

Paralelização da Execução do Modelo Matemático CROPSIM para Simulação da Cultura do Trigo em um Grid Oportunista

Vinícius C. Reck¹, Élder F. F. Bernardi¹, Alexandre T. Lazzaretti¹, João Eduardo Pedrini²,

¹Instituto Federal Sul-Rio-Grandense – Campus Passo Fundo

²Universidade de Passo Fundo – Campus Passo Fundo

Estrada Perimetral Leste, 150 - 99064-440 - Passo Fundo - RS – Brasil

contato@viniciuscr.com.br,

{[elder.bernardi](mailto:elder.bernardi@passofundo.ifsul.edu.br), [alexandre.lazzaretti](mailto:alexandre.lazzaretti@passofundo.ifsul.edu.br)}@passofundo.ifsul.edu.br,

jjepjjep@gmail.com

1. Introdução

Sistemas computacionais têm tido uma importância crescente em diversos setores, entre esses, o setor agrícola. A utilização dos modelos de simulação de desenvolvimento e crescimento de cultivares é uma das aplicações de destaque.

Tais modelos combinam informações oriundas de diversas fontes, a saber: dados de clima, coeficientes de culturas, dados de solo, dados de manejo das culturas, entre outros. Normalmente essas fontes têm registros de vários anos e são constantemente acrescidas, ampliando significativamente o conjunto de dados. Isto contribui para que esses modelos tenham sua execução onerosa em termos de tempo, quando executados com uma abordagem sequencial. Em virtude disso, se faz necessário a exploração de meios de distribuir a carga de execução de tais modelos de modo que se possa assim fazer uso de arquiteturas distribuídas e paralelas objetivando a redução do tempo de execução desses modelos.

Assim, apresenta-se uma abordagem de paralelização do modelo CROPSIM em uma arquitetura de Grid Oportunista, mostrando-se a abordagem utilizada os resultados obtidos. Tal arquitetura foi escolhida em virtude de seu baixo custo de implantação e pelo fato de que o modelo de simulação explorado se faz presente em organizações que geralmente possuem uma infraestrutura computacional subutilizada em laboratórios, ou até mesmo servidores.

2. Grids Oportunistas

Um dos possíveis ambientes formados a partir da união de computadores é chamado de *Grid* [DANT05], sendo esse composto por computadores de diferentes configurações e, até mesmo diferentes arquiteturas, unidos por um conjunto de ferramentas denominado *middleware*. Quando os computadores não são de uso exclusivo do *grid*, ou seja, possuem outras funções dentro do cenário em que se encontram, diz-se que o *grid* é oportunista, pois faz uso dos recursos ociosos da infraestrutura já existente [DANT05].

2.1. Middleware Alchemi

Alchemi [LUTH07] é um *middleware* de *grid* desenvolvido pela University of Melbourne, com o objetivo de fornecer um ambiente de HPC para aplicações

desenvolvidas na plataforma .NET sobre o ambiente Microsoft Windows, mas também permite a execução de aplicações desenvolvidas em outras plataformas, nesse caso fornecendo uma ferramenta para o envio de aplicações em lotes.

3. Modelo de Simulação CROPSIM

A ferramenta utilizada para o desenvolvimento deste trabalho é o CROPSIM [HUNT95] que se trata de um modelo matemático desenvolvido em Fortran que, baseado em dados climatológicos e ecológicos, simula o crescimento e desenvolvimento da cultura do trigo. Tal modelo é amplamente utilizado, destacando-se seu uso pela EMBRAPA e UPF.

3.1. Aplicação e Utilização do CROPSIM

O modelo de simulação da cultura do trigo CROPSIM pode ser utilizado para diversos experimentos e trabalhos, como, por exemplo, calcular o efeito que um determinado cenário causa no crescimento e desenvolvimento na cultura do trigo. A partir das saídas que o modelo gera, podem ser extraídas diversas informações, como estimar a produção da cultura do trigo em um cenário de seca em determinado período de desenvolvimento, determinar qual estágio fenológico de crescimento fica mais comprometido em um ano de influência do el-niño; estimar qual a melhor época de semeadura para a cultura em determinada região, entre outras. Estes dados são úteis para geração de cenários históricos, para prevenção de eventos futuros, e também sendo útil em um sistemas de suporte à tomada de decisão.

3.2. Funcionamento

O modelo basicamente utiliza como entrada dados de clima, solo, cultivar, e um arquivo com parâmetros gerenciais, onde são definidos os cenários, os quais serão simulados. Após a execução desses cenários, diversos arquivos com as saídas das simulações são gerados, os quais serão analisados para extração de conhecimento. Esse processo se repete milhares de vezes, onde em cada nova simulação novos dados de entrada independentes são passados.

4. Execução do CROPSIM em um *grid* oportunista

Para que os resultados produzidos pelo modelo possam ser utilizados é necessário que uma grande quantidade de dados seja utilizada em sua entrada, o modelo precisa ser executado milhares de vezes com entradas diferentes, gerando saídas diferentes a cada execução. Sendo que cada uma dessas execuções são independentes entre si, esse comportamento identifica uma aplicação da classe do tipo *parameter sweep*, sendo assim, passível de ser paralelizada.

Seguindo esse raciocínio, foi possível agilizar o processo de cálculos do modelo matemático executando em paralelo os processamentos com suas diferentes entradas. Para isso foi gerado, para o *middleware* Alchemi, uma tarefa para cada execução do modelo, onde cada tarefa recebe uma entrada de dado diferente, sem haver qualquer dependência entre as tarefas, gerando assim, um alto grau de granularidade e escalabilidade da execução paralela do modelo.

4.1. Experimentos

Os experimentos para avaliação seguiram a seguinte metodologia: foram executadas

oito baterias de testes, onde cada bateria foi constituída de 5 execuções redundantes, onde cada execução foi composta por 2400 execuções. A bateria inicial serviu como parâmetro para medição de desempenho das demais e foi baseada na utilização do *grid* com apenas uma máquina para realizar o processamento. A partir daí, seguiu-se pelo acréscimo de uma máquina a cada nova bateria, até o total de 8 máquinas. Ao final, foram levantados todos os resultados e eliminados o melhor e pior resultado de cada bateria, realizando a média dos 3 resultantes. O baixo número de repetições se deu devido à baixa disponibilidade do cenário para testes, o que delimitou o tempo dos testes.

4.2. Cenário de execução

O cenário escolhido será um dos laboratórios de informática do Instituto Federal Sul-Rio-Grandense - Campus Passo Fundo. O laboratório é constituído por 9 máquinas de configuração semelhantes, correspondentes a um processador de 3GHz com 2GB de memória RAM. No entanto, 8 máquinas serão utilizadas para realizar o processamento e a nona máquina conterá o *Manager* da ferramenta para gerenciar o *Grid*.

É importante salientar que as máquinas utilizadas não são destinadas unicamente para este propósito, sendo que, seu principal objetivo é servir como laboratório de aulas práticas para a instituição.

4.3. Resultados

Ao término das execuções e após finalizar o levantamento dos dados obtidos, foi possível reunir e analisar os dados resultantes, mensurando suas variações de tempo e realizando o cálculo de *Speed-Up* e eficiência. Com essas informações foi montado o Gráfico 1 com a exposição dos resultados, sendo o eixo X correspondente ao número de nós do Grid, o eixo Y, ao lado direito do gráfico, representa a escala de eficiência e o eixo Y, ao lado esquerdo, a escala do *Speed-Up*.

Como parâmetro para avaliação dos tempos através do Figura 1, observa-se que o tempo de execução sequencial do modelo foi de aproximadamente 12 minutos, enquanto, com 8 processos, em 8 nós, obteve-se, aproximadamente, 1 minuto e 40 segundos. Tendo-se este desempenho, pode-se inferir que numa execução real onde seriam executados 48.000 entradas de dados, podendo-se potencialmente ser executadas em 4.8000 tarefas, poderia-se, com a paralelização da execução, diminuir o tempo de execução do modelo de 228 minutos para 28 minutos, o que atesta a viabilidade da paralelização da execução do referido modelo.

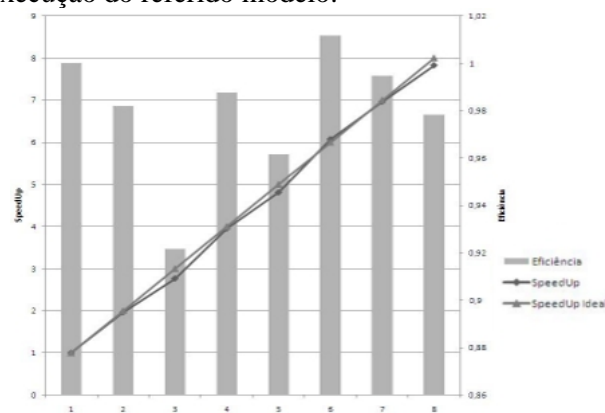


Figura 1. Gráfico de speedup e eficiência

5. Conclusão e Trabalhos Futuros

Ao fim dos experimentos e com a análise do Gráfico 1 é possível notar um crescimento linear no *Speed-Up* da aplicação à medida que novos recursos são incluídos no *Grid*.

Apesar do curto espaço de tempo para cada execução, o acúmulo de grandes quantidades da mesma, ocasiona um longo tempo de espera para obtenção dos resultados. Isso ocorre porque as execuções são realizadas sequencialmente, subutilizando os recursos computacionais disponíveis.

Em um ambiente paralelo várias execuções são realizadas simultaneamente, sendo assim, realizando as execuções do simulador nesse ambiente obteve-se uma diminuição significativa no tempo total para realização das simulações, visto que, com a paralelização das execuções foi possível amenizar o gargalo.

Uma possível continuação do trabalho seria a a expansão dos testes com novas ferramentas de análise de cultivares ou, até mesmo, testes em novos *middlewares* de *Grid* para avaliar seus desempenhos.

References

- [DANT05] DANTAS, M. (2005), “Computação Distribuídas de Alto Desempenho: Redes, Clusters e Grids Computacionais”. [S.l.]: Axcel Books do Brasil Editora.
- [CONT09] CONTI, F. de. (2009) ,“Grades computacionais para processamento de alto desempenho”. Universidade Federal de Santa Maria.
- [LUTH07] LUTHER, A. et al. , (2003), “Alchemi: A .NET-based Grid Computing Framework and its Integration into Global Grids”.
- [HUNT95] HUNT, L. A., (1995) PARARAIASINGHAM, S. “Cropsim - wheat:a model describing the growth and development of wheat”. Department of Crop Science, University of Guelph, Ontario, Canada.