A ARQUITETURA DO SUPERCOMPUTADOR SDUMONT E OS DESAFIOS DA PESQUISA BRASILEIRA NA ÁREA DE COMPUTAÇÃO DE ALTO DESEMPENHO

Carla Osthoff^{1,2}, Stiw Herrera^{1,2}, Thiago Teixeira^{1,2}, Micaella Coelho^{1,2},
Mateus Melo^{1,2}, Gabriel Costa^{1,2}, Frederico Cabral^{1,2}, Maria E. Welter^{1,2},
Bruno Fagundes^{1,2}, André Carneiro^{1,2}, Roberto Souto^{1,2}, Lucas Cruz²
Guilherme Freire², Luiz Gadelha^{1,2}, Kary Ocaña²

¹Centro Nacional de Processamento de Alto Desempenho (CENAPAD)

{osthoff, stiw, tteixeira, micaella, msmello, gcosta, fcabral}@lncc.br

{bfagundes, acarneiro, rpsouto, lgadelha}@lncc.br

²Laboratório Nacional de Computação Científica (LNCC) - Petrópolis - RJ – Brasil

{lcruz, gfreire, karyann}@lncc.br

Abstract. *The Santos Dumont supercomputer is the fourth largest in Latin America according to the TOP500 list, summarizing 5.1 petaflops. It is hosted at the National Laboratory for Scientific Computing (Petrópolis, Brazil), a research unit of the Brazilian Ministry of Science, Technology and Innovation. This work briefly presents the main architectural characteristics of the supercomputer, its areas and research projects, scientific HPC software and gateways using it, ongoing work to optimizing its execution, and finally useful information to potential users for Santos Dumont project submissions.*

Resumo. *O supercomputador Santos Dumont é o quarto maior da América Latina segundo a lista TOP500 com um total de 5.1 petaflops. Ele está hospedado no Laboratório Nacional de Computação Científica (Petrópolis, Brasil), uma unidade de pesquisa do Ministério da Ciência, Tecnologia e Inovação do Brasil. Neste trabalho apresentamos as principais características do supercomputador, suas áreas e projetos de pesquisa envolvidos, software e portais científicos acoplados nele, trabalhos em andamento do nosso grupo para otimizar as execuções e informações gerais a usuários potenciais que pretendam submeter projetos nas chamadas para o uso do Santos Dumont.*

1. Introdução

Santos Dumont (SDumont) é o maior supercomputador disponível para a pesquisa científica da América Latina. É administrado pelo Laboratório Nacional de Computação Científica, sob encomenda do Ministério da Ciência, Tecnologia, Inovação e Comunicação (MCTI) do Brasil. A plataforma SDumont atende atualmente 133 projetos científicos pertencentes a 28 áreas de pesquisa e é a primeira plataforma *petascale* da América Latina. Ele realiza chamadas para apoio a projetos científicos que pretendam usufruir da sua capacidade, como por exemplo, seu papel como principal plataforma de execução nas crises epidemiológicas causadas pelo Zika em 2016 e atualmente o COVID-19, consideradas

CPUs com arquitetura *multi-core*. Há, no entanto, uma quantidade adicional significativa de nós que além das CPUs *multi-core*, contém outros tipos de dispositivos com a chamada arquitetura *many-core* como as GPU e MIC. SDumont é dotado de um nó diferenciado, o MESCA2 que possui um número elevado de núcleos (240 núcleus) e arquitetura de memória compartilhada de grande capacidade (6 Tb em um único espaço de endereçamento). Além disso, existe um nó especialmente projetado para aplicações de Inteligência Artificial (Aprendizado de Máquinas e Aprendizado de Máquinas Profunda) que dispõe de 8 GPUs NVIDIA Tesla V100-16Gb com Nvlink, totalizando 40.960 *CUDAcore* e 5120 *Tensor-core*. Uma descrição detalhada da configuração dos nós do SDumont pode ser obtida em <https://sdumont.lncc.br/> e é apresentada a seguir:

504 nós de computação B710 (*thin node*), onde cada nó possui 2 CPU Intel Xeon E5-2695v2 Ivy Bridge (12c @2,4GHz) e 64Gb de memória RAM. 198 nós de computação B715 (*thin node*) com GPUs K40, onde cada nó possui 2 x CPU Intel Xeon E5-2695v2 Ivy Bridge (12c @2,4GHz) e 64Gb de memória RAM. 54 nós de computação B715 (*thin node*) com XEON PHI, onde cada nó possui 2 x CPU Intel Xeon E5-2695v2 Ivy Bridge (12c @2,4GHz) e 64Gb de memória RAM. 1 nó de computação MESCA 2 com memória compartilhada, 16 x CPU Intel Xeon Ivy Bridge (15c @2,4GHz) e 6TB de memória RAM. 246 nós computacionais (CPU), cada um com 2x Intel Xeon Cascade Lake Gold 6252 e 384Gb de memória RAM. 36 nós computacionais (CPU), cada um com 2x Intel Xeon Cascade Lake Gold 6252 e 768 Gb de memória RAM. 94 nós computacionais (GPUs), cada um com 2x Intel Xeon Cascade Lake Gold 6252, 4x NVIDIA Volta V100 GPU e 384Gb de memória RAM.

Os 1.134 nós do SDumont são interligados por uma rede de interconexão Infiniband proporcionando alto rendimento e baixa latência tanto para comunicação entre os processos quanto para o acesso ao sistema de arquivos. Por fim, o SDumont dispõe de um sistema de arquivos paralelo Lustre, integrado a rede Infiniband, com capacidade bruta de armazenamento da ordem de 1,7 PBytes, bem como um sistema de arquivos secundário com capacidade bruta de 640 TBytes.

3. Pesquisas desenvolvidas para uso eficiente do SDumont

As aplicações executadas em infraestruturas de CAD geralmente inserem ou produzem grande quantidade de dados heterogêneos. Isso geralmente é realizado executando operações de E/S em um *Parallel File System* (PFS). Como o desempenho de E/S é um fator limitante para muitas aplicações científicas, há uma necessidade importante de avaliar a E/S quando usada a infraestrutura do SDumont. Considerando os altos custos financeiros associados à aquisição e manutenção de um supercomputador, o uso eficiente da máquina é de suma importância. Para fornecer informações valiosas que podem ser levadas em consideração pelos administradores para promover o uso eficiente do sistema e para orientar futuras atualizações, estudamos a carga de trabalho de E/S do SDumont. Ao ajudar os usuários a obter melhor desempenho para seus aplicativos, promovemos um melhor uso da máquina [Bez et al. 2020]. Existe o desafio de ajudar os usuários a usar a máquina de maneira eficiente o que depende muitas vezes do tipo de dados, experimentos e aplicações a serem acoplados na infraestrutura do SDumont.

LNCC possui um grupo de pesquisa de CAD que colabora com os usuários do SDumont para ajudar a melhorar a eficiência do código [Cabral et al. 2018b],

[Cabral et al. 2018a] que é de extrema importância pois o uso de uma infraestrutura tão especializada como é no caso do SDumont pode ser difícil, especialmente se usado por não especialistas em CAD, acarretando em um processamento de dados ineficiente, quando ocorre um uso não otimizado das aplicações no nível de CAD e gerando sobrecarga nas filas de espera para os usuários da máquina. As publicações desenvolvidas pelo CENAPAD podem ser acessadas no site (<http://www.cenapad-rj.lncc.br/publicacoes/>). De forma a melhorar a utilização dos usuários, LNCC oferece uma Escola de Verão de "Computação de Alto Desempenho (<https://www.cenapad-rj.lncc.br/cursos-e-seminarios/>)", em colaboração com institutos de pesquisa nacionais e internacionais que tem como objetivo para disseminar conhecimento na área de programação para além dos usuários do SDumont.

3.1. Portais Científicos

O SDumont disponibiliza diversas aplicações e softwares, dentre deles, portais para a comunidade científica, que cobrem diversas áreas. Alguns exemplos desses portais são BRAMS, Gaussina, CAM-Grid, Gdick, TruRNG, Dance, Profranger, GROMACS, Docthor e Bioinfo-Portal [Ocaña et al. 2020]. Dentre os portais mais acessados, gostaríamos de destacar os portais Bioinfo-Portal e Dockthor, que serão mais detalhados a seguir. Maiores detalhes sobre os portais podem ser encontrados no site do LNCC (<https://www.lncc.br/>) ou do SDumont (<https://sdumont.lncc.br/>).

Portal Bioinfo: As instituições científicas latino-americanas têm uma comunidade de bioinformática ativa altamente diversificada. A Rede Brasileira de Bioinformática (RNBio) visa fortalecer projetos de pesquisa multi-institucionais com a formação de recursos humanos especializados. RNBio tem colaborações científicas com LNCC no desenvolvimento de portais ou chamados também de *gateways* científicos para bioinformática, como o Bioinfo-Portal (www.bioinfo.lncc.br). SDumont fornece a infraestrutura computacional combinando a adaptação de metodologias analíticas de Inteligência Artificial (IA) eficientes para análises de dados e execuções em larga escala de aplicativos e fluxos de trabalho especializados [Ocaña et al. 2019a]. Os portais científicos são soluções de *software* de infraestrutura eletrônica que trazem a integração de dados reutilizáveis e técnicas especializadas por meio de servidores da Web, ao mesmo tempo em que escondem a complexidade de uso dos recursos de CAD subjacentes para comunidades de pesquisa. No entanto, os requisitos para planejar e projetar portais científicos exigem esforços para implementar e manter infraestruturas extensíveis, escaláveis, fáceis de usar e sustentáveis. A plataforma multiusuário do Bioinfo-Portal fornece uma interface personalizada fácil de usar e que permite acessar e usufruir dos recursos de CAD especializados que suportem execuções de bioinformática paralelas e distribuídas em grande escala. Bioinfo-Portal registra o processo e a proveniência dos dados para melhor apoiar a reprodutibilidade computacional [Ocaña et al. 2019b]. A tarefa de desenho das camadas propostas do Bioinfo-Portal é um processo contínuo e iterativo, planejamos processos de melhoria contínua.

Portal Docthor: SDumont disponibiliza o servidor DockThor para a comunidade científica (www.dockthor.lncc.br) dedicado à triagem virtual de ligantes e *docking* de pequenas moléculas. DockThor usa técnicas de Inteligência Computacional para CAD desenvolvidas por pesquisadores do LNCC, que permitem experimentos de triagem virtual de alto rendimento para identificar compostos químicos que podem se ligar a de-

terminados alvos de drogas. SDumont forneceu a infraestrutura computacional para os esforços contínuos do LNCC na pesquisa de predição da estrutura de proteínas, permitindo que a instituição participasse do experimento bi anual mundial, a partir de 2016, para a Avaliação Crítica da Predição da Estrutura da Proteína (CASP). Recentemente, a expansão do supercomputador reforçou os esforços do LNCC para participar da comunidade iniciativa de modelagem de estrutura SARS-2-CoV em toda a extensão organizada pelo *Protein Structure Prediction Center*.

4. Submissão de Projetos de Pesquisa

SINAPAD é uma infraestrutura de CAD disponível para instituições brasileiras, públicas ou privadas, que visam oferecer suporte a atividades de ensino, pesquisa e desenvolvimento no Brasil. Todo pesquisador vinculado a uma instituição brasileira, com um problema relevante e que demande um sistema computacional de larga escala, pode submeter propostas para utilizar os recursos computacionais do SINAPAD. SDumont é o nó principal dessa infraestrutura e, por esse motivo, apresenta um processo de avaliação de propostas próprio, que pode ser acessado no site do LNCC (www.lncc.br). A Figura 1b apresenta as 28 áreas de pesquisa que atualmente utilizam atualmente o SDumont.

Referências

- Bez, J. L., Carneiro, A. R., Pavan, P. J., Girelli, V. S., Boito, F. Z., Fagundes, B. A., Osthoff, C., da Silva Dias, P. L., Méhaut, J.-F., and Navaux, P. O. (2020). I/o performance of the santos dumont supercomputer. *The International Journal of High Performance Computing Applications*, 34(2):227–245.
- Cabral, F., Osthoff, C., Souto, R. P., Costa, G. P., de Oliveira, S. L. G., Brandão, D. N., and Kischinhevsky, M. (2018a). An improved openmp implementation of the tvd–hopmoc method based on a cluster of points. In *International Conference on Vector and Parallel Processing*, pages 132–145. Springer.
- Cabral, F. L., Osthoff, C., Souto, R. P., Costa, G. P., de Oliveira, S. L. G., Brandão, D., and Kischinhevsky, M. (2018b). Fine-tuning an openmp-based tvd–hopmoc method using intel® parallel studio xe tools on intel® xeon® architectures. In *Latin American High Performance Computing Conference*, pages 194–209. Springer.
- Ocaña, K., Galheigo, M., Osthoff, C., Gadelha, L., Gomes, A. T. A., De Oliveira, D., Porto, F., and Vasconcelos, A. T. (2019a). Towards a science gateway for bioinformatics: experiences in the brazilian system of high performance computing. In *2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, pages 638–647. IEEE.
- Ocaña, K., Osthoff, C., Coelho, M., Galheigo, M., Canuto, I., de Oliveira, D., and de Oliveira, D. (2019b). Performance evaluation of parallel inference of large phylogenetic trees in santos dumont supercomputer: A practical approach. In *Latin American High Performance Computing Conference*, pages 448–463. Springer.
- Ocaña, K. A., Galheigo, M., Osthoff, C., Gadelha Jr, L. M., Porto, F., Gomes, A. T. A., de Oliveira, D., and Vasconcelos, A. T. (2020). Bioinfoportal: A scientific gateway for integrating bioinformatics applications on the brazilian national high-performance computing network. *Future Generation Computer Systems*, 107:192–214.