



# Unidade 4: Tecnologia, Hierarquia e Tipos de Memórias

## Apresentação

Seja bem-vindo!

Os sistemas computacionais necessitam armazenar os dados em lugares onde fiquem disponíveis para processamento a qualquer momento. Este é o papel das memórias: armazenar dados. Esse armazenamento pode ser temporário ou permanente, e vai depender do objetivo do processamento em um dado momento.

As memórias desempenham um dos papéis mais importantes nos sistemas computacionais. Cada tipo de memória funciona de uma forma diferente e apresenta características específicas.

Nesta Unidade de Aprendizagem, você vai estudar como os diversos tipos de memórias funcionam, as diferenças entre memórias de leitura e escrita e os tipos de memórias existentes.

Bons estudos.

**Ao final desta Unidade de Aprendizagem, você deve apresentar os seguintes aprendizados:**

- Descrever o conceito de memória.
- Analisar as tecnologias e a hierarquia de memória
- Listar as diferenças entre memórias de leitura e escrita, volátil e não-volátil.
- Identificar os tipos de memórias: primária, secundária, cache.

# Infográfico: características das memórias voláteis e não voláteis

---

No infográfico a seguir, você vai visualizar as principais características das memórias voláteis e não voláteis.

Conteúdo interativo disponível na plataforma de ensino!

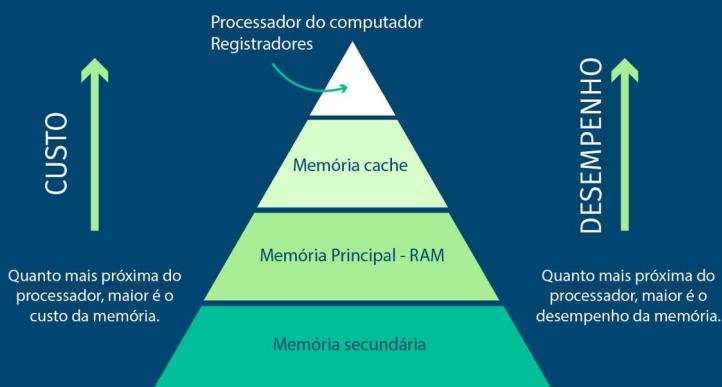
# Infográfico: Tecnologia e Hierarquia de Memórias

---

Hierarquia de memórias é um conceito curioso que, por vezes, passa despercebido quando alguém compra um computador, preocupado com o seu desempenho. O conjunto de memórias formado por registradores, cache, memória principal e memória secundária interferem de forma significativa no desempenho e no custo do computador. Registradores são a menor, mais rápida e mais cara memória, já a memória secundária é a maior, mais lenta e mais barata das memórias.

Neste Infográfico, você vai ver como funciona a hierarquia de memória, a qual está centrada no conceito de que, quanto mais próximo ao processador a memória está, mais rápida é; porém, menor é o seu tamanho.

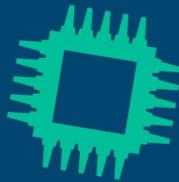
# HIERARQUIA DE MEMÓRIAS



Veja, a seguir, o que é cada uma das memórias apresentadas:

## PROCESSADOR DO COMPUTADOR REGISTRADORES

- ▶ O processador é uma das partes principais do computador, pois é responsável pelo processamento (execução) de todas as instruções executadas por este.
- ▶ O processador tem acoplado a ele os registradores, que são pequenas memórias de acesso muito rápido e custo elevado.
- ▶ Os registradores são utilizados, em geral, para entrada e saída de valores para o processador.
- ▶ Registradores são memórias voláteis.



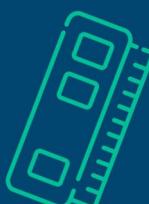
## MEMÓRIA CACHE

- ▶ A memória cache é uma memória volátil e está localizada entre a memória principal e os registradores.
- ▶ Essa memória armazena as informações que foram selecionadas na memória RAM por serem as que têm acesso mais frequente, evitando que o processador precise acessar, a todo momento, a memória RAM.
- ▶ O acesso à cache é mais rápido que o acesso à memória RAM, por ela estar mais perto do processador.
- ▶ O tamanho da cache é menor que o tamanho da memória RAM e maior do que o tamanho dos registradores.



## MEMÓRIA PRINCIPAL — RAM

- ▶ A memória RAM, conhecida como memória principal, é também uma memória volátil e tem um tamanho maior e custo menor quando comparada à memória cache e aos registradores.
- ▶ A função dessa memória é evitar que o computador precise acessar repetidas vezes a memória secundária e, para isso, no primeiro acesso à memória secundária, algumas informações que costumam ser acessada com maior frequência são transferidas para a memória RAM, para depois o processador acessar diretamente a memória RAM.



## MEMÓRIA SECUNDÁRIA

- ▶ A memória secundária é a memória de armazenamento definitivo do computador, na qual estão armazenados dados como fotos, vídeos e softwares instalados. Por isso, tem um tempo de acesso mais alto.
- ▶ As principais memórias secundárias são o Hard Disc (HD) e o Solid-State Drive (SSD).



Aponte a câmera para o código e accese o link do conteúdo ou clique no código para accesar.

# **Conteúdo do Livro: Tipos de memórias**

---

No capítulo Tipos de memórias, do livro *Fundamentos computacionais*, que serve de base teórica para esta Unidade de Aprendizagem, você aprofundará seus conhecimentos sobre os principais tipos de memórias que armazenam os dados de um sistema computacional. Você verá que esse armazenamento pode ser temporário ou permanente, dependendo do objetivo do processamento em um dado momento.

Boa leitura.

# FUNDAMENTOS COMPUTACIONAIS

Sidney Cerqueira  
Bispo dos Santos

**Revisão técnica:**

**Izabelly Soares de Moraes**

*Licenciada em Ciência da Computação  
Mestre em Ciência da Computação*



C796f Córdova Junior, Ramiro Sebastião.

Fundamentos computacionais [recurso eletrônico] /  
Ramiro Sebastião Córdova Junior, Sidney Cerqueira Bispo dos  
Santos, Pedro Kislansky; [revisão técnica: Izabelly Soares de  
Moraes]. – Porto Alegre: SAGAH, 2018.

ISBN 978-85-9502-394-9

1. Computação. 2. Tecnologia da informação. I. Santos,  
Sidney Cerqueira Bispo dos. II. Kislansky, Pedro. III. Título.

CDU 004

# Tipos de memórias

## Objetivos de aprendizagem

Ao final deste texto, você deve apresentar os seguintes aprendizados:

- Descrever o conceito de memória.
- Listar as diferenças entre memórias de leituras e escritas.
- Identificar os tipos de memórias.

## Introdução

Os sistemas computacionais precisam armazenar os dados em lugares em que estes possam ficar disponíveis para processamento a qualquer momento. Esse é o papel das memórias: armazenar dados. Esse armazenamento pode ser temporário ou permanente, dependendo do objetivo do processamento em um dado momento.

As memórias desempenham um papel muito importante — talvez um dos mais importantes nos sistemas computacionais. Cada tipo de memória funciona de forma diferente e apresenta diferentes características.

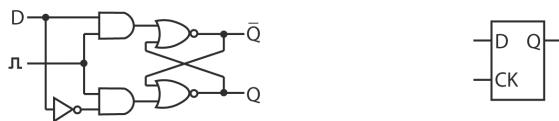
Neste capítulo, você vai compreender como os diversos tipos de memórias funcionam, bem como estudar as diferenças entre memórias de leituras e escritas e quais são os tipos de memórias existentes.

## Conceito de memória

Um sistema computacional precisa ter à sua disposição informações que ele possa utilizar para realizar os diversos processamentos que permitem a ele cumprir a sua finalidade. Essas informações (programas e dados) são guardadas nas memórias do sistema que, dependendo do nível do processamento e da sua função, são chamadas de registradores, *buffers*, memórias RAM, ROM, cache, principal, etc.

As memórias são componentes essenciais em um computador. Sem elas, os computadores como os conhecemos hoje não existiriam.

Vamos examinar o seu funcionamento, começando com as formas de armazenar um bit. Uma forma de armazenar um bit é utilizando um Latch D, composto de uma configuração especial de portas *NOT*, *AND* e *NOR*, como mostrado na Figura 1.



**Figura 1.** Latch D e representação do Latch D, respectivamente.

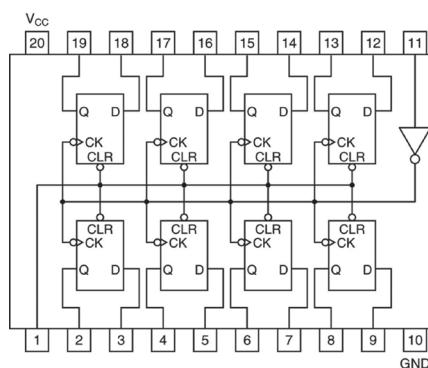
Um flip flop D nada mais é do que um Latch D para as transições que ocorrem comandadas pelas bordas dos pulsos do relógio (*clock*). Em um Latch, as transições ocorrem comandadas pelos níveis.



### Saiba mais

Relógio (*clock*) é um tipo de circuito que gera uma série de pulsos retangulares em determinada frequência.

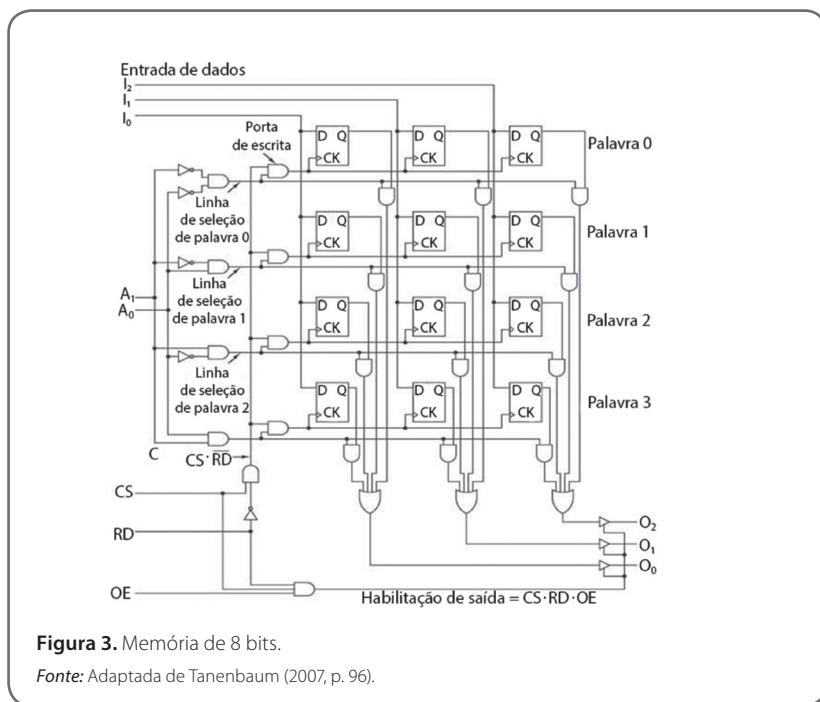
A Figura 2 apresenta uma configuração para uma memória de 8 bits, utilizando flip flops D.



**Figura 2.** Memória de 8 bits.

O arranjo mostrado na Figura 2 pode ser utilizado como um registrador de 8 bits. Se dois deles forem utilizados em paralelo, com um comando único para o pino 1 e outro para o pino 11, consegue-se um registrador de 16 bits.

Entretanto, para construir memórias muito grandes, é necessário outro arranjo. A Figura 3 mostra uma organização bastante utilizada, com quatro palavras de 3 bits.

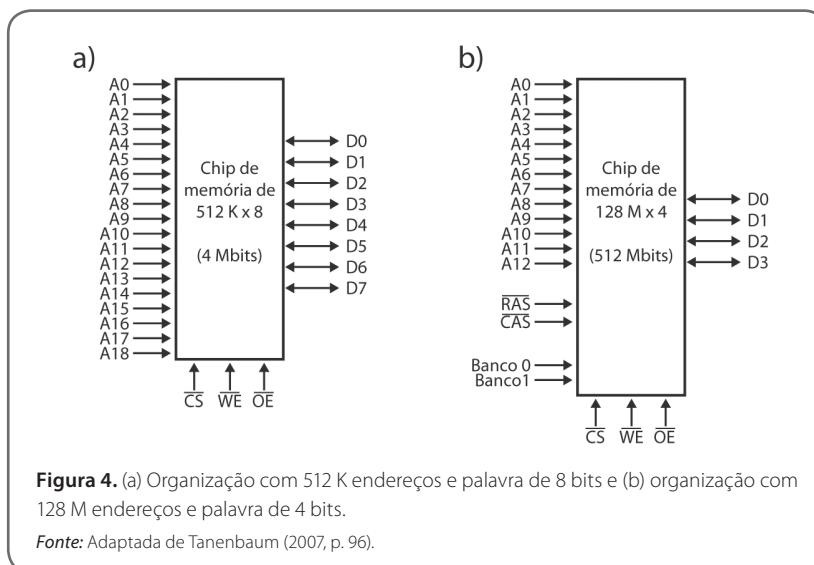


**Figura 3.** Memória de 8 bits.

*Fonte:* Adaptada de Tanenbaum (2007, p. 96).

Essa memória tem oito entradas e três saídas.  $I_0$ ,  $I_1$  e  $I_2$  são as entradas de dados,  $A_0$  e  $A_1$  são os endereços, CS (*Chip Select*) é onde se faz a seleção do chip de memória, RD serve para selecionar escrita ou leitura e OE (*Output Enable*), para habilitar as saídas  $O_0$ ,  $O_1$  e  $O_2$ .

A organização mostrada na Figura 3 é muito flexível e pode ser expandida para qualquer número de palavras de potência de 2. As Figuras 4 mostram duas organizações de memórias, uma com 4 Mbits e outra com 512 Mbits.



**Figura 4.** (a) Organização com 512 K endereços e palavra de 8 bits e (b) organização com 128 M endereços e palavra de 4 bits.

**Fonte:** Adaptada de Tanenbaum (2007, p. 96).

## Diferenças entre memórias de leitura e escrita

As memórias apresentadas anteriormente podem ser lidas e escritas. Esse tipo de memória é conhecido como memórias RAM (*Random Access Memories*), ou seja, memórias de acesso aleatório. Esse nome deve ser interpretado com cuidado, uma vez que todas as memórias que você verá têm acesso aleatório.

As memórias RAM podem ser de dois tipos: estáticas ou dinâmicas. As **memórias estáticas**, chamadas de SRAM (*Static Random Access Memories*), têm a construção muito parecida com as memórias apresentadas anteriormente com flip flops D. Essas memórias são extremamente rápidas e conseguem manter os dados enquanto houver energia fornecida. Seu principal emprego é como memória cache de segundo nível.



### Saiba mais

Memória cache é um tipo de memória que armazena temporariamente as instruções e os dados que são utilizados mais frequentemente pelo processador. Antes de buscar uma instrução ou um dado na memória RAM, o processador acessa primeiro a memória cache, para verificar se a instrução ou o dado encontra-se armazenado nela.

As memórias dinâmicas DRAM (*Dynamic Random Access Memories*) não utilizam os arranjos de flip flops apresentados anteriormente, e sim um conjunto de células que consistem em um capacitor e um transistor. O armazenamento se dá pela carga ou descarga do capacitor. Esse tipo de memória tem o inconveniente de que, de tempos em tempos (da ordem de milissegundos), os bits devem ser recarregados, devido a vazamentos da carga elétrica do capacitor. Como o controle dessa recarga é feito externamente à memória, o seu circuito de controle é mais complexo que o das memórias estáticas. Sua vantagem está na capacidade de maior quantidade de memória.

As DRAMs, por serem mais simples que as SRAMs (apenas um capacitor e um transistor), podem conter maior quantidade de bits por chip; por isso, são muito utilizadas como memórias principais dos computadores. Entretanto, o grande número de bits nesses chips de memórias as torna mais lentas. Em função disso, os computadores utilizam as memórias DRAM como memórias principais (da ordem de gigabytes), e vários níveis de cache com memórias mais rápidas, explorando assim as melhores características de cada uma delas.

As memórias apresentadas anteriormente são chamadas de **memórias voláteis**, porque, se não forem energizadas, as informações se apagam. Quando o computador é desligado, toda a informação armazenada desaparece.

Existe outro tipo de memória chamado de memória **não volátil**, na qual, após o carregamento da informação, esta permanece por bastante tempo, mesmo que a energia seja desligada — normalmente, ela não pode ter outro ciclo de escrita. Essas memórias são chamadas de memórias somente de leitura, ou ROMs (*Read Only Memories*). Elas são muito úteis em brinquedos, computadores, eletrodomésticos, carros e outros produtos nos quais a programação e os dados já devem vir carregados de fábrica e não precisam ser alterados.

Os bits nas ROMs são gravados na fábrica e não podem ser alterados. Portanto, elas precisam ser produzidas sob medida. Uma empresa, após fazer o projeto de um produto, precisa fabricar as ROMs necessárias ou encomendar os chips de um fabricante, de acordo com a sua necessidade. Isso demanda tempo e exige que a quantidade a ser encomendada seja suficiente para cobrir os custos de produção.

Essas desvantagens fizeram com que fosse desenvolvida uma ROM que fosse programável, ou seja, que permitisse que as empresas pudesse gravar essa memória. Surgiu assim a PROM (*Programmable Read Only Memory*). Esse tipo de memória é constituído de pequenos fusíveis, que podem ser queimados aplicando-se determinada tensão em um pino especial e indicando a linha e a coluna do fusível a ser queimada.

A evolução seguinte foi o desenvolvimento da EPROM (*Erasable Programmable Read Only Memory*), que pode ter os seus dados apagados ou

programados por meio da sua exposição a uma luz ultravioleta. As EPROMs podem ser reutilizadas; assim, são muito úteis em projetos que precisam de várias alterações durante o seu desenvolvimento.

A EEPROM (*Electrically-Erasable Programmable Read-Only Memory*) é mais versátil que a EPROM, porque, em vez de luz ultravioleta, utiliza pulsos de tensão para apagar o seu conteúdo e pode ser reprogramada. Entretanto, ela não admite grandes capacidades, é muita mais lenta que a EPROM, muito mais lenta ainda que a DRAM e SRAM, além de ser muito cara.

Um tipo especial de EEPROM são as atuais memórias flash, que podem ser apagadas e reescritas quando conectadas a um dispositivo.



### Link

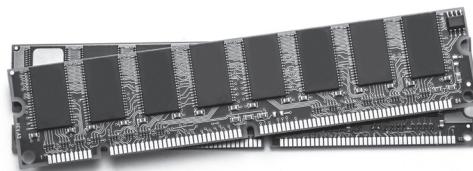
O link a seguir traz um vídeo sobre uma dúvida bastante frequente: é possível usar tipos de memórias diferentes em uma mesma placa mãe, ao mesmo tempo?

<https://goo.gl/DrSQL5>



## Identificando os tipos de memórias

As memórias RAM são vendidas no mercado em forma de pentes, conforme mostra a Figura 5, e dispostas de mais de uma maneira, cada uma com terminologia própria.



**Figura 5.** Pente de memória.

Fonte: Garsya/Shutterstock.com.

A primeira terminologia a aprender é memória DDR (*Double Data Rate*), ou memória de taxa dupla de (transferência de) dados: ela permite a transferência de dois dados ao mesmo tempo (no mesmo ciclo de relógio). A memória DDR é uma memória SDRAM (*Synchronous Dynamics Random Access Memory*), ou seja, memória RAM dinâmica de acesso síncrono. Em outras palavras, ela é uma DRAM que atualiza os dados (escrita/leitura) de forma síncrona.

A ideia da sincronia é utilizar um relógio para controlar a leitura ou gravação dos dados em ciclos. Isso exige um tempo mínimo para acessar um endereço e garantir a gravação ou leitura correta dos dados, pois a SDRAM precisa aguardar uns poucos nanosegundos antes de efetuar uma operação de leitura/escrita.

Quem controla esse retardo é a CPU. Como ela normalmente tem uma frequência de trabalho muito maior que a memória, a consequência é que ela precisará aguardar algum tempo antes de obter os dados que solicitou.

Outro detalhe que merece atenção são as siglas SIMM (*Single Inline Memory Module*), módulo de memórias em linha simples, e DIMM (*Dual Inline Memory Module*), módulo de memória em linha dupla. As SIMM possuem somente uma linha de memórias e trabalham com palavras de 32 bits, enquanto DIMM significa que os chips de memória são instalados em linha dupla e trabalham com palavras de 64 bits.

## Memórias DDR-DIMM

Encontrada em módulos com linha dupla de memórias SDRAM, são DDRs, SDRAMs e DIMMs. Devido às características do DDR, em tese, o módulo deve ser capaz de transferir dados duas vezes mais rápido que uma SDRAM comum. As memórias DDR-DIMM — ou apenas DDR — possuem 184 pinos.

Com o aumento do poder de processamento dos microprocessadores, as memórias também tiveram de acompanhar essa evolução e, assim, começaram a surgir módulos de memória com velocidades e capacidades cada vez maiores. Dessa forma, o próximo módulo a surgir foi a DDR2.

## Memórias DDR2

As DDR2 foram lançadas trabalhando com o dobro da frequência das DDR, com 240 pinos e capacidade de transferir o dobro de dados, ou seja, quatro por ciclo do relógio. Melhoraram o consumo de energia e a sensibilidade à interferência eletromagnética, mas a latência aumentou. A próxima evolução foi a DDR3.

## Memórias DDR3

As DDR3 podem trabalhar com relógios de até 2,8 GHz, com taxas de transferência um pouco inferiores ao dobro das taxas conseguidas pelas DDR2. A latência também aumentou, em relação à DDR2, e foram lançadas com 204 pinos. Existem várias versões dos módulos DDR3; o Quadro 1 apresenta um resumo dessas versões.

**Quadro 1.** Resumo das versões dos módulos DDR3.

Memória	Frequência real	Frequência DDR	Taxa de transferência máxima
DDR3-800	400 MHz	800 MHz	6,4 GB/s
DDR3-1066	533 MHz	1,066 GHz	8,533 GB/s
DDR3-1333	666 MHz	1,333 GHz	10,666 GB/s
DDR3-1600	800 MHz	1,6 GHz	12,8 GB/s
DDR3-1866	933 MHz	1,866 GHz	14,933 GB/s
DDR3-2133	1,066 GHz	2,133 GHz	17,066 GB/s
DDR3-2400	1,2 GHz	2,4 GHz	19,2 GB/s
DDR3-2600	1,3 GHz	2,6 GHz	20,8 GB/s
DDR3-2800	1,4 GHz	2,8 GHz	22,4 GB/s

## Memórias DDR4

A DDR4 oferece melhor desempenho (2 Gbps por pino) e até 50% de aumento de desempenho em relação à DDR3, maiores capacidades DIMM, maior integridade de dados (CRC) e menor consumo de energia (40%).

A Samsung já possui módulos DDR4 de alto desempenho, com 3,2 GB/s de taxa por pino. Em dezembro de 2017, a empresa anunciou o início da produção de DDR4 de 2ª geração, com 3,6 GB/s por pino e eficiência energética até 15% maior.

Atualmente, fabricantes já anunciam que estão em processo de desenvolvimento acelerado da nova geração de memórias, chamadas de DDR5, que deverá duplicar as velocidades de funcionamento em relação às memórias atuais.



## Referência

TANENBAUM, A. S. *Organização estruturada de computadores*. 5. ed. São Paulo: Pearson, 2007.

### Leituras recomendadas

MONTEIRO, M. A. *Introdução à organização de computadores*. 5. ed. Rio de Janeiro: LTC, 2007.

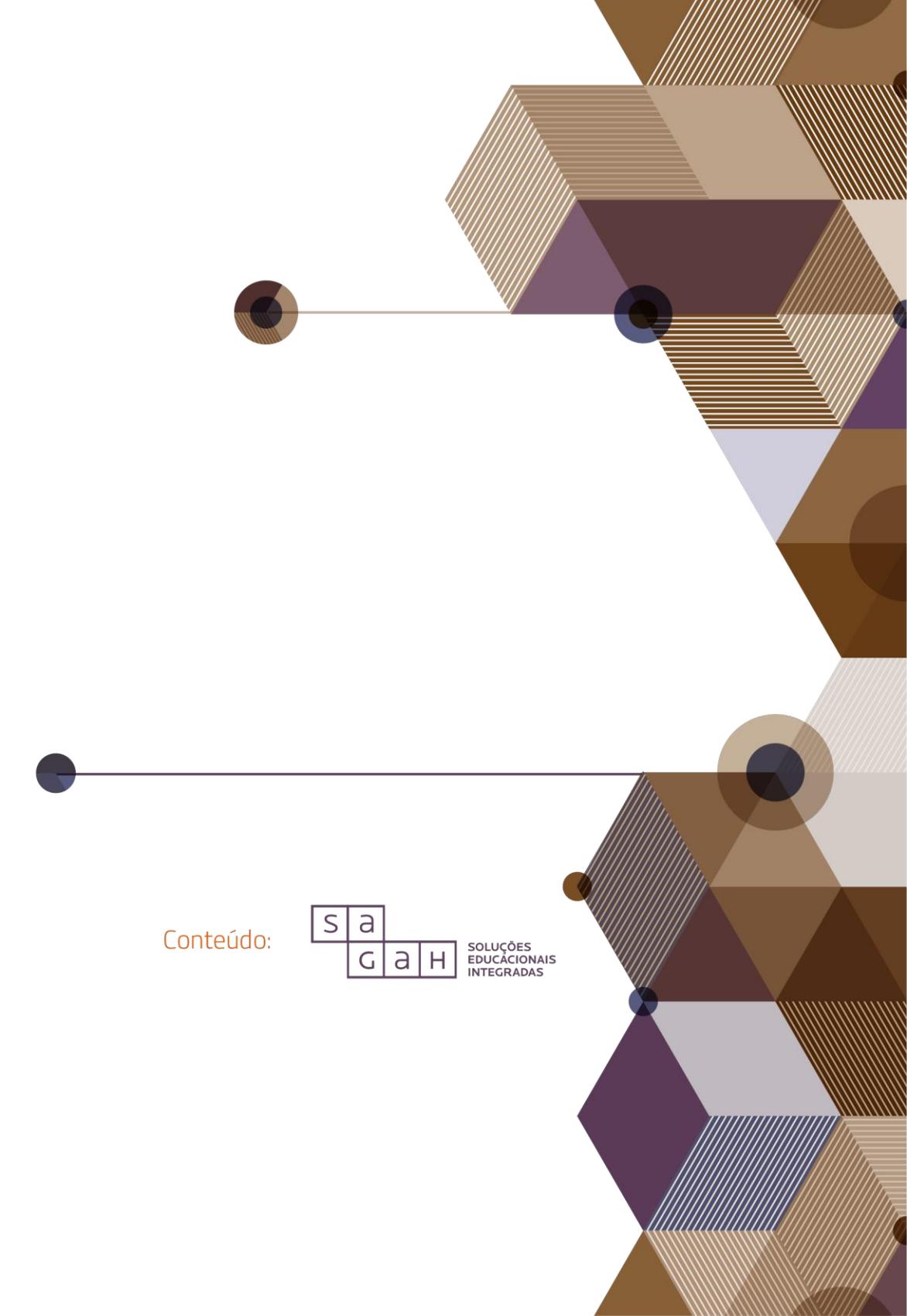
STALLINGS, W. *Arquitetura e organização de computadores*. 8. ed. São Paulo: Pearson, 2010.

TANENBAUM, A. S. *Sistemas operacionais*. 3. ed. São Paulo: Pearson, 2010.

TOCCI, R. J. et al. *Sistemas digitais*. 11. ed. São Paulo: Pearson, 2015.

WEBER, R. F. *Fundamentos de arquitetura de computadores*. 4. ed. Porto Alegre: Bookman, 2012.

Encerra aqui o trecho do livro disponibilizado para esta Unidade de Aprendizagem. Na Biblioteca Virtual da Instituição, você encontra a obra na íntegra.



A large, abstract graphic in the background consists of several overlapping hexagons in various shades of brown, tan, purple, and blue. Small black circles of different sizes are scattered across the white space, some connected by thin black lines.

Conteúdo:



SOLUÇÕES  
EDUCACIONAIS  
INTEGRADAS



# Conteúdo do Livro: Tecnologia e Hierarquia de Memórias

---

A memória do computador se divide hierarquicamente; essa hierarquia precisa ser compreendida e utilizada da melhor maneira possível, uma vez que cada tipo de memória tem uma *performance* e um custo diferente. Hierarquicamente, as memórias são organizadas conforme a sua proximidade com o processador, nesta ordem: registradores, cache, memória principal e memória secundária. Além dessas memórias, existe a memória virtual, a qual utiliza o disco rígido (HD) como uma extensão da memória principal.

No capítulo Tecnologia e hierarquia de memórias, da obra *Arquitetura e organização de computadores*, base teórica desta Unidade de Aprendizagem, você vai ter um maior detalhamento sobre todos esses tipos de memória, além de conhecer também o conceito de localidade de memória.

Boa leitura.

# ARQUITETURA E ORGANIZAÇÃO DE COMPUTADORES

The background of the book cover features a photograph of four people in an office environment. A man with curly hair and a woman with long dark hair are in the foreground, looking at a tablet device. Behind them, another man and a woman are also focused on the screen. The desk is cluttered with papers, notebooks, and a coffee cup. The overall atmosphere is one of collaboration and technology.

Aline Zanin

# Tecnologia e hierarquia de memórias

## Objetivos de aprendizagem

Ao final deste texto, você deve apresentar os seguintes aprendizados:

- Definir tecnologias e hierarquia de memória.
- Discutir o conceito de localidade de memórias.
- Definir memória cache e memória virtual.

## Introdução

Um dos elementos mais conhecidos e comentados por usuários de computadores, especialmente os leigos, é a memória. Memória nada mais é do que um conjunto de armazenadores de informações. Quando se fala em memória no contexto de computadores, pode-se comparar com a memória do ser humano. O ser humano, assim como o computador, possui memórias recentes e memórias longínquas, memórias que são rapidamente esquecidas e memórias que ficam armazenadas por muito tempo para serem lembradas ou utilizadas. No computador, toda a memória é formada por palavras, e cada palavra possui um endereço que é único e que a identifica dentro do computador. Essas palavras podem representar dados (informações salvas, por exemplo, uma foto) ou instruções (uso do sistema para, por exemplo, abrir um programa), conforme Weber (2012).

Diversas são as tecnologias utilizadas em memórias de computadores. Contudo, elas se dividem em apenas dois grandes grupos: memória RAM (*random access memory*) e memória ROM (*read only memory*). A memória RAM é chamada de memória principal do computador e é uma memória volátil, enquanto a memória ROM é a memória pequena, que não é volátil e armazena dados necessários, por exemplo, para inicializar o sistema operacional.

Dado que um computador pode possuir diversas tecnologias para memória, e as tecnologias se diferem entre si, especialmente no que

diz respeito ao desempenho, os sistemas de computação utilizam uma combinação de tipos de memória para melhorar o custo e manter o desempenho do computador. Essa abordagem é chamada de memória hierárquica, conforme Null e Lobur (2011).

Neste capítulo, você vai estudar as tecnologias e a hierarquia da memória, verificando o conceito de localidade de memórias e distinguindo memória cache e memória virtual.

## Conceitos básicos

Diversas são as tecnologias de memória que podem ser utilizadas em um computador. Elas se dividem em dois grandes grupos: memória RAM e memórias ROM. A memória RAM, também chamada de memória principal ou memória primária, é utilizada para leitura rápida e tem armazenamento volátil — os dados armazenados são perdidos quando o computador é desligado. Ela interfere de forma direta no desempenho dos computadores. É devido a essa memória, por exemplo, que, quando você abre o mesmo programa pela segunda vez, sem ter desligado o computador entre as duas inicializações do programa, ele será aberto de forma mais rápida. Isso acontece porque, na primeira execução, o computador buscou os dados na memória secundária e salvou na memória RAM todas as informações que ele precisava para abrir o programa; então, na segunda execução, pôde buscar diretamente da memória sem acessar o *hard disk* (HD).

Existem duas tecnologias bastante utilizadas de memória: DRAM (RAM dinâmica) e SRAM (RAM estática), conforme destaca Null e Lobur (2011). A memória DRAM ou RAM dinâmica é uma memória que precisa que a informação seja atualizada o tempo todo para que permaneça armazenada. Com isso, esse tipo de RAM gasta mais energia, se comparado com a SRAM. A memória SRAM ou memória estática é uma memória mais rápida e de custo maior, contudo, possui maior eficiência energética, porque consegue manter os *bytes* armazenados nela sem necessidade de atualização constante, sendo perdidos apenas após a interrupção da fonte de energia.

Em geral, as duas tecnologias são combinadas para que se possa aproveitar o melhor que elas podem oferecer. Costuma-se utilizar a DRAM para memória principal e a SRAM para cache. Os diferentes tipos de SRAM incluem SRAM assíncrona, SRAM síncrona e SRAM com *pipeline*, conforme aponta Null e Lobur (2011).

Adicionalmente à memória RAM, os computadores podem possuir uma memória ROM. Essa memória não é volátil, isto é, armazena valores de forma definitiva e é utilizada para armazenar informações fundamentais do sistema operacional. Essa memória é utilizada para o armazenamento da BIOS — *basic input/output system* (sistema básico de entrada/saída). A BIOS é o primeiro programa executado pelo computador ao ser ligado e tem como função principal preparar a máquina para que o sistema operacional possa ser executado. Essa memória também é utilizada em sistemas embarcados que possuem pequenos fragmentos de dados; nesse caso, a memória é utilizada para armazenar os dados quando a energia é desligada.

Existem cinco tipos de memória ROM, conforme Null e Lobur (2011). Além da ROM, que já conhecemos, temos as seguintes:

- PROM, *programmable read-only memory*: é uma adaptação da ROM que permite ser programada pelo usuário em equipamento apropriado, enquanto a ROM é implementada em *hardware*.
- EPROM, *erasable PROM*: é programável e reprogramável, diferentemente da PROM, que não pode ser reprogramada. A reprogramação é feita por meio de luz ultravioleta.
- EEPROM, *electrically erasable PROM*: também é reprogramável, contudo, tem sua reprogramação facilitada em relação à EPROM. Além disso, na EEPROM, é possível apagar apenas partes do *chip*, enquanto na EPROM é necessário apagar o *chip* inteiro.
- Memória flash: é uma EEPROM com a vantagem de que os dados podem ser escritos e apagados em blocos, enquanto na EEPROM esse processo precisa ser feito *byte a byte*; por isso, a memória flash é mais rápida que a EEPROM.

## Hierarquia de memória

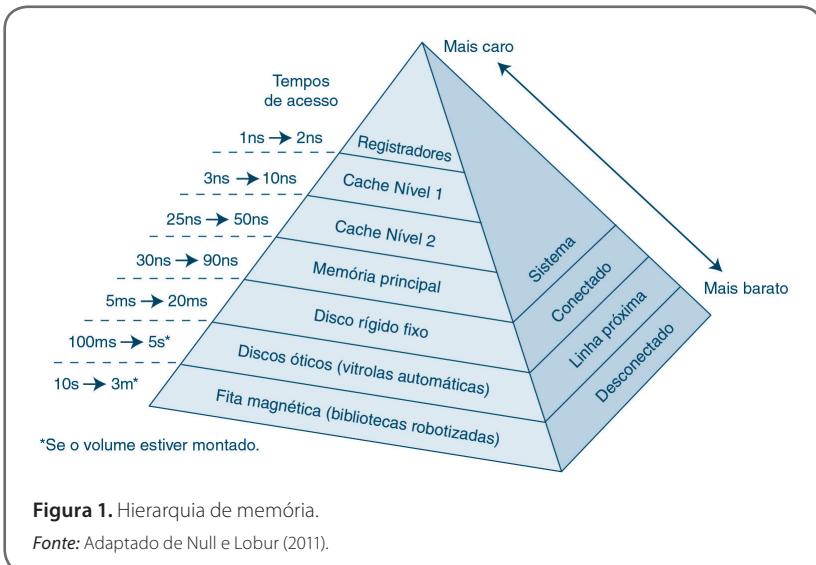
Conforme visto anteriormente, nem todas as tecnologias possuem o mesmo desempenho, embora tenham objetivos similares. Em geral, tipos de memória que possuem um desempenho melhor também possuem um custo maior, e é nesse cenário que a hierarquia de memória é utilizada. A hierarquia de memórias tem entre seus objetivos proporcionar aos computadores o melhor custo–benefício possível, utilizando uma mescla das tecnologias de memória que garanta uma redução do custo do equipamento, sem causar impactos significativamente grandes na performance da máquina.

Além da redução de custos, a hierarquia de memória também pode ser utilizada como estratégia para a melhoria de desempenho das máquinas. Os tipos básicos de recursos que, em geral, fazem parte de um sistema de memória hierárquica são: registradores, cache, memória principal e memória secundária.

Atualmente, o esquema hierárquico ideal de memória está estruturado da seguinte forma: o nível superior e mais próximo do processador é constituído pelos registradores do processador. Em seguida, os computadores possuem uma memória de tamanho reduzido e de velocidade bastante alta, chamada de **cache**, onde dados e posições de memória usados com maior frequência são temporariamente armazenados. Um exemplo de uso da cache são as páginas da *web* que sempre armazenam as informações do último acesso a elas, para que o segundo acesso seja feito de forma mais rápida ou *off-line*. Essa cache é conectada a uma memória principal, maior e de média velocidade, que normalmente usa módulos de memória dinâmica de acesso aleatório (DRAM). Essas memórias são consideradas internas ao sistema de computação. Essa memória é complementada por uma memória secundária, composta por discos rígidos e discos removíveis, chamados de memórias externas. É nessa memória secundária que são armazenados, por exemplo, fotos e arquivos de texto, conforme Stallings (2002).

Um fator que interfere diretamente no desempenho das memórias é a sua distância do processador, sendo a distância medida pelo número de ciclos de máquina requeridos para se acessar essa memória, conforme Null e Lobur (2011). Nesse sentido, quanto mais próxima do processador a memória estiver, maior é a velocidade de acesso a ela, e quanto mais afastada do processador, maior o custo de leitura por *bit* e maior o tempo de acesso.

A hierarquia de memória é exibida na Figura 1. Ela é desenhada em forma de pirâmide, fazendo alusão ao tamanho dessas memórias. Memórias próximas ao processador (topo) tendem a ser menores, contudo, possuem maior performance e um custo mais alto. Os números à esquerda representam tempos de acesso típicos, e as informações à direita dizem respeito à relação dessas memórias com o sistema operacional. Por exemplo, a memória principal faz parte do sistema, o disco rígido está conectado ao sistema, enquanto uma fita magnética é desconectada do sistema.



**Figura 1.** Hierarquia de memória.

Fonte: Adaptado de Null e Lobur (2011).

Ao falar sobre hierarquias de memórias, alguns termos são utilizados para designar o resultado de pesquisas de dados em memória, bem como para medir a eficiência da pesquisa de dados. Esses termos são descritos a seguir, com base em Null e Lobur (2011).

## Terminologia usada em hierarquia de memória

- **Acerto:** o dado requerido é localizado em um nível de memória.
- **Falha:** o dado requerido não foi localizado em um nível de memória.
- **Taxa de acertos:** percentagem de acessos de memória encontrados em um dado nível de memória.
- **Taxa de falhas:** percentagem de acessos de memória não encontrados em um dado nível de memória.
- **Tempo de acerto:** tempo requerido para acessar a informação requisitada em um dado nível de memória.
- **Penalidade de falha:** tempo requerido para processar uma falha.

## Localidade de referência

A localidade é um fator de grande relevância para o desempenho das memórias e, por consequência, para o desempenho dos computadores. O princípio da

localidade considera que todo o programa acessa uma porção relativamente pequena do espaço endereçável de memória em um instante qualquer. Ou seja, se a posição de memória X foi acessada no tempo t, existe uma grande possibilidade de a posição de memória X + 1 ser acessada em um futuro próximo, possivelmente t + 1. Esse conjunto de referências à memória que são geradas em cada acesso é chamado de **localidade de referência**, conforme Null e Lobur (2011). Em um sistema de computadores, existem três tipos de localidade de referência, os quais são descritos a seguir, com base em Null e Lobur (2011).

- **Localidade temporal:** se um item ou um endereço de memória foi referenciado, ele tende a ser referenciado novamente. Exemplo: *loops* (instruções e dados).
- **Localidade espacial:** se um item é referenciado, itens cujos endereços são próximos a este possuem grande possibilidade de serem referenciados também. Exemplo: acesso a dados de um *array*.
- **Localidade sequencial:** instruções tendem a ser acessadas sequencialmente.

O princípio da localidade melhora efetivamente a maioria dos acessos à memória, devido à oportunidade de usar uma pequena quantidade de memória muito rápida.

## Memória cache e memória virtual

Conforme descrito anteriormente, o computador possui uma hierarquia de memórias. Isso ocorre porque, quanto mais próxima do processador a memória estiver, maior é a velocidade em que o processador acessa essa memória. Entre todas as memórias, a que está mais próxima do processador é a memória cache. Esta é uma memória bem pequena, temporária e muito rápida.

Para entender o conceito de cache, vamos usar uma analogia com o mundo real. Imagine que você foi a uma mercearia para comprar alimentos. Raramente você compra apenas um item — você costuma comprar, além do item que precisa nesse momento, coisas que vai precisar no futuro. A mercearia é similar à memória principal, e a sua casa é o cache que vai armazenar os itens para uso imediato e acessos futuros. Assim funciona o cache — ele copia os dados frequentemente utilizados para a memória cache, fazendo com que o processador não precise acessar a memória principal para recuperar os dados.

O computador, entretanto, tão tem como saber antecipadamente quais informações precisarão ser acessadas ou quais serão acessadas com maior frequência. Dessa forma, a memória cache trabalha com o conceito de bloco, buscando e armazenando um bloco de informações. A memória cache é, por vezes, conhecida como memória endereçável por conteúdo ou CAM (*content addressable memory*) — isto é, a memória cache não é acessada por endereço, como são as outras memórias, e sim via dado. Para simplificar o processo de localização desses dados, alguns esquemas de mapeamento de cache precisam ser utilizados.

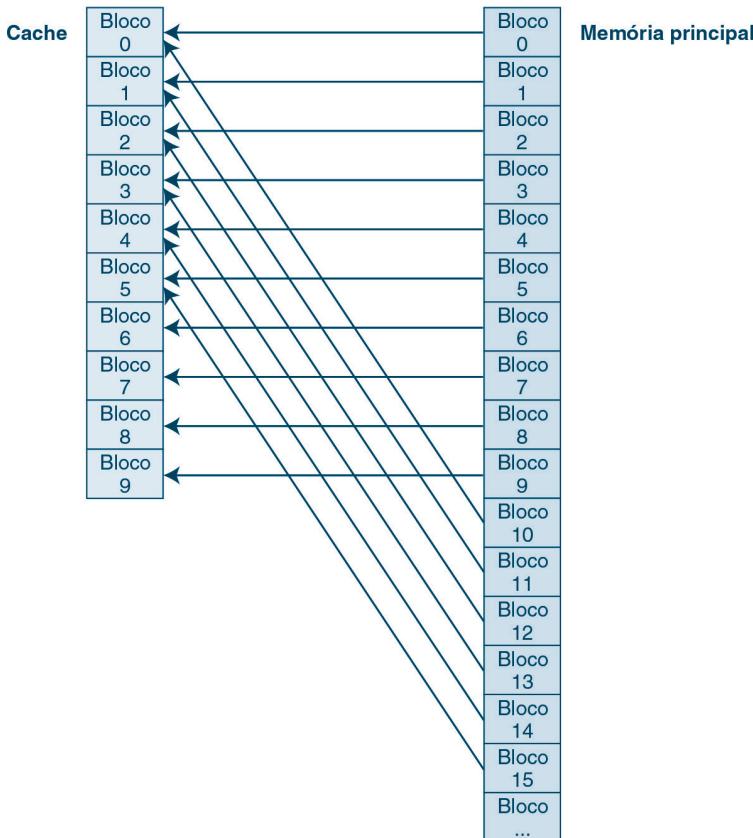
Um dado, quando originado da memória principal e transportado para a memória cache, não mantém na cache o mesmo endereço da memória principal. Dessa forma, torna-se de extrema importância a criação de um mapeamento entre a memória principal e a memória cache. Esse mapeamento serve como um mapa que vai facilitar a localização desses dados. A seguir, você vai conhecer alguns dos tipos de mapeamento que podem ser utilizados.

## Mapeamento de cache

- Cache com mapeamento direto: faz mapeamento de rede usando uma abordagem modular, ou seja, utilizando blocos. O mapeamento direto mapeia o bloco x da memória principal com o y da cache, estabelecendo uma relação entre eles. A Figura 2 demonstra essa relação de mapeamento. Nela, é possível visualizar que, por vezes, a memória principal possui mais blocos do que a cache; nesse caso, quando o último bloco de cache for alocado por um bloco da memória principal, o bloco seguinte de memória principal tentará alocar novamente o primeiro bloco de cache. Isso ocorre no bloco 10 da Figura 2, conforme Null e Lobur (2011).
- Cache totalmente associativa: nesse tipo de mapeamento, permite-se que um bloco da memória principal seja colocado em qualquer lugar da cache. A localização desse bloco se torna mais cara, uma vez que a única forma de localizar é procurar em toda a cache. Para isso, toda a cache precisa ser construída de forma associativa, para permitir ser pesquisada em paralelo, ou seja, permitir que, com uma busca, seja comparado o identificador buscado com todos os identificadores da cache. Ou seja, se a memória principal possuir um bloco A, e esse bloco contiver informações que precisam ser armazenadas na cache, a cache associativa permite que esse bloco seja salvo em qualquer local da cache. Permite também que, futuramente, o computador possa pesquisar um

dado qualquer em toda a cache, até localizar esse bloco que contém informação, conforme Null e Lobur (2011).

- Cache associativa por conjunto: é um esquema similar ao mapeamento direto, uma vez que usamos o endereço da memória principal para mapear um endereço da cache. A diferença é que, em vez de mapear um endereço para uma certa posição da cache, um endereço mapeia um conjunto de blocos, segundo Null e Lobur (2011).



**Figura 2.** Mapeamento direto de blocos da memória principal para a memória cache.

Fonte: Null e Lobur (2011, p. 318).

## Memória virtual

O objetivo da memória virtual é usar o disco rígido (HD) como uma extensão da memória RAM. Essa estratégia permite aumentar o espaço de endereçamento para que os processos possam usá-lo. A memória virtual pode ser implementada de várias formas — paginação e segmentação são alguns exemplos, conforme Weber (2012).

- **Paginação:** a paginação é o processo mais utilizado; ela permite que o programa possa ser espalhado por áreas não contínuas da memória. Dessa forma, o endereço lógico de um processo e a memória física são divididos em páginas de tamanho fixo. A partir disso, o programa é carregado por página, e cada página lógica ocupa uma página física. Em síntese, a paginação busca alocar memória física (HD) para processos em pedaços de tamanho fixo e manter o registro onde as páginas do processo residem. A organização desses endereços é feita em uma tabela de páginas.
- **Segmentação:** nessa técnica, os programas são divididos em segmentos de tamanhos variados, e cada um tem seu espaço de endereçamento. A principal diferença entre a paginação e a segmentação é a alocação da memória de maneira dinâmica, não fixa. Na segmentação, a alocação depende da lógica do programa.

O Quadro 1 sintetiza as informações descritas no capítulo sobre os principais tipos de memória.

**Quadro 1.** Tipos de memória

Cache	Principal	Secundária	Virtual
Tamanho pequeno, muita velocidade, próxima do processador, armazena cópias de fragmentos da memória principal para facilitar e agilizar o acesso.	Memória um pouco maior, volátil, lê informações da memória secundária e armazena para facilitar o acesso.	Armazena a maior parte das informações, não é volátil, armazena tudo aquilo que o usuário salva.	É a transformação de um pedaço da memória secundária em complemento da memória principal.



## Link

Nos *links* a seguir, você pode obter mais informações sobre os tipos de memória.

- Tipos de memória:

<https://goo.gl/MPFXsL>

- Memória cache:

<https://goo.gl/kjwSdP>



## Referências

NULL, L.; LOBUR, J. *Princípios básicos de arquitetura e organização de computadores*. 2.ed. Porto Alegre: Bookman, 2011.

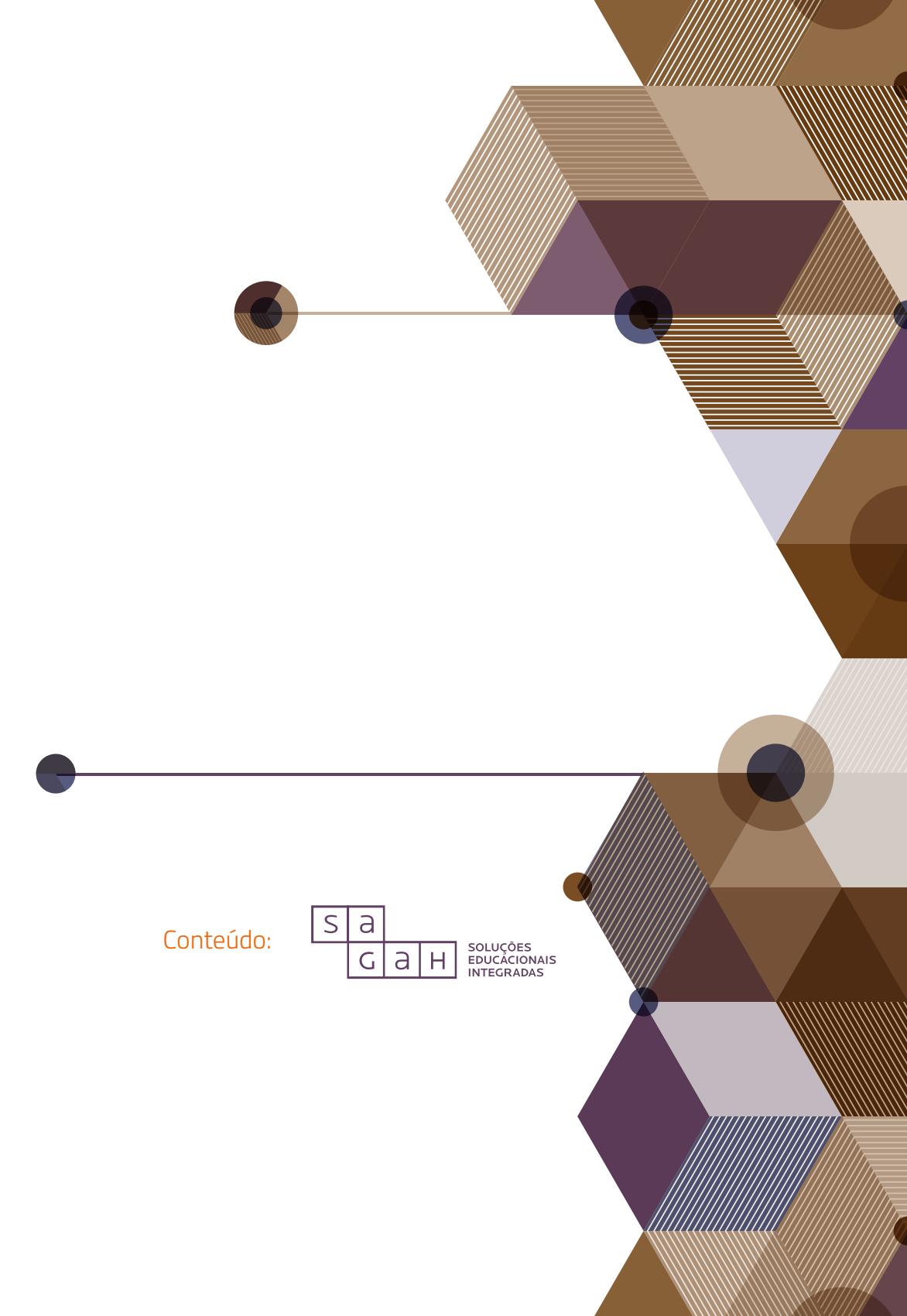
STALLINGS, W. *Arquitetura e organização de computadores: projeto para o desempenho*. ed. 5. São Paulo: Prentice Hall, 2002.

WEBER, R. F. *Fundamentos de arquitetura de computadores*. 4. ed. Porto Alegre: Bookman: 2012.

## Leitura recomendada

PANNAIN, R. *Arquitetura de computadores*. 2009. Disponível em: <[http://www.ic.unicamp.br/~ducatte/mc542/Arquitetura/arq\\_hp7.pdf](http://www.ic.unicamp.br/~ducatte/mc542/Arquitetura/arq_hp7.pdf)>. Acesso em: 10 jan. 2019.

Encerra aqui o trecho do livro disponibilizado para esta Unidade de Aprendizagem. Na Biblioteca Virtual da Instituição, você encontra a obra na íntegra.



Conteúdo:



SOLUÇÕES  
EDUCACIONAIS  
INTEGRADAS

# Dica do Professor: Tipos de memórias

---

Neste vídeo, você encontrará informações sobre como escolher a memória certa para seu computador.



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

# Dica do Professor: Tecnologia e Hierarquia de Memórias

---

As memórias são componentes de extrema importância para o computador, uma vez que são elas que permitem armazenar, de forma permanente ou temporária, todas as informações. Dentro do computador as memórias são organizadas hierarquicamente.

Nesta Dica do Professor, você vai aprender um pouco mais sobre como as memórias são organizadas hierarquicamente.



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

Os sistemas permitem manter o registro das necessidades de memória de um programa enquanto ele está rodando. Ou seja, podemos manter uma aplicação funcionando consumindo muito menos da memória RAM. Mas se uma parte da aplicação está na memória, então onde fica o resto dela?

Neste Momento Sinapse, você vai encontrar a resposta conhecendo um pouco sobre o *Backing store*.



Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

# **Na prática: Características das memórias**

---

Muitas vezes, queremos verificar qual é a memória RAM instalada em nosso computador, ou mesmo conhecer algumas características dela, como frequência de trabalho, taxa de transferência e outros aspectos, como verificar se o desempenho do computador está compatível com seus componentes ou se é possível aumentar a memória para que fique mais rápido, e assim por diante. Obter essas informações pode não ser fácil, mesmo abrindo o gabinete e fazendo uma inspeção visual.

Como o tipo e a capacidade da memória RAM são muito importantes para o desempenho do computador, assim como a frequência com que a memória trabalha está diretamente relacionada a seu desempenho, você verá a seguir algumas formas de conhecer as características das memórias que são instaladas nos computadores do tipo PC.

**Conteúdo interativo disponível na plataforma de ensino!**

# Na prática: Tecnologia e Hierarquia de Memórias

---

Memória é o local onde ficam armazenados eletronicamente instruções e dados que devem ser acessados rapidamente por um computador; ou seja, guarda informações para uso imediato.

A memória *cache* é um tipo de memória volátil que fornece acesso a dados em alta velocidade, por isso é de extrema importância em um processo computacional. Confira o tutorial abaixo para saber um pouco mais.



Aponte a câmera para o código e accese o link do conteúdo ou clique no código para accesar.

Paulo é técnico em informática de uma empresa multinacional e foi procurado por Ana, gestora financeira da empresa, porque seu setor tem encontrado problemas de lentidão para executar tarefas simples de comando em sua rede de computadores, levando muito tempo para executar tarefas simples. Paulo começou a analisar os computadores do setor financeiro e percebeu que haviam 5 anos que o setor estava trabalhando com a mesma quantia de memória em seus computadores, ou seja, estavam sobrecarregados e obsoletos.

Como solução imediata, Paulo propôs um aumento na memória RAM dos computadores, para que fosse reduzida a quantidade de acesso do processador à memória secundária e, com isso, permitisse uma *performance* melhor na máquina com um custo baixo.

Contudo, Ana achou a solução demasiadamente simples, ela acreditava que precisariam ser comprado computadores novos. Para explicar a Ana o porquê aumentar a memória aumentaria o desempenho do computador, Paulo montou uma apresentação que explica o que acontece com cada memória do computador, desde o momento que o computador é ligado até que seja desligado.

Veja o que Paulo explicou para Ana.

## QUANDO O COMPUTADOR ESTÁ DESLIGADO:

Esse é o estado inicial da máquina, não existe nenhum registro em cache, nenhum registro na memória principal; contudo, os dados armazenados no computador, por meio da memória principal, estão armazenados e estarão acessíveis quando o computador for ligado.



## QUANDO PROGRAMA X É EXECUTADO:

Nesse acesso, o processador precisa buscar as informações sobre o programa na memória secundária, aquela que estava com todas as informações salvas mesmo com o computador desligado. Essas informações são levadas para a memória principal, e a memória principal envia para o cache apenas o fundamental para a execução do programa.



## QUANDO PROGRAMA X É ENCERRADO E OUTROS PROGRAMAS SÃO ABERTOS:

As informações sobre o programa X saíram da cache porque informações sobre outros programas foram inseridas e a cache tem um espaço de armazenamento muito pequeno. Na memória principal, as informações sobre o programa X seguem acessíveis.



## QUANDO PROGRAMA X É REEXECUTADO:

Nesse acesso, o processador busca as informações sobre o programa na memória principal e a memória principal envia para a cache apenas o fundamental para a execução do programa. Esse acesso é mais rápido que o primeiro acesso, no qual as informações foram buscadas na memória secundária.



## 3D: Acesso direto à memória

"html": "

Conteúdo interativo disponível na plataforma de ensino!

"

# Saiba mais

Para ampliar o seu conhecimento a respeito desse assunto, veja abaixo as sugestões do professor:

## **Manual da placa-mãe**

Veja a seguir um exemplo de manual de placa-mãe em que se pode verificar os tipos de memórias aceitos por ela.



Aponte a câmera para o código e accese o link do conteúdo ou clique no código para accesar.

## **A memória**

Neste material, você vai conhecer algumas curiosidades sobre o funcionamento das memórias do computador. São explicadas, por exemplo, as unidades de medida utilizadas nos dispositivos de armazenamento de dados. Além do conceito de largura de dados, velocidade e frequência.



Aponte a câmera para o código e accese o link do conteúdo ou clique no código para accesar.

## **Conheça as diferenças entre as memórias RAM e ROM**

Este vídeo apresenta vários fundamentos das memórias voláteis e não voláteis.



Aponte a câmera para o código e accese o link do conteúdo ou clique no código para accesar.

# Tipos de memória de computador

Neste vídeo, você vai conhecer um pouco mais sobre os tipos de memória disponíveis nos computadores, conhecendo melhor, por exemplo, os tipos de memória ROM: PROM, EPROM, EEPROM, EAROM e Flash, e os tipos de memória RAM: SRAM, DRAM e MRAM.



Aponte a câmera para o código e accese o link do conteúdo ou clique no código para acessar.

## — ACSO - Exercícios unidade 4 —

- 1) Todo o armazenamento de dados de um computador é feito por meio de uma das memórias, seja ela de armazenamento definitivo ou volátil.

Assinale a alternativa que apresenta, sequencialmente, a memória volátil principal do sistema, ou seja, a memória que armazena informações de acesso frequente e não é parte do processador, e a memória não volátil, que armazena informações essenciais para a inicialização do sistema operacional.

- A) Memória cache e memória virtual.
- B) Memória secundária e memória ROM.
- C) Memória RAM e memória ROM.
- D) Memória RAM e memória virtual.
- E) Memória RAM e memória shell.

- 2) Entre os tipos de memória RAM disponíveis encontra-se a memória DRAM e a memória SRAM.

Sobre essas memórias, seus pontos positivos e negativos, assinale a alternativa correta.

- A) Em relação à memória ROM, a memória DRAM tem uma eficiência maior porque permite o armazenamento de mais dados.
- B) Memória SRAM tem um desempenho melhor porque a memória DRAM precisa ser reescrita a todo o momento, e a SRAM mantém os dados até que o computador seja desligado.
- C) A principal diferença entre DRAM e SRAM é que uma se comunica com a cache, a outra, não.
- D) DRAM é uma memória que pode ser utilizada para substituir a memória ROM.
- E) Memória DRAM em geral tem uma capacidade de armazenamento maior que a SRAM.

- 3) Um conceito muito importante no que diz respeito à memória é o conceito de localidade de referência. Esse conceito diz basicamente que se um valor  $a$  é acessado em um tempo  $x$ , provavelmente o valor  $a+1$  será acessado em um tempo  $x+1$ .

**Sobre localidade de referência, assinale a alternativa que apresenta os tipos de localidade que podem ser utilizados.**

- A) Localidade de memória, referência e dados.
  - B) Localidade temporal, espacial e sequencial.
  - C) Localidade relevância, registradores e cache.
  - D) Localidade virtual, sequencial e secundária.
  - E) Localidade principal, temporal e espacial.
- 4) A memória virtual é uma alternativa para auxiliar na armazenagem de dados e na eficiência do computador.

**Sobre esse tipo de memória, assinale a alternativa correta.**

- A) Esse tipo de memória converte uma parte da memória secundária para a memória principal.
  - B) Esse tipo de memória converte uma parte da memória secundária para memória cache.
  - C) Esse tipo de memória converte uma parte da memória secundária para memória ROM.
  - D) Esse tipo de memoria é usada, por exemplo, em pendrive.
  - E) Um exemplo desse tipo de memória é o Google Drive.
- 5) Quando os dados são transportados da memória principal para a memória cache, se faz necessária uma associação para que eles possam ser localizados.

**Existem diversas formas de fazer essa associação, analise as alternativas e assinale a correta.**

- A) Cache com mapeamento direto, permite que os dados sejam mapeados de forma direta em qualquer local da cache.
- B) Cache totalmente associativa realiza uma associação por blocos, em que cada bloco da memória principal é referenciado por um bloco na cache.
- C) Cache virtual.
- D) Cache com mapeamento direto um bloco da memória principal equivale a um bloco na memória cache.

- E) Cache associativa por conjunto, um conjunto de dados na memória principal é referenciado por um conjunto de informações na cache.

# — ACSO - Desafio em GRUPO unidade 4 —

O Sistema de planejamento de recurso corporativo (*Enterprise Resource Planning – ERP*) é um sistema de informação que integra todos os dados e processos de uma organização em um único sistema.

O mais novo desafio passado para você e para seus colegas de trabalho é estruturar um computador que tenha um bom desempenho e não tenha um custo muito alto. Como a equipe é formada por várias pessoas, cada um irá pesquisar uma parte do computador, para juntos efetuarem a montagem.

Considere um computador que será utilizado prioritariamente para acessar o sistema ERP da empresa e para utilizar pacotes de ferramentas de escritório.

Contudo, esse computador pode eventualmente ser utilizado pelos programadores da empresa, em caso de necessidade e, por isso, precisa ter memória suficiente para executar a ferramenta Android Studio.



Você é o profissional de TI que trabalha com arquitetura de computadores e sistemas operacionais na equipe, por isso, você deve indicar o tamanho de memória primária, secundária e cache para este cenário e explicar o motivo dessas escolhas.