



11ª Jornada Científica e Tecnológica do IFSULDEMINAS & 8º Simpósio de Pós-Graduação

ESTUDO E IMPLEMENTAÇÃO DE *CLUSTER* COMPUTACIONAL DE ALTO DESEMPENHO COM ARMAZENAMENTO DISTRIBUÍDO EM REDE.

Guilherme V. P. PEDROZA¹; Augusto M. da S. JUNIOR²

RESUMO

A evolução dos computadores e da internet, além do aumento na quantidade de usuários ao redor do mundo tem como consequência um aumento exponencial do volume de dados produzido a todo momento. Devido a isso o poder de processamento demandado aos sistemas também aumenta proporcionalmente. Tal demanda muitas vezes supera o desempenho de máquinas convencionais, pois sua capacidade de processamento é limitada. Isto leva a necessidade de se utilizar sistemas e arquiteturas diferenciadas, denominados Plataformas de Computação de Alto Desempenho, para se obter desempenho satisfatório nessas tarefas. Por conseguinte, temos o *Cluster*, uma alternativa para satisfazer a demanda por poder de processamento através da divisão de tarefas em múltiplos computadores. O objetivo do trabalho aqui descrito é a implementação de uma solução computacional de alto desempenho voltada para utilização na disciplina de Programação Paralela, como resultado visa-se verificar o benefício que tal implementação teve no aprendizado da disciplina, além do ganho de performance na utilização do *Cluster* em relação aos sistemas convencionais.

Palavras-chave: Programação Paralela; Sistemas Distribuídos; *Beowulf*; *FreeNAS*.

1. INTRODUÇÃO

A demanda por poder de processamento de certos dados e aplicações muitas vezes supera o desempenho de máquinas convencionais, pois sua capacidade de processamento é limitada. Isto leva a necessidade de se utilizar sistemas e arquiteturas diferenciadas, denominados Plataformas de Computação de Alto Desempenho.

Um dos tipos de arquitetura que pode satisfazer tais condições, é a Arquitetura Distribuída, uma de suas definições é, “uma coleção de computadores independentes entre si que se apresenta ao usuário como um sistema único e coerente” (STEEN; TANENBAUM, 2002, p. 3), corroborando com a afirmação também pode-se dizer que é coleção de computadores autônomos interligados através de uma rede de computadores e equipados com *software* que permita o compartilhamento dos recursos do sistema: *hardware*, *software* e dados (CAMACHO, 2009).

¹ Orientado, IFSULDEMINAS - Campus Muzambinho. E-mail: bherewow@gmail.com.

² Orientador, IFSULDEMINAS - Campus Muzambinho. E-mail: augusto.junior@muz.ifsuldeminas.edu.br.

Por conseguinte, temos o *Cluster*, um tipo de implementação de Arquitetura Distribuída, que, consiste em várias máquinas interconectadas em uma rede para a paralelização de processos, a fim de se conseguir um melhor desempenho e eficiência em dada tarefa (STEEN; TANENBAUM, 2002). Existem vários tipos destes, cada qual com suas vantagens, desvantagens e utilidade. O aglomerado *Beowulf* é um desses tipos, e vem sendo utilizado em diversas áreas, como alternativa para supercomputadores (SOUZA, 2001). Ele em suma é uma implementação de baixo custo utilizando-se de máquinas comuns e software *Open Source*(*Linux*), que o torna ideal para ser implementado em um ambiente acadêmico com propósitos educacionais, Yokokura (2005) e Junior (2013) observaram, que vislumbram neles a oportunidade de oferecer aos seus estudantes e pesquisadores os meios necessários para que eles possam realizar estudos e pesquisas que envolvam cálculos complexos e, normalmente, alcançados apenas com máquinas de alto poder de processamento.

Ademais, também auxiliam no aprendizado de várias disciplinas das áreas de conhecimento da Ciência da Computação, entre elas podemos citar : Programação Paralela, Redes, Sistemas Distribuídos, Arquiteturas de Computadores, entre outras (FASSBINDER, 2015). Pois permitem que alunos possam exercer na prática o conteúdo teórico visto em aula, que muitas vezes devido ao alto nível de abstração torna-se difícil de compreender, a aplicação prática de tais conceitos contribui para um melhor entendimento dos mesmos.

2. MATERIAL E MÉTODOS

A proposta da pesquisa se caracteriza como aplicada, de caráter descritiva e explicativa, de abordagem quantitativa a ser objetiva, descreve por meio de linguagem matemática as causas de um fenômeno. Em relação aos procedimentos, esta pesquisa se enquadra simultaneamente como experimental e, um estudo de caso (GERHARDT; SILVEIRA, 2009).

Toda a implementação e desenvolvimento do *Cluster* foi realizada no Laboratório de Redes e Sistemas Distribuídos (LARES) no Instituto Federal do Sul de Minas Campus Muzambinho, materiais e ferramentas utilizados, incluindo, computadores, *Switches*, roteadores, são todos pertencentes ao LARES.

2.1 PROCEDIMENTOS

Foram utilizadas 5 máquinas para formar o *Cluster*, com a adição de uma sexta máquina extra funcionando como uma *NAS*(*Network-Attached Storage*) compartilhando um diretório comum através de protocolo *NFS*(*Network File System*) entre todos os nós do *Cluster*.

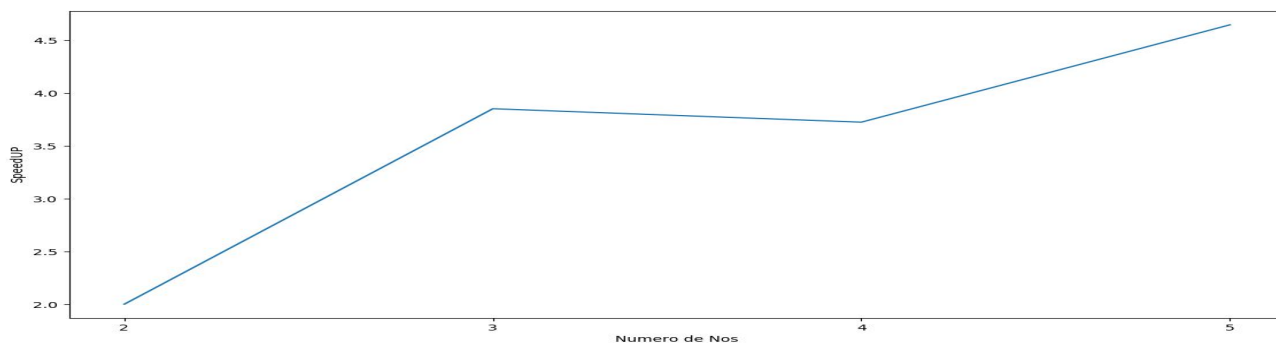
O volume da *NAS* foi montado como um diretório comum em todos os nós, desta forma qualquer alteração feita neste diretório terá efeito em todos os elementos do *Cluster*. Isso permite que apenas uma das máquinas compile o código fonte, e as outras simplesmente executam o código executável gerado, sem a necessidade de recompilar o código fonte ou de especificar o caminho de execução do mesmo, tornando a utilização do *Cluster* mais simples e rápida.

Para esta pesquisa, utilizou-se, nas máquinas, o Sistema Operacional Debian 9, todas conectadas em uma mesma rede, e ainda com um mesmo usuário em comum. A fim de melhorar a performance dos sistemas foram instalados apenas os softwares mais básicos e essenciais para o funcionamento do *Cluster* ou seja, nenhum tipo de interface Gráfica ou pacotes adicionais que pudessem atrapalhar o desempenho no momento da execução de algum algoritmo paralelo.

3. RESULTADOS E DISCUSSÕES

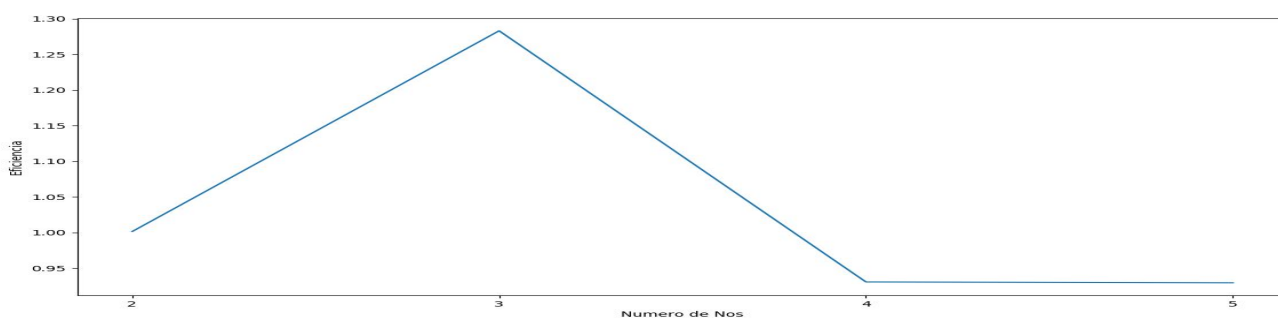
Após implementado o *Cluster* foi utilizado por 25 alunos do curso de Ciência da Computação, os gráficos a seguir demonstram um resultado satisfatório em termos de desempenho e eficiência do *Cluster* nos algoritmos testados. No primeiro gráfico observa-se o ganho de desempenho pelo aumento no número de nós utilizados. Já no segundo nota-se a eficiência da execução conforme o número de nós, ou seja, o quanto o sistema se aproveita dos recursos disponíveis para realizar a tarefa dada.

Gráfico de *SpeedUp* Médio.



Fonte: Autor, 2019.

Gráfico de Eficiência Média.



Fonte: Autor, 2019.

O maior ganho de *SpeedUp* se deu, em sua maioria, com a utilização de dois ou três e nós, acredita-se que este fato ocorre devido a heterogeneidade entre os nós do *Cluster*, pois os dois últimos nós são também os mais lentos, graças a seu hardware antigo e inferior. Em algumas execuções observa-se que o tempo do processo com um único nó é drasticamente maior que o período gasto pelo mesmo ao distribuir o código, e esta ocorrência pode ser justificada pela limitação de memória disponível ao se utilizar apenas um nó, pois neste caso, devido a esta insuficiência, o sistema deverá recorrer a memória virtual (*SWAP*) que é significativamente mais lenta que a memória primária, acarretando assim uma latência de transferência interna de memória, que implica no aumento do tempo de execução dos algoritmos.

Assim ao se utilizar dois ou mais nós, a memória disponível para o sistema passa a ser suficiente, tirando a necessidade de se utilizar o arquivo de paginação, por o sistema não depender mais da memória virtual, o intervalo de execução cai bruscamente, resultando em um surto de *SpeedUp* e *Eficiência*.

4. CONCLUSÕES

O presente trabalho buscou a implementação do *Cluster Beowulf* com *Storage* na rede, que foi realizado, e então foi disponibilizado para utilização dos alunos de Programação Paralela do Instituto Federal do Sul de Minas - Campus Muzambinho. Os resultados demonstram que o sistema foi utilizado por todos os 25 alunos do curso e que na maioria dos casos apresentou melhorias significativas na execução dos algoritmos paralelos desenvolvidos pelos usuários como retratado nos gráficos acima.

REFERÊNCIAS

- CAMACHO, C.; COULOURIS, G.; AVAGYAN, V.; MA, N.; PAPADOPOULOS, J.; BEALER, K.; MADDEN, T. L. **Blast+: architecture and applications**. BMC bioinformatics, BioMed Central, v. 10, n. 1, p. 421, 2009.
- FASSBINDER, M. **Implantação e avaliação de um cluster beowulf para fins educacionais e de pesquisa: Um estudo de caso aplicado ao IFSULDEMINAS – Câmpus Muzambinho**. 2015.
- GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de pesquisa**. [S.l.]: Plageder, 2009.
- JUNIOR, E. E. **Estudo de tecnologias para computação paralela e distribuída: Implementação de um cluster beowulf**. 2013.
- SOUZA, R. de. **Clusters Linux e suas Aplicações**. [S.l.], 2001.
- STEEN, M. V.; TANENBAUM, A. S. **Distributed systems principles and paradigms**. Network, v. 2, p. 28, 2002.
- YOKOKURA, A. **Estudo de viabilidade da implantação de técnicas de cluster ao projeto servidor de estações de trabalho (set)**. Universidade do Pará, Centro de Ciências Exatas e Naturais (Conclusão de Curso) Belém, 2005.