

Análise Paralela e Distribuída de Dados Micrometeorológicos Utilizando a Plataforma JXTA

Marcelo Veiga Neves¹, Tiago Scheid², Andrea Schwertner Charão²,
Guilherme Sausen Welter³, Osvaldo Luiz Leal de Moraes³

¹Instituto de Informática - Universidade Federal do Rio Grande do Sul (UFRGS)
²Laboratório de Sistemas de Computação (LSC)
Curso de Ciência da Computação - Universidade Federal de Santa Maria (UFSM)
³Laboratório de Micrometeorologia (LuMet)
Centro de Ciências Naturais e Exatas - Universidade Federal de Santa Maria (UFSM)

mveives@inf.ufrgs.br, (scheid, andrea)@inf.ufsm.br
gselter@yahoo.com.br, ollmoraes@mail.ufsm.br

Objetivos

- Paralelizar uma aplicação de análise de dados coletados por sensores micrometeorológicos
- Obter ganho de desempenho com a adaptação da aplicação
- Flexibilizar o uso dos recursos computacionais disponíveis
 - Aproveitar a ociosidade dos mesmos
- Explorar a convergência entre grades e P2P

Aplicação Alvo

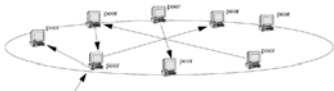
- Análise de dados coletados por sensores micrometeorológicos
- Utilizada pelo Laboratório de Micrometeorologia da UFSM (LuMet)
- Código em FORTRAN 77
- Normalmente utilizada para processar arquivos de aproximadamente 100MB
- A execução leva, em média, 24.04 min. em uma máquina Pentium 4 a 2.4 Ghz

Fluxo de dados das rotinas da aplicação



Arquitetura do Sistema Desenvolvido

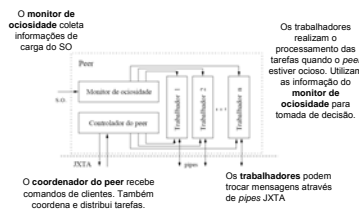
- Implementa uma plataforma dinâmica de execução



Fonte: Traversat, B. 2003, figura adaptada.

- Segue um modelo P2P puro, onde peers são os computadores que integram a rede virtual JXTA
- Monitora a ociosidade computacional dos peers
- A adaptação da aplicação micrometeorológica gerou duas versões, uma distribuída e outra paralela

Estrutura de um Peer



Avaliação do Sistema

- Objetivo: avaliar a sobrecarga imposta pela plataforma de execução implementada

	Tempo médio	Desvio padrão
Inicialização do peer	23.32	4.77
Envio de arquivo de dados	14.59	1.10
Processamento	1825.20	9.05
Retorno dos resultados	9.79	1.94

- tempo de inicialização é altamente depende da rede
- tempo de processamento aumentou (de 24.04 para 30.42 min.)
- tempo de transmissão não é muito significativo se comparado ao tempo de processamento

Avaliação da Aplicação Distribuída

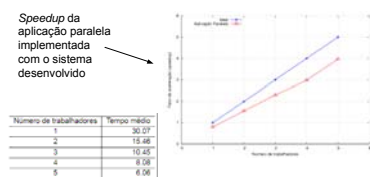
- Objetivo: avaliar o ganho de desempenho processando diferentes conjuntos de dados simultaneamente.

Número de Conj. de Dados (1 dia)	Tempo médio
1	30.76
2	31.18
3	32.66

- duplicando a quantidade de dados, o tempo de execução aumenta em 1.36 %
- comparado com a versão sequencial (24.04 min.), obteve-se um ganho de 35 % ao processar 2 arquivos e de 55 % ao processar 3 arquivos simultaneamente

Avaliação da Aplicação Paralela

- Objetivo: avaliar o ganho de desempenho ao processar um conjunto de dados de forma paralela



Conclusão

- A distribuição permitiu processar vários conjuntos de dados simultaneamente, obtendo um ganho de desempenho
- A paralelização aumentou a eficiência, reduzindo o tempo de execução
- O sistema implementado poderá melhorar o aproveitamento dos recursos disponíveis no laboratório LuMet
- A plataforma implementada pode ser adaptada para outras aplicações
- A plataforma implementada pode ser adaptada para outras aplicações, permitindo a execução em grades