



Universidade Federal de Santa Catarina

Centro Tecnológico  
Departamento de Informática e Estatística

---

## Contribuições para Melhoria de Desempenho de Aplicações Paralelas em Nuvens Computacionais

Laboratório de Pesquisa em Sistemas Distribuídos



---

Projeto submetido ao Programa Institucional de Bolsas de Iniciação em Pesquisa Científica (PIBIC/CNPq – 2024/2025).

**Proponente:** Prof. Dr. Márcio Bastos Castro

**Email:** [marcio.castro@ufsc.br](mailto:marcio.castro@ufsc.br)

**Site pessoal:** <http://www.marciocastro.com>

**Lattes:** <http://lattes.cnpq.br/6876016315737507>

Florianópolis, 13 de maio de 2024

## Resumo

Pesquisadores das áreas das Engenharias, Física, Química, Biologia, Geologia, Medicina, entre outras, desenvolvem modelos, métodos computacionais e simulações numéricas em suas pesquisas científicas. Tendo em vista o elevado poder de processamento necessário para execução destas aplicações, infraestruturas físicas de grande porte (*clusters* ou *data-centers*) para Computação de Alto Desempenho (*High Performance Computing* - HPC) são utilizadas. Estas infraestruturas estão normalmente disponíveis apenas localmente nos centros de desenvolvimento das pesquisas, com restrições de acesso externo, limitando o compartilhamento destes recursos computacionais. Além do alto custo de manutenção, os *clusters* também exigem dos pesquisados conhecimentos aprofundados sobre aspectos técnicos da infraestrutura para executarem suas aplicações ou para realizarem atualização neste sistema computacional. Por outro lado, o paradigma da Computação em Nuvem democratizou o acesso a grandes infraestruturas de *hardware* (*data-centers*) para milhões de organizações e indivíduos com poucos recursos capitais. Todavia, o ecossistema de HPC na nuvem ainda não evoluiu o suficiente para suportar o uso generalizado, não somente devido a limitações tecnológicas, mas também principalmente devido ao processo laborioso de migração de aplicações HPC legadas para um ambiente completamente diferente do ambiente ao qual foram originalmente projetadas. Este projeto visa avaliar o desempenho e custo de execução de aplicações de HPC na nuvem. A pesquisa permitirá identificar gargalos e propor novas técnicas e otimizações para as aplicações paralelas.

## Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação e Objetivo Geral da Pesquisa . . . . .	2
<b>2</b>	<b>Problemas Abordados e Metas</b>	<b>3</b>
<b>3</b>	<b>Planos de Atividades e Cronogramas</b>	<b>3</b>
3.1	Meta 1 – Bolsista IC1 . . . . .	3
3.2	Meta 2 – Bolsista IC2 . . . . .	4
<b>4</b>	<b>Metodologia</b>	<b>4</b>
<b>5</b>	<b>Resultados Esperados</b>	<b>5</b>
<b>6</b>	<b>Exequibilidade do Projeto</b>	<b>5</b>
	<b>Referências</b>	<b>6</b>

# 1 Introdução

A área de Computação de Alto Desempenho (*High Performance Computing* - HPC) faz uso de grandes infraestruturas de *hardware* (também conhecidos como *clusters* ou *data-centers*) para resolver grandes problemas computacionais extremamente complexos e que demandam um grande poder de processamento. Devido a isto, a área de HPC viabiliza pesquisas científicas inovadoras em outras áreas do conhecimento, tais como as Engenharias, a Medicina, a Biologia, a Meteorologia, entre outras. Muitas vezes, as pesquisas nessas áreas se baseiam fortemente no uso de modelos matemáticos, métodos computacionais e simulações numéricas que demandam um alto poder computacional para estudar os problemas existentes e propor novas soluções. Por exemplo, nas Engenharias, diversas pesquisas focam no estudo de materiais inteligentes, os quais podem ser aplicados na fabricação de biomateriais e na engenharia aeroespacial; na Biologia e Medicina, modelos e métodos computacionais são propostos para realizar sequenciamento de DNA (e mais recentemente para o sequenciamento genômico do SARS-CoV-2) e para o desenvolvimento de novos materiais e fármacos por meio de métodos do tipo *Machine Learning* e *Data Mining*; na Meteorologia, modelos matemáticos são implementados numericamente para prever condições climáticas ou desastres naturais com alta precisão. Nesse contexto, pesquisadores da área de HPC desenvolvem métodos, técnicas e soluções inovadoras para que os modelos matemáticos, métodos computacionais e simulações numéricas desenvolvidos em outras áreas do conhecimento possam, de fato, serem executados de forma eficiente em grandes *data-centers*.

Nos últimos anos, o paradigma da Computação em Nuvem democratizou o acesso a grandes infraestruturas de *hardware* (*data-centers*) para milhões de organizações e indivíduos com poucos recursos capitais, aplicando economia de escala de forma massiva reduzindo custos com Tecnologia da Informação (TI) (1). Nuvens computacionais apoiam-se no mesmo modelo de renovação e manutenção da infraestrutura física de grandes centros de computação, porém diferem no modelo de negócio. Os provedores têm seus negócios focados na otimização do uso da infraestrutura física, independentemente do foco da aplicação ou do usuário. Os serviços são executados na plataforma sob demanda com pagamentos da infraestrutura física de acordo com o que foi alocado (*pay-as-you-go*) ou com o que utilizado (*pay-as-you-use*). A diversidade de usuários/serviços é possível em nuvens computacionais devido à elasticidade da plataforma. A virtualização dos recursos garante às nuvens o compartilhamento de recursos físicos com isolamento entre usuários/aplicações. Entretanto, no contexto de HPC, a característica marcante da Computação em Nuvem, que é a abstração de *hardware*, é um forte ponto negativo para aplicações HPC tradicionais. Nesse contexto, geralmente é necessário um conhecimento granular acerca da infraestrutura de *hardware* subjacente para realização de ajustes otimizados para a execução de cargas de trabalho intensivas (2).

Em contraste às aplicações de HPC tradicionais, que geralmente são executadas em *batches* utilizando *Message Passing Interface* (MPI) em *hardware* especializado, as plataformas de Computação em Nuvem surgiram e evoluíram para suportar aplicações Web comerciais, executadas em *hardware* virtualizado sobre uma infraestrutura física compartilhada. Dessa forma, o paradigma da Computação em Nuvem é voltado para permitir a rápida alocação e de-alocação de recursos virtualizados, permitindo que o usuários escale suas infraestruturas horizontalmente de maneira dinâmica, efetivamente permitindo a redução de custos de infraestrutura ociosa e facilitando a implantação de sistemas tolerantes a falhas e com *zero-downtime* (3).

No contexto de HPC, melhorias recentes nas tecnologias de virtualização e redes, combinadas ao desenvolvimento de ferramentas chave de *software*, podem permitir que a comunidade de HPC aproveite recursos de Computação em Nuvem de baixo custo sob-demanda, reduzindo drasticamente o desperdício de recursos e custos de manutenção de *clusters* tradicionalmente implantados da forma *on-premise* (4, 5). Entretanto, o ecossistema de HPC na nuvem ainda não evoluiu o suficiente para suportar o uso generalizado, não somente devido a limitações tecnológicas, mas também principalmente devido ao processo laborioso de migração de aplicações HPC legadas para um ambiente completamente diferente do ambiente ao qual foram originalmente projetadas (2). Grande parte da pesquisa existente tanto na academia quanto na indústria em relação à HPC na nuvem é focada no entendimento dos custos-benefícios de migrar aplicações legadas de *clusters on-premise* para nuvem (2) com uso de instâncias<sup>1</sup> do tipo *on-demand*. Estas instâncias são usualmente Máquinas Virtuais (VMs) que possuem um valor fixo por unidade de tempo de utilização.

## 1.1 Motivação e Objetivo Geral da Pesquisa

Atualmente, os provedores de nuvem têm oferecido outros tipos de instâncias de mais baixo custo aos usuários. Por exemplo, a Amazon AWS oferece dois tipos de instâncias que tem um potencial de redução de custos para os usuários: *spot* e *burstable*. Os detalhes destes tipos de instâncias são apresentados a seguir:

**Instâncias *Spot*.** O mercado *spot* permite que os provedores possam otimizar o uso de recursos alugando infraestrutura sobressalente dos provedores de nuvem com descontos significativos. Instâncias *spot* custam consideravelmente menos por hora do que as instâncias *on-demand* padrão e são executadas no mesmo tipo de infraestrutura com desempenho e características idênticas. No entanto, o provedor de nuvem pode revogar máquinas *spot* quando a capacidade precisa ser recuperada para instâncias persistentes, tornando responsabilidade do usuário salvar dados preventivamente e recuperar aplicativos em caso de reintegração de posse da instância. Este fator tem impacto significativo para aplicações de HPC, pois estas aplicações podem ser executadas por horas. Logo, algum mecanismo de tolerância a falhas precisa ser utilizado para garantir que parte da computação já realizada pela aplicação não seja totalmente perdida.

**Instâncias *Burstable*.** Estas instâncias que fornecem um desempenho base de CPU e podem expandir acima dessa taxa base por quanto tempo for necessário. Na AWS o custo de uma instância com características expansíveis é calculado a partir de créditos, gerados e gastos, durante a execução. Cada instância, pertencente a uma classe *burstable*, gera uma quantidade de créditos de CPU ao longo de sua execução (instâncias mais robustas podem gerar mais créditos). Cada instância também possui um tamanho máximo de armazenamento de créditos (instâncias mais robustas podem armazenar mais créditos e por mais tempo). Cada unidade de crédito significa 100% de utilização de uma vCPU por um minuto. Cada instância tem um limiar de utilização total onde a quantidade de créditos gerados é a mesma de créditos consumidos (a.k.a linha de base ou *baseline*). Quando uma instância está com utilização de CPU abaixo de sua linha de base ela acumula créditos (estado

<sup>1</sup>Instâncias são máquinas virtuais ou físicas oferecidas pelos provedores de nuvem.

de *baseline*). Os créditos são consumidos quando a instância está com utilização de CPU acima do limiar (estado de *burst*).

Este projeto de pesquisa visa avaliar o custo e o desempenho de instâncias *spot* e *burstable* para execução de aplicações de HPC e propor novas técnicas e otimizações para que as aplicações possam ser executadas nestes tipos de instâncias com menor custo financeiro sem afetar drasticamente seu desempenho.

A presente proposta está diretamente relacionada ao projeto de pesquisa existente intitulado “*Sustainable High Performance Computing on AWS*”, financiado através da Chamada CNPq/AWS Nº 64/202 - Acesso às Plataformas de Computação em Nuvem da AWS (AWS Cloud Credits for Research)<sup>2</sup>. Este projeto, também coordenado pelo Prof. Márcio Castro, disponibilizará créditos na AWS para a realização das atividades de pesquisa dos bolsistas PIBIC.

## 2 Problemas Abordados e Metas

Tendo em vista a recente tendência no uso de nuvens computacionais para execução de aplicações de HPC, as pesquisas propostas no presente projeto buscarão atingir as seguintes metas:

**Meta 1** Propor e desenvolver uma solução inteligente que utilize instâncias *burstable* e/ou *spot* para reduzir o custo financeiro de execução de aplicações de HPC na AWS;

**Meta 2** Propor e desenvolver técnicas e otimizações para melhorar o desempenho de aplicações de HPC na AWS.

## 3 Planos de Atividades e Cronogramas

Os dois bolsistas de Iniciação Científica, doravante denominados **IC1** e **IC2**, trabalharão com pesquisa e desenvolvimento, porém, com enfoques bastante distintos. O bolsista **IC1** terá como foco a **Meta 1**. Por outro lado, o bolsista **IC2** terá como foco a **Meta 2**. As seções seguintes destacam as atividades e o cronograma de cada bolsista.

### 3.1 Meta 1 – Bolsista IC1

O enfoque do bolsista **IC1** será na **proposta e desenvolvimento de uma solução inteligente que utilize instâncias *burstable* e/ou *spot* para reduzir o custo financeiro de execução de aplicações de HPC na AWS (Meta 1)**. A solução proposta pelo bolsista será implementada na biblioteca HPC@Cloud (6), a qual é desenvolvida pelo grupo de pesquisa do proponente. O bolsista **IC1** realizará as seguintes atividades:

---

<sup>2</sup>O resultado final da Chamada CNPq/AWS Nº 64/202 está disponível [neste link](#).

### Meta 1 – Bolsista IC1

- A1.1. Levantamento bibliográfico: HPC e computação em nuvem
- A1.2. Levantamento bibliográfico: AWS
- A1.3. Proposta e desenvolvimento da solução
- A1.4. Execução e avaliação de experimentos
- A1.5. Elaboração de artigos e relatório final contendo os resultados obtidos

A Figura 1a apresenta o cronograma proposto para a realização das atividades do bolsista IC1.

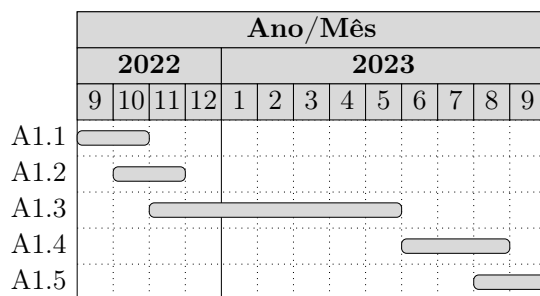
### 3.2 Meta 2 – Bolsista IC2

O enfoque do bolsista IC2 será na **proposta e desenvolvimento de técnicas e otimizações para melhorar o desempenho de aplicações de HPC na AWS (Meta 2)**. A solução proposta pelo bolsista será implementada na biblioteca HPC@Cloud (6), a qual é desenvolvida pelo grupo de pesquisa do proponente deste projeto. O bolsista IC2 realizará as seguintes atividades:

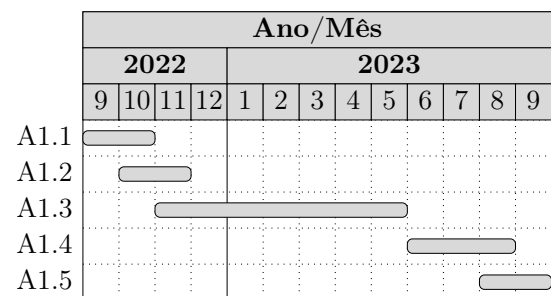
### Meta 2 – Bolsista IC2

- A2.1. Levantamento bibliográfico: HPC e computação em nuvem
- A2.2. Levantamento bibliográfico: AWS
- A2.3. Proposta e desenvolvimento das soluções
- A2.4. Execução e avaliação de experimentos
- A2.5. Elaboração de artigos e relatório final contendo os resultados obtidos

A Figura 1b apresenta o cronograma proposto para a realização das atividades do bolsista IC2.



(a) Atividades do bolsista IC1



(b) Atividades do bolsista IC2

Figura 1: Cronograma de atividades dos bolsistas.

## 4 Metodologia

As atividades previstas na Seção 3 para atingir as metas deste projeto serão executadas pelos bolsistas de IC, os quais irão colaborar com as atividades de pesquisa que estão sendo atualmente realizadas por alunos de pós-graduação orientados pelo proponente.

Reuniões periódicas serão realizadas com os bolsistas para realizar o acompanhamento das atividades e definir os direcionamentos das pesquisas e encaminhamentos futuros.

Para fins de reprodutibilidade, todos os códigos-fonte e experimentos produzidos no contexto deste projeto serão disponibilizados de maneira livre no repositório da ferramenta HPC@Cloud<sup>3</sup>. O desenvolvimento das soluções propostas será feito, principalmente, com as linguagens C/C++ e Python. Os códigos desenvolvidos serão integrados na ferramenta HPC@Cloud desenvolvida pelo LaPeSD.

Artigos científicos contendo as contribuições e resultados obtidos ao longo da execução do presente projeto serão submetidos a periódicos e conferências nacionais e internacionais nas áreas de Computação Paralela e Distribuída. Dentre as possibilidades de periódicos na área de Computação Paralela e Distribuída destaca-se: *IEEE Transactions on Parallel and Distributed Systems*, *Concurrency and Computation: Practice and Experience*, *Parallel Computing*, *Journal of Parallel and Distributed Computing* e *International Journal of Parallel Programming*. As seguintes conferências da área de Computação Paralela e Distribuída também serão foco de submissões: *IEEE International Parallel and Distributed Processing Symposium* (IPDPS), *International Conference on Parallel Processing* (ICPP), *European Conference on Parallel Computing* (Euro-Par), *International Conference on High Performance Computing* (HiPC), *Euromicro International Conference on Parallel, Distributed and Network-Based Computing* (PDP), *International Symposium on Computer Architecture and High Performance Computing* (SBAC-PAD) e Simpósio em Sistemas Computacionais de Alto Desempenho (WSCAD).

## 5 Resultados Esperados

Os principais resultados esperados do projeto de pesquisa são:

- Capacitação dos bolsistas de IC envolvidos na execução do projeto na área de Computação de Alto Desempenho, cuja a demanda está em constante crescimento nos âmbitos acadêmico e industrial;
- Despertar o interesse dos bolsistas de IC em seguir uma carreira acadêmica e/ou de pesquisador, aumentando as chances de atração de bons alunos para o Programa de Pós-Graduação em Ciência da Computação (PPGCC) do Centro Tecnológico;
- Contribuir significativamente no desenvolvimento de soluções que permitam a adoção de HPC na nuvem;
- Disseminação do conhecimento e da qualidade da pesquisa feita pela UFSC em diferentes meios, assim como através de publicações em conferências e em periódicos.

## 6 Exequibilidade do Projeto

A pesquisa proposta neste projeto se enquadra nas atividades exercidas pelo proponente no **Laboratório de Pesquisas em Sistemas Distribuídos (LaPeSD)** do Departamento de Informática e Estatística (INE). O LaPeSD possui diferentes linhas de pesquisa, dentre elas a linha de pesquisa em Sistemas Distribuídos de Alto Desempenho.

<sup>3</sup><<http://github.com/lapesd/hpcac-toolkit>>



O proponente é atualmente o coordenador do LaPeSD, o qual está cadastrado no Diretório dos Grupos de Pesquisa do CNPq<sup>4</sup>. Informações mais detalhadas sobre as linhas de pesquisa, professores e alunos envolvidos no LaPeSD poderão ser obtidas no site do laboratório<sup>5</sup>.

Com relação aos recursos humanos é importante ressaltar que a concessão de bolsas de IC será de extrema importância para viabilização e execução do projeto de pesquisa proposto. Os bolsistas desenvolverão suas pesquisas juntamente com outros alunos de mestrado e de doutorado orientados (ou co-orientados) pelo proponente. Esses alunos já realizam pesquisas em assuntos fortemente relacionados às atividades propostas aos bolsistas de IC. Isso permitirá uma interação positiva entre os discentes de pós-graduação e os bolsistas de IC, despertando possivelmente o interesse nos bolsistas de IC em seguirem carreiras acadêmicas ou de pesquisa.

Com relação aos recursos materiais, cada um dos bolsistas de IC terá à disposição uma estação de trabalho (mesa e computador) no LaPeSD. Os créditos para utilização de recursos da AWS serão obtidos do projeto de pesquisa “*Sustainable High Performance Computing on AWS*” aprovado por meio da Chamada CNPq/AWS Nº 64/202 - Acesso às Plataformas de Computação em Nuvem da AWS (AWS Cloud Credits for Research).

## Referências

- 1 BUYYA, R.; VARGHESE, B. Next Generation Cloud Computing: New Trends and Research Directions. *Future Generation Computer Systems*, Elsevier Science Publishers B. V., NLD, v. 79, n. 3, p. 849–861, sep 2017. Disponível em: <<https://doi.org/10.1016/j.future.2017.09.020>>.
- 2 NETTO, M. A. S. *et al.* HPC Cloud for Scientific and Business Applications: Taxonomy, Vision, and Research Challenges. *ACM Computing Surveys*, Association for Computing Machinery, New York, NY, USA, v. 51, n. 8, jan 2018. ISSN 0360-0300. Disponível em: <<https://doi.org/10.1145/3150224>>.
- 3 BUYYA, R. *et al.* Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility. *Future Generation Computer Systems*, Elsevier Science Publishers B. V., NLD, v. 25, n. 6, jun 2009. ISSN 0167-739X. Disponível em: <<https://doi.org/10.1016/j.future.2008.12.001>>.
- 4 BROWNE, P. *et al.* OpenStack and Network Fabrics. In: *The Crossroads of Cloud and HPC: Exploring OpenStack Cloud Computing for Scientific Workloads*. CreateSpace Independent Publishing Platform, 2017. p. 14–22. Disponível em: <<https://object-storage-ca-ymq-1.vexxhost.net/swift/v1/6e4619c416ff4bd19e1c087f27a43eea/www-assets-prod/science/CrossroadofCloudandHPC.pdf>>.
- 5 BETHWAITE, B.; LU, X. OpenStack and Virtualised HPC. In: *The Crossroads of Cloud and HPC: Exploring OpenStack Cloud Computing for Scientific Workloads*. CreateSpace Independent Publishing Platform, 2017. p. 2–12. Disponível em: <<https://object-storage-ca-ymq-1.vexxhost.net/swift/v1/6e4619c416ff4bd19e1c087f27a43eea/www-assets-prod/science/CrossroadofCloudandHPC.pdf>>.
- 6 MUNHOZ, V.; CASTRO, M. HPC@Cloud: A Provider-Agnostic Software Framework for Enabling HPC in Public Cloud Platforms. In: *XXIII Simpósio em Sistemas Computacionais de Alto Desempenho (WSCAD)*. Florianópolis: Brazilian Computer Society, 2022. p. 157—168.

<sup>4</sup>LaPeSD no CNPq: <<http://dgp.cnpq.br/dgp/espelhogrupo/1234490143448197>>

<sup>5</sup>Site do LaPeSD: <<http://lapesd.inf.ufsc.br>>