

Desempenho e Eficiência de HPC em Ambientes Serverless: Uma Revisão Integrativa

Adriano Prado Cavalheiro, Brenda Medeiros Lopes, Claudio Schepke

¹Laboratório de Estudos Avançados em Computação (LEA)
Universidade Federal do Pampa (UNIPAMPA) – Alegrete – RS – Brazil

{adrianocavalheiro.aluno,brendalopes.aluno,claudioschepke}@unipampa.edu.br

Abstract. *This paper presents an integrative literature review on the application of High-Performance Computing (HPC) in serverless environments, highlighting the current state of research and identifying possible future directions. The systematic review was conducted in several academic databases, using a search string that combined terms related to integrating HPC with serverless, focusing on performance, scalability, and efficiency. We applied inclusion and exclusion criteria to select the most relevant studies. The results indicate that combining HPC with serverless offers benefits such as elasticity and cost optimizations but still faces challenges such as remote invocation latency and efficient resource management. Technologies such as RDMA and I/O optimizations have the potential to mitigate some of these issues. The integration of HPC with serverless shows potential opportunities for optimization and innovation but requires additional advances to overcome existing limitations.*

Resumo. *Este artigo apresenta uma revisão integrativa da literatura sobre a aplicação de High Performance Computing (HPC) em ambientes serverless, destacando o estado atual das pesquisas e identificando possíveis direções futuras. A revisão foi conduzida em diversas bases de dados acadêmicas, utilizando uma string de pesquisa que combinou termos relacionados à integração de HPC com serverless, com foco em desempenho, escalabilidade e eficiência. Critérios de inclusão e exclusão foram aplicados para selecionar os estudos mais relevantes. Os resultados apontam que a combinação de HPC com serverless oferece benefícios como elasticidade e otimização de custos, mas ainda enfrenta desafios, como latência de invocações remotas e gerenciamento eficiente de recursos. Tecnologias como RDMA e otimizações de I/O têm potencial para mitigar alguns desses problemas. A integração de HPC com serverless apresenta potencial, com oportunidades para otimização e inovação, mas requer avanços adicionais para superar as limitações existentes.*

1. Introdução

A Computação de Alto Desempenho (*High-Performance Computing* - HPC) agrega recursos computacionais para processar grandes volumes de dados e executar cálculos complexos em alta velocidade. Essa abordagem é determinante em aplicações como simulações de sistemas de energia, previsão meteorológica e análise financeira [Falcão 1996]. HPC viabiliza a resolução de problemas que seriam inviáveis em infraestruturas convencionais, utilizando supercomputadores e *clusters* de computadores [NetApp 2024, Cloud 2024].

No entanto, essa abordagem enfrenta desafios, como a complexidade no gerenciamento de sistemas e o alto consumo de energia [Navaux and da Silva Serpa 2021]. Entre os principais desafios de HPC estão a necessidade de uma infraestrutura robusta, que exige investimentos substanciais em *hardware* e manutenção. A Computação *Serverless* surge como uma alternativa potencial para mitigar esses problemas, automatizando a alocação de recursos e eliminando a necessidade de gerenciamento direto de servidores [Cavalheiro and Schepke 2023]. A capacidade de ajustar dinamicamente os recursos conforme a demanda faz de *serverless* uma solução promissora, aumentando a eficiência operacional e reduzindo os custos associados à infraestrutura física.

Com a crescente adoção dessa tecnologia em diversas áreas, torna-se relevante investigar seu potencial no contexto da HPC. Este estudo é importante para avaliar como a integração dessas tecnologias pode otimizar operações em ambientes de nuvem, oferecendo fundamentação para futuras implementações e avanços científicos.

A metodologia deste estudo utiliza uma revisão integrativa da literatura para analisar a viabilidade da integração de HPC com Computação *Serverless*. Essa abordagem permite reunir e avaliar estudos teóricos e empíricos relevantes, a fim de sintetizar as evidências existentes [de Souza et al. 2010]. O foco é identificar padrões, desafios e oportunidades nessa interseção tecnológica, proporcionando compreensão das tendências e implicações futuras dessa combinação em ambientes de nuvem. A questão de pesquisa que guia este estudo é: **Como a Computação *Serverless* pode potencializar o desempenho e a eficiência da Computação de Alto Desempenho (HPC) em ambientes de nuvem?** Esse método é aplicado para compreender as tendências atuais e antecipar direções futuras na interação entre HPC e Computação *Serverless*.

Este estudo tem como principal propósito compreender como a integração entre HPC e a Computação *Serverless* pode influenciar o desempenho e a eficiência de operações em ambientes de nuvem. Para alcançar esse objetivo, o trabalho foi desdobrado em objetivos específicos, que visam abordar os diferentes aspectos dessa integração, desde os desafios tecnológicos até a aplicação prática e a formulação de recomendações baseadas nas evidências obtidas. Entre os objetivos específicos, tem-se:

- Identificar os principais desafios da integração de HPC com *serverless*.
- Investigar casos de uso de HPC implementado em ambientes *serverless*.
- Examinar as limitações tecnológicas na aplicação de *serverless* em HPC.
- Sintetizar as evidências encontradas para propor recomendações sobre a adoção de *serverless* em HPC.

Esses objetivos buscam responder à questão de pesquisa e apoiar futuros desenvolvimentos na interseção dessas tecnologias.

A estrutura do artigo inclui a A apresentação dos resultados da revisão bibliográfica ocorre na Seção 3, analisando as contribuições e limitações dos estudos sobre a viabilidade de HPC em ambientes *serverless*. Por fim, a Seção 4 sintetiza as descobertas e sugere direções para trabalhos futuros, que serão detalhadas na Seção 5.

2. Metodologia

Este estudo utiliza uma revisão integrativa da literatura para reunir e sintetizar evidências de pesquisas teóricas e empíricas sobre a integração entre *serverless* e HPC. Segundo

Broome et al. 2000 e Whittemore and Knafl 2005, essa abordagem permite o desenvolvimento conceitual em temas complexos ao combinar diferentes tipos de evidências. O estudo segue um processo sistemático que inclui a aplicação de critérios de inclusão e exclusão, estratégias de busca, e métodos de análise e síntese dos dados.

2.1. Critérios de Inclusão e Exclusão

Foram incluídos estudos focados na integração entre HPC e Computação *Serverless*, que apresentassem dados empíricos ou teóricos relevantes. A opção por publicações predominantemente em inglês deve-se à predominância da literatura internacional sobre o tema. Foram considerados artigos em português, mas nenhum relevante foi encontrado em bases brasileiras. Excluíram-se estudos em outros idiomas, publicados antes de 2010, ou que não abordassem HPC e *serverless* nos resumos ou títulos, bem como artigos indisponíveis na íntegra.

2.2. Estratégias de Busca

As buscas foram realizadas nas bases ACM Digital Library, Google Scholar, IEEE Xplore, Science Direct e Springer Link. A mesma *string* de busca foi utilizada em todas as bases, conforme descrito a seguir: (“*High Performance Computing*” OR “*HPC*”) AND (“*serverless*” OR “*serverless computing*”). Os critérios de inclusão e exclusão foram aplicados, sempre que possível, por meio dos filtros oferecidos pelas próprias bases. Isso incluiu a aplicação de filtros para idioma e data de publicação em bases como Google Scholar e IEEE Xplore, onde se pôde limitar os resultados a apenas a artigos em inglês e publicados após 2010. Dessa forma, assegurou-se que a busca fosse alinhada aos objetivos da revisão.

2.3. Processo de Seleção de Estudos

A busca inicial resultou em 751 artigos. Após a análise de títulos e resumos, foram selecionados 60 estudos. Em uma segunda triagem, com verificação de duplicatas e relevância, chegaram-se a 23 estudos finais. Este processo visou a inclusão de estudos que contribuíssem para os objetivos da revisão.

2.4. Limitações do Estudo

Este estudo pode apresentar vieses na seleção dos artigos, devido à natureza subjetiva da triagem. Para mitigar esses vieses, utilizou-se a ferramenta StArt¹, que facilitou a aplicação consistente dos critérios e trouxe transparência ao processo. Embora a ferramenta tenha ajudado na organização e seleção, limitações relacionadas à qualidade e disponibilidade dos estudos podem ter influenciado as conclusões, dado que a literatura sobre HPC e *serverless* ainda está em desenvolvimento.

2.5. Estruturação e Análise dos Dados

Os estudos foram analisados quanto à qualidade metodológica e à contribuição para o tema, conforme critérios estabelecidos por Whittemore and Knafl 2005. As evidências foram integradas para construir uma visão compreensiva sobre a viabilidade da combinação de HPC e Computação *Serverless*. Essa abordagem estruturada visa avaliar a aplicabilidade e os desafios dessa integração no contexto de computação avançada.

¹<https://www.lapes.ufscar.br/resources/tools-1/start-1>

3. Resultados da Revisão

A análise dos 23 estudos selecionados nesta revisão integrativa explora como a Computação *Serverless* pode contribuir para o desempenho e a eficiência da HPC em ambientes de nuvem. Os resultados encontrados a partir dos estudos analisados são discutidos a seguir, agrupados em quatro categorias principais: desempenho e eficiência de recursos, escalabilidade e elasticidade, desafios técnicos e limitações, e viabilidade econômica.

3.1. Desempenho e Eficiência de Recursos

Os estudos indicam que a adoção de plataformas *serverless* pode melhorar o desempenho e a eficiência de recursos em certos cenários de HPC. Copik et al. 2024 evidenciam que a desagregação de recursos em ambientes *serverless* permite uma utilização mais eficiente de núcleos e aceleradores inativos em supercomputadores, atingindo uma melhor alocação dinâmica.

Além disso, Al-Haboobi and Kecskemeti 2021 demonstram que a redução de transferências de dados duplicadas em *workflows* científicos pode diminuir o tempo total de execução em até 10%, melhorando a eficiência no processamento de grandes volumes de dados. Estudos como o de Basu Roy and Tiwari 2024 destacam o papel do gerenciamento de *I/O* para otimizar o uso de recursos em ambientes de HPC, o que é essencial para a execução eficiente de *workloads* intensivos.

Para lidar com limitações de recursos, como memória e largura de banda, Liu et al. 2022 propõem um *framework* baseado em pipelines, otimizando o uso de memória e largura de banda, o que resulta em melhorias significativas na eficiência do treinamento de modelos de grande escala.

3.2. Escalabilidade e Elasticidade

A escalabilidade elástica é um dos principais atributos da Computação *Serverless* que beneficia *workloads* de HPC, especialmente em ambientes híbridos. Kehrer and Blochinger 2020 mostraram que o uso de controladores de elasticidade permite ajustar dinamicamente os recursos de acordo com a carga de trabalho, evitando o provisionamento estático e melhorando a utilização em sistemas de busca paralela.

Além disso, Petrosyan and Astsatryan 2022 demonstraram que a integração de *serverless* com *containers* e Kubernetes aumenta a escalabilidade e reduz a complexidade na gestão de *workloads* intensivos em processamento e dados, permitindo que as aplicações escalem automaticamente em resposta a picos de demanda.

3.3. Desafios Técnicos e Limitações

Apesar dos benefícios, muitos desafios técnicos precisam ser resolvidos para a adoção de *serverless* em HPC. Spillner et al. 2018 identificam que invocações remotas em plataformas *serverless* introduzem *overhead*, particularmente em cargas de trabalho paralelas e intensivas em dados, que exigem coordenação contínua entre funções.

Outro desafio está na virtualização de (*Graphics Processing Units* - GPUs), como discutido por Naranjo et al. 2020, que mostraram que a elasticidade oferecida por essas plataformas é promissora, mas limitada pela latência e infraestrutura. Estudos também apontam a variabilidade na alocação de *Central Processing Units* (CPUs) em plataformas *Function-as-a-Service* (FaaS) como um fator que afeta o desempenho de tarefas sensíveis à latência [Malla and Christensen 2020].

3.4. Viabilidade Econômica

A eficiência de custos é uma vantagem central de *serverless* para *workloads* de HPC. Malla and Christensen 2020 indicam que FaaS pode ser mais barato que IaaS para *workloads* paralelos, embora essa economia possa variar com a demanda por recursos. Em treinamento de *deep learning*, o uso de pipelines *serverless* reduz custos em até 77%, destacando o *serverless* como uma alternativa econômica em cenários específicos [Liu et al. 2022].

3.5. Síntese dos Resultados

Os resultados sugerem que a Computação *Serverless* oferece vantagens em termos de escalabilidade, eficiência de recursos e redução de custos para certos *workloads* de HPC. No entanto, desafios técnicos persistem, especialmente em latência e *overhead* de invocações remotas. Tecnologias como RDMA, que permitem acesso direto à memória de servidores remotos, podem reduzir a latência [Copik et al. 2021]. A integração com *containers* e Kubernetes também pode melhorar a virtualização de hardware [Petrosyan and Astsatryan 2022].

Assim, a adoção de *serverless* em HPC depende de avanços contínuos, especialmente para adaptar *workloads* à arquitetura. A Tabela 1 resume esses aspectos, relacionando-os aos objetivos do estudo, incluindo desafios, casos de uso, limitações tecnológicas e recomendações sobre a adoção de *serverless* em HPC.

Esses aspectos são analisados para ilustrar como a Computação *Serverless* atende ou desafia os objetivos de integração com HPC. Com essa visão, o estudo fornece uma base para identificar áreas de melhoria e explorar recomendações para adoção em contextos de HPC.

4. Conclusão

Este estudo revelou que a integração de HPC com *serverless* apresenta tanto desafios quanto oportunidades. A revisão dos estudos discutiu a escalabilidade, eficiência de custos e o desempenho dessas tecnologias em ambientes de nuvem, proporcionando uma base para avaliar a viabilidade dessa integração.

Embora a elasticidade e o modelo de pagamento por uso ofereçam vantagens claras, limitações como a latência de invocações remotas e o gerenciamento de dados em ambientes heterogêneos ainda precisam ser abordadas. As abordagens propostas em alguns estudos indicam que é possível mitigar esses desafios, mas a aplicação prática do *serverless* em HPC ainda depende de avanços contínuos em otimizações tecnológicas.

A maioria dos estudos revisados encontra-se em fase experimental, com poucos casos de uso em larga escala. A variabilidade no desempenho sugere que *serverless* é atualmente mais adequado para *workloads* menos sensíveis à latência, o que ressalta a necessidade de mais pesquisas práticas para validar essas tecnologias em cenários reais.

Assim, a integração de HPC com Computação *Serverless* demonstra potencial em termos de flexibilidade e redução de custos, mas superar desafios relacionados à latência e desempenho é essencial para viabilizar plenamente essa combinação em HPC. Com mais pesquisas focadas em cenários reais, o *serverless* poderá expandir suas possibilidades em contextos de computação avançada.

Tabela 1. Síntese dos Resultados

| Aspecto | Descrição | Objetivo Relacionado |
|--|--|------------------------------|
| Escalabilidade | A Computação <i>Serverless</i> oferece escalabilidade automática que se ajusta à demanda dos <i>workloads</i> , vantajosa para HPC em cenários variáveis de carga. Tecnologias como Kubernetes melhoram a elasticidade e a escalabilidade. | Casos de uso e recomendações |
| Eficiência de Recursos | Proporciona uma utilização eficiente de recursos com alocação sob demanda, reduzindo ociosidade. Estudos demonstram a otimização de recursos em <i>workflows</i> científicos e processamento de dados. | Casos de uso e recomendações |
| Custos Operacionais | Redução de custos operacionais, com cobrança conforme o uso, evitando custos fixos. O uso de FaaS mostra-se economicamente viável em comparação com IaaS em certos <i>workloads</i> . | Viabilidade econômica |
| Latência | A latência é um desafio em HPC de baixa latência, especialmente para tarefas de comunicação intensiva. Soluções como RDMA podem mitigar esse problema. | Desafios e limitações |
| <i>Overhead</i> de Invocações Remotas | O <i>overhead</i> de invocações remotas afeta o desempenho de <i>workloads</i> sensíveis ao tempo, sendo necessário otimizar a comunicação entre funções. | Desafios e limitações |
| Limitações na Virtualização de <i>Hardware</i> | Limitações na virtualização de <i>hardware</i> impactam a performance. A integração com <i>containers</i> e Kubernetes pode reduzir essas limitações em HPC. | Limitações e recomendações |
| Viabilidade Econômica | Adoção de <i>serverless</i> reduz custos em até 77% para certos <i>workloads</i> , especialmente em treinamentos de <i>Deep Learning</i> com pipelines. | Casos de uso e recomendações |

5. Trabalhos Futuros

Este estudo identifica áreas que podem ser exploradas em pesquisas futuras, principalmente relacionadas à melhoria e adaptação das tecnologias *serverless* para ambientes de HPC.

A Tabela 2 apresenta as principais áreas para pesquisas futuras em computação *serverless* aplicadas a HPC. Entre as áreas destacadas, estão a **Otimização de Desempenho**, que sugere o desenvolvimento de novas estratégias para reduzir a latência e aprimorar o gerenciamento de dados, especialmente em *workloads* paralelizados. A **Integração com Arquiteturas Híbridas** é também uma área promissora, pois explora a combinação de computação *serverless* com arquiteturas de IaaS e PaaS para aumentar a escalabilidade e a flexibilidade, mantendo desempenho e eficiência.

Outra área importante é a realização de **Estudos Empíricos de Grande Escala** que avaliem a viabilidade da computação *serverless* em contextos produtivos, como supercomputadores, focando em métricas de desempenho, custos e eficiência energética. Adicionalmente, o uso de **Inteligência Artificial para Gerenciamento de Recursos** pode aprimorar a previsão de demanda e a alocação de recursos, reduzindo custos operacionais em ambientes *serverless* dinâmicos.

As questões de **Segurança e Privacidade** também são destacadas, especialmente em ambientes *multitenant*, onde há necessidade de isolamento de *workloads* e proteção de dados sensíveis. Por fim, a tabela explora a aplicação de **Novas Aplicações para Serverless** em áreas como simulações em tempo real, processamento de grandes volumes de dados e treinamento de modelos de *Deep Learning*, abrindo caminho para o uso dessas tecnologias em diversos setores científicos e industriais. Embora algumas soluções, como o uso de RDMA e pipelines, já tenham mostrado melhorias no desempenho e na eficiência de *serverless* em HPC, ainda há necessidade de novas estratégias de otimização. Estudos futuros podem concentrar-se em desenvolver soluções para mitigar a latência de invocações remotas e otimizar o gerenciamento de dados em ambientes heterogêneos, particularmente para *workloads* altamente paralelizados e sensíveis à latência.

Uma área promissora é a integração de modelos híbridos, combinando Computação *Serverless* com outras arquiteturas de computação em nuvem, como (*Infrastructure as a Service* - IaaS) e (*Platform as a Service* - PaaS). Esses modelos poderiam aproveitar o melhor dos dois mundos, oferecendo escalabilidade e flexibilidade de *serverless*, enquanto lidam com limitações de desempenho e armazenamento de dados. Essa combinação pode ser particularmente útil em cenários onde *workloads* variam entre fases intensivas de computação e períodos de baixa utilização.

A maior parte dos trabalhos analisados está em fase experimental, e há uma lacuna considerável de estudos empíricos em grande escala que demonstram a viabilidade de *serverless* para HPC em cenários reais. Investigações futuras devem focar na aplicação dessas tecnologias em contextos produtivos, como supercomputadores, e na avaliação de seu impacto em termos de desempenho, custos e eficiência energética.

Outra área para pesquisas futuras é o uso de técnicas de inteligência artificial e automação para o gerenciamento de recursos em plataformas *serverless*. Ferramentas que utilizem aprendizado de máquina para prever a demanda de recursos e ajustar a alocação de acordo com as necessidades de HPC poderiam aumentar a eficiência e reduzir custos

operacionais, especialmente em cenários dinâmicos.

Estudos futuros também poderiam investigar questões de segurança e privacidade em ambientes *serverless* aplicados a HPC. A natureza de *multitenant* dessas plataformas pode levantar preocupações sobre isolamento de *workloads* e proteção de dados sensíveis, especialmente em ambientes compartilhados de nuvem pública. Soluções que garantam o isolamento eficiente de funções e a proteção de dados críticos seriam uma contribuição importante para a adoção mais ampla dessas tecnologias.

Além das cargas de trabalho convencionais de HPC, estudos futuros podem explorar a aplicação de *serverless* em novas áreas, como simulações em tempo real, processamento de grandes volumes de dados para análises em escala, e treinamento de modelos avançados de *Deep Learning*. A integração dessas tecnologias poderia abrir caminho para novos casos de uso e soluções inovadoras em diversas áreas científicas e industriais. Sugere-se que futuras pesquisas explorem maneiras de enfrentar os desafios relacionados à latência, escalabilidade e gerenciamento de recursos, além de investigar novas oportunidades para expandir o uso da Computação *Serverless* em contextos e aplicações variados. A continuidade dessas investigações tem o potencial de promover avanços na computação em nuvem, favorecendo o desenvolvimento de soluções mais eficientes e acessíveis para HPC em ambientes *serverless*.

Referências

- Al-Haboobi, A. and Kecskemeti, G. (2021). Execution Time Reduction in Function Oriented Scientific Workflows. *Acta Cybernetica*, 25(2):131–150.
- Basu Roy, R. and Tiwari, D. (2024). StarShip: Mitigating I/O Bottlenecks in Serverless Computing for Scientific Workflows. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 8(1):1–29.
- Broome, M. E. et al. (2000). Integrative Literature Reviews for the Development of Concepts. *Concept development in nursing: foundations, techniques and applications*. Philadelphia (USA): WB Saunders Company, pages 231–50.
- Cavalheiro, A. P. and Schepke, C. (2023). Exploring the serverless first strategy in cloud application development. In *2023 International Symposium on Computer Architecture and High Performance Computing Workshops (SBAC-PADW)*, pages 89–94.
- Cloud, G. (2024). What is High Performance Computing? <https://cloud.google.com/discover/what-is-high-performance-computing>. Acesso em: 23 de agosto de 2024.
- Copik, M., Chrapek, M., Schmid, L., et al. (2024). Software Resource Disaggregation for HPC with Serverless Computing. *arXiv*.
- Copik, M., Taranov, K., Calotoiu, A., and Hoefler, T. (2021). rFaaS: RDMA-Enabled FaaS Platform for Serverless High-Performance Computing. *Preprint*.
- de Souza, M. T., da Silva, M. D., and de Carvalho, R. (2010). Revisão integrativa: o que é e como fazer. *Einstein (São Paulo)*, 8:102–106.
- Falcão, D. M. (1996). High performance computing in power system applications. In *International Conference on Vector and Parallel Processing*, pages 1–23. Springer.

Tabela 2. Áreas para Pesquisas Futuras em Serverless e HPC

| Área de Pesquisa | Descrição |
|--|---|
| Otimização de Desempenho | Explorar novas estratégias para mitigar latência de invocação e otimizar o gerenciamento de dados, especialmente para <i>workloads</i> altamente paralelizados e sensíveis à latência. |
| Integração com Arquiteturas Híbridas | Investigar modelos híbridos que combinam Computação Serverless com <i>Infrastructure as a Service</i> (IaaS) e <i>Platform as a Service</i> (PaaS), para melhorar a escalabilidade e flexibilidade, mantendo desempenho e eficiência em armazenamento. |
| Estudos Empíricos de Grande Escala | Conduzir estudos em grande escala que avaliem a viabilidade de <i>serverless</i> para HPC em cenários produtivos, como supercomputadores, e analisar o impacto em desempenho, custos e eficiência energética. |
| Inteligência Artificial para Gerenciamento de Recursos | Explorar o uso de aprendizado de máquina e automação para prever demandas de recursos em plataformas <i>serverless</i> , otimizando a alocação de acordo com as necessidades de HPC para aumentar a eficiência e reduzir custos operacionais. |
| Segurança e Privacidade | Examinar questões de segurança e privacidade, como o isolamento de <i>workloads</i> e proteção de dados em ambientes de <i>serverless</i> multitenant, especialmente em nuvens públicas. |
| Novas Aplicações para Serverless | Investigar a aplicação de <i>serverless</i> em áreas como simulações em tempo real, processamento de grandes volumes de dados e treinamento de modelos de <i>Deep Learning</i> , visando ampliar seu uso em diferentes setores científicos e industriais. |

- Kehrer, S. and Blochinger, W. (2020). Equilibrium: An Elasticity Controller for Parallel Tree Search in the Cloud. *Journal of Supercomputing*, 76.
- Liu, Y., Jiang, B., Guo, T., Huang, Z., Ma, W., Wang, X., and Zhou, C. (2022). FuncPipe: A Pipelined Serverless Framework for Fast and Cost-Efficient Training of Deep Learning Models. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 6(3):1–30.
- Malla, S. and Christensen, K. (2020). HPC in the Cloud: Performance Comparison of Function as a Service (FaaS) vs Infrastructure as a Service (IaaS). *Internet Technology Letters*, 3.
- Naranjo, D. M., Risco, S., de Alfonso, C., Pérez, A., Blanquer, I., and Moltó, G. (2020). Accelerated Serverless Computing Based on GPU Virtualization. *Journal of Parallel and Distributed Computing*, 139:32–42.
- Navaux, P. O. A. and da Silva Serpa, M. (2021). Desafios do Processamento de Alto Desempenho. In *Anais do XLVIII Seminário Integrado de Software e Hardware*, pages 39–49. SBC.
- NetApp (2024). What is High Performance Computing? <https://www.netapp.com/data-storage/high-performance-computing/what-is-hpc/>. Acesso em: 23 de agosto de 2024.
- Petrosyan, D. and Astsatryan, H. (2022). Serverless High-Performance Computing over Cloud. *Cybernetics and Information Technologies*, 22(3):82–92.
- Spillner, J., Mateos, C., and Monge, D. A. (2018). Faaster, Better, Cheaper: The Prospect of Serverless Scientific Computing and HPC. In *High Performance Computing: 4th Latin American Conference, CARLA 2017, Buenos Aires, Argentina, and Colonia del Sacramento, Uruguay, September 20-22, 2017, Revised Selected Papers 4*, pages 154–168. Springer.
- Whittemore, R. and Knafl, K. (2005). The integrative review: updated methodology. *Journal of advanced nursing*, 52(5):546–553.