

# Uma Linguagem Específica de Domínio com Geração de Código Paralelo para Visualização de Grandes Volumes de Dados

Cleverson L. Ledur, Dalvan Griebler, Luiz G. Fernandes, Isabel Manssour

<sup>1</sup>PPGCC – Pontifícia Universidade Católica do Rio Grande do Sul  
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

{cleverson.ledur, dalvan.griebler}@acad.pucrs.br

{luiz.fernandes, isabel.manssour}@pucrs.br

***Resumo.** Este artigo apresenta uma análise sobre linguagens específicas de domínio para a criação de visualizações. Ao final, propõe uma nova linguagem específica de domínio para geração de visualizações de quantidades massivas de dados, paralelizando não só a geração e a interação da visualização, mas também o pré-processamento dos dados brutos.*

## 1. Introdução

Nos últimos 10 anos, ocorreu um crescimento exponencial de volume de dados gerados em todo o planeta [Senbalci et al. 2013]. A maior parte destes dados tem origem em atividades diárias, tais como registros de ligações, redes sociais, reclamações de consumidores e sensores. Essa quantidade de dados pode fornecer informações preciosas se bem analisada. No entanto, a análise destes dados brutos é complexa, abrindo campo para o uso de técnicas de visualização para percepção de informações.

Assim, a utilização de visualizações permite uma apresentação de dados que torna a percepção das informações mais eficiente. Isso se caracteriza devido a interpretação de uma imagem poder ser realizada em paralelo pelo sistema de percepção humano, enquanto a análise de texto ou dados é limitada ao processo sequencial de leitura [Ward et al. 2010].

No entanto, além das visualizações se tornarem lentas devido ao grande volume de dados, cientistas de áreas pouco contempladas pela computação e leigos em programação encontram desafios em desenvolver visualizações, pois é necessário se preocupar com o pré-processamento dos dados e a programação da visualização. Isso é uma tarefa complexa que pode exigir um grande esforço de desenvolvimento e exigir a paralelização dos algoritmos [Bethel et al. 2012].

## 2. Trabalhos Relacionados

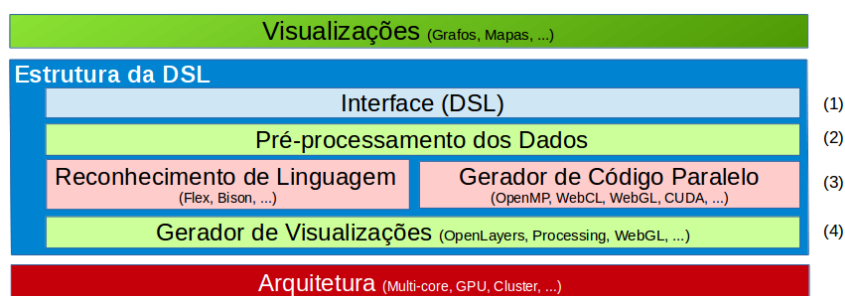
Linguagens Específicas de Domínio (DSL) são linguagens que procuram resolver um domínio de problema particular [Fowler 2010]. Recentemente, foram criadas DSLs para problemas de visualização, fornecer um ambiente de alto nível de programação, incluir o paralelismo e reduzir o tempo necessário de desenvolvimento.

A DSL-POPP (baseada em padrões paralelos) foi proposta com o objetivo de reduzir o esforço de programação paralela [Griebler and Fernandes 2013, Griebler et al. 2014]. Em 2012, a DSL Diderot focada em análise e visualização de dados de biomedicina foi anunciada [Chiw et al. 2012]. Em 2013, duas DSLs voltadas para visualizações de grandes quantidades de dados foram apresentadas: uma que demonstra uma metodologia de criação de visualizações interativas e de alto desempenho

[Meyerovich et al. 2013]; e outra voltada para a criação de uma plataforma para Big Data no setor de Telecom [Senbalci et al. 2013]. Em 2014, foram demonstradas a ViSlang [Rautek et al. 2014], uma DSL para visualização científica, e a Vivaldi [Choi et al. 2014] para processamento volumétrico e de visualizações em ambientes heterogêneos.

### 3. Proposta

Considerando o que foi exposto, propõe-se a criação de uma DSL (Figura 1) que permita a geração de visualizações a partir de Big Data. Pretende-se fornecer uma interface de alto nível para a programação da visualização e inserção dos dados brutos, realizando não só a geração do código paralelo para a visualização mas também o pré-processamento dos dados utilizando técnicas já utilizadas em visualização e mineração de dados.



**Figure 1. Escopo da DSL Proposta**

A estrutura da DSL (Figura 1) possuirá quatro camadas, sendo: (1) interface; (2) pré-processamento dos dados; (3) reconhecimento e geração de código; (4) Gerador de Visualizações. Inicialmente pretende-se utilizar um tipo de visualização para validar a solução. A interface deverá possuir apenas parâmetros da DSL, pois será uma DSL Externa. O pré-processamento irá organizar os dados brutos e, em conjunto com a terceira camada, criar o gerador de visualizações para executar na arquitetura paralela.

### References

- Bethel, E., Childs, H., and Hansen, C. (2012). *High Performance Visualization: Enabling Extreme-Scale Scientific Insight*. Chapman & Hall/CRC Computational Science. Taylor & Francis.
- Chiw, C., Kindlmann, G., Reppy, J., Samuels, L., and Seltzer, N. (2012). Diderot: a parallel dsl for image analysis and visualization. In *ACM SIGPLAN Notices*, volume 47, pages 111–120. ACM.
- Choi, H., Choi, W., Quan, T., Hildebrand, D. G., Pfister, H., and Jeong, W.-K. (2014). Vivaldi: A domain-specific language for volume processing and visualization on distributed heterogeneous systems.
- Fowler, M. (2010). *Domain-Specific Languages*. Addison-Wesley, Boston, USA.
- Griebler, D., Adornes, D., and Fernandes, L. G. (2014). Performance and Usability Evaluation of a Pattern-Oriented Parallel Programming Interface for Multi-Core Architectures. In *The 26th International Conference on Software Engineering & Knowledge Engineering*, pages 25–30, Vancouver, Canada. Knowledge Systems Institute Graduate School.
- Griebler, D. and Fernandes, L. G. (2013). Towards a Domain-Specific Language for Patterns-Oriented Parallel Programming. In *17th Brazilian Symposium - SBLP*, volume 8129, pages 105–119, Brasilia, Brazil. Springer Berlin Heidelberg.
- Meyerovich, L. A., Torok, M. E., Atkinson, E., and Bodik, R. (2013). Superconductor: A language for big data visualization.
- Rautek, P., Bruckner, S., Groller, M., and Hadwiger, M. (2014). Vislang: A system for interpreted domain-specific languages for scientific visualization.
- Senbalci, C., Altuntas, S., Bozkus, Z., and Arsan, T. (2013). Big data platform development with a domain specific language for telecom industries. In *HONET-CNS, 2013*, pages 116–120.
- Ward, M., Grinstein, G., and Keim, D. (2010). *Interactive data visualization: foundations, techniques, and applications*. AK Peters, Ltd.