Memória externa

MAC 344 - Arquitetura de Computadores
Prof. Siang Wun Song

Baseado parcialmente em W. Stallings

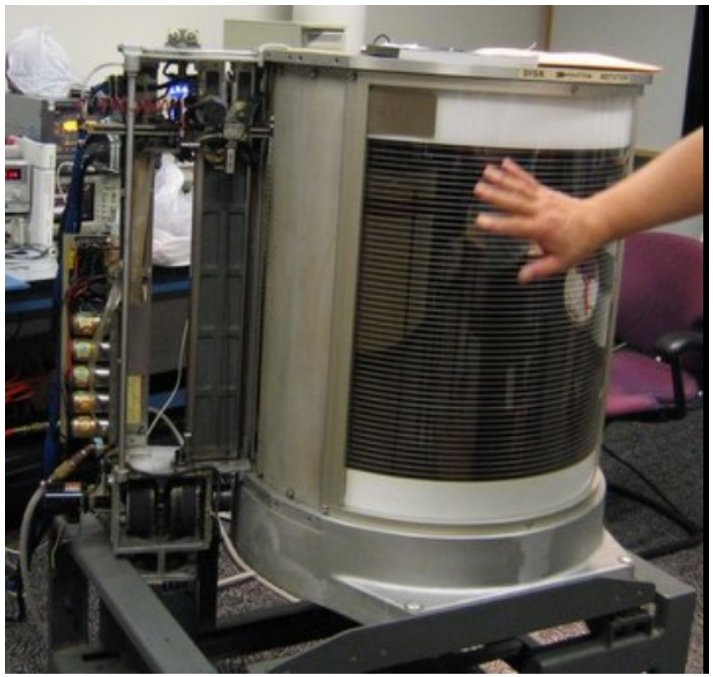
Computer Organization and Architecture

Disco da IBM em 1956

Em 1956, IBM 305 inventou primeiro disco magnético de cabaça móvel RAMAC - Random Access Method of Access and Control (Fonte: *Newsweek*, Aug 14, 2006, p. 8.)

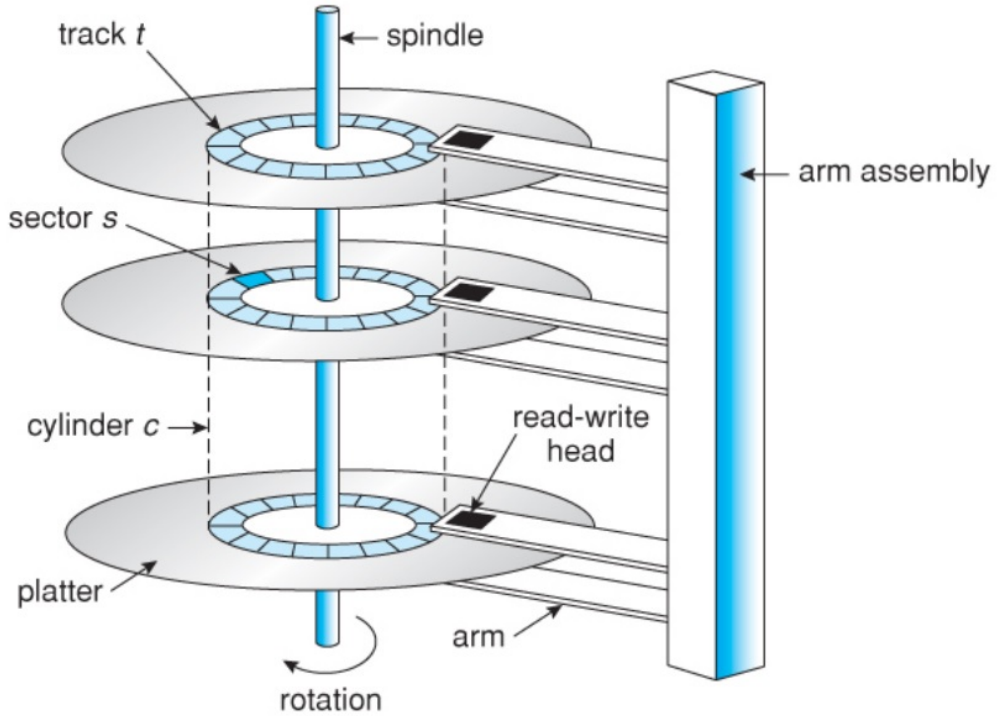
- Pesava uma tonelada
- Era alugado por US$ 250.000,00 por ano
- Tinhas capaciade de 5 Megabytes

Source: Computer History Museum



Disco magnético

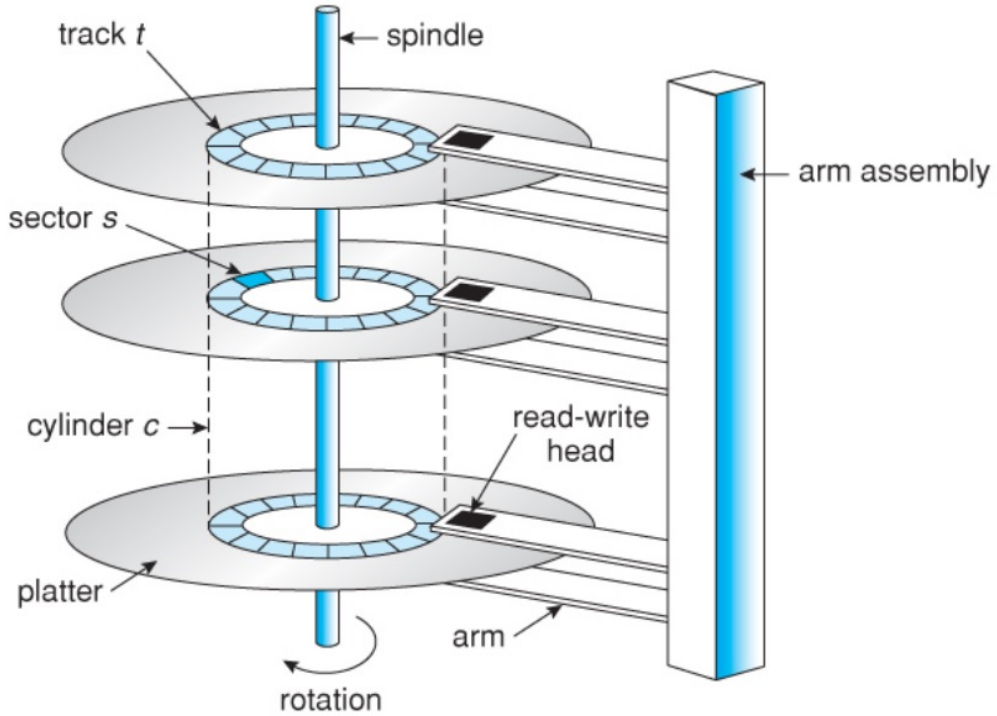
- O disco magnético consiste de fatias circulares de substrato formado de alumínio ou de vidro coberto por uma camada magnética.
- O disco é dividido em trilhas que, por sua vez, é organizada em setores. Cada setor contém tipicamente 512 bytes da dado mais alguns de controle.



Source: A. Silberschatz, G. Gagne, and P. B. Galvin. Operating System Concepts.

Disco magnético

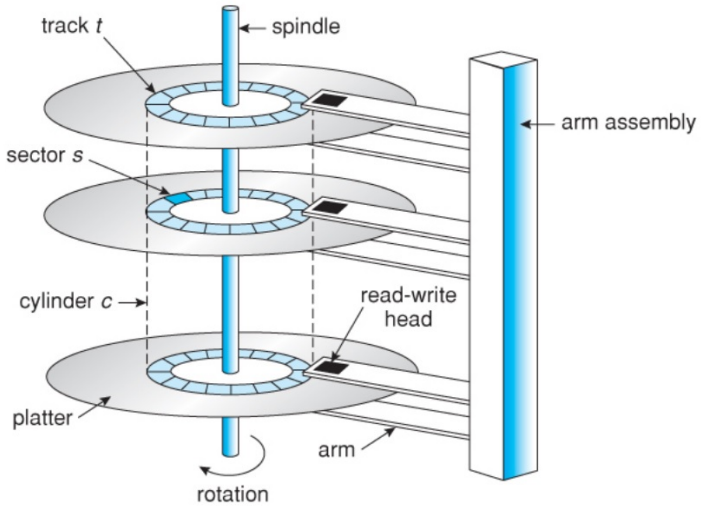
- As cabeças de leitura/gravação podem ser do tipo móvel (ver figura): primeiro a cabeça é posicionada em cima da trilha desejada antes de proceder o acesso.
- Discos mais modernos possuem cabeças fixas: uma cabeça em cima de cada trilha, dispensando a movimentação das mesmas.



Source: A. Silberschatz, G. Gagne, and P. B. Galvin. Operating System Concepts.

Parâmetros de desempenho do disco magnético

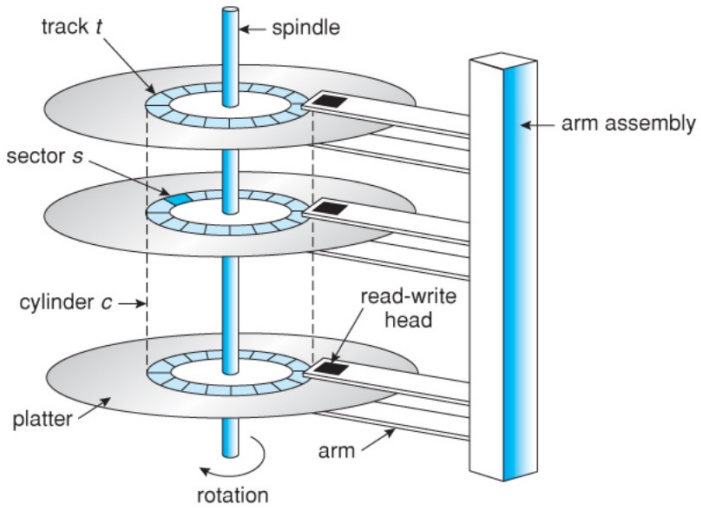
- Para acessar dados em um disco de cabeça móvel, é necessário primeiro posicionar a cabeça em cima da trilha desejada.
- Esse tempo é denomindo *seek time*. O valor típico do seek time é de 3 a 12 ms.
- Posicionada a cabeça na trilha desejada, é necessário ainda esperar que o setor desejado chegue em baixo da cabeça.
- Esse tempo é denominado *latência rotacional*. O valor típico é de 4 a 8 ms.



Source: A. Silberschatz, G. Gagne, and P. B. Galvin. Operating System Concepts.

Parâmetros de desempenho do disco magnético

- O melhor caso para a *latência rotacional* é o setor desejado já está junto à cabeça. O pior caso é ter que esperar uma volta inteira. O caso médio é esperar meia rotação.
- A soma de *seek time* mais *latência rotacional* é denominada *tempo de acesso*: a cabeça está pronta para acessar o setor.
- Tempo médio de acesso = seek time + $\frac{1}{2r}$ onde r é a velocidade em rotações por segundo.
- O *tempo de transferência* depende de quantidade de bytes a acessar.



Source: A. Silberschatz, G. Gagne, and P. B. Galvin. Operating System Concepts.

RAID - Redundant Array of Independent Disks

- O acesso a disco magnético leva tipicamente de 10 ms ou mais.
- Por essa razão projeto de estruturas de dados que residem em disco deve levar isso em consideração. Um exemplo é a *B-árvore*.
- Melhorias no desempenho do disco magnético é substancialmente menor que melhorias no desempenho do processador e memória interna.
- Isso levou a projetos de arranjos de múltiplos discos que operam independentemente e em paralelo.
- Com múltiplos discos, demandas separadas de entrada e saída podem ser atendidas em paralelo, desde que dados desejados residam em discos separados.
- Mesmo uma mesma requisição de entrada e saída pode ser executada em paralelo, desde que blocos de dados a serem acessados estejam distribuídos em múltiplos discos.

RAID - Redundant Array of Independent Disks

- RAID (*Redundant Array of Independent Disks*) é um conjunto de discos físicos visto pelo sistema operacional como uma unidade lógica.
- Dados são distribuídos nos múltiplos discos para viabilizar acesso simultâneo a dados de múltiplos discos.
- O uso de múltiplas cabeças de leitura/gravação aumenta a vazão de transferência, mas também aumenta a probabilidade de falhas.
- RAID usa redundância de dados que permite a recuperação de dados em falhas.
- Um artigo em 1988 define as configurações RAID em sete níveis.

RAID - Redundant Array of Independent Disks

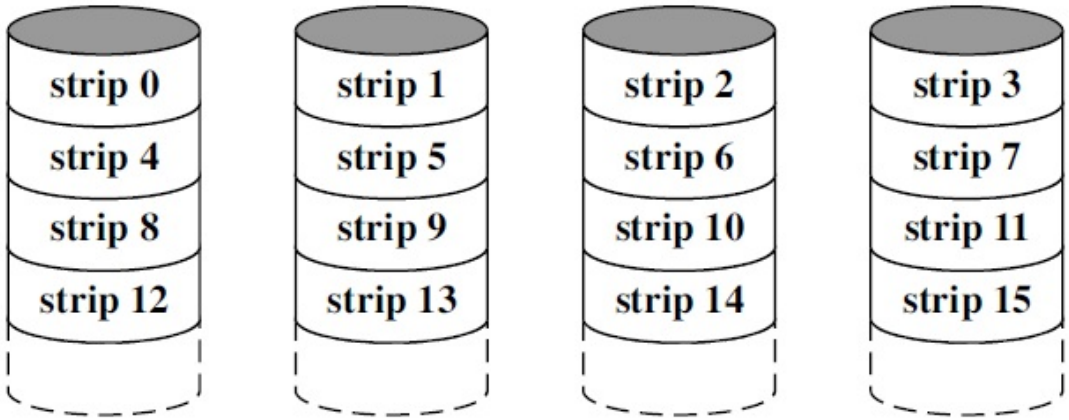
Sete níveis de RAID:

Level	Description	Minimum number of drives[b]	Space efficiency	Fault tolerance	Read performance	Write performance
RAID 0	Block-level striping without parity or mirroring	2	1	None	$n\times$	$n\times$
RAID 1	Mirroring without parity or striping	2	$\frac{1}{n}$	$n-1$ drive failures	$n\times$[a][15]	$1\times$[c][15]
RAID 2	Bit-level striping with Hamming code for error correction	3	$1-\frac{1}{n}\log_2(n-1)$	One drive failure[d]	Depends	Depends
RAID 3	Byte-level striping with dedicated parity	3	$1-\frac{1}{n}$	One drive failure	$(n-1)\times$	$(n-1)\times$[e]
RAID 4	Block-level striping with dedicated parity	3	$1-\frac{1}{n}$	One drive failure	$1-(1-r)^n-nr(1-r)^n$ [1]	$(n-1)\times$
RAID 5	Block-level striping with distributed parity	3	$1-\frac{1}{n}$	One drive failure	$n\times$[e]	$(n\quad 1)\times$[e] [citation needed]
RAID 6	Block-level striping with double distributed parity	4	$1-\frac{2}{n}$	Two drive failures	$n\times$[c]	$(n-2)\times$[e] [citation needed]

Source: Wikipedia

RAID 0 - Sem redundância, com strips round robin

- Sem redundância. Distribui *strips* ou blocos de dados logicamente contíguos em discos em forma de *round robin* ou *rodízio*: i.e. para n discos,
  strip i é armazenado no disco $i \mod n$.
- Essa distribuição permite acesso paralelo de *strips* logicamente contíguos, pois residem em discos diferentes.

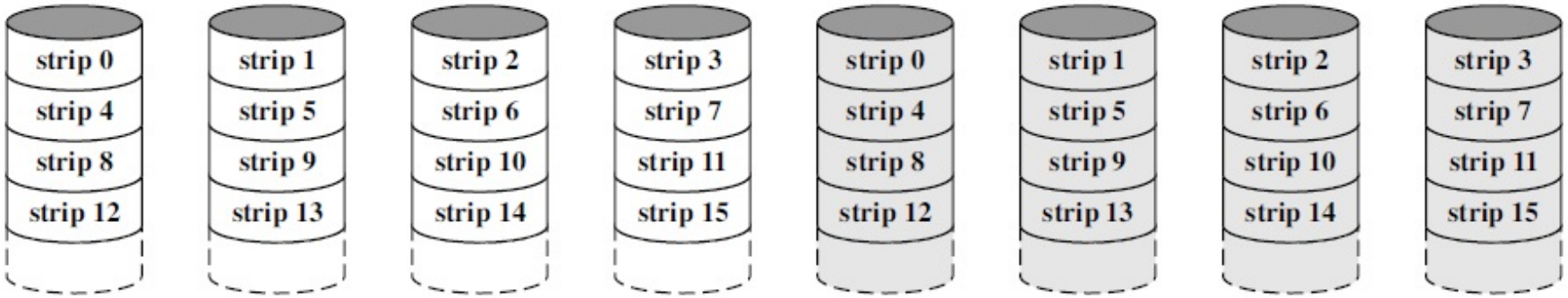


(a) RAID 0 (Nonredundant)

Source: W. Stallings

RAID 1 - Redundância por duplicação de dados

- A redundância consiste em duplicar cada strip de dado em dois discos. Apesar da simplicidade, a desvantagem é o custo.
- Recuperação de erro é simples: quando um disco falha, pega-se o dado no disco que o espelha. Escrita deve ser feita em ambos os discos replicados.

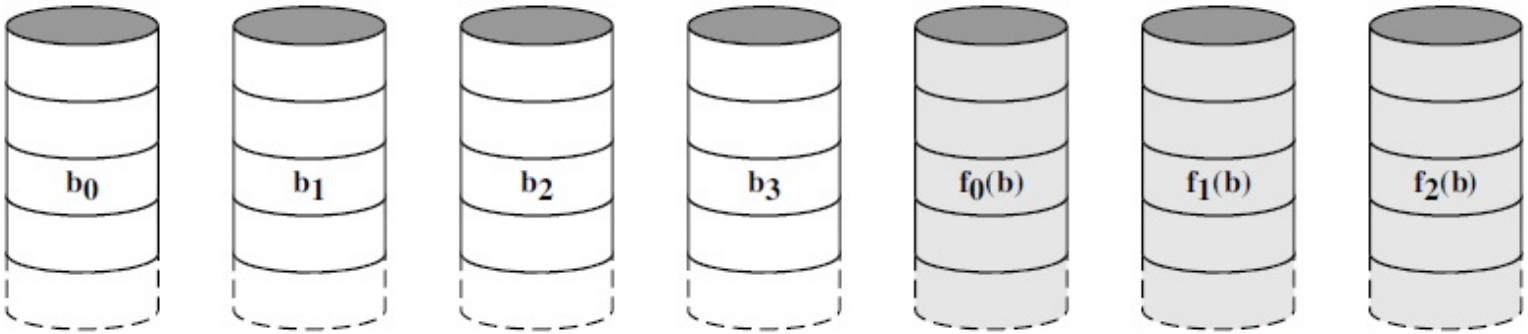


(b) RAID 1 (Mirrored)

Source: W. Stallings

RAID 2 - Redundância usando Hamming code

- Todos os discos posicionam a sua cabeça na mesma posição. Os strips são pequenos (um byte ou uma palavra). Hamming code estendido é usado para correção de erro de 1 bit e deteção de erros de 2 bits.
- RAID 2 requer menos disco que RAID 1. Mas ainda é custoso: o número de discos redundantes é proporcional ao logaritmo do número de discos de dados. É usado quando erros são frequentes. Caso contrário não justifica.

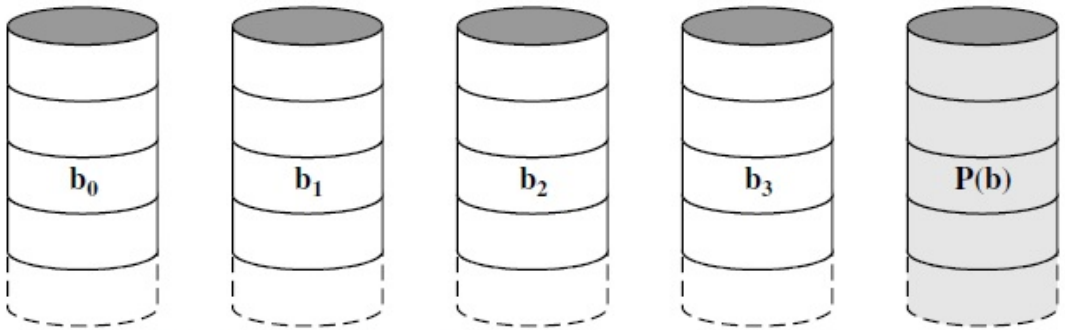


(c) RAID 2 (Redundancy through Hamming code)

Source: W. Stallings

RAID 3 - Redundância usando bit de paridade

- Todos os discos posicionam a sua cabeça na mesma posição. O strip é pequeno, no nível de byte. Usa apenas um disco redundante, contendo o bit paridade dos bits correspondentes dos discos de dados.
- Se um disco da dado falhar, ele pode ser substituído por um novo disco cujo conteúdo é facilmenete calculado como o *ou-exclusivo* de todos os bits dos discos de dados e o disco redundante. (Vale para RAID 3 até 6.)

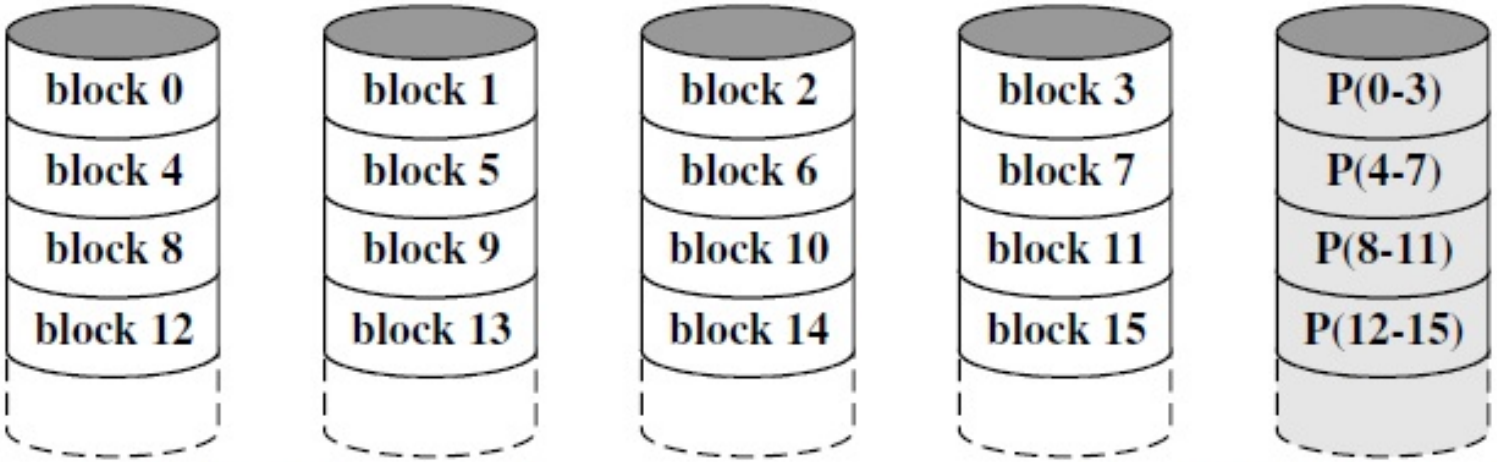


(d) RAID 3 (Bit-interleaved parity)

Source: W. Stallings

RAID 4 - Paridade em nível de bloco

- Em RAID 4, 5 e 6 os discos operam de modo independente. Pedidos de acessos a dados podem ser atendidos em paralelo. Strips são blocos grandes. Um disco redundante contém bits paridades dos bits de blocos correpondentes.
- Ao escrever um bit em um dos discos de dados, o bit paridade precisa ser atualizado. Isso não precisa envolver dados de todos os discos. O novo bit de paridade é igual ao anterior caso o bit escrito é igual ao bit antigo. Caso contrário, é o complemento da paridade antiga.

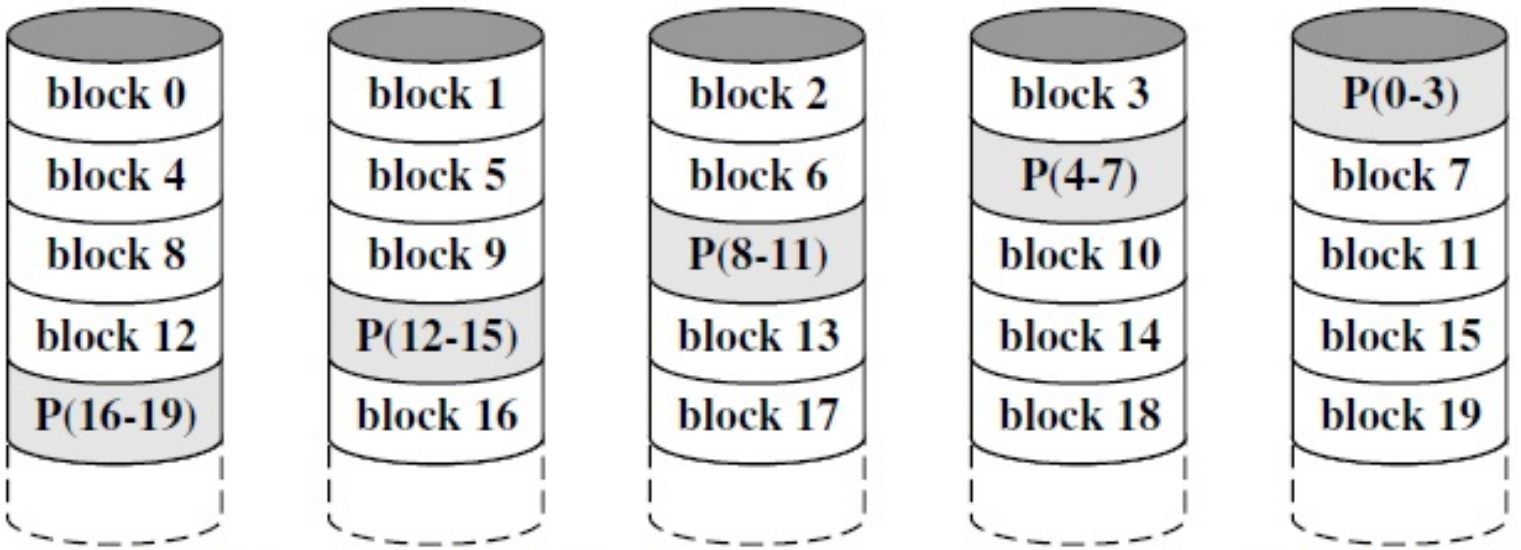


(e) RAID 4 (Block-level parity)

Source: W. Stallings

RAID 5 - Paridade em nível de bloco distribuído

- Em RAID 4, toda escrita envolve o disco redundabte de paridade. Esse disco pode se tornar gargalo.
- Em RAID 5, os blocos paridade não estão concentrados em um único disco, mas distribuídos entre os discos de dados, e.g. em forma de *round robin* ou *rodízio*.

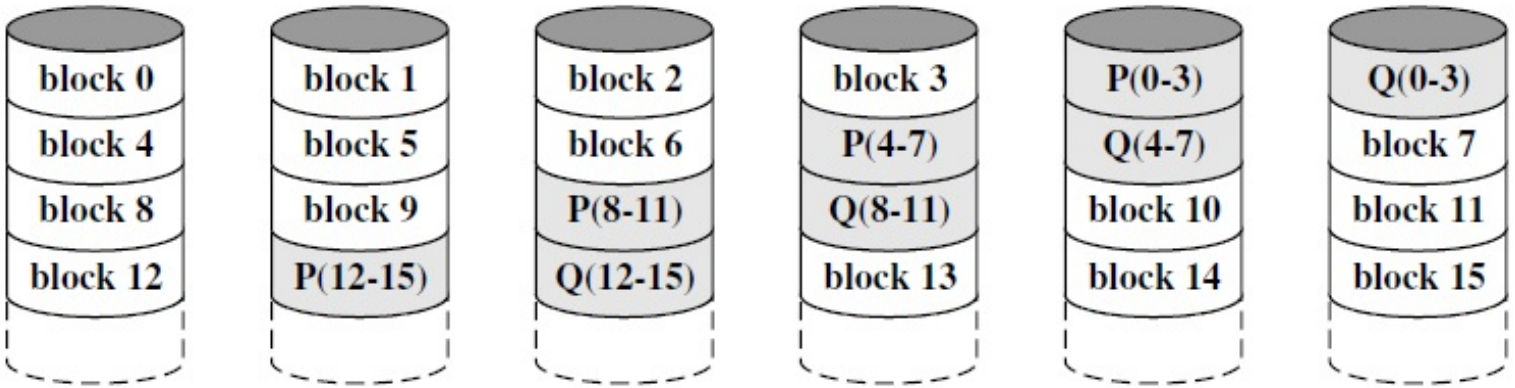


(f) RAID 5 (Block-level distributed parity)

Source: W. Stallings

RAID 6 - Redundância dual

- Usando paridade, se um disco falhar, já vimos como solucionar. O problema é quando dois discos falharem. RAID 6 usa redundância dual com dois cálculos diferentes para verificação. Um é o tradicional bit paridade calculado por *ou-exclusivo*. O outro usa outro cálculo independente (e.g. Reed-Solomon).
- Em RAID 6, a falha de dois discos pode ser recuperado. Só com a falha de três discos ou mais é que dados são perdidos.



(g) RAID 6 (Dual redundancy)

Source: W. Stallings

Como foi o meu aprendizado?

- Vamos fazer uma pequena brincadeira: previsão do futuro do disco: Se o disco magnético HD vai ser substituído por SSD (*Solid State Drive*).
- Cada aluno(a), procura descobrir uma vantagem/desvantage de um dos dois tipos. Vamos enumerar em classe todas essas vantagens/desvantagens. (Isso poderá gerar um bom material didático.)
- No final da discussão, vamos ver se chegamos a uma conclusão:
  - Se SSD vai derrubar completamente HD.
  - Caso positivo, em que ano isso irá ocorrer.
- Participação voluntária, quem não quiser, pode só assistir. Mas é mais divertido tomar algum partido nessa briga.

Disco rígido versus disco de estado sólido

- “Disco” em estado sólido ou *SSD - Solid State Drive* usa a tecnologia de memória flash para servir de memória externa.
- SSD é mais rápido que HD (*Hard Drive*), e também mais caro em termos de dólar por Gigabyte.
- HD funciona melhor quando arquivos grandes ocupam blocos contíguos do disco. Com o tempo de uso, pode ser necessário alocar arquivos grandes em blocos não contíguos espalhados ao longo do disco e fica fragmentado. SSD não apresenta esse problema.
- SSD não apresenta partes móveis e não está vulnerável a vibrações como o HD.
- Com preços mais acessíveis e capacidades cada vez maiores, SSD está se tornando um competidor sério do HD. Resta ver como será o futuro do HD.
- Para complicar a equação, não podemos também deixar de considerar também armazenamento na nuvem.

Disco rígido versus disco de estado sólido

- Em setembro de 2005, ao lançar a 16 GBytes NAND flash memory, o dono da Samsung prevê o fim do disco rígido.

  `http://www.techworld.com/storage/news/index.cfm?NewsID=4387&inkc=0`

  Samsung boss predicts death of hard drives.
- Confirmação preliminar pela notícia de 15/03/2007: "Memória flash começa a substituir HDs e promete deixar PC mais rápido."

  `http://tecnologia.uol.com.br/especiais/cebit/2007/ultnot/2007/03/15/ult4473u17.jhtm`

  SanDisk lança SSD (solid state drive) de 32GB, 100 vezes mais rápido que o HD.

Avanço do SSD

- Em 2009, Kingston lançou um flash drive (Kingston DataTraveler 300) de 256GB.
- Em 2013, Kingston anunciou o lançamento de DataTraveler HyperX Predator (USB 3.0) de 1 TB.
- (Em 2015 voce pode comprar esse drive pela Amazon por US$ 772,74 :-)

Dimensão: $2,8 \times 1,1 \times 0,8$ polegadas.

Avanço do SSD

Em agosto de 2015, na Flash Memory Summit, Samsung anunciou o SSD (solid state drive) de 16 Tbytes, chamado PM1633a.

Samsung mostrou um servidor com 48 desses drives, totalizando 758 Tbytes.

`http://www.dpreview.com/articles/5938341907/samsung-introduces-pm1633a-world-first-2-5-16tb-ssd`

Em agosto de 2016, na Flash Memory Summit, Seagate anunciou o lançamento de um SSD de 60 Tbytes.

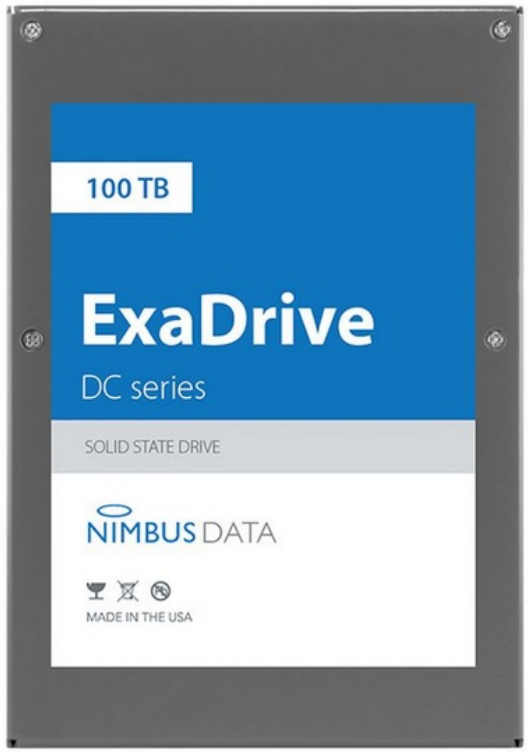
Source: Seagate, Flash Memory Summit

`https://arstechnica.com/gadgets/2016/08/`

Avanço do SSD

Em março de 2018, foi anunicado um SSD de 100 TB da Nimbus Data: ExaDrive DC 100, com garantia de cinco anos.

`https://www.theverge.com/circuitbreaker/2018/3/19/17140332/worlds-largest-ssd-nimbus-data-exadrive-dc100-100tb`



Source: Nimbus Data

SSDs existentes tipicamente custam US$ 500 por TB.

Quanto será que vai custar esse SSD de 100 TB? Ainda não se sabe.

SSD x HD

Comparação entre SSD e HD. Vamos montar uma tabelinha comparativa.
Vamos discutir os seguintes itens na próxima aula. Procurem levantar alguns dessees dados sobre SSD e HD.

- Velocidade de acesso.
- Maior capacidade.
- Preço por TB.
- Tolerância a falhas.
- Durabilidade.
- Que mais?