

Simulação da Propagação de Onda 2D em um meio Acústico – Aferição da Performance de um Cluster Beowulf

Lidiane Nazaré Monteiro Penha¹ - lidianepenha@gmail.com

Alessandra Natasha Alcântara Barreiros¹ e Francisco de Assis Silva Neto²

¹ Centro Universitário do Estado do Pará ² Secretaria de Educação do Estado do Pará

Copyright 2012, SBGf - Sociedade Brasileira de Geofísica

Este texto foi preparado para a apresentação no V Simpósio Brasileiro de Geofísica, Salvador, 27 a 29 de novembro de 2012. Seu conteúdo foi revisado pelo Comitê Técnico do V SimBGF, mas não necessariamente representa a opinião da SBGf ou de seus associados. É proibida a reprodução total ou parcial deste material para propósitos comerciais sem prévia autorização da SBGf.

Resumo

Este trabalho objetiva propor um modelo computacional para simular aquisição sísmica 2D em um meio acústico. Tal proposta será baseada na aferição do tempo de resposta de um *Cluster* do tipo *Beowulf* – *Gauss*.

Em virtude das altas demandas computacionais requeridas em levantamento sísmico, a utilização de ambientes de programação paralela e distribuída vem sendo adotada como proposta atraente. Isso se dá pelo fato desses ambientes mostrarem-se como uma solução para a redução dos custos, bem como para o aumento da velocidade com que as informações podem ser obtidas.

Para isto, este trabalho apresenta um referencial teórico introdutório dos conceitos necessários ao entendimento e uso do processamento paralelo, do *cluster Beowulf*, assim como da modelagem sísmica. Ainda, serão apresentados resultados baseados nos tempos de processamento obtidos em testes realizados em *Cluster Beowulf* com 10 nós.

Introdução

Antigamente, o óleo exudava para a superfície. Já no século 19, poços precisavam ser perfurados com algumas dezenas de metros de profundidade. Com o passar do tempo e o excesso de extrações, atualmente está cada vez mais difícil encontrar rochas impregnadas por hidrocarbonetos.

Entretanto, apesar deste cenário de progressiva escassez natural, o Brasil tem vivido um momento privilegiado no que diz respeito à exploração do petróleo, sobretudo em termos de avanços tecnológicos. Para atender a este setor que trabalha com grande volume de dados, é fundamental a utilização de recursos computacionais que atendam demandas com alta eficiência como, os *Clusters* (Buyya, 1999).

Um dos processos largamente utilizados na indústria de petróleo e gás é a modelagem sísmica, na qual se consome grandes recursos computacionais, exigindo, ainda, algoritmos eficientes e processadores de dados de grande performance.

Um dos procedimentos para a modelagem sísmica é a utilização das soluções aproximadas da equação de onda sísmica 2D em meio acústico por meio de métodos como o de Diferenças Finitas (MDF), visando obter modelagem do levantamento sísmico.

Dentro deste contexto, encontrar um modelo computacional que mais rápido atenda ao processamento de uma aquisição sísmica 2D em um meio acústico, vem chamando atenção das empresas petrolíferas.

Referencial Teórico

Para uma maior compreensão do cenário montado para análise do ambiente em processamento paralelo e distribuído, faz-se necessário entender que segundo Tanenbaum (2003), as máquinas que possuem múltiplo fluxo de instruções e de dados também podem dividir-se em: multiprocessadores (fortemente acoplados) e multicomputadores (fracamente acoplados).

Desta forma, um Sistema Distribuído (Coulouris et al. 2007, Ribeiro 2005) baseado em um *Cluster Beowulf* foi utilizado para simular aquisições sísmicas 2D em meio acústico visando encontrar o menor tempo de resposta e agilizar assim, através de um alto poder de processamento, a aferição do ambiente em questão.

A modelagem sísmica, aplicação proposta no estudo de caso deste trabalho, é uma técnica utilizada pelas indústrias do petróleo e gás para auxiliar na compreensão do fenômeno da propagação de ondas sísmicas no subsolo, através das diferentes formações geológicas, com o objetivo de aumentar o nível de sucesso na perfuração em busca de reservas de óleo e de gás em regiões de geologia complexa.

Metodologia

Para o estudo apresentado, foi utilizado como modelo de velocidade uma seção retirada de um modelo geológico sintético em 3D – Figura 01.

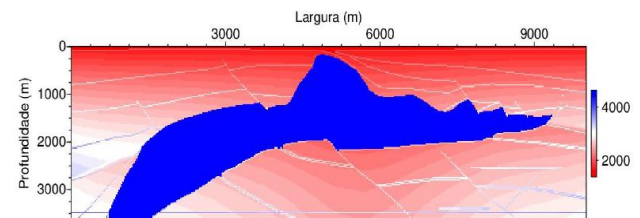


Figura 01 – Seções verticais do modelo sintético no plano xz

Esse modelo apresenta um corpo salino (cor azul) dentro de uma estrutura sedimentar, escolhido por apresentar um desafio para a Geofísica atual, visto que o contraste de velocidade da onda compressional entre o corpo e os sedimentos dificultam o processo de imageamento, e conseqüentemente a interpretação sísmica (Yilmaz, 1987). As regiões abaixo e laterais ao corpo salino são consideradas atrativas para a

exploração sísmica, e são conhecidas como trapas estratigráficas.

Com este modelo foi realizada a simulação de uma aquisição sísmica 1D, em linha posicionada na superfície ($z=0$), e calculada através do processamento paralelo.

O procedimento de cálculo consiste em simular explosões em pontos determinados ao longo da superfície, e que geram frentes de ondas compressionais que se propagam através do modelo, sendo medidas na superfície através de receptores.

Esta modelagem baseia-se na solução da equação escalar da onda acústica simulando uma propagação onde a velocidade compressional é variável e a densidade é constante.

Neste caso, a equação da onda acústica é dada por (Aki e Richards, 1980):

$$\frac{\partial^2 P}{\partial t^2} = c^2 \left(\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial z^2} \right); \quad \text{Equação 1}$$

caso haja uma perturbação no meio, a equação 1 ganhará a seguinte forma:

$$\frac{\partial^2 P}{\partial t^2} = c^2 \left(\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial z^2} \right) + f(x, z, t); \quad \text{Equação 2}$$

em que P é o campo de pressão, c é a velocidade da onda compressional e $f(x, z, t)$ representa uma função fonte do tipo explosiva ou de injeção de volume.

Existem vários métodos numéricos para se obter soluções aproximadas da equação da onda na forma acústica. Sendo que no presente trabalho, a solução da equação de onda acústica é obtida pelo método de diferenças finitas (MDF), onde se substitui as derivadas parciais no espaço e no tempo por suas aproximações em diferenças finitas que utiliza a expansão em série de Taylor (Etgen, 1986) para obter a solução aproximada.

O método MDF (Brito, 2011) é útil para modelagem sísmica por substituir a equação diferencial por um conjunto de equações lineares algébricas que são aproximações da equação diferencial (McDonald & Harbaugh, 1988), principalmente quando se trata de estruturas complexas (formas dômicas, tipo sal halogênico) devido às dificuldades encontradas em obter soluções analíticas, mostrando-se um método satisfatório na validação/comparação dos resultados.

Ainda para o estudo foram levantados pontos importantes, tais como: linguagem de programação, processamento paralelo em ambientes distribuídos, arquiteturas distribuídas, tipos de processamento, bibliotecas padrão, gerenciador de processos e tecnologias de rede, com objetivo de escolher quais destes recursos serveriam e adequariam-se para montar um ambiente paralelo e distribuído que trabalhasse de forma otimizada.

Além disso, pretendia-se, também, que a estrutura fosse ainda útil para o gerenciamento, monitoração, avaliação e comparação dos resultados deste trabalho. Estes resultados serão obtidos em cinco etapas: cálculo do modelo sintético, paralelização dos processos, distribuição, utilização de recursos gráficos e comparação dos sistemas paralelos e distribuídos, a fim de classificar o comportamento computacional do modelo frente ao melhor resultado.

Resultados Numéricos do processamento em sistemas: Paralelos e distribuídos

Os resultados obtidos do processamento realizado no *Cluster Beowulf Gauss* para comparar a diferença entre a variação do número de nós quando submetidos à simulação de uma aquisição sísmica 2D em meio Acústico será apresentado nesta seção.

Assim, tem-se que a primeira etapa que consiste na solução da equação de onda com programas escritos na linguagem de programação Fortran 90.

A segunda etapa consiste em paralelizar os processos (emissão de ondas pela fonte) no *Cluster Beowulf Gauss* utilizando os vários nós.

A terceira etapa refere-se a utilização de recursos gráficos do Excel para plotagem dos resultados além do sistema Seismic Un*x – SU (Forel *et al.* 2005) para gerar a seção sísmica após o processamento paralelo e a propagação da frente de onda no modelo.

Após concluir estas três etapas, será possível realizar a última parte que é aferir os resultados do funcionamento *Cluster Beowulf Gauss* quando submetido à geração da propagação de ondas 2D em meio acústico.

O esquema apresentado na Figura 02 corresponde a primeira, a segunda e a terceira etapas descritas acima, que ilustram, de forma simplificada, o processo de aquisição sísmica realizada no modelo sintético deste trabalho.

Como mostrado na Figura 02, as ondas emitidas por uma(s) fonte(s) propagam-se pelas diversas camadas do subsolo e são medidas através de receptores na superfície.

A informação gravada em cada receptor resultará em um traço sísmico. Dessa forma, um conjunto de traços sísmicos gerados para cada fonte é chamado de seção sísmica que receberá do ponto de vista da computação a denominação: processo.

A partir deste esquema é possível entender como serão obtidos os resultados após o término do processamento. Observa-se que cada fonte acionada na superfície do modelo sintético – Figura 01 – produzirá uma seção sísmica.

As simulações serão distribuídas nos diversos nós de processamento do *Cluster Beowulf Gauss*. Ao término da execução destes processos, as seções serão organizadas uma ao lado da outra, produzindo uma família de seções fonte-comum representado esquematicamente na Figura 02.

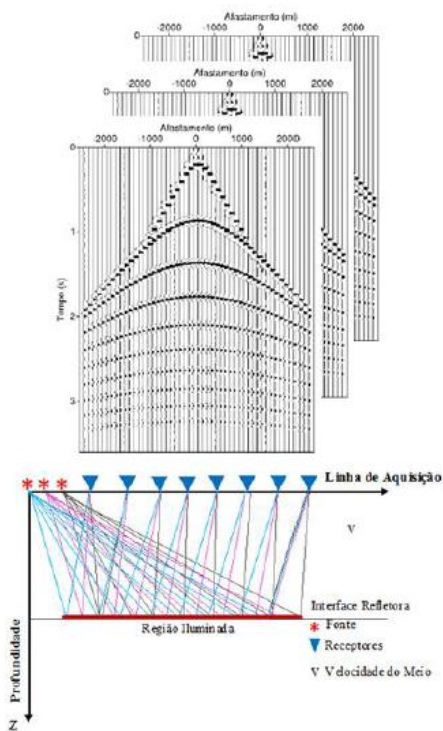


Figura 02 - Esquema mostrando o resultado da aquisição sísmica do modelo sintético considerando várias fontes e 144 receptores.

Apresenta-se na Figura 03 uma seção sísmica obtida a partir do modelo sintético utilizado, sendo possível obter também snapshots da propagação da onda em instantes específicos.

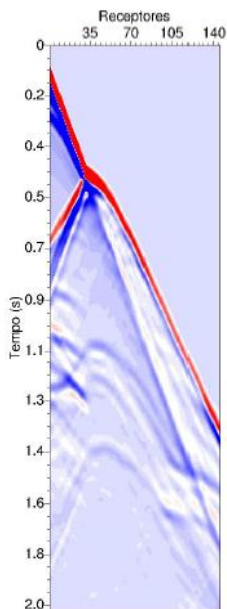


Figura 03 – Seção sísmica do modelo sintético

Neste trabalho será utilizado o tempo total de processamento, como medida de desempenho, para avaliar e comparar os resultados obtidos com a aplicação da análise em processamento paralelo no ambiente distribuído. O tempo total de processamento corresponde ao somatório dos tempos parciais de cada nó, sendo que a forma de medição do tempo realizada em cada nó consiste na determinação da diferença

entre a hora do início e do final do processo, representada, portanto, pelo processamento somado a troca de mensagens somado ao tráfego de rede somado ao tempo de espera em filas somado ao overhead, tempo real transcorrido e não apenas no tempo de utilização do processador.

A seguir, apresentam-se os experimentos realizados em um Cluster do tipo *Beowulf Gauss* para verificar em qual das etapas dos experimentos se obteve o melhor tempo de resposta. Foram realizados 2 experimentos, em média 12 repetições em cada.

Experimento 1: variações no número de nós com número limitado de núcleos

O primeiro experimento foi realizado com um modelo sintético da Figura 01, foram produzidas 288 famílias de tiros comuns, sendo que para cada tiro tem-se 144 receptores espaçados de 10m na direção do eixo X e 10m de profundidade, com afastamento mínimo de 100m e máximo de 3000m. O intervalo de aquisição é de 2ms, sendo obtidas 1501 amostras por traço, e o intervalo de amostragem da modelagem para garantir estabilidade é 0,5 ms.

No experimento realizado em ambiente paralelo, foram gerados 288 famílias de tiros em 1 nó. Sendo que este nó tem 2 processadores e cada um contem 4 núcleos – Figura 04. Dessa forma, esse nó executará sempre 8 processos de forma concorrente, pertencendo a um ou mais processadores que vão resolver uma única tarefa.

Após a execução dos 288 tiros verificou que o tempo total de processamento foi de 1h e 25 mim. Como o valor deste tempo foi um pouco maior que os outros, pode-se concluir que para ambiente de paralelismo puro, não é uma boa solução, já que os processos foram executados no mesmo nó.

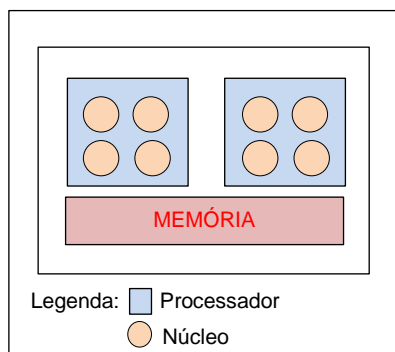


Figura 04 – Estrutura de um nó – Ambiente de Paralelismo

O ambiente distribuído em 2 nós rodando processamento paralelo foi outra opção utilizada para analisar o tempo total de processamento.

Numa segunda situação foram adotados 2 nós e 4 núcleos cada, sendo assim, neste caso o número de núcleos vai ser limitado. Ou seja, de cada um destes 2 nós será utilizado 2 núcleos de um processador e 2 núcleos do outro processador, conforme mostra a Figura 05, totalizando e permanecendo os mesmos 8 núcleos do primeiro teste, porém em novo arranjo computacional.

Dessa forma, indo da esquerda para direita, utilizar-se-á 2 núcleos no primeiro processador, 2 núcleos no segundo processador, assim por diante.

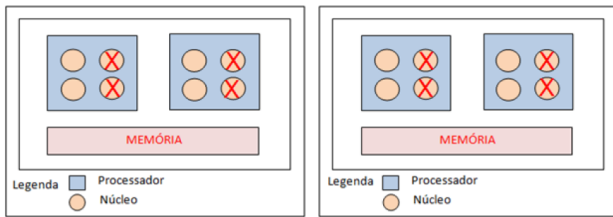


Figura 05 – Estrutura de 2 nós – Ambiente distribuído para processamento paralelo

Após a execução dos 288 tiros verificou-se que o tempo total de processamento foi de 1h e 20 min. Esse valor foi 5 minutos menor que o de 1 único nó, mesmo distribuindo os processos entre os 2 nós ainda não representa o melhor resultado da série de testes, ficando abaixo do resultado encontrado com 1 nó.

Numa terceira situação foram adotados 4 nós e 2 núcleos, sendo que neste caso o número de núcleos será novamente limitado a 8.

Assim, cada um destes 4 nós utilizará somente 2 núcleos de um dos 2 processadores, conforme mostra a Figura 06.

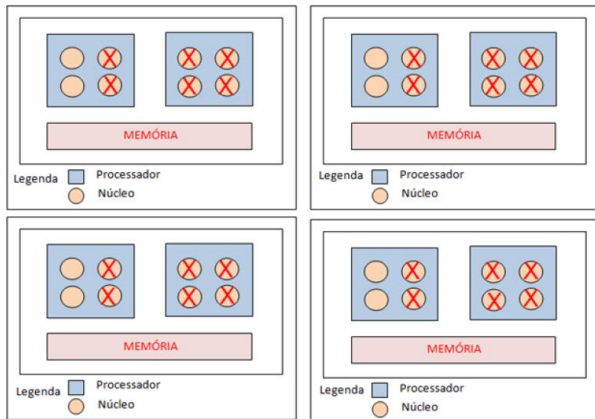


Figura 06 – Estrutura de 4 nós – Ambiente distribuído para processamento paralelo

Após a execução dos 288 tiros verificou-se que o tempo total de processamento foi de 1h e 15 min. Esse resultado mostrou-se melhor ao de 1 único nó e ao de 2 nós, pois nesse teste houve mais distribuições dos processos entre os 4 nós.

Numa quarta disposição, foram adotados 8 nós e 1 processador em cada, sendo que neste caso o número de núcleos também será limitado a 8. Ou seja, de cada um destes 8 nós será utilizado 1 núcleo de um dos 2 processadores, conforme mostra a Figura 07.

Após a execução dos 288 tiros verificou-se que o tempo total de processamento foi de 1h e 10 min. Este foi o melhor resultado encontrado, demonstrado que o comportamento do modelo sintético foi melhor atendido, sobretudo pelo fato de ter distribuído mais os processos entre os 8 nós.

De um modo geral, o tempo total de processamento tende a diminuir com a distribuição dos processos através dos nós do *Cluster Beowulf Gauss*, mesmo mantendo-se fixo o número de processos.

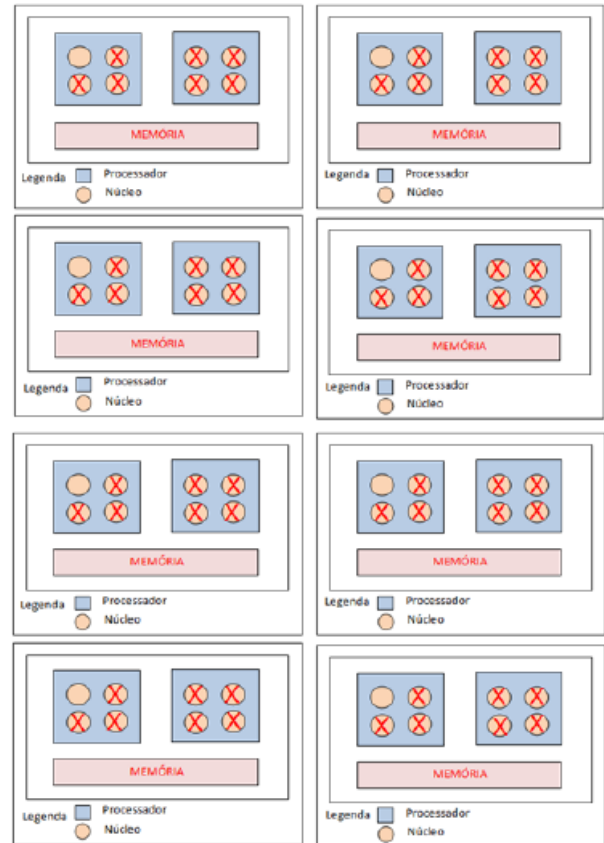


Figura 07 – Estrutura de 8 nós – Ambiente distribuído para processamento paralelo

Um resumo de todos os resultados obtidos quanto ao tempo total de resposta neste primeiro experimento é mostrado na Tabela 1. Pode-se perceber, a pequena variação no tempo de resposta face a alteração no número de nós e no número de processos por nó.

Tabela 1 - Resultados obtidos no experimento 1

Número de Nós	Número de Processos por Nó	Número de Tiros	Tempo Total de Processamento
1	8	288	1h e 25 min
2	4	288	1h e 20 min
4	2	288	1h e 15 min
8	1	288	1h e 10 min

Experimento 2: variações no número de nós com número limitado de núcleos, porém mantendo-se o modelo e reduzindo a discretização

O objetivo deste experimento é validar os resultados obtidos no primeiro a partir da observação de coerência proporcional nos resultados ali aferidos face a mudança da discretização de 10m para 5m, conseqüentemente, dobrando o número de pontos em cada dimensão do modelo sintético agora submetido.

Desta forma, neste segundo experimento, observou-se um comportamento similar ao experimento 1, como, de fato, houve uma aferição de tempo de resposta quatro vezes maior que os resultados realizados no experimento 1.

A tabela 2 apresenta todos os resultados obtidos do tempo total de resposta, verificando-se que o comportamento do modelo sintético foi melhor atendido à medida que aumentava a distribuição dos processos entre os nós do *Cluster Beowulf Gauss*.

Tabela 2 - Resultados obtidos no experimento 2

Número de Nós	Número de Processos por Nó	Número de Tiros	Tempo Total de Processamento
1	8	288	5h e 40 min
2	4	288	5h e 20 min
4	2	288	5h
8	1	288	4h e 40 min

Discussões e Conclusões

Referente ao primeiro experimento, dos quatro testes realizados em oito núcleos, distribuídos conforme a tabela 1, constata-se que o melhor modelo computacional obtido pelo menor tempo de resposta encontrado é conferido ao arranjo computacional inerente a 8 nós, cada um com 1 núcleo.

O aumento do tempo de resposta observado em 1 único nó com 8 núcleos, mostrou-se em sobrecarga de execução devido não haver distribuição dos processos entre os nós do *Cluster Beowulf Gauss*; enquanto que os tempos de resposta obtidos com 2 e 4 nós não foram os esperados para atender o modelo computacional.

Neste sentido, quanto mais distribuídos estiverem os processos entre os nós do *Cluster Beowulf Gauss* menor será o tempo de acesso, visto que o *overhead* de acesso à memória será menor.

Como exposto anteriormente, o segundo experimento serviu como forma de validação dos resultados, a partir do aumento no número de pontos em cada direção do modelo (discretização), de forma similar ao método utilizado na matemática para comprovação de conclusões empíricas através de indução finita.

Neste sentido, esse segundo experimento foi bem sucedido, conforme atestam os resultados da tabela 2, mantendo-se como melhor modelo computacional, através do menor tempo de resposta aferido, o arranjo topológico de 8 nós, 1 núcleo cada quando submetido a mesma carga de trabalho (288 famílias de tiros).

Finalmente, foi traçado no gráfico a seguir, todos os resultados obtidos pelos 2 experimentos, fazendo um contraponto do número de nós versus núcleos com o tempo – Figura 8.

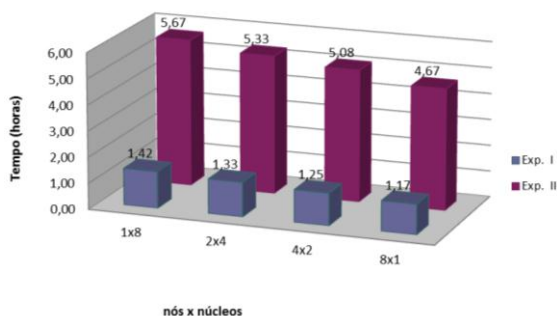


Figura 08 – Comparação dos resultados obtidos nos 2 experimentos

Neste trabalho, comparou-se o modo de execução de cada análise a partir de um estudo aprofundado de processamento paralelo em ambiente distribuído no *Cluster* do tipo *Beowulf Gauss*. Concluída a aferição do tempo de resposta deste *Cluster*, quando submetido ao processamento de aquisição sísmica 2D em meio acústico, pode-se constatar que esta técnica, pode ser vantajosa, por tratar-se de medição realizada sobre um ambiente real.

O presente trabalho mostrou-se fundamental para uma visão clara dos autores acerca dos estudos em modelo computacional baseado em sistemas distribuídos e processamento paralelo, servindo ainda, como forma de constatar a importância fundamental da Tecnologia da Informação como ferramenta de trabalho de diversas áreas, sobretudo a geofísica, no processamento de modelos sísmicos, foco deste trabalho.

Agradecimentos

Um dos autores agradece a contribuição do prof. Dr. Lourenildo Leite, Coordenador do Laboratório de Processamento Sísmicos (PROSIS) da UFPA, pela oportunidade de reabertura de conta no PROSIS e e ao Coordenador da Pós- Graduação em Geofísica Marcos Welby pela oportunidade de abertura de conta no Cluster para realizar o processamento paralelo.

Referências

- Aki, K. & Richards, P. G., 1980. Quantitative Seismology. [S.l.]: W. H. Freeman & Co.
- Brito, F. B. C., 2011. Capítulo 2: Método de Diferenças Finitas. Disponível em: <<http://pt.scribd.com/doc/50378223/Metodo-das-Diferencas-Finitas>> Acesso em: 12 Set. 2011.
- Buyya, R., 1999. High Performance Cluster Computing: Architecture and Systems, vol 1. pag 45, New Jersey: Prentice Hall.
- Coulouris, G., Dollimore, J. & Kindberg, T., 2007. Sistemas Distribuídos: Conceitos e Projeto. 4ª ed. Bookman, Porto Alegre.
- Etgen, J. T., 1986. High-order finite-difference reverse time migration with the 2-way nonreflecting wave equation. SEP-48: 133–146.
- Forel, D., Benz T. & Pennington W., 2005. Seismic Data Processing with Seismic Un*x. [S.l.]: Society of Exploration Geophysicists.
- Jájá, J., 1992. An Introduction to Parallel Algorithms. Redwood City, USA: Addison Wesley Longman Publishing Co.
- Mcdonald, M. G. & harbaugh, A. W., 1988. A modular three-dimensional finite-difference ground-water flow model. Technical report, U.S. Geol. Survey, Reston, VA.
- Ribeiro, U., 2005. Sistemas Distribuídos: Desenvolvendo Aplicações de Alta Performance no Linux. Editora Axcel Books do Brasil.
- Tanenbaum, A. S., 2003. Sistemas Operacionais Modernos. 2ª ed., Prentice Hall, São Paulo.
- Yilmaz, O., 1987. Seismic data processing. [S.l.]: Society of Exploration Geophysicists.