

# Cluster: Computação de Alto Desempenho

Hilário Viana Bacellar  
Instituto de Computação, Universidade  
Estadual de Campinas  
Av. Albert Einstein 1251, Cidade  
Universitária, CEP 13083-970  
Campinas, SP, Brasil  
hvbacellar@gmail.com

## RESUMO

Este artigo apresenta um estudo sobre a computação de alto desempenho de sistemas fracamente acoplados – clusters. Apresenta a arquitetura comum utilizada na construção de clusters de computadores, também apresenta soluções de cluster de software livre, sendo um paralelo e outro que funciona por migração de tarefas, assim como as bibliotecas de comunicação paralela, por fim apresenta as vantagens e desvantagens da computação de alto desempenho cluster.

## Palavras-chave

Cluster; Beowulf; OpenMosix; MPI; PVM.

## 1. INTRODUÇÃO

Atualmente pode-se observar um grande numero de aplicações que exigem cada vez mais uma grande quantidade de processamento de dados, aplicações de mapeamento genético, computação gráfica, previsões metrológicas e até mesmo programas que exigem um grande numero de variáveis de entrada.

Para solucionar essa demanda por poder computacional foram criados os supercomputadores. Os supercomputadores que são sistemas fortemente acoplados possuem um custo elevado e, portanto não é usual a sua implementação em termos de custo. Uma alternativa viável para essa demanda é a utilização de cluster, também conhecido como sistema fracamente acoplado.

Os clusters de computadores processam as tarefas paralelas de forma transparente, ou seja, aparentando ser um único sistema para o usuário.

Para a implementação de um cluster, deve-se levar em consideração a aplicação que será executada, haja vista que, em problemas de granulosidade fina, o cluster não é tão eficiente quanto a um supercomputador, entretanto, para os problemas de granulosidade grossa, o cluster se mostra tão eficiente quanto um supercomputador.

Conforme explica Kalinka Regina, ao se tratar de problemas de granulosidade fina a aplicação de cluster não se mostra tão eficiente, devido à necessidade de troca de informações dos nós com o nó mestre. Essa atividade causa alto tráfego na rede, e isso interfere diretamente no tempo de execução das tarefas, logo o cluster perde sua finalidade que é ganhar tempo com o paralelismo das tarefas [2].

Para o problema de granulosidade grossa, que exige um número considerável de processamento, a solução via cluster torna-se uma alternativa viável, uma vez que, cada nó recebe uma grande

quantidade de carga que passa a ser processada de maneira independente, limitando a troca de informação entre os nós dos clusters.

O objetivo geral desse artigo é fazer um levantamento sobre os tipos de clusters de alto desempenho, e listar as vantagens e desvantagens dessa arquitetura.

Para apresentar as vantagens e desvantagens dos clusters o artigo está organizado da seguinte forma: capítulo 2, fundamentação teórica, neste capítulo é abordado os principais componentes de um cluster de alto desempenho. No capítulo 3 é apresentada a arquitetura de clusters de computadores. No capítulo 4 são apresentados os tipos mais comuns de clusters de alto desempenho. No capítulo 5 são escritos as considerações finais sobre os cluster de computadores abordado nesse artigo.

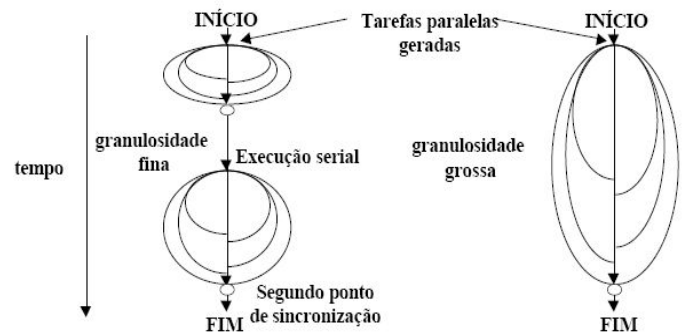


Figura 1 - Granulosidade

## 2. FUNDAMENTAÇÃO TEORICA

Na fundamentação teórica são apresentados conceitos técnicos e teóricos para o entendimento sobre os clusters de computadores.

### 2.1 Sistemas Operacionais

A principal função de um sistema operacional é controlar o funcionamento de um computador, gerenciar a utilização e o compartilhamento dos seus diversos recursos como processadores, memórias e dispositivos de entrada e saída [1].

Em clusters o sistema operacional é o principal sistema, e através dele o sistema deve garantir o funcionamento primordial das seguintes tarefas:

- Controle de recursos: Nessa tarefa o sistema operacional realiza um controle de entrada e saída de cada nó do cluster envolvido na realização da tarefa, e assim, garantir a consistência dos dados executados.
- Monitoração: Tarefa responsável por verificar a disponibilidade de cada nó do cluster, assim como a quantidade de carga que cada nó poderá receber.
- Contabilidade: Essa tarefa funciona como um medidor de desempenho do cluster seja ela, para calcular custos da operação ou até mesmo medir desempenho.
- Consulta. Nesta tarefa o sistema operacional armazena em fila de prioridades todos os processos que serão executados, levando em consideração as tarefas executadas por diferentes níveis de usuários.
- Planificação: Funciona como um planejador, e é através dele que o sistema operacional organiza da melhor maneira quais tarefas serão distribuídas entre os nós do cluster.

## 2.2 Hardware

Atualmente existem muitas implementações de clusters no mercado, isso ocorre devido à grande variedade de opções de arquitetura que um cluster de alto desempenho pode proporcionar.

Na parte de hardware clusters pode ser dividido em duas partes principais, execução e transmissão de dados.

### 2.2.1 Execução

Execução ou processamento de dados é realizado por computadores usuais, usadas para operações domésticas. Como podemos destacar o cluster é um aglomerado de computadores que funciona de forma distribuída. Os recursos que influenciam no desempenho desses computadores são:

- Processador: Responsável pelo processamento de dados, atualmente são desenvolvidos processadores de múltiplos núcleos, tornando-o, a tarefa mais paralelizável possível. O destaque para os processadores são suas memórias caches cada vez maiores.
- Memória RAM: Também denominada como memória principal, são responsáveis por armazenar os dados dos programas que estão em execução, são memórias voláteis e de rápido acesso se comparada ao disco.
- Placa Mãe: Principal item do computador, nela contem todos os chips e ligações entre os componentes de um computador. O barramento de uma placa mãe é a principal via de comunicação entre memória, disco e processador.
- Disco rígido: Armazena os dados dos programas de maneira permanente, é uma memória não volátil, o acesso é lento, porém é a memória que possui maior capacidade de armazenamento.

- Memória Virtual (Swap): Memória utilizada para gerenciar o uso de programas que exijam mais memória do que a RAM possa fornecer. Através da memória virtual, é alocada uma área em disco e assim evitar que programas travem por falta de recursos.

### 2.2.2 Transmissão de dados

Em clusters, a transmissão de dados é o principal gargalo, nas operações. Isso ocorre devido à latência que há nas informações que são trocadas entre os nós.

Hoje o principal meio de comunicação de dados é a rede, sendo composta por meios físicos e mecanismos de controle para o transporte, além de uma política de sincronização de dados.

Na seqüência são apresentadas algumas das principais tecnologias de redes existentes usadas em clusters:

- Ethernet: Existem dois tipos de Eternet, a Fast Ethernet e a Gigabit Ethernet, a mais utilizada em clusters ainda é a Fast que tem velocidade entre 10 Mbps e 100 Mbps, porém com uma rede Gigabit a velocidade pode chegar a casa dos 1 Gbps
- Myrinet: rede de alta velocidade, sua velocidade pode chegar à faixa de 2,4 Gbps, seu funcionamento se deve através de interruptores no qual são traçadas as rotas de transmissão de dados [7].
- cLan: rede apropriada para os clusters, seu funcionamento assim como o Myrinet também se deve através de interruptores, alcança a faixa de 2,5 Gbps, entretanto, com a falta de especificações e documentações apropriadas fazem com que esse tipo de rede não seja usadas freqüentemente.

## 2.3 Bibliotecas de comunicação

As bibliotecas de comunicação paralela são responsáveis pela comunicação entre os nós do cluster. Cada tipo de biblioteca de comunicação tem suas particularidades, ou seja, elas implementam de maneiras diferentes as soluções para os problemas de comunicação paralela.

Atualmente existem duas bibliotecas que se destacam, PVM (Parallel Virtual Machine) e o MPI (Message Passing Interface).

### 2.3.1 PVM

O PVM é uma biblioteca de comunicação que emula computação concorrente heterogênea de propósitos gerais em computadores interconectados, no qual pode se trabalhar com diversas arquiteturas. A idéia do PVM é montar uma máquina virtual de n processadores e usá-los para enviar tarefas e receber os resultados, de maneira cooperativa. Tudo isso é realizado de modo simplificado, utilizando apenas rotinas básicas, enviando e recebendo mensagens.

### 2.3.2 MPI

O surgimento do MPI teve como objetivo padronizar a troca de mensagem em ambientes paralelos de memória distribuída. Além da padronização, o MPI também procura otimizar a comunicação e aumentar o desempenho de aplicações paralelas ou distribuídas.

O MPI surgiu da necessidade de se resolver alguns problemas relacionados à portabilidade existentes entre as diferentes plataformas e características peculiares de algumas arquiteturas paralelas [2].

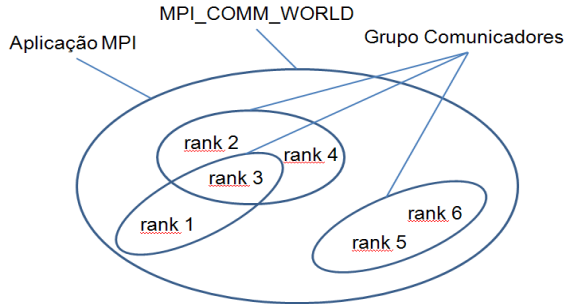


Figura 2 – Estrutura MPI

A eficiência e a generalidade do MPI são garantidas por meio da disponibilidade de diversas implementações para uma mesma funcionalidade. Por exemplo, para o envio de mensagens há funções que implementam comunicação ponto a ponto e coletiva.

Uma das grandes vantagens, do ponto de vista da engenharia de programas, é que MPI suporta programação modular. Por meio desse conceito, o comunicador é capaz de identificar um grupo de processos, no qual uma determinada operação deve ser efetuada.

## 3. ARQUITURA CLUSTER

Cluster é um sistema distribuído de computadores independentes e interligados, cujo o objetivo é suprir a necessidade de um grande poder computacional com um conjunto de computadores de forma transparente ao usuário.

Conforme apresentado no capítulo 2, o cluster possui três condições primordiais pra o seu devido funcionamento, Sistema Operacional, Hardware e biblioteca de comunicação.

Na figura 3 é descrito um ambiente cluster com um nó controlador e cinco nós escravos, todos eles conectados a um hub.

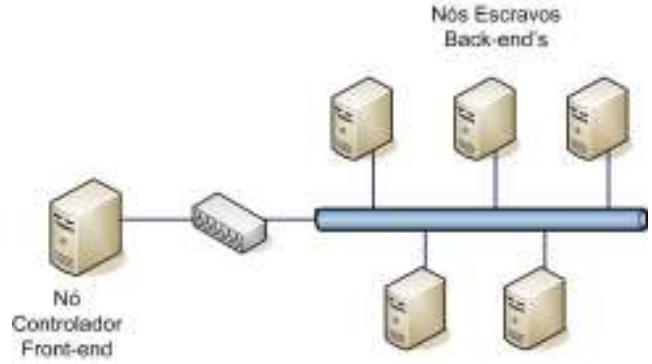


Figura 3 - Arquitetura de cluster com um nó controlador

Sistemas de computação em cluster tornaram-se populares quando a razão preço/desempenho de computadores pessoais e estações de trabalho melhorou [3]. Outra razão para o surgimento e rápida aceitação de sistemas em cluster foi à maior necessidade de tolerância a falhas e a alta disponibilidade [1].

“Os clusters fornecem desempenho e tolerância a falhas, não encontrados em qualquer sistema com multiprocessamento simétrico”[5].

### 3.1 Desempenho

Aumentar o desempenho é um dos principais objetivos de clusters de computadores, entende-se desempenho como carga ou tempo de execução de tarefas.

Os clusters exploram explicitamente o paralelismo como principal ferramenta, e para comprovar ganhos existem métricas que são utilizadas a fim de estruturar uma melhor solução para os problemas paralelizados.

O Speedup ou Aceleração ( $S_p$ ) é uma métrica usada para medir o desempenho de uma aplicação paralela. É definida pela razão do tempo gasto para a execução seqüencial ( $T_s$ ) e o tempo gasto da execução paralela em  $p$  processadores idênticos ( $T_p$ ), ou seja,  $S_p = T_s/T_p$  [2].

A Eficiência é uma medida que mostra quanto da potência computacional disponível que foi utilizada no processamento. É apresentada como a razão entre a Aceleração e o número de processadores, logo  $E_f = S_p/p$ .

Tabela 1 – Aceleração e eficiência

Caso	Aceleração ( $S_p$ )	Eficiência ( $E_f$ )
Ideal	$= p$	$= 1$
Real	$< p$	$< 1$
Excepcional	$> p$	$> 1$

### 3.2 Vantagens

A utilização de clusters de computadores tem inúmeras vantagens, abaixo segue as principais levantadas:

- **Expansibilidade:** A utilização de clusters de computadores deixa o sistema computacional facilmente expansível, uma vez que, para aumentar o poder de processamento, basta apenas incluir um novo no ao cluster.
- **Baixo custo:** Como os clusters de computadores podem ser configurados e utilizados por computadores convencionais, o custo não fica limitado a apenas um único fornecedor, dessa forma clusters de computadores são bem mais econômicos que sistemas específicos.
- **Alta disponibilidade:** Um nó que está desativado não prejudica o sistema como um todo, levando em consideração que também, para a manutenção não é necessário tirar cluster inteiro de funcionamento, apenas o nó com problemas
- **Tolerância a falhas:** Como os clusters de computadores funcionam de maneira paralela e distribuída, caso uma nó venha a falhar, é possível que os dados das operações não se percam totalmente, pois há a possibilidade de redirecionamento da carga de processamento para outro nó que esteja ativo.
- **Balanceamento de carga:** Cluster de computadores também podem ser formados de forma heterogêneas ( com máquinas de configurações diferentes), sendo assim é possível realizar o balanceamento de carga para as máquinas com maior e menor processamento para executar tarefas distintas.

### 3.3 Desvantagens

Como toda tecnologia, clusters de computadores também possuem desvantagens e cabe ao pesquisador ou projetista fazer o levantamento de prós e contras na hora de implementar um sistema desse tipo, abaixo segue as principais desvantagens levantadas:

- **Manutenção de equipamento:** Por o cluster ser facilmente expansível, o sistema computacional pode se tornar muito grande, e a manutenção do sistema pode se tornar uma tarefa imensamente grande pois cada máquina em um clusters devem ter todos os seus componentes em perfeito estado de funcionamento.
- **Monitoração dos nós:** Monitorar as informações trocadas em cada nó pode ser um problema dependendo como foi configurado o cluster, levando em consideração a expansibilidade do cluster.
- **Gargalos de troca de informações:** Como a comunicação de clusters de computadores ocorrerem por uma tecnologia de rede, a troca de informação se transforma no principal gargalo, uma vez que a transmissão de rede é bem lenta se comparada à troca de informação com um barramento de um sistema de memória compartilhada, entretanto, é possível realizar ajustes de granulosidade para diminuir esse problema.

## 4. TIPOS DE CLUSTERS

O cluster estudado nesse artigo é o de alto poder de processamento, que tem como foco aumentar o desempenho das aplicações, particularmente às de grandes tarefas computacionais. Uma grande tarefa computacional pode ser dividida em pequenas tarefas que são distribuídas entre as estações, como se fosse um supercomputador paralelo.

O cluster de processamento paralelo é muito usado para computação científica ou análises financeiras, tarefas típicas para exigência de alto poder de processamento.

### 4.1 Beowulf

O cluster Beowulf nasceu em 1994, idealizado pelos pesquisadores da NASA Thomas Sterling e Donald J. Becker, e é voltado à computação paralela, com a finalidade de processar as informações espaciais que a entidade recolhia. O que motivou a criação do cluster Beowulf foi a necessidade de processamento dessas informações espaciais na ordem de gigaflops , e um supercomputador com esse desempenho custava em torno de um milhão de dólares.

O cluster Beowulf, diferente de outros tipos empresariais e mais específicos, não exige uma arquitetura específica tão pouco máquinas homogêneas. Além disso, deve satisfazer as seguintes premissas para ser considerado um cluster dessa classe:

- Conexão entre os nós, que pode ser feita por meio de ethernet.
- Deve haver um ou mais nós mestres (front-end) para realizar o controle dos nós escravos (back-end).
- O sistema operacional deve ser baseado em código aberto, sendo que o mesmo deve conter todas as ferramentas necessárias para a configuração do cluster.

É necessário que haja um nó mestre (servidor) que realiza toda a distribuição das tarefas e o monitoramento do desempenho do cluster. Este front-end é responsável pelo monitoramento das falhas que possivelmente podem ocorrer e o direcionamento da carga de processamento, caso haja alguma indisponibilidade.

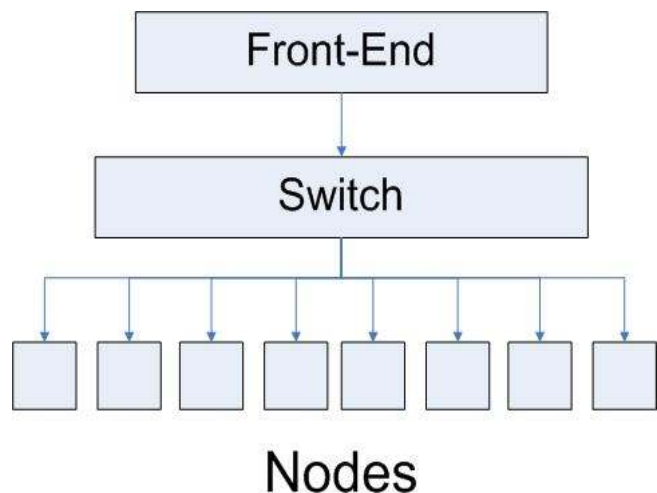


Figura 4 – Esquema de um cluster Beowulf

## 4.2 OpenMosix

Os clusters OpenMosix trabalham com distribuição de processos, que ao detectar o alto volume de processamento, migram as instâncias entre as máquinas do cluster, sendo processadas simultaneamente, sem a necessidade de adequação do código. A principal diferença é essa, ao invés de quebrar os processos como em clusters Beowulf, o Mosix realiza esta migração.

“A migração de processos permite que um usuário inicie uma aplicação em qualquer nó (denominado nó nativo), e se mover transparentemente para outros nós a fim de, por exemplo, fazer uso eficiente de recursos”, [3].

Caso possua três micros homogêneos, ao abrir uma instância de programa, após o OpenMosix atuar, cada um ficaria com uma instância e a tarefa total seria concluída em pouco mais de um terço do tempo original.

Mas o que torna ele pouco flexível é quando o desejo é executar uma única tarefa ou instância do programa, pois para agilizar esta tarefa ele move a mesma para um nó mais rápido e não a divide. Após detectar a ociosidade de um dos nós do cluster, por meio de troca de informações de carga a instância do processo todo é migrada para a máquina do cluster.

Apesar de o sistema possuir uma das maiores bases de usuários dentre todos os gerenciadores, os responsáveis pelo projeto anunciaram que descontinuaram suas atividades e o projeto em março de 2008 [4].

## 5. TRABALHOS CORRELATOS

Nesta seção são apresentados trabalhos acadêmicos com foco em pode de processamentos desenvolvidos com cluster de computadores.

### 5.1 Multipingüim

No trabalho realizado por Marcos Pitanga [5], o cluster Multipingüim é usado para a computação paralela constituído por um computador mestre e sete computadores escravos interligados por uma rede local de alto desempenho. O computador mestre é responsável por distribuir as tarefas entre os computadores escravos que realizam o processamento. O projeto é uma implementação de cluster do tipo Beowulf, cujo objetivo é criar um supercomputador acessível a laboratórios de pesquisa e instituições de ensino superior.

### 5.2 Microwulf

O professor americano Joel Adams e o estudante Tim Brom da Faculdade Calvin em Michigan (Estados Unidos) desenvolveram um projeto de pesquisa chamado Microwulf [6]. O objetivo do projeto era atingir a menor relação custo/benefício de cluster de alta performance portátil, visando quebrar a barreira dos 100 dólares por gigaflop. Ao final do projeto com um custo de 2570 dólares o supermicro computador alcançou a marca de 26,25 gigaflops.

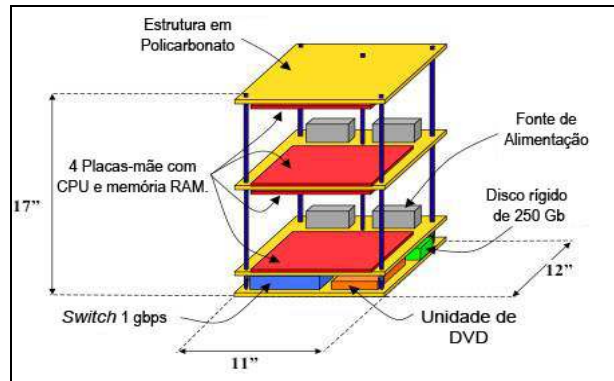


Figura 5 – Arquitetura Microwulf

## 6. CONCLUSÃO

Os clusters de computadores possuem inúmeras vantagens, porém como todo sistema computacional, também possui desvantagens, cabe ao projetista analisar as opções e escolher a melhor tecnologia para resolver sua tarefa da melhor forma possível.

Com o baixo custo de implementação clusters de computadores atualmente são usados com bastante frequência nos mundos acadêmico e empresarial devido a sua grande aplicabilidade em diversas áreas científicas e tecnológicas.

Também é importante destacar que os cluster de computadores não são bons para resolver problemas que exijam constante troca de informações, pois o tempo, limita-se pela tecnologia de rede, entretanto, o programador pode aumentar a carga e assim diminuir a troca de informação entre os nós, diminuindo assim necessidade de troca de informações reduzindo o tempo de espera.

A disponibilidade dos serviços e tolerância a falhas e escalabilidade também são vantagens presentes em cluster de computadores, uma vez que, sistemas em cluster são formados por micros subsistemas independentes.

## 7. REFERENCIAS

- [1] MACHADO, Francis Berenger. MAIA, Luiz Paulo. Arquitetura de Sistemas Operacionais. 4ª ed. Editora LTC, São Paulo, 2007.
- [2] JAQUIE, Kalinka Regina Lucas. Extensão da Ferramenta de Apoio à Programação Paralela (F.A.P.P.) para ambientes paralelos virtuais. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-08022001-095456/>>, visitado em: 23 de mai.2010.
- [3] TANENBAUN, Andrew S. Sistemas Distribuídos: Princípios e Paradigmas. 2ª ed. Editora Pearson Prentice Hall, São Paulo, 2007
- [4] Site Oficial OPENMOSIX.ORG. Disponível em: <<http://openmosix.sourceforge.net/>>, visitado em: 25 de mai.2010.
- [5] PITANGA, Marcos. Construindo supercomputadores com Linux. 2ª ed. Editora Brasport. Rio de Janeiro, 2004.

[6] ADAMS, Joel. BROM Tim. Layton Jeff. Microwulf: Breaking the \$100/GFLOP Barrier. Disponível em: <<http://www.clustermonkey.net/content/view/211/1/>>, visitado em: 30 de mai.2010.

[7] Site Myrinet. Disponível em: <<http://www.myri.com/myrinet/overview/>>, visitado em: 29 de mai.2010.