

Componentes do Computador: Processador e Memória

Conceitos Gerais

1. Processador (CPU)

- O CPU (Central Processing Unit), também chamado de processador, é o componente principal de um computador. Por isso, muitas vezes é chamado de cérebro do computador ou do sistema!
- A atividade principal do CPU é processar dados, executar cálculos e controlar as operações de outros componentes do sistema, como operações de E/S (teclado, mouse, monitor etc.).
- O processador interpreta e executa as instruções armazenadas na memória do computador, realizando tarefas como operações matemáticas (adição, subtração etc.), lógicas (AND, OR etc.) e de controle de fluxo.
- Os processadores modernos possuem diferentes arquiteturas, podendo ser RISC, CISC e principalmente híbridos (combinação das duas anteriores).
- Além disso, os processadores podem ter vários núcleos (multicore), permitindo a realização de múltiplas tarefas simultaneamente.

1. Processador (CPU)

- Os componentes principais que formam o processador são:

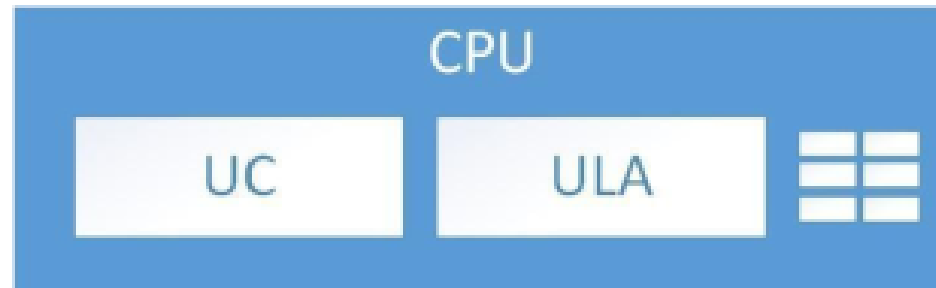
1. **Unidade de Controle (UC).**

2. **Unidade Lógica e Aritmética (ULA).**

3. **Registradores.**

4. **Interconexão da CPU:** barramentos que conectam a UC, ULA e os registradores.

5. **Cache:** Este não é um componente essencial, mas está presente na grande maioria das CPU's.



2. Unidade de Controle (UC)

- A **Unidade de Controle (UC)** é responsável por receber instruções da memória pelo barramento de instruções. A UC não executa as instruções, apenas faz a leitura, decodifica e passa os comandos para a UCD (Unidade de Ciclo de Dados) determinando como as instruções devem ser executadas e com quais dados.
- Baseada nesses comandos, a UCD pode buscar os dados necessários na memória, executar as devidas operações e enviar o resultado de volta para a memória para ser armazenado.
- Tudo é controlado de acordo com os comandos internos enviados pela UC, que por sua vez se baseia na instrução decodificada.
- **Observação:** Alguns autores desconsideram a existência da UCD, descrevendo que a UC é a responsável pela execução.

2. Unidade de Controle (UC)

- Todo este processo de execução é guiado por um sinal síncrono de relógio (clock). A cada clock a unidade sabe que deve executar um passo, passar os dados para quem deve, e se preparar para o próximo passo.
- Obviamente que quanto mais rápido for o relógio, mais operações por segundo o processador consegue executar. A velocidade do relógio é medida em frequência, utilizando a unidade Hertz.
- Um Hertz significa um passo por segundo. Os processadores atuais trabalham na faixa média de 2 até 5GHz (bilhão de Hertz, ou de passos por segundo).

3. Unidade Lógica Aritmética (ULA)

- A **Unidade Lógica e Aritmética (ULA)** faz as operações matemáticas do processador, como adição, subtração, “E” lógico (AND), “OU” lógico (OR) etc.
- Estas operações são efetuadas sobre os valores de entrada, produzindo um resultado no registrador de saída. Depois, esse valor pode ser escrito na memória principal, caso necessário.
- Grande parte das instruções pode ser classificadas como registrador-memória ou registrador-registrador.
- Instruções **registrador-memória** permitem que palavras (unidades de dados movimentadas entre a memória e os registradores.) sejam buscadas diretamente da memória para a ULA ou armazenadas na memória diretamente a partir da ULA.
- Instruções **registrador-registrador** buscam dois operandos nos registradores para a ULA, efetua a operação e armazena o resultado em um dos registradores.

4. Registradores

- Os **registradores** são pequenas memórias (8 a 128 bits) de alta velocidade que ficam dentro do CPU. Eles armazenam resultados temporários e dados de controle, sendo que alguns são de uso geral e outros de uso específico.
- O registrador mais importante é o **PC (Program Counter)**, que indica a próxima instrução a ser buscada para execução. Outro importante é o **IR (Instruction Register)**, que mantém a instrução que está sendo executada no momento.
- Os registradores são memórias do tipo **SRAM** (Static RAM), diferente da memória principal, que é do tipo DRAM (Dynamic RAM, mas comumente chamada apenas de RAM).

5. Processador 32 x 64 bits

- A grande diferença entre processadores de 32 e 64 bits está na quantidade de dados que podem ser processados por vez. Vejamos algumas características abaixo:

1. Tamanho de Palavra:

- Processador de 32 bits: Pode processar dados em blocos (“pedaços”) de 32 bits por vez, ou seja, é capaz de gerenciar números inteiros de até 32 bits = 4.294.967.296 valores possíveis
- Processador de 64 bits: Pode processar dados em blocos (“pedaços”) de 64 bits por vez, ou seja, pode lidar com números inteiros de até 64 bits = 18.446.744.073.709.551.616 valores possíveis.

2. Capacidade de Endereçamento de Memória RAM:

- Processador de 32 bits: Pode endereçar até 4 GB de memória RAM
- Processador de 64 bits: Possui uma capacidade de endereçamento muito maior, teoricamente permitindo acessar até 18,4 milhões de terabytes de RAM

5. Processador 32 x 64 bits

3. Desempenho e Eficiência:

- Processador de 64 bits: Em aplicações que envolvem grandes conjuntos de dados, processadores de 64 bits podem oferecer um desempenho superior.

4. Compatibilidade de Software:

- Processador de 32 bits: Pode executar software compilado para sistemas de 32 bits;
- Processador de 64 bits: Pode executar tanto software de 32 bits quanto de 64 bits.

5. Segurança:

- Processador de 64 bits: Introduziu alguns recursos de segurança aprimorados (como registradores extra para proteção de instruções que envolve hardware) que ajudam a prevenir certos tipos de ataques de software malicioso.

6. Características Gerais de CPU

Multicore:

- Um processador multicore possuem dois ou mais núcleos independentes no mesmo chip. Com isso é possível executar várias tarefas simultaneamente.
- Cada núcleo pode executar suas próprias instruções, processar dados e realizar operações separadamente dos outros núcleos, o que resulta em um aumento geral no desempenho.

Interrupção:

Mecanismo pelo qual é possível interromper a sequência normal do processamento, ou seja, pode interromper o ciclo de instrução. As interrupções podem ser causadas por:

- Programa, ex.: overflow, divisão por zero;
- Timer, gerado pelo temporizador interno do processador;
- Controlador de E/S;
- Falha de hardware, ex.: erro de paridade de memória.

6. Características Gerais de CPU

Coprocessador:

- É uma CPU complementar usada para suplementar as funções do processador principal. As operações realizadas pelo coprocessador podem ser para cálculos de ponto flutuante, computação gráfica, processamento de cadeias de caracteres e criptografia.

GPU (Graphics Processing Unit):

Processador composto por muitos núcleos menores e mais especializados. Trabalhando em conjunto, os núcleos oferecem desempenho massivo quando uma tarefa pode ser paralelizada.

- Uma CPU é adequada para processamento de tarefas gerais. Ela possui um menor número de núcleos focados em tarefas individuais e em executá-las rapidamente.
- Já as GPUs, além de processamento gráfico, evoluíram para se tornarem também processadores paralelos para tarefas que podem ser feitas em paralelo. São muito usadas para renderização de gráficos 3D, processamento de vídeo, IA e cálculos **paralelos** de alto volume.

6. Características Gerais de CPU

ACPI (Advanced Configuration and Power Interface):

- ACPI é um padrão de interface entre o sistema operacional e o hardware do computador. Ele foi criado para permitir que o sistema gerencie energia, configurações de dispositivos e controle de hardware de forma mais eficiente e centralizada.

Socket (soquete) do processador:

- É a entrada onde o processador é instalado na placa mãe. Cada geração de processadores possui compatibilidade com um tipo específico de socket. Por exemplo, os chips Intel Core i3 e i5 são compatíveis com o socket LGA1150. Já os AMD A4, A6 e A8 são compatíveis com o socket FM2.

7. Multiprocessamento

- A arquitetura de computação clássica (von Neumann) foi elaborada com base em sistemas que executam de forma sequencial (uma instrução de cada vez).
- Contrariando o que dizia von Neumann, atualmente temos processadores que são capazes de executar instruções em paralelo (mais de uma instrução por vez). O paralelismo tem duas formas gerais:
 1. **Nível de instrução:** procura-se executar o maior número de instruções por segundo (pipelining);
 2. **Nível de processador:** mais de uma CPU trabalhando em conjunto no mesmo problema (multiprocessamento).

A capacidade de ampliar o poder computacional do sistema, apenas adicionando novos processadores, é denominada **escalabilidade**.

8. Taxonomia de Flynn

- Essa classificação se refere à maneira como instruções e dados são organizados em um determinado tipo de processamento. São quatro classificações de processadores:

	Single Data	Multiple Data
Single Instruction	SISD	SIMD
Multiple Instruction	MISD	MIMD

8. Taxonomia de Flynn

- **SISD (Single Instruction Single Data):** há apenas um conjunto de instruções e um conjunto de dados. Esses processadores executam uma instrução completa por vez, sequencialmente, cada uma manipulando um dado específico. Era mais utilizado em processadores antigos, como por exemplo, o 8080/8085, da Intel;
- **MISD (Multiple Instruction Single Data):** tipo de arquitetura que pode utilizar múltiplas instruções para manipular apenas um conjunto de dados, como por exemplo, um vetor;
- **SIMD (Single Instruction Multiple Data):** o processador opera de maneira que uma única instrução acessa e manipula um conjunto de dados simultaneamente.
- **MIMD (Multiple Instruction Multiple Data):** trata-se da categoria mais avançada, produzindo um elevado desempenho do sistema de computação. Múltiplas instruções lidam com múltiplos dados, ao mesmo tempo.

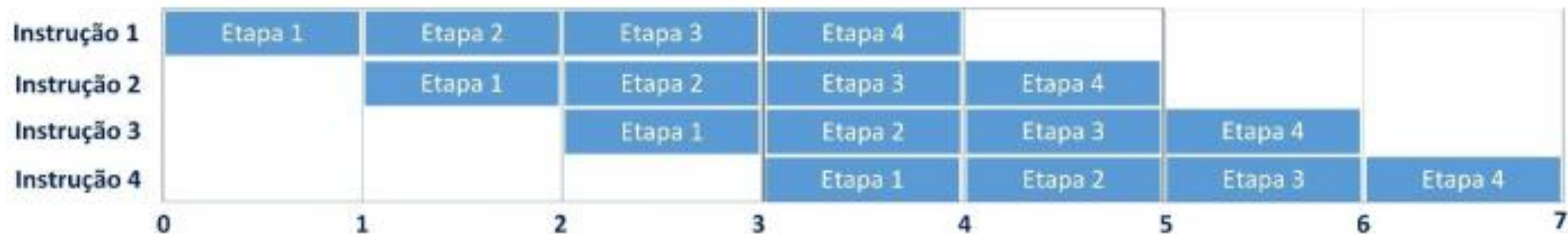
9. Processamento Superescalar e Superpipelining

- O processamento **superescalar** e **superpipelining** são duas formas diferentes de buscar o paralelismo na execução de instruções. Ambos tentam otimizar a técnica de pipeline, porém com metodologias diferentes.
- Pipeline é uma técnica que busca aceleração do processamento com a execução de múltiplas instruções ao mesmo tempo, independente se tem múltiplos CPU's.
- Em um determinado instante de tempo, cada instrução está em uma etapa diferente de sua execução. A ideia é que uma instrução use o CPU em uma fatia de tempo, siga para uma próxima etapa (buscar dados por exemplo) e libere o CPU para que outra instrução inicie a execução

9. Processamento Superescalar e Superpipelining

- Cada instrução é dividida em etapas. Como exemplo, vamos dividir em 4 etapas:

1. Busca da informação (fetch cycle);
2. Decodificação da instrução (instruction decode);
3. Execução da operação (execute cycle);
4. Escrita do resultado (write back).



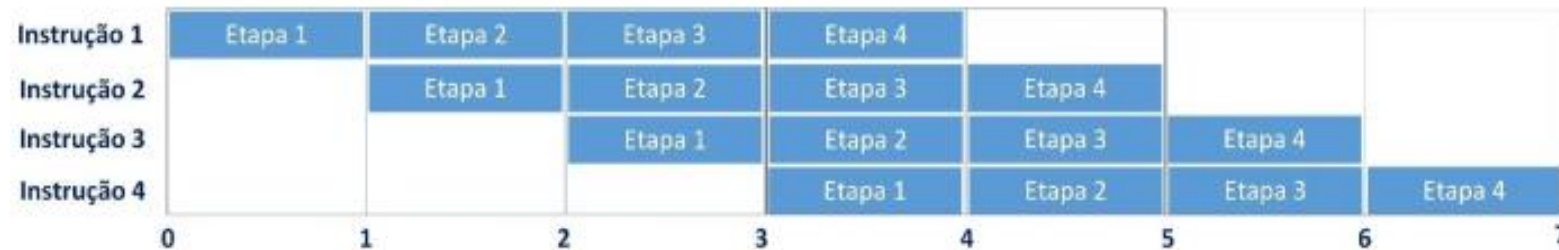
9. Processamento Superescalar e Superpipelining

- Em um **processamento escalar** (sequencial, sem pipeline), ao final do instante de tempo 7, apenas apenas uma instrução completa e 3/4 da segunda teriam sido executadas. Com o pipeline, no mesmo tempo são executadas 4 instruções completas.
- No **processamento superescalar** procura-se obter um grau pleno de paralelismo, ou seja, via hardware (vários núcleos) são criados dois (ou mais) pipelines separados e, então, várias instruções podem ter sua execução em paralelo de fato. As estruturas superescalares possuem paralelismo de instrução e de hardware.

Instrução 1	Etapa 1	Etapa 2	Etapa 3	Etapa 4			
Instrução 2	Etapa 1	Etapa 2	Etapa 3	Etapa 4			
Instrução 3		Etapa 1	Etapa 2	Etapa 3	Etapa 4		
Instrução 4		Etapa 1	Etapa 2	Etapa 3	Etapa 4		
Instrução 5			Etapa 1	Etapa 2	Etapa 3	Etapa 4	
Instrução 6			Etapa 1	Etapa 2	Etapa 3	Etapa 4	
	0	1	2	3	4	5	6

9. Processamento Superescalar e Superpipelining

- Em resumo, **Pipeline** é a técnica de dividir as instruções em estágios e executar esses estágios (etapas) de instruções diferentes em paralelo. Abaixo um exemplo com 4 estágios:



- Uma arquitetura **superescalar** utiliza pipelines com mais componentes físicos (hardware), ex.: um duplo pipeline de cinco estágios, sendo o primeiro deles compartilhado:

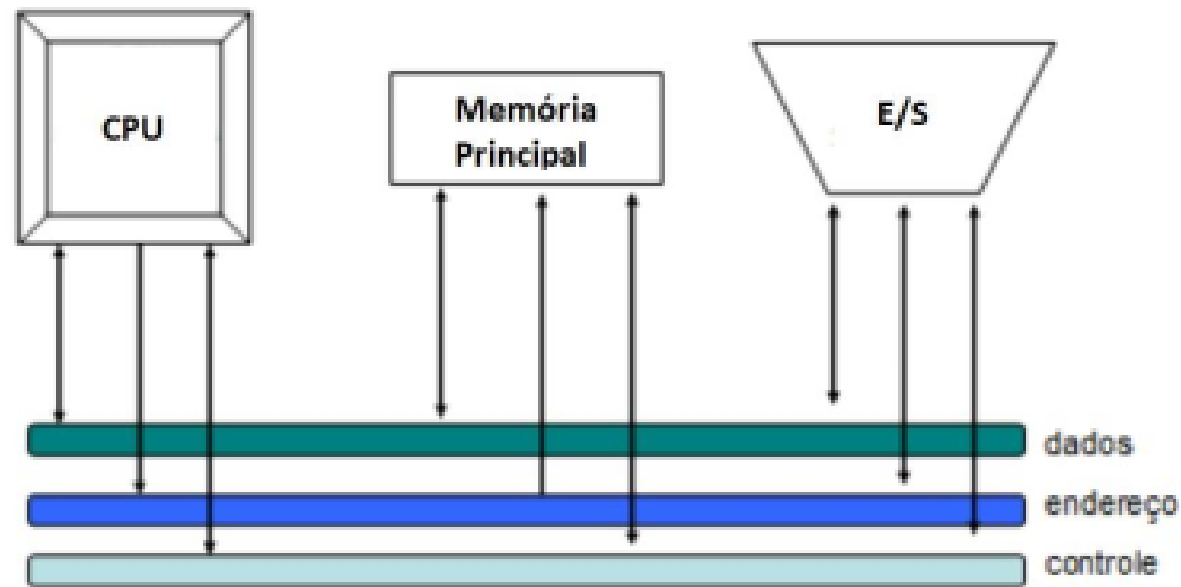


10. VLIW (Very Long Instruction Word)

- A técnica VLIW (instruções em palavras muito longas) consiste em criar “instruções de grande tamanho”, as quais podem “juntar” mais de uma das instruções de máquina de verdade.
- Entretanto, essa estrutura é criada pelo compilador. Ao criar o código-objeto (a ser executado), o compilador utiliza um formato de instrução que inclui mais de uma instrução de máquina na mesma “instrução” a ser buscada e decodificada pela CPU (processador).
- Nas máquinas VLIW o emprego do paralelismo é decidido antecipadamente (compilador), ao contrário dos processadores superescalares, nos quais a escolha dos caminhos é decidida em tempo de execução (run time).
- Desta forma, os programas já devem ser pensados para serem executados em paralelo durante o seu processo de desenvolvimento, o que torna esta atividade mais complexa.

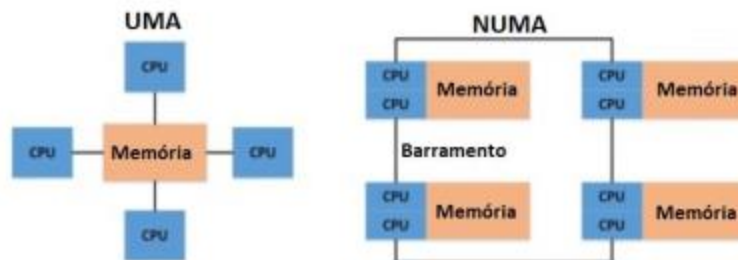
11. Memória

- A imagem abaixo representa um resumo da estrutura de um computador, onde podemos visualizar a memória principal (memória RAM), dispositivos de entrada/saída, incluindo a memória secundária (HD).
- Também é possível observar o processador (CPU), onde se localiza a memória cache, que tem o objetivo de melhorar o desempenho, deixando o que é mais acessado mais próximo do CPU.



12. Acesso a Memória (UMA x NUMA)

- Existem diferentes abordagens de acesso a memória pelo processador:
- UMA (Uniform Memory Access):** todos os processadores possuem acesso a todas as partes da memória principal (leituras e escritas). O tempo de acesso à memória de um processador para todas as regiões da memória é o mesmo
- NUMA (Non-Uniform Memory Access):** todos os processadores possuem acesso a todas as partes da memória (leituras e escritas). O tempo de acesso à memória de um processador difere dependendo de qual região da memória é acessada. Isso ocorre para todos os processadores, porém, para processadores diferentes, as regiões de memória mais lentas e mais rápidas diferem;



13. Hierarquia de Memória

- Existem vários tipos diferentes de dispositivos de armazenamento, com diferentes características relacionadas ao tempo de acesso, capacidade e aplicabilidade. Em conjunto, podemos formar uma hierarquia, a qual podemos resumir na figura abaixo:



13. Hierarquia de Memória

- Quanto mais próxima ao processador uma memória está, mais rápida e cara ela tende a ser. A capacidade geralmente é menor, pois se os registradores e a cache forem grandes, não sobraria dentro do chip para o próprio núcleo de processamento.
- Observe que há uma linha tracejada separando os discos das mídias óticas. Dependendo do autor, pode haver uma diferenciação na nomenclatura entre memória secundária e memória terciária.
- A memória **secundária** não necessita de operações de montagem (inserção de uma mídia em um dispositivo de leitura/gravação) para acessar os dados. A memória **terciária** depende das operações de montagem, como discos óticos, fitas magnéticas, pendrive, etc.
- Só como exemplo, a média de tamanho das memórias são: Registrador – 64 bits; Memória cache – 12 MB (para o L3, o nível L2 tem menos e o L1 menos ainda); Memória RAM – 64 GB; Discos – HD (4 TB), SSD (1TB); Mídias óticas – 650 MB (CD-ROM), 4,7 GB (DVD-ROM).

14. Memória Principal

- A memória principal é indispensável para o funcionamento do computador, pois é onde ficam os programas e dados que devem ser executado pelo processador.
- Quando um usuário abre o Excel e cria uma planilha, tanto o processo do Excel (programa em execução) quanto o texto digitado ficam na memória principal. Existem dois tipos de memória principal (RAM e ROM).
- Um endereço de memória pode ser referenciado por byte, ou seja, se um computador com arquitetura de 32 bits possui 4 GB de memória RAM instalados, é possível acessar os endereços (em hexadecimal) 00000000, 00000001, 00000002, ..., FFFFFFFF.

14. Memória Principal

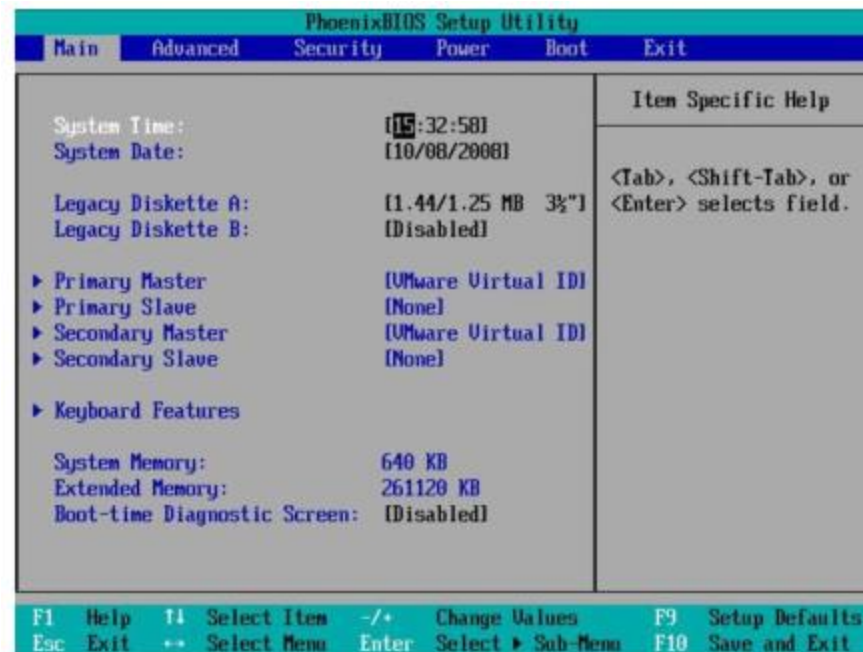
- Os tipos de memória RAM são a DRAM e a SRAM:
- **DRAM (Dynamic Random Access Memory):** O termo dinâmico indica que a memória deve ser constantemente atualizada, ou perderá seu conteúdo. Normalmente é utilizada para a memória principal dos computadores.
- **SRAM (Static Random Access Memory):** é mais utilizada para o cache do sistema

14. Memória Principal

- A **memória ROM** (Read Only Memory – Memória somente para leitura) também é um tipo de memória principal. As suas informações não podem ser apagadas (já vem de fábrica).
- Geralmente estas informações são características de hardware, de forma que desligando o computador, nada é perdido. Observe que ao ligar o computador, uma tela preta é mostrada com algumas informações relacionadas ao hardware. Estes dados estão na ROM da placa mãe.
- Os tipos de memória ROM são:
 1. **PROM** (Programmable Read-Only Memory): pode ser escrita com dispositivos especiais, mas não podem mais ser apagadas ou modificadas;
 2. **EPROM** (Erasable Programmable Read-Only Memory): pode ser apagada pelo uso de radiação ultravioleta, permitindo sua reutilização;
 3. **EEPROM** (Electrically Erasable Programmable Read-Only Memory): pode ter seu conteúdo modificado eletricamente, mesmo quando já estiver funcionando em um circuito eletrônico;

14. Memória Principal

- A **BIOS** (Basic Input/Output System - Sistema Básico de Entrada e Saída) é um aplicativo responsável pela execução das várias tarefas enquanto o computador é ligado e o SO iniciado.
- A BIOS fica armazenada em uma ROM na placa mãe, e possui um **SETUP** para que configurações no computador possa ser feita (ordem do boot, data e hora, overclock, etc).



15. Memória Cache

- Antes de analisar o conceito de memória cache é importante entender dois princípios:

1. Princípio da Localidade Temporal: um dado acessado recentemente tem mais chances de ser usado novamente do que um dado usado há mais tempo. Isso ocorre porque as variáveis de um programa tendem a ser acessadas diversas vezes durante a execução de um programa

2. Princípio da Localidade Espacial: há uma maior probabilidade de acesso para dados e instruções em endereços próximos àqueles acessados recentemente. Isso ocorre porque os programas são sequenciais e usam laços.

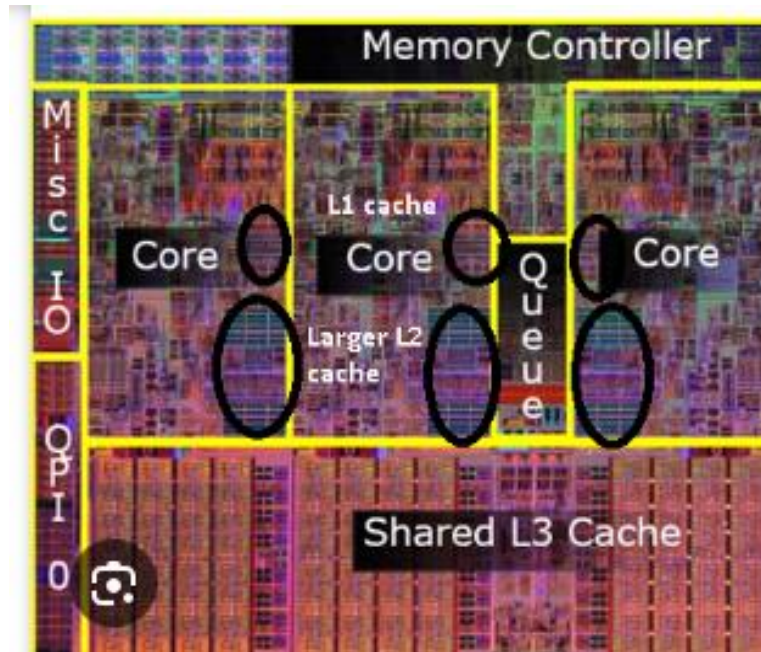
15. Memória Cache

- Diante desses princípios, podemos entender o porquê a memória cache fica entre a memória principal (DRAM) e o processador, sendo que a cache é bem menor e armazena as instruções e dados que possuem uma maior probabilidade de serem utilizados em breve.
- A cache pode ser dividida em níveis. Na figura aparecem apenas dois níveis (L1 e L2), mas pode haver mais, dependendo do processador. Também podemos ver que a L1 está dividida em duas partes. Trata-se de uma parte para **dados** e outra para **instruções**, o que torna o desempenho ainda melhor, pois é possível buscar dado e instrução em **paralelo**.



15. Memória Cache

- A cache L1 é mais próxima ao núcleo de processamento, e armazena as instruções mais uteis ao CPU no momento. A L2 pode estar dentro ou fora do processador (geralmente dentro) e a L3 é na grande maioria das vezes fora do processador.
- Como exemplo de proporção de tamanho, um processador i7 possui: L1 = 64 KB por núcleo, L2 = 256 KB por núcleo, L3 = 12 a 20 MB.



16. Memória Secundária

- Esse tipo de memória armazena dados de forma “permanente”, ou seja, mesmo que a máquina seja desligada, os dados não são perdidos. Só serão perdidos caso o usuário exclua ou ocorra algum dano físico na mídia de armazenamento.
- Geralmente possuem maior capacidade de armazenamento (tamanho) mas com velocidade de leitura e escrita muito inferiores as demais memórias do computador. Apesar de importantes, um computador funciona sem memória secundária.
- O exemplo mais conhecido de memória secundária é o HD.

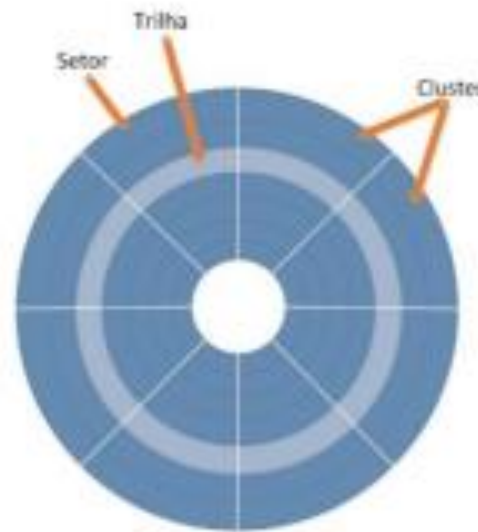
16. Memória Secundária

- Os HD's possuem em seu interior um ou mais pratos (discos) com uma cabeça de leitura/gravação para cada face, que se movimentam presas a um braço. A superfície desses pratos é coberta por um material magnético, possibilitando a leitura e gravação pelas cabeças.
- A imagem abaixo mostra o interior de um HD:



16. Memória Secundária

- A figura a seguir apresenta a distribuição lógica de um HD, formado por Setor (é a menor unidade de armazenamento física do dispositivo) e pelo Cluster (é a menor unidade de armazenamento lógica de dados em um dispositivo).
- Se um arquivo possuir o tamanho maior do que um cluster, ele será distribuído em tantos clusters quanto forem necessários. Entretanto, um mesmo cluster não poderá armazenar mais de um arquivo.



16. Memória Secundária

- Atualmente os HD's estão em desuso e sendo substituídos pelos discos SSD (Solid State Disk). Esta nova tecnologia de armazenamento não possui partes móveis e é construído em torno de um circuito integrado semicondutor.
- Com a eliminação das partes mecânicas (utilizadas em um HD), há redução de vibrações, além de velocidade de acesso aos dados muito maior.
- O SSD também é mais resistente que os HDs comuns devido à ausência de partes mecânicas, algo considerado muito importante quando se trata de computadores portáteis. A imagem mostra o padrão SATA e M2:

