

## CPU

Na aula anterior aprendemos um pouco sobre como se constrói uma Unidade Central de Processamento (UCP ou CPU em inglês). Nesta aula veremos como uma CPU funciona.

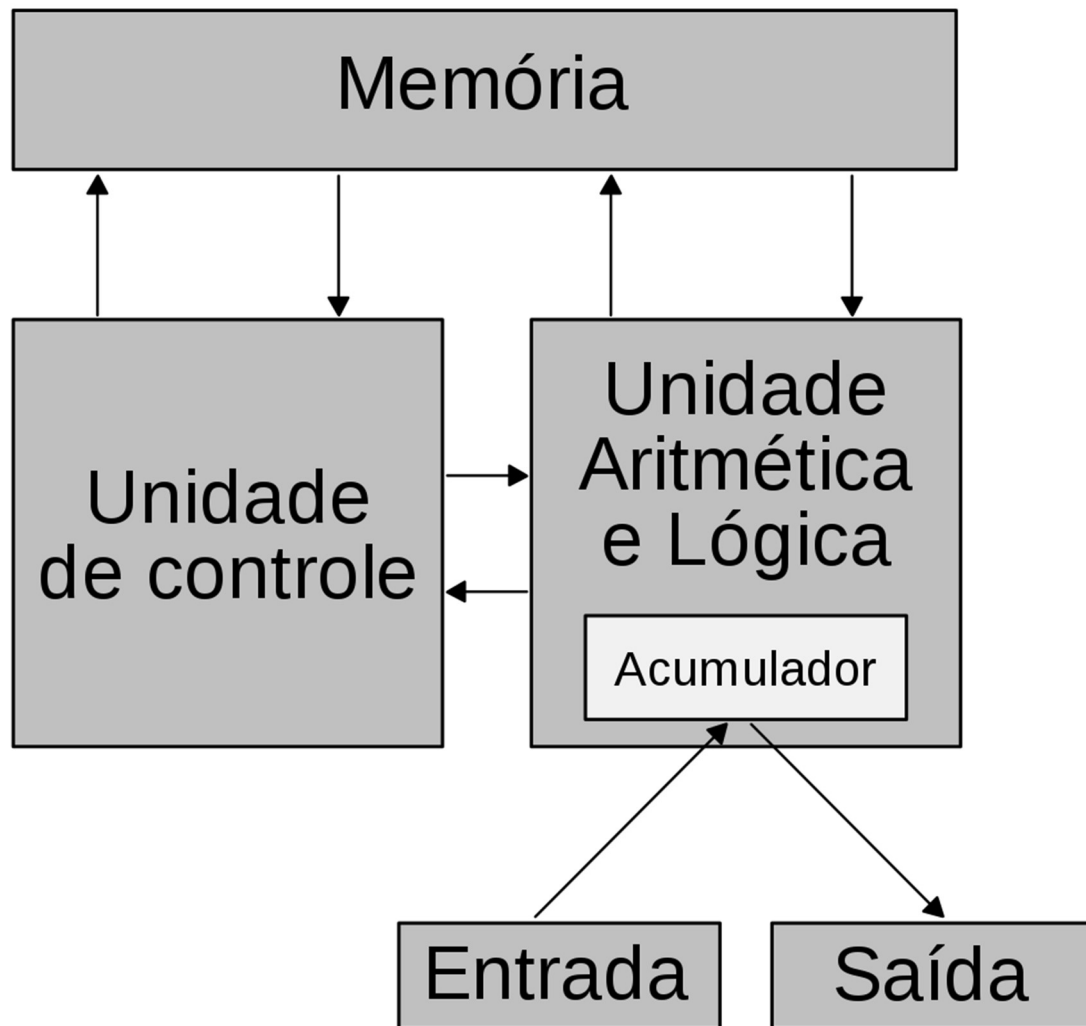
## Modelos de CPU

### Arquitetura de von Neumann

A Arquitetura de von Neumann (de John von Neumann, pronunciado Nóimánn) é uma arquitetura de computador que se caracteriza pela possibilidade de uma máquina digital armazenar seus programas no mesmo espaço de memória que os dados, podendo assim manipular tais programas. Esta arquitetura é um projeto modelo de um computador digital de programa armazenado que utiliza uma unidade de processamento (CPU) e uma de armazenamento ("memória") para comportar, respectivamente, instruções e dados.

A máquina proposta por Von Neumann reúne os seguintes componentes:

Uma memória  
Uma unidade aritmética e lógica (ALU)  
Uma unidade central de processamento (CPU), composta por diversos registradores, e  
Uma Unidade de Controle, cuja função é a mesma da tabela de controle da Máquina de Turing universal: buscar um programa na memória, instrução por instrução, e executá-lo sobre os dados de entrada.



Arquitetura de von Neumann  
Gargalo de Von Neumann

O canal de transmissão de dados entre a CPU e a memória leva ao gargalo de von Neumann, a troca de dados limitada (taxa de transferência) entre a CPU e a memória em relação à quantidade de memória. Na maioria dos computadores modernos, a troca de dados entre o processador e a memória é muito menor do que a taxa com que o processador pode trabalhar. Isso limita seriamente a velocidade eficaz de processamento, principalmente quando o processador é exigido para realizar o processamento de grandes quantidades de dados. A CPU é constantemente forçada a esperar por dados que precisam ser transferidos para, ou a partir da, memória. Como a velocidade da CPU e o tamanho da memória têm aumentado muito mais rapidamente que a taxa de transferência entre eles, o gargalo se tornou mais um problema, um problema cuja gravidade aumenta com cada geração de CPU.

O termo “gargalo de von Neumann” foi cunhado por John Backus em sua palestra Award 1977 ACM Turing. Segundo Backus:

Certamente deve haver uma maneira menos primitiva de se fazer grandes alterações na memória, do que empurrando um grande número de palavras, de um lado a outro, do gargalo de von Neumann. Não é somente um gargalo literal para o tráfego de dados, mas, o mais importante, é um gargalo intelectual que nos tem mantido atados a um pensamento de “uma palavra por vez” em vez de encorajar-nos a pensar em termos unidades conceituais maiores. Assim, a programação é basicamente o planejamento e detalhamento do enorme tráfego de palavras através do gargalo de von Neumann e grande parte desse tráfego não diz respeito aos dados propriamente ditos, e sim onde esses dados são encontrados.

O problema de desempenho pode ser aliviado (até certo ponto) por diversos mecanismos. Colocando uma memória cache entre o processador e a memória principal, proporcionando caches separados com os caminhos de acesso separado para dados e instruções (a chamada arquitetura Harvard Modificada), utilizando um algoritmo preditor de salto e lógica. O problema também pode ser contornado usando um pouco de computação paralela, por exemplo a arquitetura NUMA, esta abordagem é geralmente utilizada em supercomputadores.

### **Arquitetura de Havard**

A Arquitetura de Harvard baseia-se em um conceito mais recente que a de Von Neumann, tendo surgido da necessidade de se pôr o microcontrolador para trabalhar mais rápido. É uma arquitetura de computador que se distingue das outras por possuir duas memórias diferentes e independentes em termos de barramento e ligação ao processador. É utilizada nos microcontroladores PIC. Tem, como principal característica, o acesso à memória de dados de modo separado em relação à memória de programa.

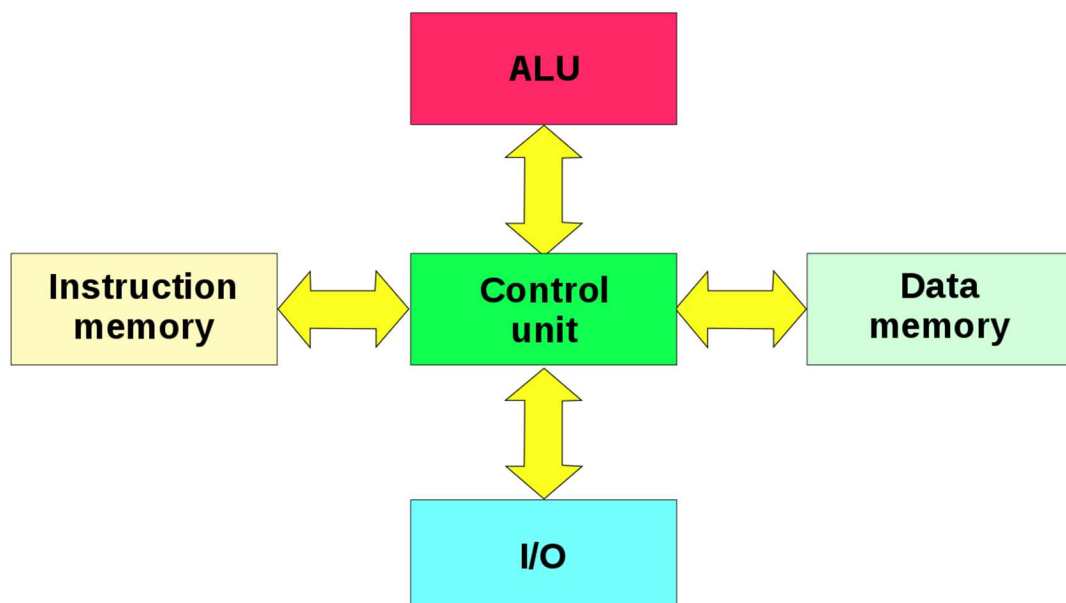
Baseada também na separação de barramentos de dados das memórias onde estão as instruções de programa e das memórias de dados, permite que um processador possa acessar as duas simultaneamente, obtendo um desempenho melhor do que a da Arquitetura de von Neumann, pois pode buscar uma nova instrução enquanto executa outra.

A principal vantagem dessa arquitetura é que a leitura de instruções e de alguns tipos de operandos pode ser feita ao mesmo tempo em que a execução das instruções (tempo  $T_{cy}$ ). Isso significa que o sistema fica todo o tempo executando instruções, o que acarreta um significativo ganho de velocidade. Enquanto uma instrução está sendo executada, a seguinte está sendo lida. Esse processo é conhecido como pipelining (canalização).

A arquitetura Harvard também possui um repertório com menos instruções que a de von Neumann, e essas são executadas apenas num único ciclo de relógio.

Arquiteturas de Harvard são normalmente utilizadas em quaisquer sistemas especializados ou para usos específicos. É utilizado em processamento de sinal digital especializados (DSP), normalmente por produtos de áudio e vídeo de transformação. Ele também é usado em muitos pequenos microcontroladores utilizados em aplicações eletrônicas, tais como os processadores ARM (sigla em inglês para Advanced RISC Machine) base de muitos fabricantes.

Os microcontroladores com arquitetura Harvard são também conhecidos como “microcontroladores RISC” (Computador com Conjunto Reduzido de Instruções), e os microcontroladores com uma arquitetura Von-Neumann, de “microcontroladores CISC” (Computador com um Conjunto Complexo de Instruções). Porém, atualmente as linhas CISC e RISC é muito tênue, não existindo mais essa diferenciação, visto que os processadores atuais têm instruções complexas ao mesmo tempo que buscam minimizar o tempo de execução.



#### Arquitetura de Harvard CPUs CISC

O projeto do Conjunto de Instruções inicia com a escolha de uma entre duas abordagens, a abordagem RISC e a CISC. O termo RISC é a abreviação de Reduced Instruction Set Computer, ou Computador de Conjunto de Instruções Reduzido e CISC vem de Complex Instruction Set Computer, ou Computador de

Conjunto de Instruções Complexo. Um computador RISC parte do pressuposto de que um conjunto simples de instruções vai resultar numa Unidade de Controle simples, barata e rápida. Já os computadores CISC visam criar arquiteturas complexas o bastante a ponto de facilitar a construção dos compiladores, assim, programas complexos são compilados em programas de máquina mais curtos. Com programas mais curtos, os computadores CISC precisariam acessar menos a memória para buscar instruções e seriam mais rápidos.

A Tabela “Arquiteturas RISC x CISC” resume as principais características dos computadores RISC em comparação com os CISC. Os processadores RISC geralmente adotam arquiteturas mais simples e que acessam menos a memória, em favor do acesso aos registradores. A arquitetura Registrador-Registrador é mais adotada, enquanto os computadores CISC utilizam arquiteturas Registrador-Memória.

Arquiteturas RISC x CISC

Características	RISC	CISC
Arquitetura	Registrador-Registrador	Registrador-Memória
Tipos de Dados	Pouca variedade	Muito variada
Formato das Instruções	Instruções poucos endereços	Instruções com muitos endereços
Modo de Endereçamento	Pouca variedade	Muita variedade
Estágios de Pipeline	Entre 4 e 10	Entre 20 e 30
Acesso aos dados	Via registradores	Via memória

Como as arquiteturas RISC visam Unidades de Controle mais simples, rápidas e baratas, elas geralmente optam por instruções mais simples possível, com pouca variedade e com poucos endereços. A pouca variedade dos tipos de instrução e dos modos de endereçamento, além de demandar uma Unidade de Controle mais simples, também traz outro importante benefício, que é a previsibilidade. Como as intrusões variam pouco de uma para outra, é mais fácil

para a Unidade de Controle prever quantos ciclos serão necessários para executá-las. Esta previsibilidade traz benefícios diretos para o ganho de desempenho com o Pipeline. Ao saber quantos ciclos serão necessários para executar um estágio de uma instrução, a Unidade de Controle saberá exatamente quando será possível iniciar o estágio de uma próxima instrução.

Já as arquiteturas CISC investem em Unidades de Controle poderosas e capazes de executar tarefas complexas como a Execução Fora de Ordem e a Execução Superescalar. Na execução Fora de Ordem, a Unidade de Controle analisa uma sequência de instruções ao mesmo tempo. Muitas vezes há dependências entre uma instrução e a seguinte, impossibilitando que elas sejam executadas em Pipeline. Assim, a Unidade de Controle busca outras instruções para serem executadas que não são as próximas da sequência e que não sejam dependentes das instruções atualmente executadas. Isso faz com que um programa não seja executado na mesma ordem em que foi compilado. A Execução Superescalar é a organização do processador em diversas unidades de execução, como Unidades de Pontos Flutuante e Unidades de Inteiros. Essas unidades trabalham simultaneamente. Enquanto uma instrução é executada por uma das unidades de inteiros, outra pode ser executada por uma das unidades de Pontos Flutuantes. Com a execução Fora de Ordem junto com a Superescalar, instruções que não estão na sequência definida podem ser executadas para evitar que as unidades de execução fiquem ociosas.

Atualmente não se pode afirmar com 100% de certeza que um processador utiliza apenas a arquitetura CISC ou RISC, pois os modelos atuais de processadores abrigam as características de ambas as arquiteturas. Processadores ARM usados em celulares são um exemplo de uso da arquitetura RISC, outro exemplo de uso dessa arquitetura é em consoles como o Nintendo 64 e o Playstation.

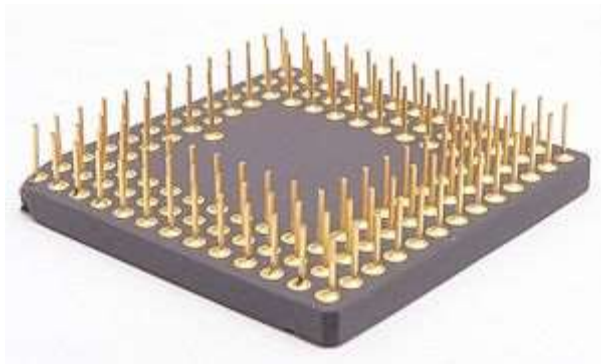
Com o passar dos anos, tanto a Intel quanto a AMD perceberam que usar alguns conceitos da arquitetura RISC em seus processadores poderia ajudá-las a criar processadores mais rápidos. Porém, ao mesmo tempo, existia a necessidade de continuar criando processadores compatíveis com os antigos. A ideia então passou a ser construir chips híbridos, que fossem capazes de executar as instruções x86, sendo compatíveis com todos os programas, mas ao mesmo tempo comportando-se internamente como chips RISC, quebrando estas instruções complexas em instruções simples, que podem ser processadas por seu núcleo RISC.

## Encapsulamento dos processadores

Já aprendemos sobre o processo de fabricação de um microprocessador e estudamos como o microprocessador é pequeno e que para proteger e tornar

possível a manipulação segura do microprocessador o die (nome da patilha de silício que contém as instruções que a transformam em um processador). Na prática o processo de encapsulamento afeta diretamente o técnico de computadores pois, este encapsulamento irá determinar que tipo de placa mãe pode-se encaixar o microprocessador. Abaixo, lista-se os formatos utilizados para microprocessadores comerciais ao longo do tempo.

**PGA:** sigla de Pin Grid Array (algo como “matriz de pinos”), esse é um tipo de encapsulamento que faz com que o processador utilize pinos de contato que devem ser inseridos em um encaixe adequado na placa-mãe do computador (ver soquete, logo abaixo). Seu material básico pode ser cerâmico (Ceramic Pin Grid Array – CPGA) ou plástico (Plastic Pin Grid Array – PPGA). Há também um tipo chamado Flip Chip Pin Grid Array (FC-PGA) onde a pastilha fica parcialmente exposto na parte superior do chip.



Pin Grid Array

**SECC:** sigla para Single Edge Contact Cartridge, este tipo faz com que o processador utilize um encaixe linear (ligeiramente semelhante aos slots de memória, por exemplo) ao invés de contatos em formato de pinos. Para isso, o processador é montado dentro de uma espécie de cartucho.

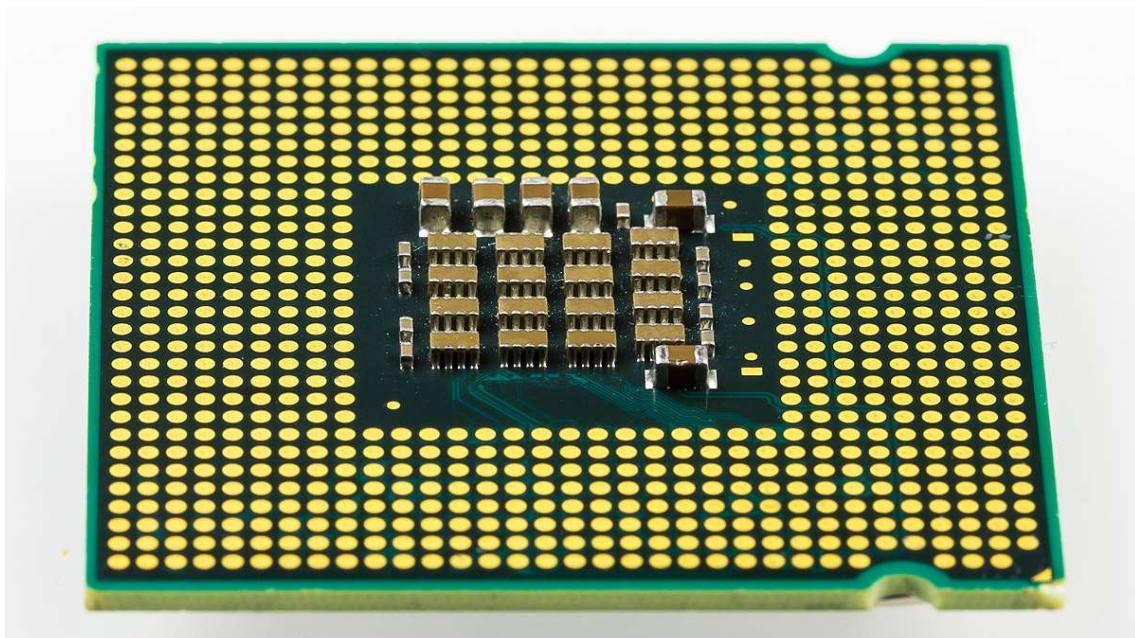




[www.cpu-world.com](http://www.cpu-world.com)

### Single Edge Contact Cartridge

**LGA:** sigla para Land Grid Array, esse é um padrão recente da Intel. Tem alguma semelhança com os padrões PGA, tendo como principal diferença o fato de que os processadores não utilizam pinos de contato em sua parte inferior, mas sim pontos metálicos. Quando o processador é encaixado na placa-mãe, esses pontos ficam em contato com pinos existentes no soquete (lembrando que nos padrões PGA há furos ao invés de pinos no soquete).



### Land Grid Array

Abaixo exibe-se uma tabela retirada da Wikipédia que ilustra bem a grande quantidade de encapsulamentos que existem/existiram nos microprocessadores usados em computadores.



## Dissipador de energia térmica

Um dissipador térmico, dissipador de energia térmica ou promotor de calor, mais conhecido – de forma pouco adequada – por dissipador de calor, é um objeto de metal geralmente feito de cobre ou alumínio, que, pelo fenômeno da condução térmica, busca maximizar, via presença de uma maior área por onde um fluxo térmico possa ocorrer, a taxa de dissipação térmica – ou seja, de calor – entre qualquer superfície com a qual esteja em contato térmico e o ambiente externo. Dissipadores térmicos têm por objetivo garantir a integridade de equipamentos que podem se danificar caso a expressiva quantidade de energia térmica gerada durante seus funcionamentos não seja deles removida e dissipada em tempo hábil.

Um dissipador térmico é essencialmente usado nos casos em que a fonte de energia térmica implique por si só uma elevada radiância térmica, a exemplo em circuitos eletrônicos com elevado grau de integração ou em componentes de hardware de equipamentos que satisfazem o requisito, como as unidades centrais de processamento de computadores e vídeo games, processadores gráficos, e outros. Em essência, o dissipador busca estabelecer uma maior condutividade térmica entre os sistemas integrados e o ambiente externo de forma que a taxa de dissipação de energia térmica requisitada ao componente não implique, entre o ambiente externo e o interno, uma diferença de temperaturas que possa comprometer a estrutura interna do componente.

Aos dissipadores dotados de uma ventoinha acoplada em suas estruturas dá-se o nome de cooler, sendo essas soluções ativas de refrigeração, enquanto os dissipadores sem ventoinha são passivos nesse aspecto. Os dissipadores dotados de ventoinhas propiciam a dissipação de energia térmica de forma muito mais eficiente que os dissipadores passivos, que contam apenas com o fenômeno de convecção térmica para auxiliá-los na tarefa.

### **Dissipadores**

### **passivos**

Os dissipadores passivos não são dotados de ventoinhas e por isso não tem a capacidade de resfriar superfícies que gerem grande quantidade de calor. Em equipamentos de hardware são usados em chips que geram pouco calor, como chipsets e controladoras. Os mesmos possuem vantagens como não gerar ruído e não consumir eletricidade.



Dissipador Térmico Passivo

### Dissipadores

### ativos

Dissipadores ativos ou coolers tem uma capacidade de refrigeração muito melhor que o dissipador passivo, já que combinando uma maior área de dissipação e uma corrente de ar passando por essa área, é possível o calor a uma taxa maior. O fluxo intenso de ar junto às lâminas impõe temperaturas mais baixas em suas superfícies e por tal gradientes de temperatura e taxas de calor mais acentuadas do que as obtidas nos dissipadores passivos, que contam apenas com o fluxo de ar induzido pelo fenômeno de convecção térmica para tal propósito. Tem seu uso destinado a componentes que exigem grandes taxas de calor, como os processadores.

O aumento excessivo da temperatura de muitos equipamentos pode fazer os mesmos queimarem. Em processadores, por exemplo, um aumento excessivo de temperatura pode através dos diferentes índices de expansão dos metais, causar microrrupturas na superfície do chip e em seus circuitos, ou em casos extremos sua fusão. O acúmulo de energia térmica e a elevação associada na temperatura podem literalmente derreter os minúsculos circuitos do processador caso não exista um cooler instalado na maioria quase absoluta dos casos, salvo aqueles chips que desenvolvem reduzida potência elétrica, e que por isso não demandam uma solução de refrigeração muito eficiente, como chipsets, processadores de baixíssimo consumo, controladoras etc.

Entre a superfície de onde origina o calor e o dissipador deve-se utilizar algum elemento que facilite a existência de calor, geralmente uma pasta conhecida como pasta térmica – dado que nenhuma das superfícies são perfeitamente planas. Em virtude da rugosidade microscópica das superfícies, sem a presença da pasta há um grande número de pequenos pontos onde o contato entre as duas superfícies não ocorre de forma eficiente, o que, em termos práticos,

diminui a área efetiva de contato, e assim o calor, para o dissipador. A pasta térmica é utilizada com frequência em componentes de hardware, assim como a fita térmica autocolante. Estes recursos preenchem as microfraturas existentes do processo de fabricação, tanto do cooler quanto dos shim (capa protetora do die), evitando qualquer espaçamento entre a superfície do chip e a superfície do dissipador de calor.



Dissipador Térmico Ativo

### **Refrigeração Líquida (Assunto Opcional Não Entra em Provas)**

Basicamente, um Water Cooler é um sistema de refrigeração a água, utilizado em computadores que seguem o mesmo esquema dos sistemas de resfriamento geralmente utilizados em motores de automóveis (radiadores). O dispositivo também colabora no resfriamento de demais partes da placa-mãe como reguladores de tensão e chipsets além de memórias e partes da placa de vídeo como a GPU (processador gráfico). Literalmente, podemos traduzir Water Cooler como “resfriamento por água”.

Concluindo, é um sistema de resfriamento muito utilizado por entusiastas de overclocking. Este sistema veio concorrer com o já existente aircooling (resfriamento a ar), permitindo assim um melhor desempenho dos componentes, conseguindo-se maiores valores de clock nos diferentes tipos de hardware, já que a água possui capacidade calorífica maior que o ar, obtendo assim resultados superiores ao de um Air Cooler.

Um watercooler consiste nas seguintes partes:

- Bomba: força a circulação de água através de um circuito fechado composto por reservatório, mangueiras, ‘waterblocks’ e pelo radiador;

- Reservatório: responsável por alocar a “água” (fluido) do watercooler e a transição do líquido para os demais componentes por meio de mangueiras;
- Fittings/Nozzles (Os conectores): normalmente feitos de plástico ou níquel ou algum metal recoberto por níquel ou uma liga de níquel-cromo. Os diâmetros mais utilizados são os de 3/4-3/8" ou 1/2-1/4" polegadas e são conectados às mangueiras;
- Mangueiras: são responsáveis por interligar as peças do watercooler e fazer a passagem do líquido até os componentes;
- Fluido/Coolant (A Água!): é através do líquido que o sistema watercooler refrigera os componentes do computador, também chamado de fluido, é recomendável usar água destilada e se possível adicionar aditivos para radiadores que são usados em carros;
- Waterblock ou Block: são peças fabricadas em cobre ou alumínio e são responsáveis pela transferência de calor dos componentes do computador para a água. É a peça que fica sobre o processador. É ela que vai trocar calor com o processador e com a água, ou seja, o processador esquentando o block e a água esfria o block, mantendo assim o processador em uma baixa temperatura;
- Radiador (Ou dissipador): responsável pelo resfriamento da água quente que vem do block. Para otimizar a dissipação do calor, são utilizadas ventoinhas que forçam o ar contra as aletas e dutos do radiador, fazendo a troca de calor água com o ambiente;

As partes que compõem o watercooler são ligadas por mangueiras, formando um circuito que é percorrido pelo líquido, que por sua vez é impulsionado pela bomba. Assim o Block, em contato com o processador, transfere o calor do mesmo para o líquido, que circula através de seu interior.