

Memória externa

MAC0344 - Arquitetura de Computadores
Prof. Siang Wun Song

Slides usados: <https://www.ime.usp.br/~song/mac344/slides06-disks.pdf>

Baseado parcialmente em W. Stallings
Computer Organization and Architecture

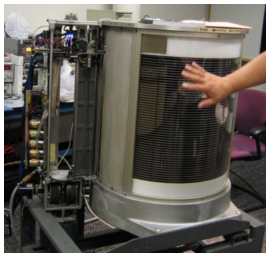
- Veremos memória externa.
- Ao final dessas aulas, vocês saberão
 - O funcionamento de disco magnético.
 - O acesso a dados de disco é substancialmente mais demorado que o acesso à memória.
 - O que vem a ser RAID (Redundant Array of Independent Disks).
 - SSD (Solid State Drive) versus HD. Vantagem e desvantagem de cada um.

Disco da IBM em 1956

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

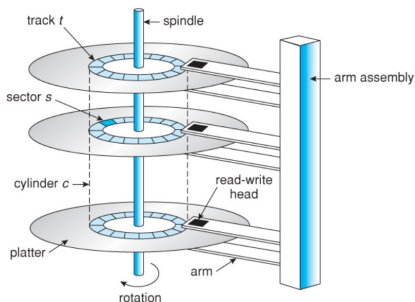
- Pesava uma tonelada
- Era alugado por US\$ 250.000,00 por ano (Era comum locação de computadores de grande porte e discos.)
- Tinha capacidade de 5 Megabytes

Source: Computer History Museum



Disco magnético

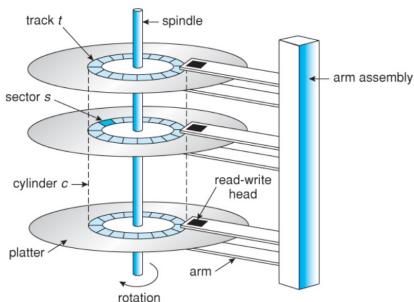
- O disco magnético consiste de fatias circulares de substrato formado de alumínio ou de vidro coberto por uma camada magnética.
- O disco é dividido em **trilhas** que, por sua vez, é organizada em **setores**. Cada setor contém tipicamente 512 bytes da dado mais alguns de controle.



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

Disco magnético

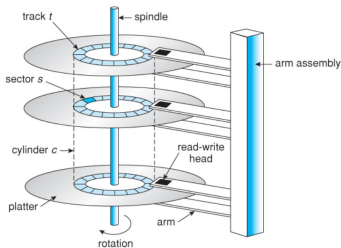
- As cabeças de leitura/gravação podem ser do tipo **móvel** (ver figura): primeiro a cabeça é posicionada em cima da trilha desejada antes de proceder o acesso.
- Discos mais modernos possuem **cabeças fixas**: uma cabeça em cima de cada trilha, dispensando a movimentação das mesmas.



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

Parâmetros de desempenho do disco magnético

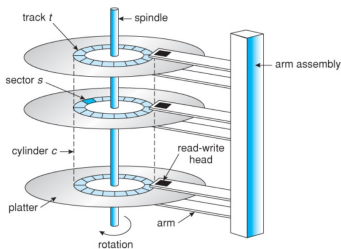
- Para acessar dados em um disco de cabeça móvel, primeiro posicionamos a cabeça na trilha desejada.
- Esse tempo é denominado *seek time*. O valor típico do seek time é de 3 a 12 ms.
- Posicionada a cabeça na trilha desejada, é necessário ainda esperar que o setor desejado chegue em baixo da cabeça.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



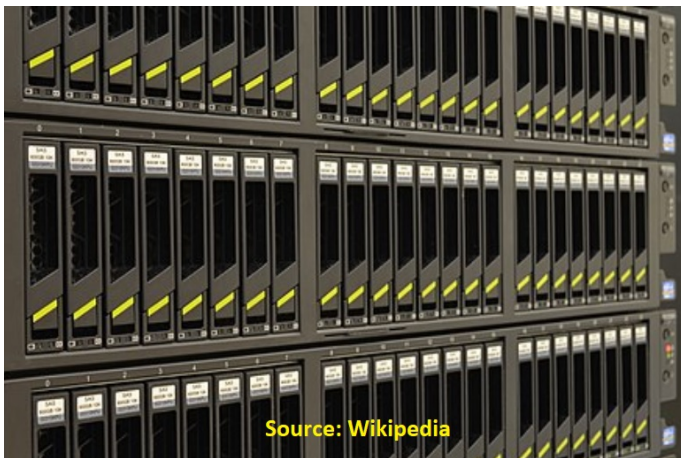
Source: A. Silberschatz, G. Gagne, and P. B. Galvin. Operating System Concepts.

RAID - Redundant Array of Independent Disks

- O acesso a disco magnético leva tipicamente de 10 ms ou mais.
- Assim, projeto de estruturas de dados que residem em disco deve levar isso em consideração. Exemplo: *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** (RAID) que operam independentemente e em paralelo.
- Com os blocos de um arquivo distribuídos em vários discos, podemos ler (ou escrever) os blocos do arquivo em paralelo.

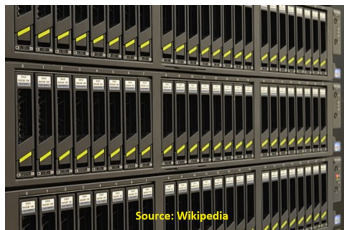
RAID - Redundant Array of Independent Disks

- **RAID** (*Redundant Array of Independent Disks*) é um conjunto de discos físicos visto pelo sistema operacional como uma unidade lógica.



RAID - Redundant Array of Independent Disks

- 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.
- Com redundância de dados e técnicas de detecção ou correção de erros, RAID permite a recuperação de dados em falhas.
- Um artigo em 1988 define as configurações RAID em sete níveis.



RAID - Redundant Array of Independent Disks

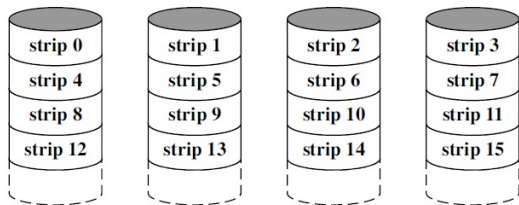
Sete níveis de RAID:

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]}	$1\times$ ^[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]	$(n - 1)\times$ ^[e] <small>[citation needed]</small>
RAID 6	Block-level striping with double distributed parity	4	$1 - \frac{2}{n}$	Two drive failures	$n\times$ ^[c]	$(n - 2)\times$ ^[e] <small>[citation needed]</small>

Source: Wikipedia

RAID 0 - Sem redundância, com strips round robin

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

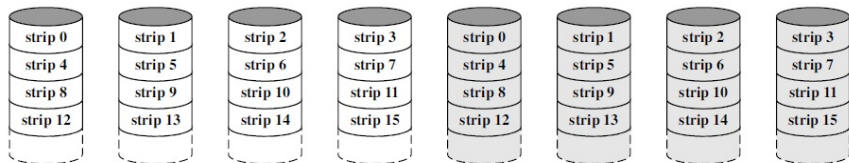


(a) RAID 0 (Nonredundant)

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.

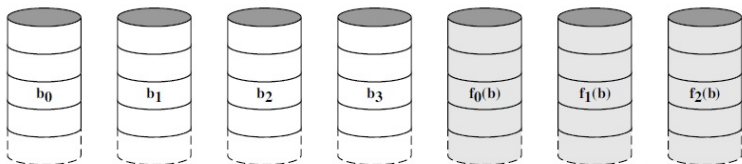


(b) RAID 1 (Mirrored)

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 detecção 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.

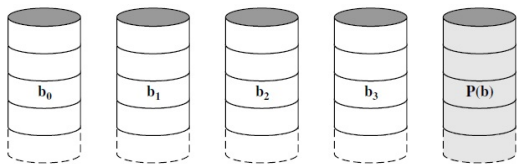


(c) RAID 2 (Redundancy through Hamming code)

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 facilmente calculado como o *ou-exclusivo* de todos os bits dos discos de dados e o disco redundante. (Vale para RAID 3 até 6.)

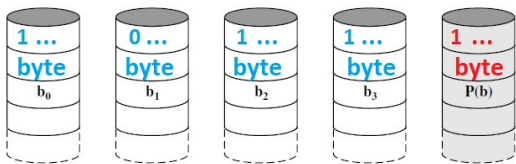


(d) RAID 3 (Bit-interleaved parity)

Source: W. Stallings

RAID 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 facilmente calculado como o *ou-exclusivo* de todos os bits dos discos de dados e o disco redundante. (Vale para RAID 3 até 6.)

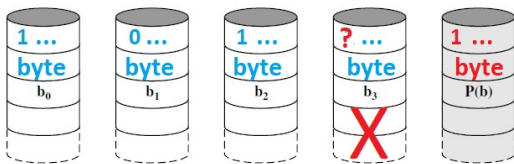


(d) RAID 3 (Bit-interleaved parity)

Source: W. Stallings

RAID 3 - Redundância usando bit de paridade

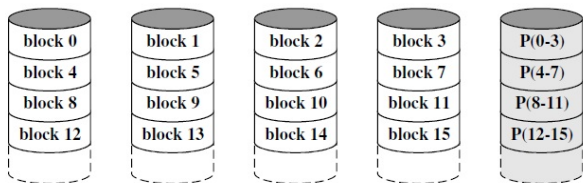
- Todos os discos posicionam a sua cabeça na mesma posição. O strip é pequeno, no nível de **byte**. Usa apenas um disco redundante, contendo o **bit paridade** dos bits correspondentes dos discos de dados.
- Se um disco da dado falhar, ele pode ser substituído por um novo disco cujo conteúdo é facilmente 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) **Fácil recuperar** 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 correspondentes.
- 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.

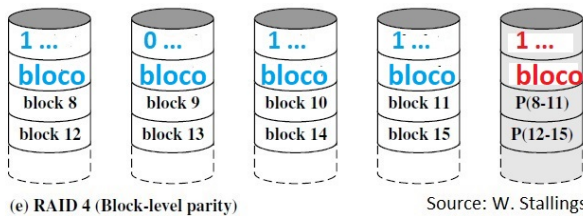


(e) RAID 4 (Block-level parity)

Source: W. Stallings

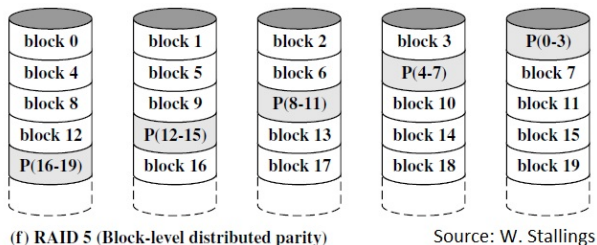
RAID 4 - Paridade em nível de bloco

- 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 correspondentes.
- 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.



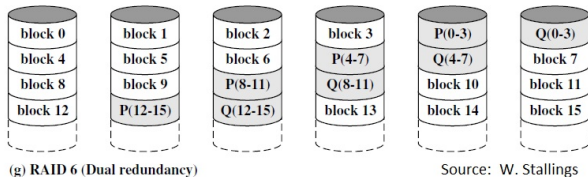
RAID 5 - Paridade em nível de bloco distribuído

- Em RAID 4, toda escrita envolve o disco redundante 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*.
- Segundo [PCMag: RAID Levels Explained](#) RAID 5 é o mais comumente usado.



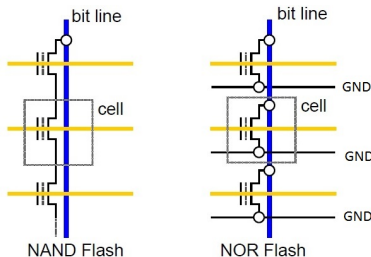
RAID 6 - Redundância dual

- 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.



Memória flash ou *flash memory* - SSD

- Recebe o nome *flash* devido à velocidade com que pode ser alterada: uma memória flash por ser apagada em poucos segundos.
- É possível apagar blocos de memória, mas não no nível de byte.
- Dois tipos: NOR e NAND.
- Flash memory usa um transistor por bit, portanto é bastante densa.
- Solid State Drive ou SSD (“Disco de Estado Sólido”) usa a tecnologia de memória flash.



Memória flash ou *flash memory* - SSD

- Há um limite no número de ciclos de escrita de uma memória flash.
- Esse limite é entre 10.000 a 100.000 para memória flash do tipo NOR e de 100.000 a 1.000.000 para o tipo NAND.
<https://focus.ti.com/pdfs/omap/diskonchipvsnor.pdf>
- Em 2012, usando uma técnica de *auto-cura*, Macronix relata a invenção de uma memória flash que sobrevive 100 milhões de ciclos de escrita.
<https://spectrum.ieee.org/semiconductors/memory/flash-memory-survives-100-million-cycles>
- Memória flash é usada como armazenamento externo (SSD Solid State Drive).

Disco rígido versus Solid State Drive - HD × SSD

- Em setembro de 2005, ao lançar a 16 GBytes NAND flash memory, o CEO da Samsung prevê o fim do disco rígido. [Samsung predicts death of hard drives. Clique aqui para ver a notícia.](#)

Veremos perigos de fazer previsões erradas.

Previsões erradas

- Perigo de fazer previsões erradas:
 - *“I think there is a world market for maybe five computers.”* (Thomas Watson, Presidente da IBM, 1943.)
 - *“Remote shopping, while entirely feasible, will flop ...”* (Time Magazine, 1966.)
 - *“There is no reason anyone would want a computer in their home.”* (Ken Olsen, fundador da DEC, 1977.)
 - *“No one will need more than 637KB for a personal computer. 640KB ought to be enough for anybody.”* (Bill Gates, fundador da Microsoft, 1981.)
 - *“Apple is already dead.”* (Nathan Myhrvold, CTO Microsoft, 1997.)
 - *“Two years from now, spam will be solved.”* (Bill Gates, fundador da Microsoft, 2004.)

Previsões erradas

- Perigo de fazer previsões erradas:
 - *“I think there is a world market for maybe five computers.”* (Thomas Watson, Presidente da IBM, 1943.)
 - *“Remote shopping, while entirely feasible, will flop ...”* (Time Magazine, 1966.)
 - *“There is no reason anyone would want a computer in their home.”* (Ken Olsen, fundador da DEC, 1977.)
 - *“No one will need more than 637KB for a personal computer. 640KB ought to be enough for anybody.”* (Bill Gates, fundador da Microsoft, 1981.)
 - *“Apple is already dead.”* (Nathan Myhrvold, CTO Microsoft, 1997.)
 - *“Two years from now, spam will be solved.”* (Bill Gates, fundador da Microsoft, 2004.)

Previsões erradas

- Perigo de fazer previsões erradas:
 - *“I think there is a world market for maybe five computers.”* (Thomas Watson, Presidente da IBM, 1943.)
 - *“Remote shopping, while entirely feasible, will flop ...”* (Time Magazine, 1966.)
 - *“There is no reason anyone would want a computer in their home.”* (Ken Olsen, fundador da DEC, 1977.)
 - *“No one will need more than 637KB for a personal computer. 640KB ought to be enough for anybody.”* (Bill Gates, fundador da Microsoft, 1981.)
 - *“Apple is already dead.”* (Nathan Myhrvold, CTO Microsoft, 1997.)
 - *“Two years from now, spam will be solved.”* (Bill Gates, fundador da Microsoft, 2004.)

Previsões erradas

- Perigo de fazer previsões erradas:
 - *“I think there is a world market for maybe five computers.”* (Thomas Watson, Presidente da IBM, 1943.)
 - *“Remote shopping, while entirely feasible, will flop ...”* (Time Magazine, 1966.)
 - *“There is no reason anyone would want a computer in their home.”* (Ken Olsen, fundador da DEC, 1977.)
 - *“No one will need more than 637KB for a personal computer. 640KB ought to be enough for anybody.”* (Bill Gates, fundador da Microsoft, 1981.)
 - *“Apple is already dead.”* (Nathan Myhrvold, CTO Microsoft, 1997.)
 - *“Two years from now, spam will be solved.”* (Bill Gates, fundador da Microsoft, 2004.)

Previsões erradas

- Perigo de fazer previsões erradas:
 - *“I think there is a world market for maybe five computers.”* (Thomas Watson, Presidente da IBM, 1943.)
 - *“Remote shopping, while entirely feasible, will flop ...”* (Time Magazine, 1966.)
 - *“There is no reason anyone would want a computer in their home.”* (Ken Olsen, fundador da DEC, 1977.)
 - *“No one will need more than 637KB for a personal computer. 640KB ought to be enough for anybody.”* (Bill Gates, fundador da Microsoft, 1981.)
 - *“Apple is already dead.”* (Nathan Myhrvold, CTO Microsoft, 1997.)
 - *“Two years from now, spam will be solved.”* (Bill Gates, fundador da Microsoft, 2004.)

Previsões erradas

- Perigo de fazer previsões erradas:
 - *“I think there is a world market for maybe five computers.”* (Thomas Watson, Presidente da IBM, 1943.)
 - *“Remote shopping, while entirely feasible, will flop ...”* (Time Magazine, 1966.)
 - *“There is no reason anyone would want a computer in their home.”* (Ken Olsen, fundador da DEC, 1977.)
 - *“No one will need more than 637KB for a personal computer. 640KB ought to be enough for anybody.”* (Bill Gates, fundador da Microsoft, 1981.)
 - *“Apple is already dead.”* (Nathan Myhrvold, CTO Microsoft, 1997.)
 - *“Two years from now, spam will be solved.”* (Bill Gates, fundador da Microsoft, 2004.)

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.



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.

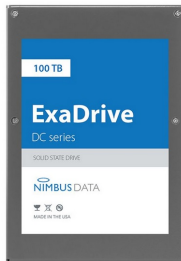
Source: Seagate, Flash Memory Summit



<https://arstechnica.com/gadgets/2016/08/seagate-unveils-60tb-ssd-the-worlds-largest-hard-drive/>

Em março de 2018, foi anunciado 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-exadrives-dc100-100tb>



Source: Nimbus Data

Preço (pesquisado em março de 2021): US\$ 40.000,00

<https://www.techradar.com/best/large-hard-drives-and-ssds>

Disco rígido versus Solid State Drive - HD × SSD

- SSD é mais rápido e ainda mais caro que HD (*Hard Drive*).
- 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.

Comparação SSD x HD

Para uma comparação entre SSD e HD, pode-se montar uma tabelinha comparativa.

	HD	SSD
Velocidade de acesso	mais lenta	mais rápida
Capacidade	maior	menor
Preço por TB	mais barato	mais caro
Fragmentação	sim	não
Problema de vibração	sim	não
Durabilidade	?	?
Que mais?	?	?

Comparação SSD x HD

O seguinte artigo apresenta uma comparação SSD x HD.

[What Is a Solid-State Drive \(SSD\)?](#)

SSD	vs	HDD
faster	✓	✗ slower
longer lifespan	✓	✗ shorter lifespan
more expensive	✗	✓ cheaper
non-mechanical (flash)	✓	✗ mechanical (moving parts)
shock-resistant	✓	✗ fragile
best for storing operating systems, gaming apps, and frequently used files		best for storing extra data, such as movies, photos, and documents

Disco rígido versus Solid State Drive - HD × SSD



- SSD vem evoluindo: aumento da capacidade e diminuição do preço.
- Uma pergunta cuja resposta saberemos no futuro:
 - Se SSD vai derrubar completamente HD.
 - Caso positivo, quando isso irá ocorrer.

Próximo assunto: Arquitetura CISC - Microprogramação

- Próximo assunto: Arquitetura CISC - Microprogramação
- CISC (Complex Instruction Set Computer): Conjunto grande e complexo de instruções de máquina.
- A execução de cada instrução de máquina envolve dezenas ou centenas de microinstruções contidas no microprograma.
- Possibilita um hardware simples que no entanto pode executar instruções complexas.
- Contrapondo a CISC veremos RISC (Reduced Instruction Set Computer)
- Não percam!