

# Proposta de Provisionamento Elástico de Recursos com MPI-2 para a DSL SPar

Cassiano Rista<sup>1</sup>, Luiz Gustavo L. Fernandes<sup>1</sup>

<sup>1</sup> Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)  
Grupo de Modelagem de Aplicações Paralelas (GMAP), Porto Alegre – RS – Brasil

`luis.rista@acad.pucrs.br, luiz.fernandes@pucrs.br`

**Resumo.** *Este artigo apresenta uma proposta para desenvolvimento de um módulo de provisionamento elástico e autônomo a ser integrado em uma linguagem específica de domínio (DSL) voltada para o paralelismo de stream. O módulo deverá explorar a elasticidade como uso de MPI-2 em um ambiente de cluster de computadores, permitindo a criação de processos em tempo de execução, serialização, ordenamento e balanceamento de carga.*

## 1. Introdução

O paradigma de computação em *cluster* se tornou um recurso computacional usual em muitas organizações, universidades e centros de pesquisa para atingir requisitos de computação de alto desempenho [Dantas and Rista 2005]. No entanto, tem-se verificado atualmente que a utilização desses ambientes para a execução de aplicações de *stream*, que demandam alta capacidade de processamento paralelo, tem sido evitada. Isso ocorre, devido a dificuldade de aproveitar a elasticidade, visto que esses ambientes são usualmente projetados para um número fixo de processos.

No cenário atual, aplicações que permitem o processamento de fluxos contínuos de dados, como aplicações de *stream*, vem tornando-se cada vez mais relevantes. Somado a isso, a elasticidade possibilita ainda a aquisição de recursos de forma dinâmica com base na variação da carga de trabalho do sistema. Dessa forma, explorar a elasticidade para aplicações de *stream* a partir de uma linguagem específica de domínio (DSL) em um ambiente de *cluster* pode ser um diferencial de pesquisa. Nesse sentido, o artigo tem como objetivo principal a implementação de um módulo de elasticidade autônomo para a DSL SPar [Griebler and Fernandes 2017], permitindo o provisionamento dinâmico de recursos para aplicações de *stream* paralelas com uso de MPI-2, a partir dos modelos de *pipeline* e *farm*.

## 2. Trabalhos Relacionados

O interesse por pesquisas voltadas para o desenvolvimento de DSLs para a abstração de paralelismo tem aumentado consideravelmente nos últimos anos. Nesse sentido, o Grupo de Modelagem de Aplicações Paralelas (GMAP), desenvolveu originalmente uma DSL em C++ projetada para fornecer abstrações de alto nível para paralelismo de *stream* em arquiteturas *multi-core*. Recentemente, Griebler et al. [Griebler and Fernandes 2017] propôs em seu trabalho, o desenvolvimento de pesquisas direcionadas ao suporte de programação paralela distribuída para ambientes de *cluster* na DSL SPar. O objetivo é preservar a semântica original, enquanto a transformação de código *source-to-source* resultante

deverá gerar código paralelo baseado na interface de troca de mensagens (MPI). Os resultados preliminares dos experimentos demonstram uma melhor programabilidade sem significativa perda de desempenho.

Por sua vez, Righi et al. [d. R. Righi et al. 2016] apresenta neste artigo o AutoElastic, um modelo de elasticidade em nível de plataforma como serviço (PaaS) para aplicações de alto desempenho na nuvem. O diferencial desta abordagem consiste em proporcionar elasticidade para aplicações de alto desempenho sem intervenção do usuário ou modificação do código fonte. Além disso, o modelo AutoElastic assemelha-se com o estilo de programação baseado em soquetes que é oferecido pelo MPI-2, em que os novos processos podem ser facilmente criados ou removidos, a partir da computação paralela. A avaliação do protótipo usando o *middleware* OpenNebula mostrou ganhos de até 26% de desempenho no tempo de execução de uma aplicação com o gerenciador AutoElastic.

### 3. Proposta

Este artigo propõe o desenvolvimento de um módulo que seja capaz de habilitar o suporte a elasticidade na DSL SPar [Griebler and Fernandes 2017], permitindo assim, o provisionamento dinâmico de recursos para aplicações de *stream* paralelas, a partir de modelos como *pipeline* e *farm*. Para isso, considerou-se nessa fase do projeto de pesquisa uma estratégia de elasticidade autônoma em nível de *cluster*, permitindo assim uma maior flexibilidade para a integração junto a DSL SPar. O módulo proposto tem como objetivo assegurar que os nodos de computação do *cluster* recebam recursos suficientes para manter a execução de aplicações paralelas de *stream* dentro de um determinado nível de qualidade de serviço (QoS). Esse nível de QoS é determinado por um controlador principal de modo autônomo, não havendo necessidade de definição de *thresholds* por parte dos usuários.

Por fim, propõe-se a criação e remoção de processos em tempo de execução, serialização, ordenamento e balanceamento de carga, preservando a semântica original e gerando transformações de código *source-to-source* resultantes em MPI-2, devido a sua característica dinâmica. Como resultado, pretende-se analisar o impacto que a estratégia de elasticidade autônoma pode trazer sobre métricas como desempenho, portabilidade e eficiência, quando do processamento paralelo de aplicações de *stream*.

### Referências

- [d. R. Righi et al. 2016] d. R. Righi, R., Rodrigues, V. F., da Costa, C. A., Galante, G., de Bona, L. C. E., and Ferreto, T. (2016). Autoelastic: Automatic resource elasticity for high performance applications in the cloud. *IEEE Transactions on Cloud Computing*, 4(1):6–19.
- [Dantas and Rista 2005] Dantas, M. A. R. and Rista, C. (2005). A wireless monitoring approach for a ha-oscar cluster environment. In *19th International Symposium on High Performance Computing Systems and Applications (HPCS'05)*, pages 302–306.
- [Griebler and Fernandes 2017] Griebler, D. and Fernandes, L. G. (2017). Towards Distributed Parallel Programming Support for the SPar DSL. In *Parallel Computing: On the Road to Exascale, Proceedings of the International Conference on Parallel Computing, ParCo'17*, Bologna, Italy. IOS Press.