

## Resumo

Fichamento do artigo Desafios do Processamento de Alto Desempenho, por Philippe Olivier Alexandre Navaux e Matheus da Silva Serpa da Universidade Federal do Rio Grande do Sul - UFRGS.

## HPC

High Performance Computing (HPC), ou Computação de Alto de Desempenho, diz respeito ao ramo encarregado do desenvolvimento, manutenção e arquitetura das máquinas com o maior poder de processamento do mundo. Um nicho antes populado por uma minoria de cientistas entusiastas, respondendo a demandas extremamente específicas, agora tem um papel central na construção dos paradigmas modernos da computação.

Especialmente, em resposta às demandas de poder de processamento da informática relativa ao desenvolvimento de tecnologias correlatas a Inteligência artificial e no sempre crescente Big Data.

## Demandas do Big Data

O conceito de Big Data tem sua origem atada ao aprimoramento e popularização de mecanismos de telemetria atrelados a sistemas computacionais. Entretanto, a própria definição deste corpo reconhece seu volume massivo como produto de ser constituído por dados brutos, sem qualquer tratamento. Este aspecto atrelado à centralidade que as informações coletadas assumiram na rentabilidade e competitividade de qualquer produto moderno.

Contudo, por definição, o Big Data não é diretamente utilizável sendo necessária alguma etapa de tratamento desses dados para quaisquer uso. Desta maneira, qualquer aplicação popular necessita, em algum nível, administrar volumes massivos de dados para implementar qualquer tipo de *feature*.

Por fim, o Big Data exige poder computacional compatível com volumes de dados em escala massiva, uma vez que ele é alçada como paradigma característico da computação contemporânea, urge-se a fronteira dos sistemas especializados em computar toda essa matéria prima como um protagonista na informática contemporânea.

## Demandas da Inteligência Artificial

Tecnologias de I.A., ou Inteligência Artificial, tem sido protagonistas na aceleração do desenvolvimento de conhecimento científicos nos últimos anos. O aperfeiçoamento de técnicas de Deep Learning permitiram análises de dados e simulações físico-químicas em velocidades inéditas, sendo essas capacidades expandidas para diversos segmentos científicos.

Todavia, essa variedade de aplicações é viável em razão dos processos intermediário característicos dos modelos de Deep Learning, mais especificamente, o processo que constitui o maior gargalo na implementação dessas tecnologias: o treinamento dessas I.A.s. Uma etapa altamente exigente em termos de poder computacional baseada também no processamento de dados já tratados, na maioria dos casos.

## Soluções para o Alto Desempenho

### Arquiteturas Heterogêneas

Uma tendência crescente que deve implicar em uma substancial integração entre tecnologias proprietárias é a elaboração de arquiteturas heterogêneas, estas são, os modelos que incorporam aceleradores, GPUs, memórias, CPUs e outros módulos incorporados nos mesmos chips, ou até, em sistemas SOC (System on a Chip), todos esses módulos em um mesmo chip. Esta paradigma arquitetônico, apesar de apresentar eficiência evidente em termos de incrementos de processamento, tem se chocado com a dificuldade de integrar modelos proprietários e de difícil interoperabilidade de diferentes fabricantes. Apesar disto, esta tendência arquitetônica tem apenas se popularizado no meio HPC nos últimos anos.

## Bibliotecas para Paralelismo

Ferramentas antigas e essenciais para HCP, as bibliotecas como CUDA e OpenMP, viabilizam paralelismo no processamento permitindo aproveitamento mais completo de todo potencial de processamento de processadores dos diversos tipos. Dado o contexto HCP, essa centralidade é agravada com os sistemas baseados em múltiplos núcleos, GPUs e placas interconectadas. Desta forma bibliotecas de paralelismo tem agora não só o trabalho de realizar o potencial de um único módulo mas sim da capacidade de múltiplas partes integradas realizarem trabalho maior que a soma direta de suas capacidades individuais.

Assim, são integrados diferentes bibliotecas responsáveis pela realização de módulos multinucleares de forma a viabilizar a cooperação e integração inteligente destes núcleos e também a intercomunicação com sistemas de memória distribuída, por exemplo, e as outras variadas partes como GPUs.

## Computação Quântica

Uma área de emergência insipiente, tem si caracterizado pela implementação de unidades de processamento de escala atômica, baseada em transistores de germânio, também denominados de processadores Qubits.

## Eficiência Energética

A implementação de sistemas de HPC é, comumente, baseada em clusters de processadores, um fator central que contribui com um consumo significativo de energia.

Assim, são implementadas soluções em favor de maior eficiência energética. Essas otimizações encopassam reestruturações de arquiteturas inteiras ou dinamização de clocks e na ativação de módulos.

Entretanto, o maior desafio de otimização no gasto de energia elétrica é, coincidentemente, encontrado no aspecto mais ineficiente: o gasto de energia estática, aquela necessária para manter o sistema ligado, mas que não é consequência direta do processamento.

## Localidade dos dados

Um outro obstáculo ao processamento de dados, é verificado na transmissão destes dentre os diferentes níveis de memória, onde verifica-se um dos maiores desperdício de energia em um computador.

Assim, existe um esforço considerável atrelado a métodos de aprimoramento de localidade destes dados. Para que não seja gasta mais energia na captação dos dados armazenados do que na computação destes dados.

## Resiliência

A Resiliência de um dado sistema é sua capacidade de perseguir com processamento apesar de variações que ocorram. Esta característica não só se limita a não interrupção do sistema em caso de uma falha, mas sim engloba toda uma cadeia de operações complexas que isolam, diagnosticam e reparam uma dada interrupção ou flutuação de desempenho.

No ambiente de HCP, a resiliência de um sistema é uma métrica especialmente importante dada a escala das operações efetuadas e das integrações que compõem um dado super-computador, que implicam em certeza matemática de um dado número de falhas. ESSa complexidade das maquinas de ponta também significa que essas falhas em sistemas de baixa resiliência podem ser especialmente danosas não só em termos de prejuízo do processamento mas também da função atrelado aos dados e processos encarregados a máquina.

## Cloud Computing

Cloud Computing, ou Computação em Nuvem, tem também se expandido no segmento de HCP. Com a demanda pelas capacidades desse segmento da computação em crescimento, verifica-se agora serviços de Cloud Computing com opções de alternativas HCP.