

BOINC + R: Executando rotinas de bioinformática em grades oportunistas

Rodrigo L. M. Flores
flores@ime.usp.br

Instituto de Matemática e Estatística
Universidade de São Paulo

16 de novembro de 2009

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Introdução

Ferramentas utilizadas

Discussão e resultados

Conclusão

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Bioinformática

- ▶ Aplicação de técnicas computacionais e matemáticas para geração, gerenciamento e análise de bioinformação
- ▶ Análise de expressão de genes de dados, Clustering
- ▶ Algoritmos combinatórios e custosos

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Computação em grade

- ▶ Clusters dedicados: caros e específicos;
- ▶ Computadores pessoais em um laboratório ociosos na maior parte do tempo
- ▶ Não há necessidade de se comprar novos equipamentos

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Computação voluntária

- ▶ Slogan: “Doe seus ciclos de CPU para um projeto”;
- ▶ Normalmente projetos com apelo “humanitário” (pesquisas de enovelamento de proteínas, malária, descoberta de vida extraterrestre, resolução de instâncias grandes do N-Rainhas, etc);
- ▶ Computação em grade cuja rede em comum é a Internet;
- ▶ Vontade de entrar na grade é do dono do computador;
- ▶ É possível configurar para que os ciclos de CPU sejam doados se o computador estiver ocioso, se estiver funcionando na energia elétrica, ou programar perfis diferentes em horários ou dias da semana
- ▶ Exemplos de projetos: SETI@HOME, Folding@Home, N Queens At Home, World Community Grid
- ▶ Projetos consolidados e funcionando a anos

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Uso de computação voluntária como uma grade

- ▶ Baseado no artigo [GGdVS08]
- ▶ Computadores são de uma empresa ou instituição, e a decisão de como eles devem ser utilizados, não são apenas mas sim de uma instituição
- ▶ Utilização de uma rede de laboratórios para isto
- ▶ Possibilidade de executar projetos próprios ou projetos já existentes

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão



- ▶ Um produto do projeto SETI@HOME (que infelizmente falhou em achar vida inteligente fora da terra)
- ▶ Middleware que possui suporte a vários projetos
- ▶ Software Livre (tanto no servidor como no cliente)

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão



- ▶ Linguagem de programação interpretada, baseada na linguagem S e com enfoque estatístico
- ▶ Muito utilizada na área de bioinformática
- ▶ Possibilidade de se compartilhar módulos (há mais de mil pacotes disponíveis para download)

Hardware

Servidor

- ▶ Hardware: AMD Athlon X2 4600+, 4GB de RAM
- ▶ Sistema Debian GNU/Linux
- ▶ BOINC Server versão 6.7.0

Clientes

- ▶ 1 máquina Windows
- ▶ 3 máquinas Linux
- ▶ Em processo de expansão (rede em “reforma”)

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Aplicações Legadas

Aplicações legadas

- ▶ Aplicações já compiladas
- ▶ Não há a necessidade de conhecer ou alterar o código
- ▶ Uso do wrapper

Wrapper

- ▶ Programa feito com a API do Boinc que lê um XML (chamado job.xml) com informações de como executar um executável
- ▶ Disponível para Linux e Windows

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

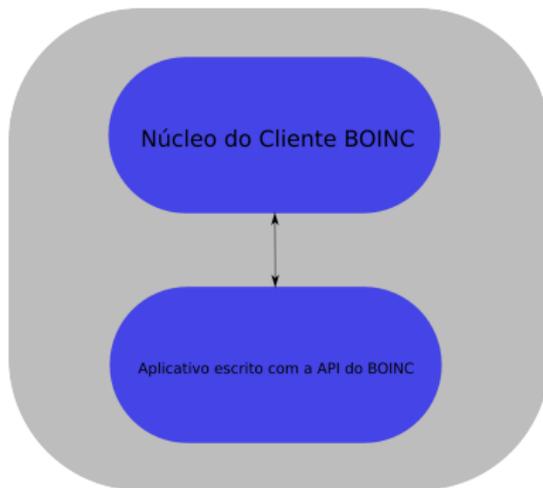
Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Wrapper



BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

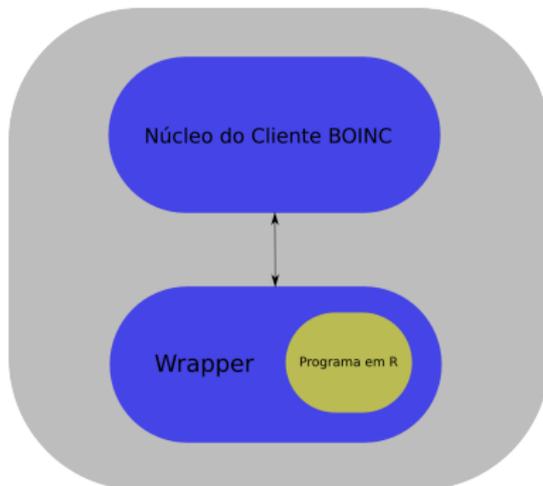
Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Wrapper



BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Implementações da solução

Sistema Linux

- ▶ Processo simples de compilação (via makefile)
- ▶ Possibilidade de usar o truque do *shebang* e usar o script como binário

Sistema Windows

- ▶ Processo de compilação mais complexo (via Microsoft Visual C++)
- ▶ Configuração de compilação do wrapper com problema
- ▶ Uso do interpretador como binário
- ▶ Pequenas dificuldades relacionadas ao sistema

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Instalação na rede do CEC

- ▶ Instalação simples no sistema Linux e no Windows
- ▶ Reforma na rede do CEC
- ▶ No momento 3 máquinas Linux e 1 máquina Windows instalada
- ▶ Instalação não finalizada (mas em andamento)

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Vantagens e desvantagens

Vantagens

- ▶ Facilidade de se adicionar novos nós;
- ▶ Processamento multiplataforma;
- ▶ Código aberto;
- ▶ Usuário não percebe que há uma tarefa executando;
- ▶ Solução funciona para qualquer linguagem de script;
- ▶ Solução boa para muitos workunits que gastam pouco tempo.

Desvantagens

- ▶ Necessidade de se ter o R instalado;
- ▶ Não há checkpoint;

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Conclusões

- ▶ Solução funciona;
- ▶ Único computador inutilizado foi o servidor;
- ▶ Custo baixo;
- ▶ Facilidade de instalação e configuração;
- ▶ Dificuldade do uso do Wrapper e mensagens de erro pouco amigáveis

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Próximos passos

- ▶ Benchmark da grade e comparação com a grade com o mesmo propósito descrita no artigo [RAD09]
- ▶ Tentar o uso da GPU para aumentar o processamento;
- ▶ Comparação com o gasto energético deste tipo de grade com o gasto de processamento de uma grade “alugada”;
- ▶ Analisar o desempenho dos dois sistemas utilizados
- ▶ Escrita de um artigo para ficar disponível na webpage do BOINC como alternativa da computação em grade;
- ▶ Uso de máquinas virtuais
- ▶ Disponibilizar a grade para submissões da comunidade científica;

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Agradecimentos

- ▶ CNPq
- ▶ Prof. Dr. Roberto Hirata Jr.
- ▶ Prof. Dr. Alfredo Goldman
- ▶ Prof. Dr. Roberto Marcondes
- ▶ Rodrigo Assirati Dias
- ▶ Rodrigo Bernardo Pimentel
- ▶ Nicolás Alvarez e Yoyo (desenvolvedores do BOINC)
- ▶ Gislaine Olivi e Janio Matsuura (administradores do CEC)

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão

Referências

 D.L. Gonzalez, G.G. Gil, F.F. de Vega, and B. Segal, *Centralized boinc resources manager for institutional networks*, April 2008, pp. 1–8.

 Roberto Hirata Jr. Rodrigo A. Dias, *Middle-r - a user level middleware for statistical computing*, VII Workshop on Grid Computing and Applications - Recife, Pernambuco - Brazil (2009).

BOINC + R:
Executando rotinas
de bioinformática
em grades
oportunistas

Rodrigo L. M.
Flores
flores@ime.usp.
br

Introdução

Ferramentas
utilizadas

Discussão e
resultados

Conclusão