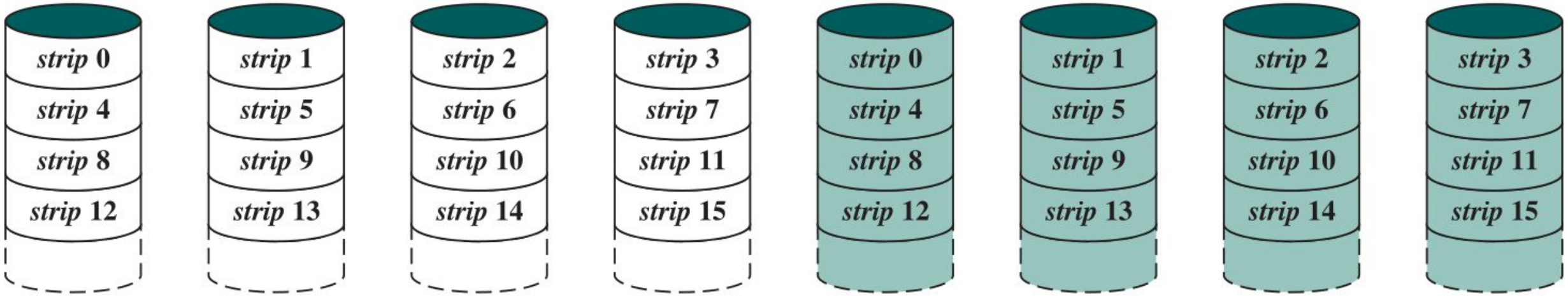
Exemplo: Discos usando RAID 1

- 2 discos de 120GB em RAID 1 →Capacidade de 120GB



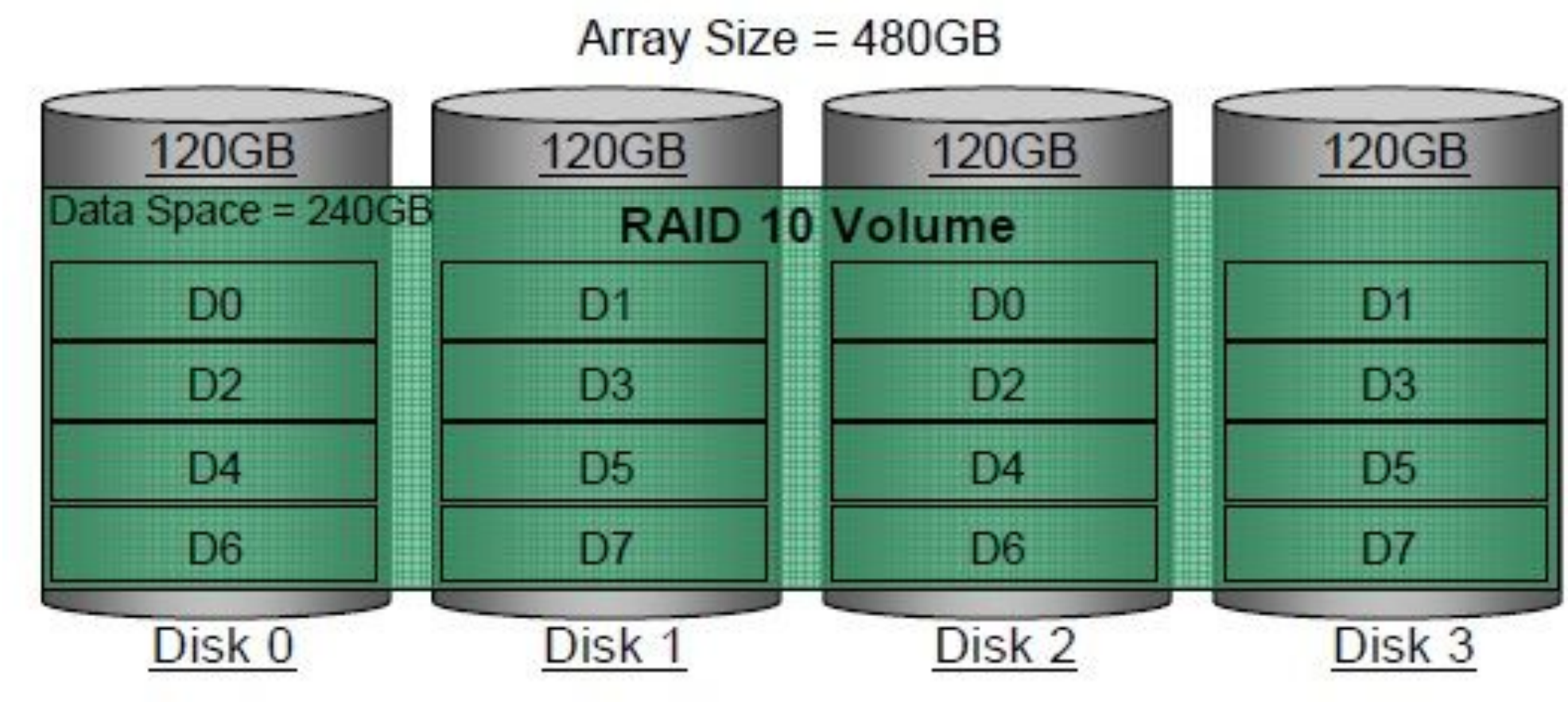
RAID 1+0 (Espelhamento e *stripping*)

- Os dados são distribuídos entre os discos (***stripping***)
- 2 cópias de cada tira em discos separados (**espelhamento**)
- Projeto **caro**, mas possui vantagens:
  - Leitura de qualquer um (aquele com menor procura)
  - Gravação em ambos: Sem penalidade de gravação
  - Recuperação simples: Troque o disco defeituoso e refaça o espelhamento



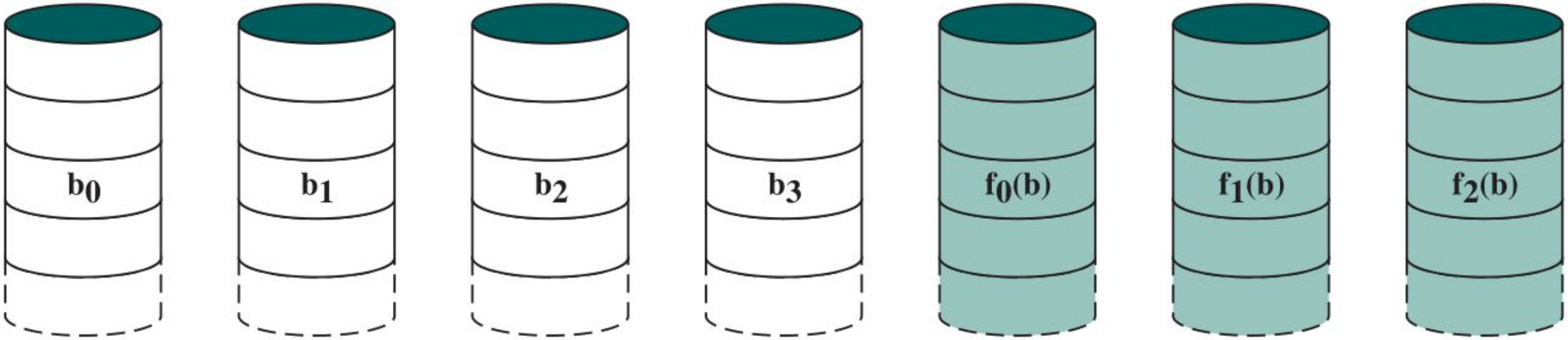
Exemplo: Discos usando RAID 1+0

- 4 discos de 120GB em RAID 1+0 →Capacidade de 240GB



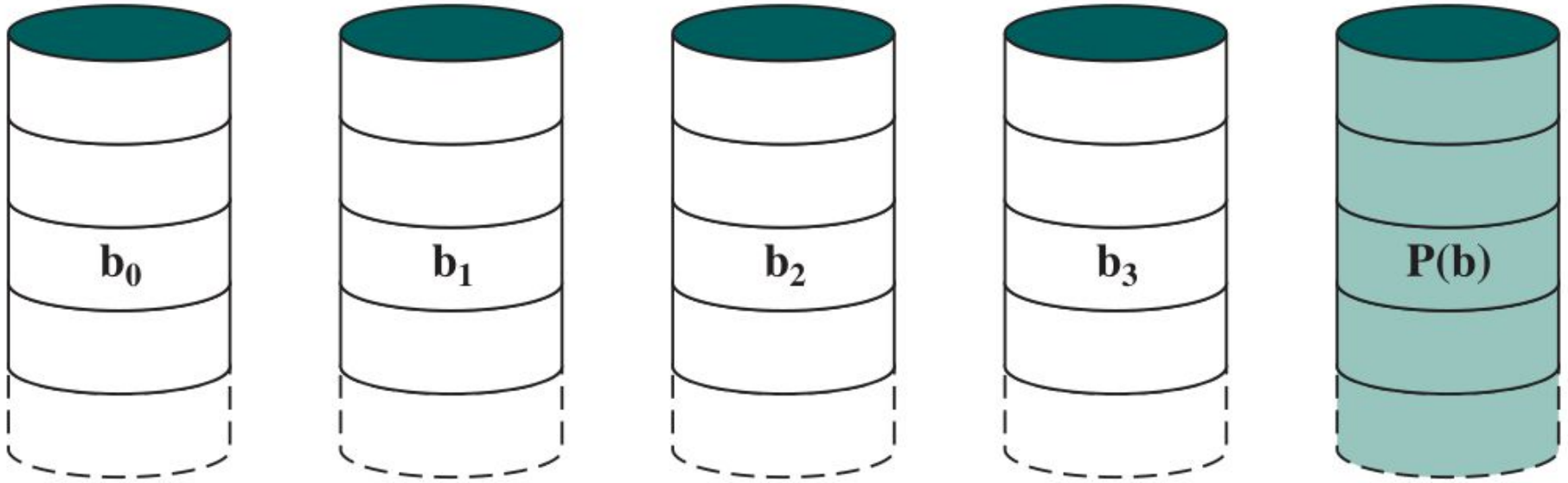
RAID 2 (*Bit-stripping* com Código de Hamming)

- Normalmente, os discos são **sincronizados**: Cabeça na mesma posição
- Listras muito pequenas: Geralmente um único byte/palavra
- Correção de erro calculada em bits correspondentes.
- Em uma única gravação: **Todos** os dados nos discos de **paridade** devem ser acessados
- Muita redundância: **Caro**
  - Somente eficaz se ocorrerem muitos erros de disco: Não usado atualmente



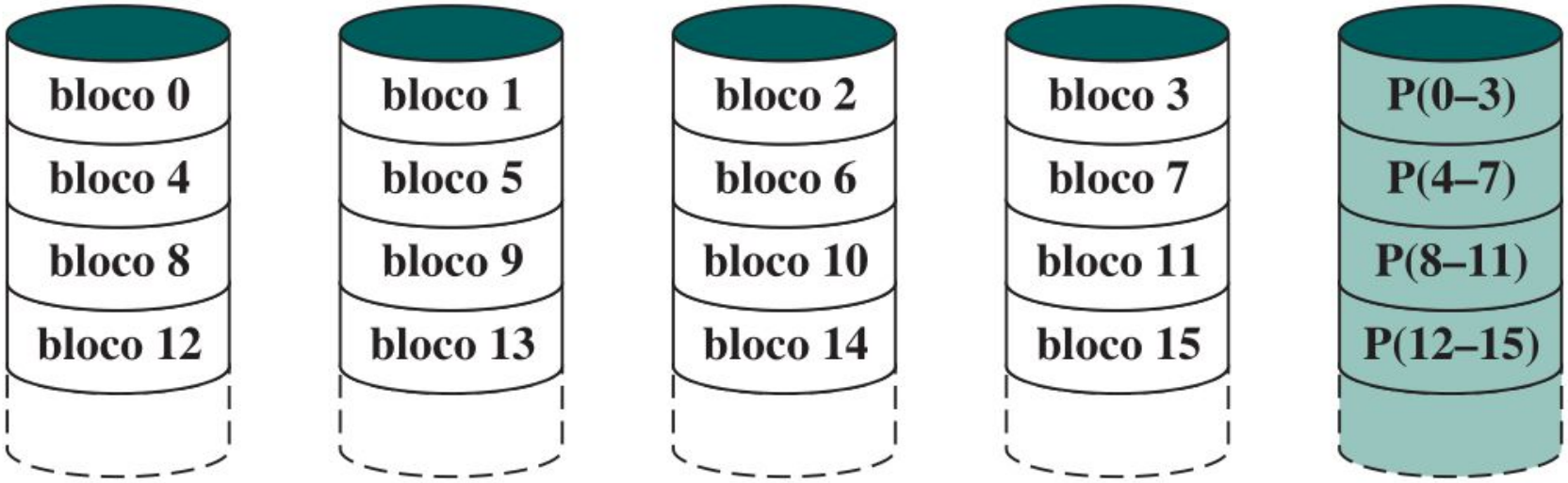
RAID 3 (*Bit-stripping* com Paridade Intercalada)

- Semelhante ao RAID 2
- Apenas **um disco redundante**, não importa o tamanho do array
  - Bit de paridade simples para cada conjunto de bits correspondentes
- Os dados da unidade com falha podem ser reconstruídos a partir dos dados sobreviventes e informações de paridade
- Taxas de transferência muito altas



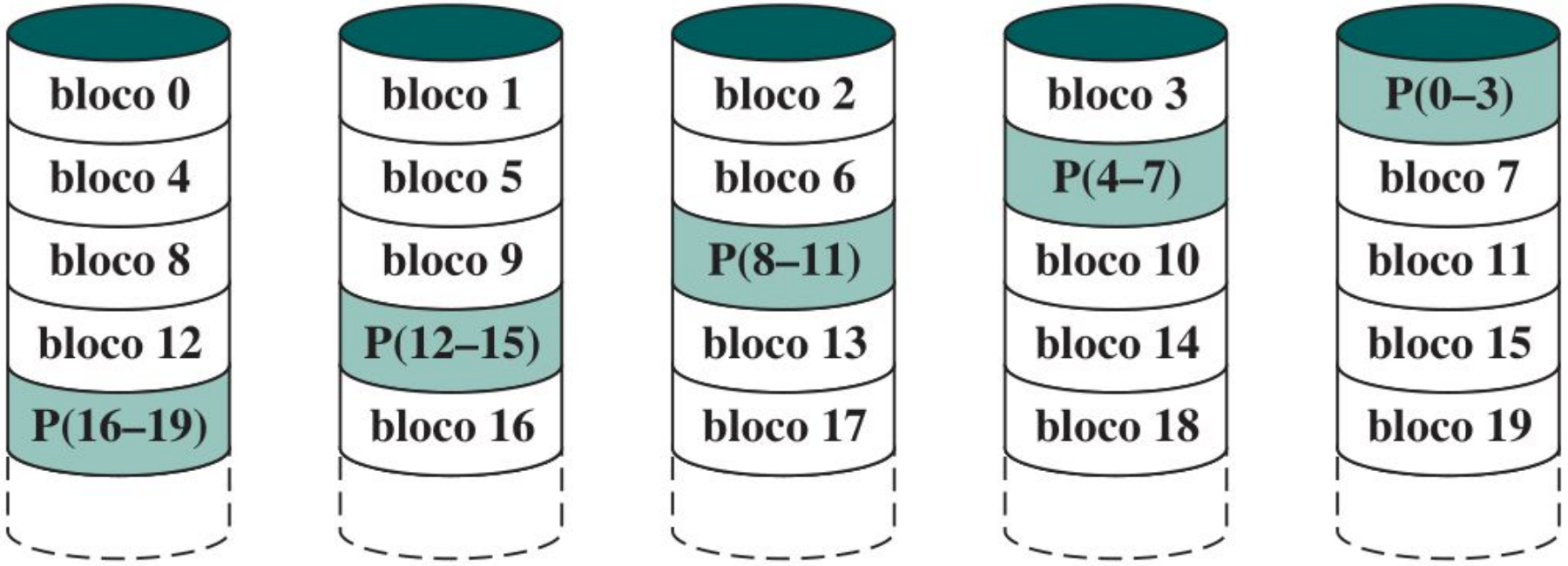
RAID 4 (Paridade a Nível de Bloco)

- Cada disco opera de forma independente
  - Solicitações de E/S separadas podem ser satisfeitas em paralelo.
  - Bom para **alta taxa de solicitação de E/S**
- Listras grandes (blocos)
- Paridade bit-a-bit calculada em faixas em cada disco: Armazenada no **disco de paridade**
- Escritas envolvem 2 leituras e gravações: Faixas de dados e paridade.
  - O disco de paridade se torna um **gargalo** e fica **sobrecarregado**.



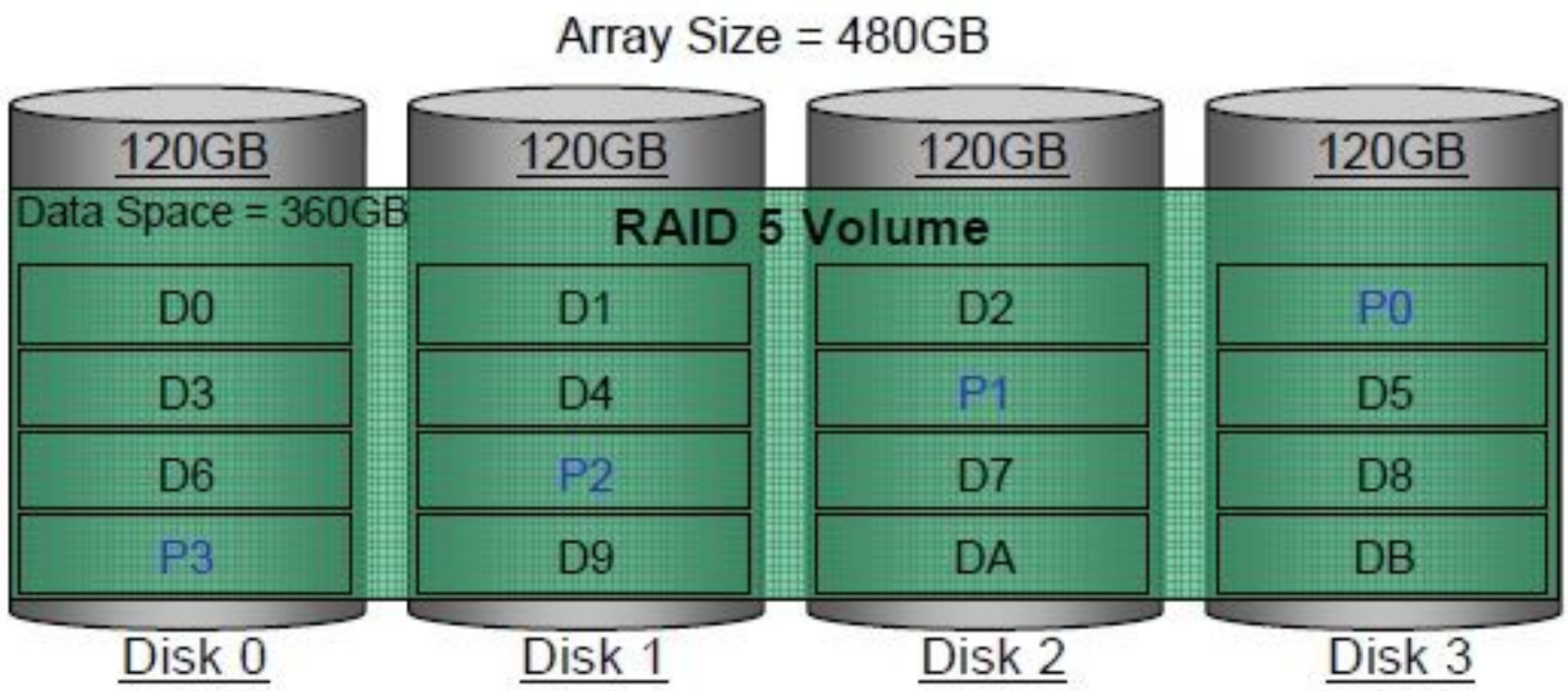
RAID 5 (Paridade Distribuída a Nível de Bloco)

- Melhoria do RAID 4: Evita **gargalos em discos de paridade**
- **Paridade distribuída** em todos os discos
  - Alocação *round robin* para faixas de paridade
- Comumente usado em servidores de rede



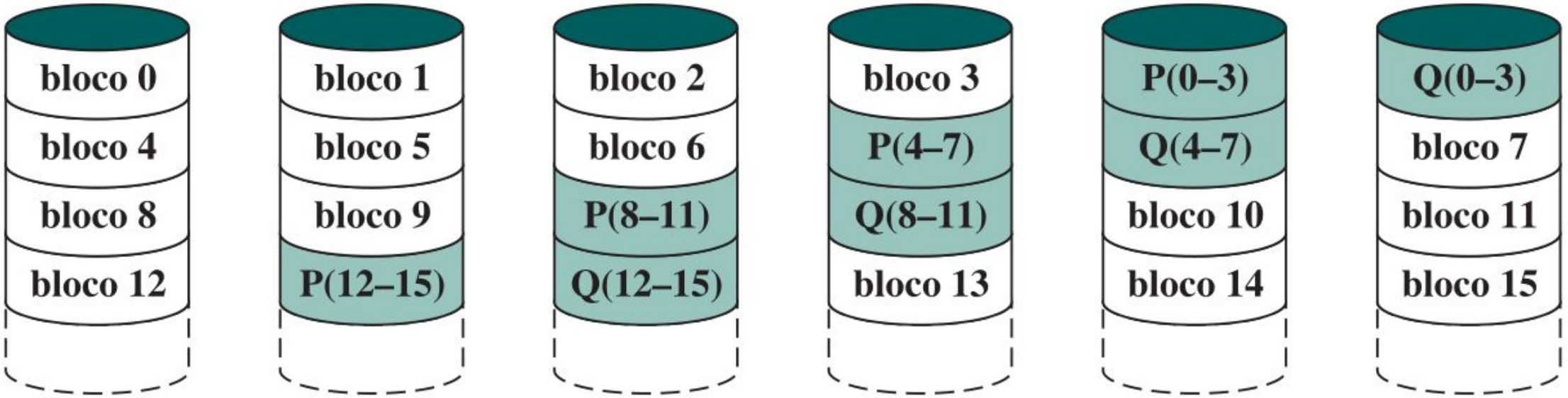
Exemplo: Discos usando RAID 5

- 4 discos de 120GB em RAID 5 →Capacidade de 360GB
  - 120GB usado para **paridade**



RAID 6 (Redundância Dupla)

- **Dois cálculos de paridade**: Armazenados em **blocos separados** em **discos diferentes**
  - Requisito do usuário de N discos precisa de N+2
- Alta disponibilidade de dados
  - Três discos precisam falhar para perda de dados
  - Penalidade de gravação significativa (30% em comparação com RAID 5)



Exemplo: Discos usando RAID 6

- 4 discos de 120GB em RAID 5 →Capacidade de 240GB
  - 240GB usado para paridade

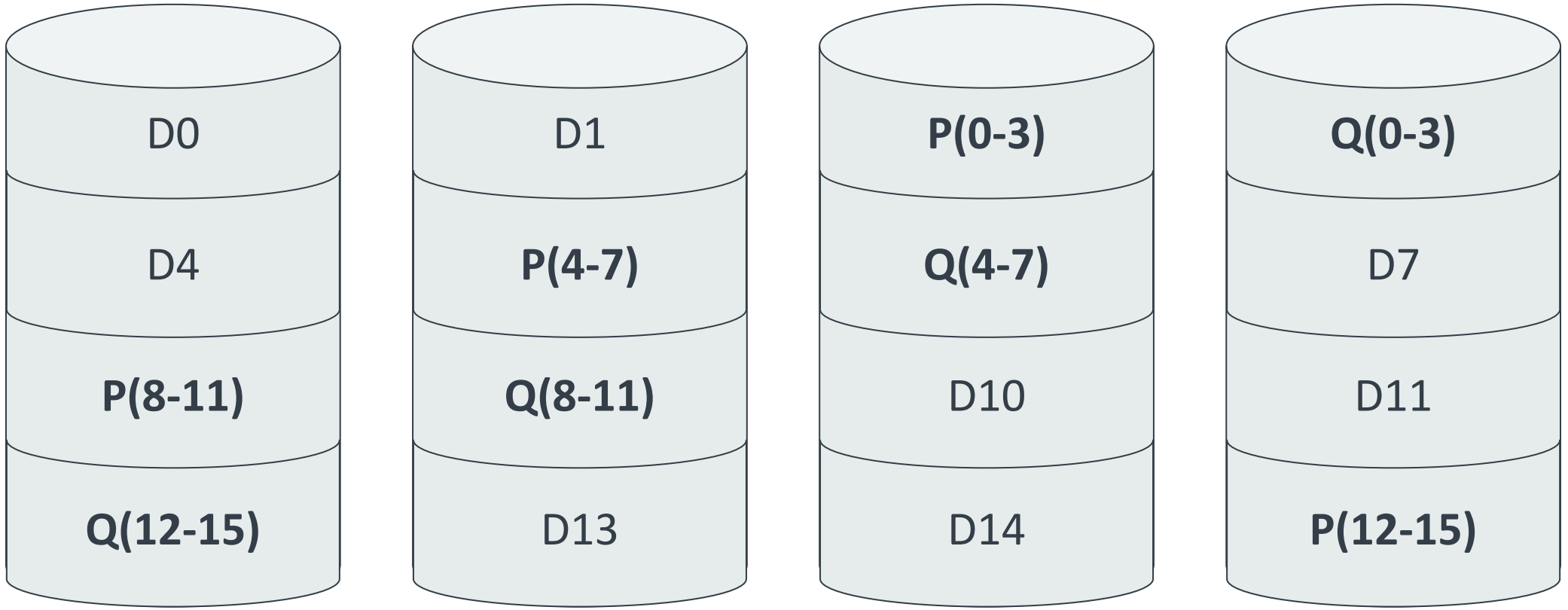


Tabela Resumo dos Níveis de RAID

Categoria	Nível	Descrição	Discos exigidos	Disponibilidade de dados	Grande capacidade de transferência de dados de E/S	Pequena taxa de solicitação de E/S
Striping	0	Não redundante	*N*	Inferior a um único disco	Muito alta	Muito alta tanto para leitura como para gravação
Espelhamento	1	Espelhado	2*N*	Mais alta que a RAID 2, 3, 4 ou 5; inferior ao RAID 6	Mais alta que o disco rígido para leitura; similar a um único disco para gravação	Até o dobro de um único disco para leitura; similar a um disco único para gravação
Acesso paralelo	2	Redundante via código de Hamming	*N* + *m*	Mais alta que um único disco; comparável ao RAID 3, 4 ou 5	Mais alta de todas as alternativas listadas	Aproximadamente o dobro de um único disco
	3	Paridade intercalada por bit	*N* + 1	Mais alta que um único disco; comparável ao RAID 2, 4 ou 5	Mais alta de todas as alternativas listadas	Aproximadamente o dobro de um único disco
Acesso independente	4	Paridade intercalada por bloco	*N* + 1	Mais alta que um único disco; comparável ao RAID 2, 3 ou 5	Similar ao RAID 0 para leitura; significativamente menor que um único disco para gravação	Similar ao RAID 0 para leitura; significativamente inferior a um único disco para gravação
	5	Paridade distribuída intercalada por bloco	*N* + 1	Mais alta que um único disco; comparável ao RAID 2, 3 ou 4	Similar ao RAID 0 para leitura; inferior a um único disco para gravação	Similar ao RAID 0 para leitura; geralmente inferior a um único disco para gravação
	6	Paridade dupla distribuída intercalada por bloco	*N* + 2	Mais alta de todas as alternativas listadas	Similar ao RAID 0 para leitura; inferior ao RAID 5 para gravação	Similar ao RAID 0 para leitura; significativamente inferior ao RAID 5 para gravação

Obs.: *N* = número de discos de dados; *m* proporcional a log *N*.

Comparativo dos Níveis de RAID

Nível	Vantagens	Desvantagens	Aplicações
0	O desempenho de E/S é bastante melhorado ao distribuir a carga de E/S por muitos canais e drives Não há *overhead* de cálculo de paridade envolvido Projeto muito simples Fácil de implementar	A falha de apenas um drive resultará na perda de todos os dados em um array	Produção e edição de vídeo Edição de imagens Aplicações de pré-impressão Qualquer aplicação exigindo grande largura de banda
1	100% de redundância de dados significa que não é preciso reconstruir em caso de falha do disco, apenas uma cópia para o disco substituto Sob certas circunstâncias, o RAID 1 pode sustentar múltiplas falhas simultâneas Projeto mais simples do subsistema de armazenamento RAID	*Overhead* de disco mais alto de todos os tipos de RAIDs (100%) — ineficaz	Contabilidade Folha de pagamento Financeiras Qualquer aplicação exigindo disponibilidade muito alta
2	Taxas de transferência de dados extremamente altas são possíveis Quanto mais alta a taxa de transferência de dados exigida, melhor a razão entre discos de dados e discos de ECC Projeto de controlador relativamente simples em comparação com os RAIDs de nível 3, 4 e 5	Razão muito alta entre discos de ECC e discos de dado com menores tamanhos de palavra — ineficazes Custo muito alto para cada nível — necessita requisitos de taxa de transferência muito altos para justificar	Nenhuma implementação comercial; inviável comercialmente

Comparativo dos Níveis de RAID

Nível	Vantagens	Desvantagens	Aplicações
3	Taxa de transação de dados muito alta para leitura Taxa de transferência de dados para gravação muito alta A falha de disco tem um impacto insignificante sobre o *throughput* Baixa razão entre discos de ECC (paridade) e discos de dados significa alta eficiência	Taxa de transação igual a de um único drive no máximo (se os eixos forem sincronizados) Projeto de controlador muito complexo	Produção de vídeo e *streaming* ao vivo Edição de imagens Edição de vídeo Aplicações de pré-impressão Qualquer aplicação exigindo alta taxa de transferência
4	Taxa de transação de dados muito alta para leitura Baixa razão entre discos de ECC (paridade) e discos de dados significa alta eficiência	Projeto de controlador muito complexo Pior taxa de transação de gravação e taxa de transferência de gravação agregada Reconstituição de dados difícil e ineficaz no caso de falha de disco	Nenhuma implementação comercial; inviável comercialmente

Comparativo dos Níveis de RAID

Nível	Vantagens	Desvantagens	Aplicações
5	Mais alta taxa de transação de dados para leitura Baixa razão entre discos de ECC (paridade) e discos de dados significa alta eficiência Bom tempo de transferência agregado	Projeto de controlador mais complexo de todos Difícil reconstituição no evento de uma falha de disco (comparado com RAID nível 1)	Servidores de arquivo e aplicação Servidores de banco de dados Servidores Web, de e-mails e de notícias Servidores de Intranet Nível RAID mais versátil
6	Oferece uma tolerância a falhas extremamente alta e pode sustentar múltiplas falhas de drives simultâneos	Projeto de controlador mais complexo *Overhead* do controlador para o endereço de paridade do computador extremamente alta	Solução perfeita para aplicações de tarefa crítica

OBS: Níveis 2, 3, e 4 são muito específicos: Pouco usados

Comparativo dos Níveis de RAID a Nível Empresarial

RAID LEVEL	METHOD	HARDWARE / SOFTWARE	MINIMUM # OF DISKS	COMMON USAGE	PROS	CONS
JBOD	SPANNING		2	INCREASE CAPACITY	COST-EFFECTIVE STORAGE	NO PERFORMANCE OR SECURITY BENEFITS
0	STRIPING		2	HEAVY READ OPERATIONS	HIGH PERFORMANCE (SPEED)	DATA IS LOST IF ONE DISK FAILS
1	MIRRORING		2	STANDARD APP SERVERS	FAULT TOLERANCE, HIGH READ PERFORMANCE	LAG FOR WRITE OPS, REDUCED STORAGE (BY 1/2)
5	STRIPING & PARITY		3	NORMAL FILE STORAGE & APP SERVERS	SPEED + FAULT TOLERANCE	LAG FOR WRITE OPS, REDUCED STORAGE (BY 1/3)
6	STRIPING & DOUBLE PARITY		4	LARGE FILE STORAGE & APP SERVERS	EXTRA LEVEL OF REDUNDANCY, HIGH READ PERFORMANCE	LOW WRITE PERFORMANCE, REDUCED STORAGE (BY 2/5)
10 (1+0)	STRIPING & MIRRORING		4	HIGHLY UTILIZED DATABASE SERVERS	WRITE PERFORMANCE + STRONG FAULT TOLERANCE	REDUCED STORAGE (1/2), LIMITED SCALABILITY

Resumo das Métricas dos Níveis de RAID

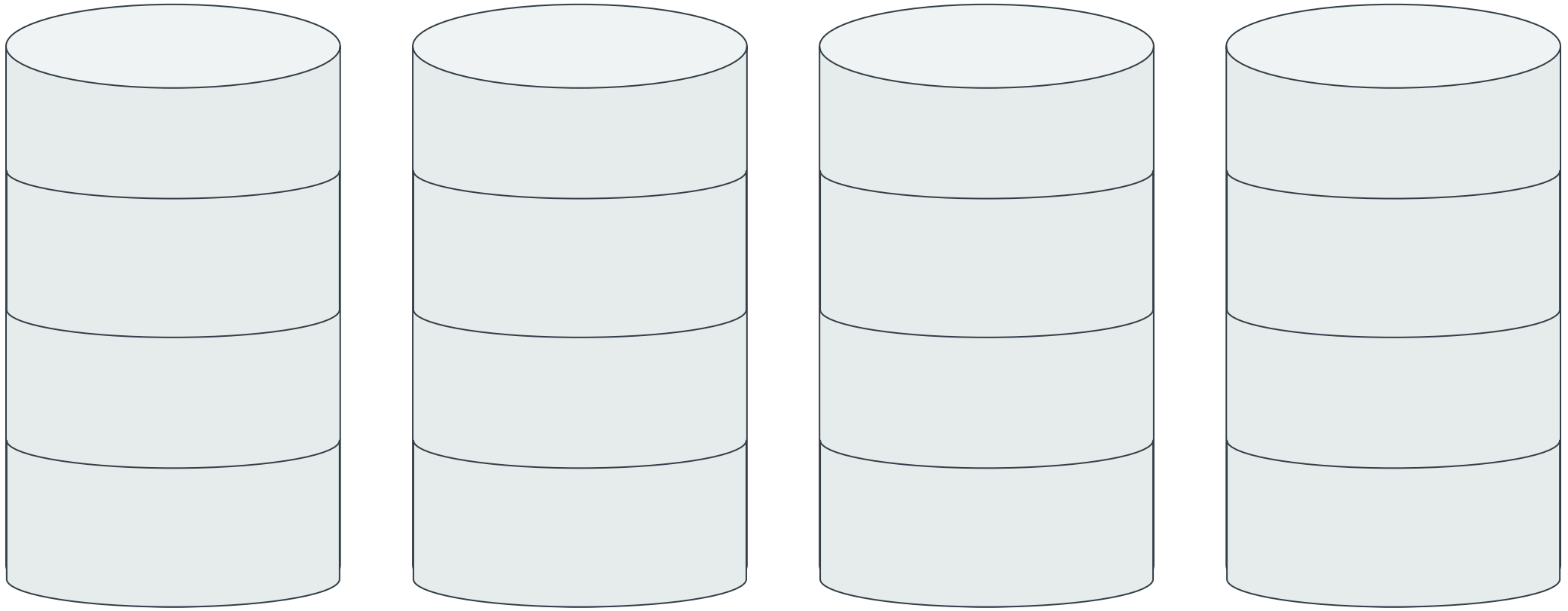
RAID 0	RAID 1	RAID 4	RAID 5	RAID 6	RAID 1+0 (10)
Blocks Striped. No Mirror. No Parity.	Blocks Mirrored. No Stripe. No Parity.	Blocks Striped and Dedicated Parity.	Blocks Striped. Distributed Parity.	Blocks Striped. Two Distributed Parity.	Blocks Mirrored and Striped.
Capacity 	Capacity 	Capacity 	Capacity 	Capacity 	Capacity 
• Fastest RAID • No protection from disk failure • Best for scratch storage when editing digital video/photos/ media • Requires 2 or more disks	• Safest RAID • Most disk failure protection • Best for critically important data where access speed is not an issue • Requires 2 or more disks	• Fast and safe • Best for general use on SSDs • Super-fast read/write of large files used for video, animation, photography, and graphics • Requires 3 or more disks	• Fast and safe • Best for general use on HDDs • Super-fast read/write of large files used for video, animation, photography, and graphics • Requires 3 or more disks	• Similar to RAID 5 with an additional parity block of recovery information • Allows for the failure of 2 disks • Slightly slower than RAID 5 on writes, no added delays on reads • Requires 4 or more disks	• Fastest and safest RAID option • Best for businesses or those needing high performance with increased reliability • Requires 4 or more disks

Arquitetura RAID

Exercícios

Exercício 1: Capacidade de Armazenamento em RAID

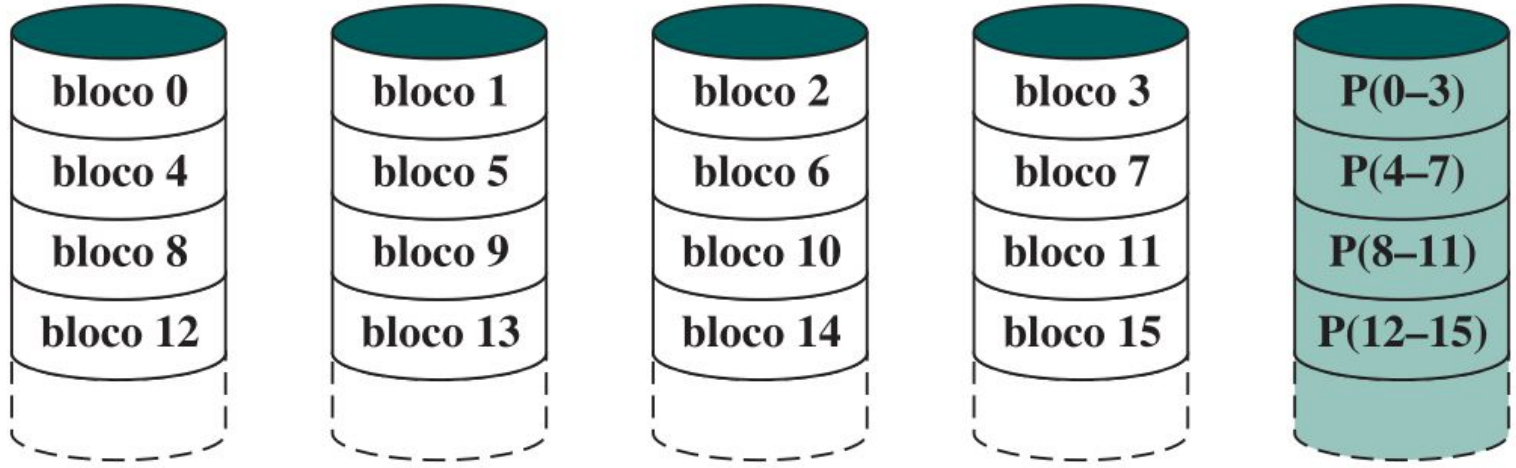
Considere um array RAID com **4 drives**, com **200 GB** cada e com strip de **50 GB**. Qual é a capacidade de armazenamento de dados disponível para cada um dos níveis de RAID 0, 1, 3, 4, 5 e 6? É possível implementar RAID 2 nesse cenário? Se não, como implementar?



Exercício 2: Recuperação de Dados

Considere um sistema de armazenamento baseado em **RAID 4**, composto por **4 discos** de blocos de dados (HD_0, HD_1, HD_2, HD_3) e 1 disco de paridade (HD_P). O sistema utiliza blocos de **2 bytes** (16 bits) para armazenar as informações. O conteúdo dos blocos de número 0, 1, 2 e do disco de paridade estão listados abaixo (em hexadecimal):

Bloco	Conteúdo
0	$00FF_{16}$
1	$F00F_{16}$
2	$?_{16}$
3	$A5A5_{16}$
P(0-3)	$AA55_{16}$



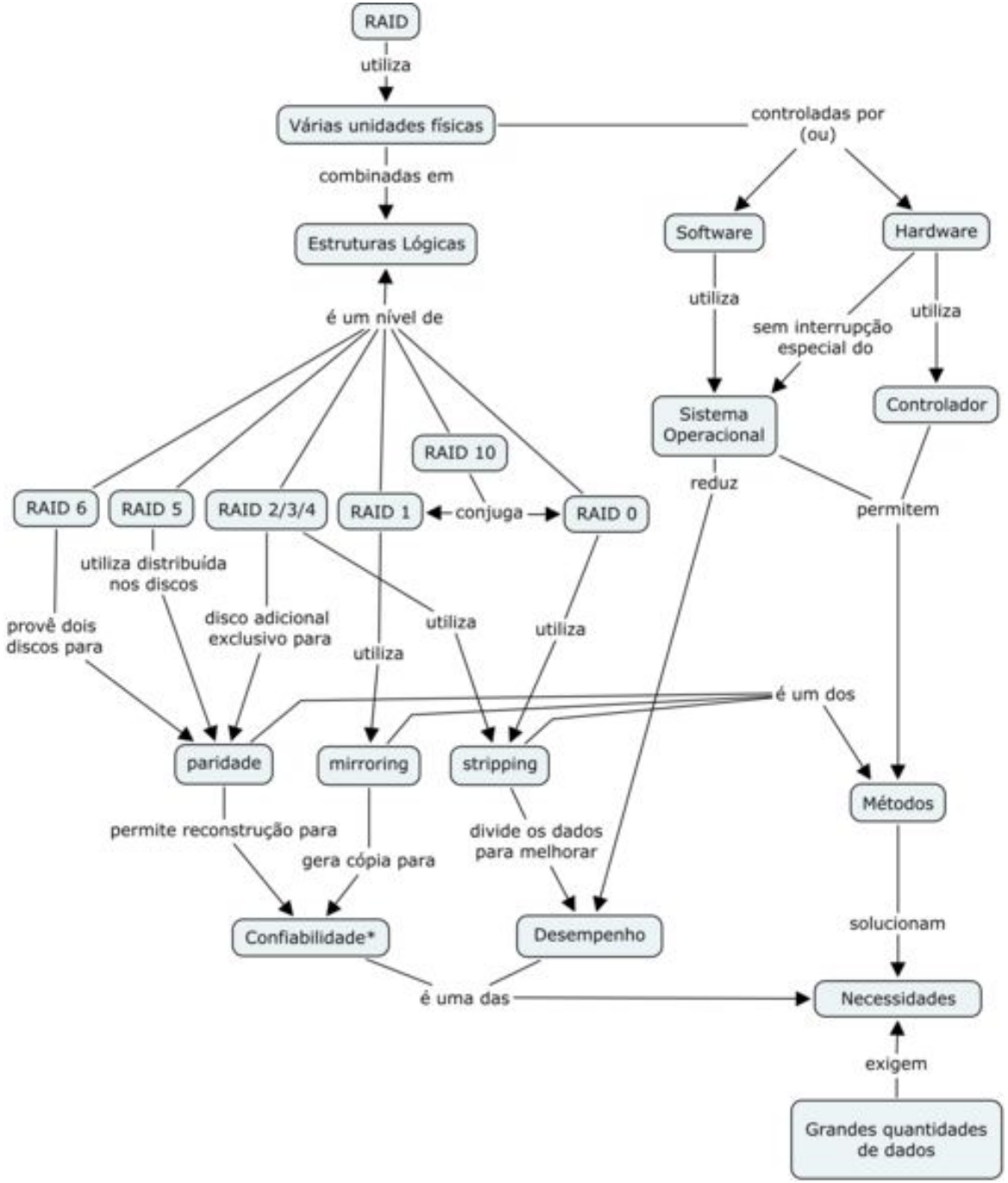
Sabe-se que o disco HD_2 falhou (queimou). Com base nas informações acima, recupere os dados perdidos no bloco 2, guardado no HD_2. Qual nível de RAID permitiria a recuperação em um cenário em que o disco de paridade queime também, além do HD_2?

Arquitetura RAID

Conclusão

Resumo da Aula

- **Arquitetura RAID**: Organização de discos de forma a oferecer (diferente do JBOD):
  - **Redundância** (Recuperação dos Dados)
  - **Performance** (Acesso Paralelo)
- Níveis RAID:
  - **0**: Stripping
  - **1**: Espelhamento
  - **1+0**: Stripping & Espelhamento
  - **2**: Bit-stripping com Código Hamming
  - **3**: Bit-stripping com Paridade Intercalada
  - **4**: Block-stripping com Paridade Centralizada
  - **5**: Block-stripping com Paridade Distribuída
  - **6**: Block-stripping com Paridade Dupla



Conclusão

- Nessa Aula:
  - Arquitetura RAID
- Bibliografia Principal:
  - Arquitetura e Organização de Computadores; Stallings, W.; 10ª Edição (Capítulo 6)
- Próxima Aula:
  - Entrada/Saída