1. As características de Hardware de Sistemas de Alto Desempenho são:

* Quanto à Memória

A memória principal de um computador paralelo pode ser:

[**compartilhada**](https://pt.wikipedia.org/wiki/Mem%C3%B3ria_compartilhada) (compartilhada entre os elementos de processamento num único [espaço de endereçamento](https://pt.wikipedia.org/wiki/Espa%C3%A7o_de_endere%C3%A7amento)) ,

ou [distribuída](https://pt.wikipedia.org/w/index.php?title=Mem%C3%B3ria_distribu%C3%ADda&action=edit&redlink=1) (cada elemento de processamento possui seu próprio espaço de endereçamento).

**Memória distribuída** se refere ao fato da memória ser logicamente distribuído, mas frequentemente isso também implica na distribuição física.

A **memória distribuída e compartilhada** é uma combinação das duas abordagens, em que cada elemento de processamento possui sua memória local e acessa a memória de outros processadores remotos. Acesso à memória local é geralmente mais rápido que acesso à memória remota.

* Quanto a CPU

A comunicação dos processadores entre si e com as memórias podem ser implementadas em hardware de várias formas, incluindo por memória compartilhada, barramento compartilhado ou uma rede interconectada. Para o último caso, deve haver uma forma de [roteamento](https://pt.wikipedia.org/wiki/Roteamento) para permitir a troca de mensagens entre nós que não estão diretamente conectados.

* Quanto a GPU

As GPUs realizam muito mais trabalho para cada unidade de energia do que as CPUs. Isso os torna essenciais para supercomputadores que, de outra forma, ultrapassariam os limites das redes elétricas de hoje. Na IA, as GPUs se tornaram a chave para uma tecnologia chamada "aprendizado profundo". O aprendizado profundo despeja grandes quantidades de dados por meio de redes neurais, treinando-as para executar tarefas muito complicadas para qualquer codificador humano descrever.

1. Esquema 1

Tipo de Classificação de Flynn segundo o esquema dado – SIMD

SIMD: Único fluxo de instruções, múltiplo fluxo de dados. Classe que representa os processadores matriciais, paralelos e associativos uma única unidade de controle que envia um fluxo de instruções para vários processadores. Os processadores recebem a mesma instrução ao mesmo tempo e atuam sobre diferentes fluxos de dados

Esquema 2

Classificação de Flynn segundo o esquema dado – MISD

MISD: Múltiplo fluxo de instruções, único fluxo de dados

Vários processadores, onde cada um recebe instruções distintas mas operam sobre o mesmo conjunto de dados.

Esquema 3

Classificação de Flynn segundo o esquema – MIMD

Múltiplo fluxo de instruções, múltiplo fluxo de dados. Vários processadores, cada um controlado por uma unidade de controle. Processadores recebem instruções diferentes e operam sob fluxo de dados diferentes, podem ser síncronos ou assíncronos.

1. No algoritmo da Migração, os dados compartilhados são enviados para o local onde eles foram requeridos permitindo assim que operações subsequentes sobre estes mesmos dados sejam executadas localmente diminuindo assim a tempo de latência.

Este algoritmo impõe a seguinte restrição: uma, e somente uma, thread pode executar operações, tanto de leitura quanto de escrita, sobre um determinado dado. Este tipo de regra é também chamado de “single reader/single writer” justificando assim a sigla que geralmente está associada a este método SRSW.

A maior intenção do algoritmo de replicação é reduzir o custo das operações de leitura partindo do pressuposto que esta é a mais comum do que a operação de escrita. Para conceder tal privilégio nas operações de leitura utiliza-se a replicação.

O primeiro problema encontrado é se existe replicação dos dados, a operação de escrita será penalizada pois exigirá um maior fluxo na rede com mensagens para invalidar as cópias que se tornaram obsoletas (ou no mínimo avisar os portadores das réplicas que tais dados não refletem os dados mais atuais).

O segundo problema, que também deve receber igual atenção, é que o algoritmo é obrigado a saber onde estão as réplicas de uma determinada página para então, no caso de uma escrita as invalidar ou informar que elas não são mais atuais. Este papel pode ser implementado por uma entidade que pode ser chamar de servidor de mapeamento.

Este algoritmo deve ser utilizado se a quantidade de leituras feitas supera em muito a quantidade de gravações.

1. A arquitetura Híbrida fornece a mesma latência geral para os processadores que acessam a memória. Bem como, a possibilidade de cada processador ter a sua memória dedicada que elimina a latência quando a memória local é acessada.