



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

**ANÁLISE ESTATÍSTICA DE SIMULAÇÕES IONOSFÉRICAS USANDO
LINGUAGEM R**

Gabriela Luisa Eckel

Relatório Final de Iniciação
Científica do programa PIBIC,
orientada pelo Dr. Adriano Petry.

INPE
Santa Maria
2021

RESUMO

Com a evolução contínua da tecnologia, a computação de alto desempenho tornou-se uma das principais áreas de pesquisa no ambiente computacional. Os aplicativos que precisam processar grandes quantidades de dados em um curto período de tempo geralmente usam ambientes de alto desempenho. Para isso, os sistemas distribuídos podem ser usados para melhorar a eficiência e a velocidade de execução. Em um ambiente de alto desempenho, há uma grande quantidade de recursos de computação e utiliza OAR, SLURM e outros sistemas de gerenciamento de recursos para otimizar o uso. Ação científica para simular o comportamento da ionosfera terrestre na América do Sul, gerando mapas de conteúdo eletrônico total todos os dias. A computação de alto desempenho é amplamente utilizada nos mais diversos campos de pesquisa, como previsão do tempo e simulação diária, pois fornece processamento em larga escala e pode obter resultados com rapidez e eficiência. Em um ambiente de alto desempenho, o uso de arquitetura de cluster fornece recursos de processamento altamente distribuídos para aplicativos, pois eles possuem computadores (geralmente chamados de nós) interconectados por redes de alta velocidade e trabalham juntos para resolver problemas, na maioria dos casos. Não é viável para ser executado por um computador a tempo. Para gerenciar todos os recursos de hardware e software em um ambiente de alto desempenho, um sistema de gerenciamento de recursos (SGR) é necessário. Buscar aumentar a utilização de recursos e a velocidade de obtenção de resultados depende de diversos fatores, além da escolha de um SGR eficiente, como OAR e SLURM Simple Linux Utility for Resource Management, é extremamente importante. Reitero que devido a pandemia a pesquisa realizada não foi a mesma que a que estava escrita na proposta da bolsa, então decidimos continuar a proposta do ano anterior.

SUMÁRIO

1. INTRODUÇÃO	3
2. DADOS E METODOLOGIA.....	4
3. RESULTADOS.....	5
4. CONCLUSÃO.....	7
5. REFERÊNCIAS.....	8

1. INTRODUÇÃO

A ionosfera é a atmosfera, que contém elétrons livres e íons carregados. O sistema de previsão de dinâmica ionosférica é desenvolvido e operado todos os dias, fazendo previsões com 24 horas de antecedência. O conteúdo total de elétrons na ionosfera pode interferir nos dados de posicionamento do sistema de navegação global por satélite (GNSS), então a simulação é muito importante para evitar erros de posicionamento. Parte do sistema de predição ionosférica é desenvolvida em linguagem C++ no INPE e tem como foco a interpolação de dados SUPIM. [Prabhu 2008]

Atualmente, a HPC (High Performance Computing) fornece os recursos de computação necessários para aplicativos que requerem altos recursos de computação nos mercados financeiros, previsão do tempo e simulação de fluidos físicos. O poderoso poder de processamento fornecido pelo ambiente de alto desempenho garante o tempo necessário para obter os mais diversos resultados seja reduzido de maneira exponencial. [Sloan 2004]

Um cluster é um exemplo de arquitetura de alto desempenho, que fornece alto poder de processamento por meio da interconexão de computadores (geralmente chamados de nós) por meio de uma rede de alta velocidade, e é projetado para distribuir o processamento necessário para a solução de problemas, o que é impossível na maioria dos casos. Use um computador tradicional. SGR (Resource Management System) é responsável por gerenciar jobs, usuários e recursos computacionais distribuídos no cluster para obter resultados com rapidez e eficiência. A escolha do SGR é a base para otimizar o uso dos recursos disponíveis, e alternativas como SLURM e OAR são muito comuns nesses ambientes. Nicholas et al. 2016] [Yoo et al. 2003]

Este estudo tem como objetivo avaliar e comparar SLURM e OAR SGR na aplicação científica real da previsão ionosférica. A aplicação gera mapas de TEC (Conteúdo Eletrônico Total) para a região da América do Sul que são disponibilizados em <http://www2.inpe.br/climaespacial/portal/tec-supim-previsao/>. [Petry et al. 2014]

2. DADOS E METODOLOGIA

Durante o período da bolsa foram realizadas duas pesquisas, onde a segunda é uma evolução da primeira, por este motivo optamos por detalhar a pesquisa que foi aceita para publicação na Escola Regional de Alto Desempenho (ERAD 2020).

Os experimentos foram realizados em um cluster de uso dedicado localizado no Centro Regional Sul de Pesquisas Espaciais do Instituto Nacional de Pesquisas Espaciais (CR-CRS/INPE). O cluster utilizado possui 6 nós de processamento e um nó de controle. Cada nó de processamento possui 8 CPUs Intel Xeon rodando a 2.40 GHz, com 74 GB de memória RAM, sistema operacional CentOS 6.4 e os Sistemas Gerenciadores de Recursos SLURM e OAR instalados. Foram realizadas simulações ionosféricas para 10 dias consecutivos, iniciando em 1 de dezembro de 2019. Para cada um dos dias foram realizadas duas simulações ionosféricas, uma utilizando o SLURM e outra o OAR. Para os experimentos foi utilizado o algoritmo FIFO -First In First Out. Em ambos os SGRs e todos os jobs foram configurados com a mesma prioridade na fila de execução. Para análise comparativa foram analisados a utilização de CPU, de memória e o tempo total de execução das simulações. Para isso foram utilizadas as ferramentas: mpstat para medição da taxa de utilização de CPUs, vmstat para taxa de utilização de memória, e o tempo de execução foi medido no log da simulação. As informações de CPU e memória foram coletadas a cada 3 segundos e armazenadas em arquivo. Após a coleta de dados foram calculadas as medidas de utilização.

3. RESULTADOS

A Figura 1 ilustra a taxa de utilização das CPUs quando utilizados os SGRs SLURM e OAR, representados nas cores azul e vermelho, respectivamente. Com os dados obtidos através das simulações é possível observar que o SLURM obteve um desempenho superior ao OAR em todos os dias simulados. Este resultado é devido a forma de controle dos jobs na fila de execução. No SLURM é possível realizar a submissão de todos os jobs e o SLURM irá realizar o gerenciamento dos jobs pendentes, agilizando a submissão do próximo job da fila assim que recursos computacionais estiverem ociosos. Já no OAR é preciso realizar a busca no banco de dados dos jobs pendentes e aguardar até que recursos computacionais estejam disponíveis, o que acaba atrasando a submissão de jobs na fila e consequentemente deixando recursos computacionais ociosos.

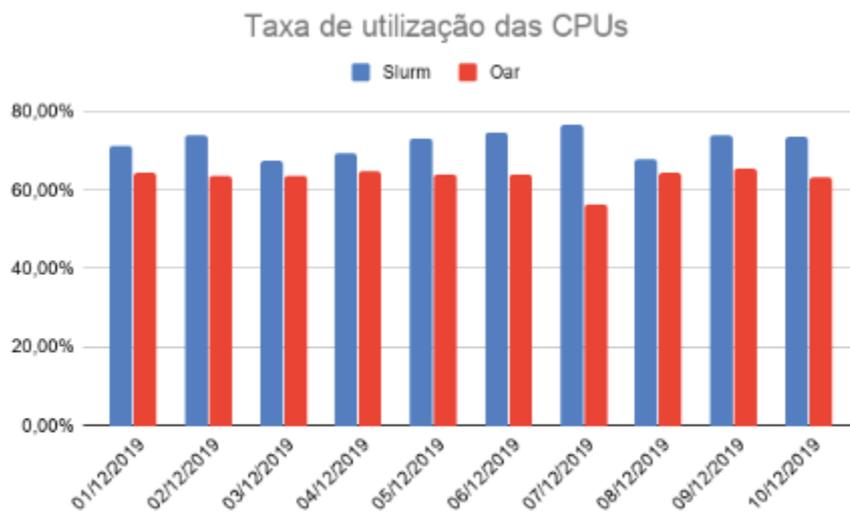


Figura 1. Média de uso da CPU

. A Figura 2 mostra a taxa de utilização de memória quando utilizados os SGRs. Mesmo que com uma baixa taxa de utilização de memória, é possível observar uma pequena melhora quando utilizado o SLURM. A baixa taxa de utilização de memória é resultado da aplicação ionosférica utilizada para os experimentos, onde na sua maioria os jobs não realizam solicitação de memória elevada, mesmo assim em alguns momentos da simulação a taxa de utilização chega próximo aos 10%.



Figura 2. Média de uso da memória

Na Tabela 1 é possível observar o tempo de execução do SLURM comparado ao OAR nos diferentes dias de teste. Podemos verificar que a simulação diminuiu seu tempo de execução consideravelmente utilizando o SGR SLURM. Explica-se que o motivo disto é a forma de controle dos jobs na fila de execução, e a sua eficiência distribuição destes quando alguns nos estiverem ociosos.

DATA	SLURM	OAR
01/02/2019	02:26	02:56
02/12/2019	02:35	03:29
03/12/2019	02:24	03:56
04/12/2019	02:34	02:32
05/12/2019	02:35	02:45
06/12/2019	02:33	02:33
07/12/2019	02:32	03:33
08/12/2019	02:36	03:00
09/12/2019	02:35	02:45
10/12/2019	02:34	02:54

Tabela 1. Análise Comparativa de tempo de execução

4. CONCLUSÃO

Este trabalho teve como objetivo analisar diferentes SGRs para um ambiente científico que executa diariamente um sistema de previsão ionosférica. Após a realização dos experimentos utilizando o OAR e o SLURM, constatou-se um melhor desempenho do SGR SLURM nas três comparações realizadas: a utilização da CPU, memória e a comparação de tempo de execução. Na taxa de utilização da CPU o SLURM foi superior em todas as análises. A taxa de utilização de memória, apesar de muito semelhante ao OAR, também se mostrou mais eficiente com o SLURM. Nota-se que a utilização do SGR SLURM tornou o ambiente propício a melhoras na velocidade da execução, o que podemos perceber com o ganho de cerca de 30 minutos, em média, no tempo de processamento.

É importante destacar que os experimentos foram realizados em um ambiente real onde uma aplicação científica executa diariamente e mesmo o OAR apresentando resultados inferiores, quando comparado ao SLURM, o tempo total de execução da simulação de ambos SGRs é aceitável para disponibilização diária dos resultados. Como pesquisas futuras pretende-se testar os dois SGRs em um ambiente formado por nós com configuração heterogênea e com algoritmos de escalonamento distintos aplicados aos SGRs, onde será possível analisar outras métricas de desempenho de sistemas gerenciadores de recursos em um ambiente real.

O período da bolsa resultou em uma publicação [Eckel 2020] na Escola Regional de Alto Desempenho da Região Sul e um artigo aprovado para publicação no Simpósio em Sistemas Computacionais de Alto Desempenho.

5. REFERÊNCIAS

Gvozdetka, N., Globa, L., and Prokopets, V. (2019). Energy-efficient backfill-based scheduling approach for slurm resource manager. In 2019 IEEE 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), pages 1–5.

Petry, A., de Souza, J. R., de Campos Velho, H. F., Pereira, A. G., and Bailey, G. J.(2014). First results of operational ionospheric dynamics prediction for the brazilian space weather program. *Advances in Space Research* , 54(1):22–36.

Prabhu, C. (2008). *Grid and cluster computing*. PHI Learning Pvt. Ltd.

Sloan, J. D. (2004). *High Performance Linux Clusters with OSCAR, Rocks, OpenMosix, and MPI: A Comprehensive Getting-Started Guide*. O'Reilly Media, Inc.

Yoo, A. B., Jette, M. A., and Grondona, M. (2003). Slurm: Simple linux utility for resource management. In Feitelson, D., Rudolph, L., and Schwiegelshohn, U., editors, *Job Scheduling Strategies for Parallel Processing*, pages 44–60, Berlin, Heidelberg. Springer Berlin Heidelberg.

Eckel, Gabriela L., Petry, Adriano, Puntel, Fernando E. (2020). Comparação de Desempenho para Gerenciadores de Recursos OAR e SLURM em Aplicação Científica de Simulação Ionosférica. ESCOLA REGIONAL DE ALTO DESEMPENHO DA REGIÃO SUL (ERAD-RS), 20. , 2020, Santa Maria. Anais da XX Escola Regional de Alto Desempenho da Região Sul. Porto Alegre: Sociedade Brasileira de Computação, apr. 2020 . p. 57-60. ISSN 2595-4164.