

Um sistema web para execução remota de aplicações de alto desempenho

Otávio Migliavacca Madalosso
Universidade Federal de Santa Maria

19 de setembro de 2015

1 Identificação

Resumo: Algumas áreas de pesquisa utilizam constantemente algoritmos que demandam alto desempenho dos seus ambientes de execução. Ocasionalmente, surgem algoritmos novos, com diferentes propriedades, que se propõem a resolver um problema de forma mais eficiente e/ou completa. Infelizmente, é comum que esses algoritmos fiquem restritos a ambientes institucionais, limitando muito a sua visibilidade para a comunidade de pesquisa. Este trabalho tem como objetivo criar um portal que permita ao usuário solicitar a execução remota de um algoritmo de acordo com as configurações que ele desejar.

Período de Execução: Setembro de 2015 até Dezembro de 2015

Unidades Participantes:

Curso de Ciência da Computação

Laboratório de Sistemas de Computação

Área do Conhecimento: Ciência da Computação

Título do Projeto: Trabalho de Conclusão de Curso

Participantes:

Prof^a Andrea Schwertner Charão - Orientadora

Otávio Migliavacca Madalosso - Orientando

2 Introdução

Algoritmos com grande custo computacional são facilmente encontrados em áreas como meteorologia, biologia e astronomia. Esses algoritmos possuem a característica de utilizar um nível elevado de processamento para concluir sua execução, e consequentemente, seus tempos de execução podem variar dependendo da máquina aonde estão sendo executados.

É comum pesquisadores destas e de outras áreas desenvolverem novas implementações de algoritmos utilizados por seus colegas. Implementações essas que podem trazer muitos benefícios para outros pesquisadores que necessitam deste tipo de solução. Infelizmente, é comum essas implementações ficarem restritas a ambientes privados, não por questões de licença, mas simplesmente pela ausência de um método prático para disponibilizar a nova ferramenta ao público.

Com isso, surge a ideia de desenvolver um portal web que permita ao usuário o cadastro de um experimento, no qual ele poderá ditar os dados de entrada desse experimento, e qual algoritmo (disponível no sistema) ele deseja utilizar para processar os dados. Depois de requisitar o experimento, o sistema deve providenciar sua execução e quando finalizar, retornar o resultado do experimento ao usuário que o requisitou.

Para este portal, será utilizado um algoritmo desenvolvido para a área de astronomia, uma versão do algoritmo Friends-of-Friends de complexidade $n \cdot \log(n)$ paralelizada através do framework OpenMP.

3 Objetivos

3.1 Objetivo Geral

O objetivo deste trabalho é criar um portal web que possibilite aos usuários cadastrados no sistema executar algoritmos presentes no sistema segundo suas configurações e disponibilizar o resultado da execução após o término da mesma.

3.2 Objetivos Específicos

- Estudo de frameworks web para ser utilizado no desenvolvimento.
- Desenvolvimento front-end do servidor.
- Administração das execuções requisitadas.
- Atualização dos estados das requisições no sistema.

4 Justificativa

O projeto é capaz de gerar benefícios significativos para a comunidade de pesquisa de diversas áreas, criando um modelo de ambiente que facilite a divulgação e teste de resultados de algoritmos alternativos para resolução de problemas comuns.

Além de servir como modelo, o projeto disponibilizará um algoritmo que se enquadra na categoria alvo do projeto: a versão de complexidade $n \cdot \log(n)$ e paralela do friends-of-friends.

5 Revisão de Literatura

5.1 Trabalhos Relacionados

Os trabalhos citados a seguir se relacionam com a proposta deste trabalho pois compartilham características como:

- Garantir ao usuário a capacidade de fazer seus experimentos em um ambiente que está além da capacidade normal de processamento de um equipamento pessoal.
- Disponibilizar softwares já configurado para o ambiente aonde será executado.
- Disponibilizar soluções para problemas que não são oferecidas ao público geral.

5.1.1 New Zeland eScience Infrastructure - NeSI

O NeSI[3] provê plataformas de grande capacidade computacional para auxiliar pesquisadores na Nova Zelândia. Atualmente eles possuem 5 ambientes disponíveis em diferentes instalações. Cada um desses ambientes possui hardware e software capazes de resolver problemas relacionados a áreas de pesquisa, especialmente relacionados a química e bioinformática.

Um desses ambientes por exemplo é o FitzRoy, que dispõe de várias ferramentas para o ambiente de programação como compiladores para as linguagens C/C++, Fortran e Python, além de *debuggers*, ferramentas de *profiling*. Além disso também conta com bibliotecas úteis que auxiliam em questões como entrada e saída de dados e problemas matemáticos. Fora da área de programação, o ambiente oferece também aplicações de simulações que requerem alto desempenho do hardware.

Cada um dos ambientes disponíveis possui seu próprio site institucional com sua respectiva documentação e formas de contato, além disso o próprio NeSI oferece suporte a usuários através da sua equipe.

5.1.2 MediGRID

O MediGRID [2] é um portal focado em pesquisa na área da biomedicina, ele oferece aplicações e infraestrutura para os pesquisadores cadastrados no portal realizarem experimentos conforme a sua necessidade e a disponibilidade do sistema.

O objetivo principal do MediGRID quando foi desenvolvido era ser uma plataforma de integração middleware ligando serviços eScience com as pesquisas de biomedicina. Porém hoje ele executa tarefas em 3 áreas majoritárias: Biomedicina, processamento de imagens e pesquisa clínica.

5.2 Framework Django

Django[1] é um framework para criação de aplicações web que encoraja o desenvolvimento ágil, em alto nível e com design pragmático. Foi criado inicialmente para manter o portal de notícias online do Lawrence Journal World pelos programadores Adrian Holovaty, Simon Willison e Jacob Kaplan-Moss.

Por se tratar de um portal inicialmente desenvolvido para administrar notícias, o framework lida muito bem com gerenciamento de conteúdo e agilidade quando é necessário fazer alterações no sistema.

5.3 Friends of Friends

Simulações de N-corpos têm sido utilizadas para promover vários avanços na compreensão de questões relevantes em astrofísica, como por exemplo o processo de formação e evolução de estruturas do Universo. Este tipo de simulação tem um papel fundamental[4, 5] no estudo da evolução cósmica em tópicos como a distribuição de matéria escura em grande escala, a formação de halos de matéria escura, e a formação e evolução de galáxias e aglomerados.

A manipulação e análise da grande quantidade de dados produzidos em tais simulações também é algo desafiador. Neste contexto, é essencial o desenvolvimento de técnicas computacionais eficientes para extrair informação significativa a partir dessas fontes de dados, em um período apropriado de tempo.

Etapas importantes neste tipo de análise são a identificação de halos de matéria escura e o estudo do espectro da energia potencial gravitacional de tais objetos. Uma abordagem para este tipo de análise consiste em usar o algoritmo de percolação Friends-of-Friends (FoF) [6]. A ideia básica deste algoritmo é a seguinte: considere uma esfera de raio R ao redor de cada partícula do conjunto total; se dentro desta esfera existirem outras partículas, elas serão consideradas pertencentes ao mesmo grupo e serão chamadas de amigas. Em seguida, toma-se uma esfera ao redor de cada amiga e continua-se o procedimento usando a regra "qualquer amigo de meu

amigo é meu amigo”. O procedimento para quando nenhuma amiga nova puder ser adicionada ao grupo.

Na 15ª edição da Escola Regional de Alto Desempenho do Rio Grande do Sul (XV ERAD-RS)(referenciar), foram publicados resultados de execuções de uma nova implementação do Friends-of-Friends[7] cuja complexidade computacional era reduzida em relação as versões anteriores, resultando em redução do tempo de processamento.

6 Metodologia

O trabalho será realizado segundo metodologia de pesquisa aplicada, na qual será feito um estudo de diversos *frameworks* que podem ser úteis para o sistema proposto. Além disso também serão analisadas soluções concorrentes ou parecidas constadas na revisão de literatura para compreensão de recursos necessários para o portal disponibilizar aos usuários.

Após a seleção do framework a ser utilizado e um levantamento de requisitos será iniciada a etapa de desenvolvimento do portal, resultando em um protótipo que tende a suprir todos os requisitos discutidos na sessão de objetivos.

Quando este protótipo foi obtido, será feito a instalação do mesmo no servidor aonde o front-end do portal será mantido. Na sequência, os esforços serão direcionados para os ajustes finos que podem ser necessários por causa de rede da instituição na qual o servidor se encontra.

7 Plano de Atividades e Cronograma

1. **Estudo de Frameworks Web para desenvolvimento** Será realizado um estudo dos frameworks mais conhecidos no mercado e selecionado o que mais apresentar benefícios para o desenvolvimento da aplicação proposta
2. **Desenvolvimento front-end** Após o estudo e seleção do framework para implementação, inicia-se o início do desenvolvimento do servidor front-end, criando o portal com as funções CRUD para gerenciar contas de usuários e os experimentos.
3. **Administração das Execuções requisitadas** Por se tratar de um portal que permitirá a execução remota de aplicações de alto desempenho, sabe-se que será necessário um estudo de como implementar a administração das requisições dos experimentos, visto que não deve existir concorrência de 2 ou mais experimentos no sistema.

4. **Atualização de estados dos experimentos** Quando resolver-se a questão de execução dos algoritmos pelo sistema, também será necessário criar um método de verificar quando essas execuções finalizaram, bem como disponibilizar os resultados obtidos para o usuário no portal.

Cronograma

Etapa	Agosto	Setembro	Outubro	Novembro	Dezembro
1	✓				
2	✓	✓			
3		✓	✓	✓	
4			✓	✓	✓

Tabela 1: Cronograma de Atividades

8 Recursos

Para a realização do projeto será utilizado o equipamento pessoal do orientando para a criação do servidor front-end. Além disso, será utilizado um servidor disponibilizado pelo Laboratório de Sistemas de Computação (LSC) da Universidade Federal de Santa Maria para manter o portal quando este estiver pronto para ser testado por requisições de usuários e para executar o algoritmo de alto desempenho Friends-of-Friends.

9 Resultados Esperados

Ao término do trabalho, espera-se obter um sistema funcional compatível com o que foi explorado na sessão de Objetivos, um portal web que permita cadastro de usuários e no qual, cada usuário pode requisitar a execução de algoritmos existentes no sistema, utilizando para tal suas próprias configurações de execução (Dados de entrada, propriedades de execução).

Referências

- [1] Django project. <https://www.djangoproject.com/>. Accessed: 2015-08-14.
- [2] Medigrid. http://www.medigrid.de/index_en.html. Accessed: 2015-08-14.
- [3] New zeland escience infrastructure. <https://www.nesi.org.nz/>. Accessed: 2015-08-14.

- [4] E. Bertschinger. Simulations of structure formation in the universe. *Annu. Rev. Astron. Astrophys* 36, 1998.
- [5] S. D. M White C. S Frenk G. Efstathiou, M. Davis. Numerical techniques for large cosmological n-body simulations. *Astrophysical Journal Supplement Series (ISSN 0067-0049)*, vol. 57, 1985.
- [6] Geller M. J. Huchra J. P. Groups of galaxies. i - nearby groups. *Astrophysical Journal, Part 1*, vol. 257, June 15, 1982, p. 423-437. *Research supported by Cambridge University*, 1982.
- [7] Andrea Schwertner Charão Otávio Migliavacca Madalosso. Implementação do algoritmo friends of friends de complexidade $n \cdot \log(n)$ para classificação de objetos astronômicos. *ERAD-RS XV*, 2015.