# Interface para Gerenciamento e Uso de Clusters para Processamento Paralelo

Elaine Patricia Quaresma Xavier[1], Gonzalo Travieso[2]

[1] Instituto de Física de São Carlos, Universidade de São Paulo
Av. Trabalhador Sãocarlense 400, São Carlos, SP.
{exavier@if.sc.usp.br}
[2] Instituto de Física de São Carlos, Universidade de São Paulo
Av. Trabalhador Sãocarlense 400, São Carlos, SP.
{g.travieso@ieee.org}

*Sumario*—
**Este trabalho descreve um sistema simples de gerenciamento de clusters que apresenta uma interface de usuário para as tarefas mais comuns de uso e gerenciamento de um cluster utilizado como máquina paralela. O sistema é baseado em páginas HTML e scripts CGI. O uso de HTML e CGI se demonstrou apropriado para o desenvolvimento desse tipo de sistemas.**

*Keywords*— **processamento paralelo, clusters, sistemas de gerenciamento de clusters.**

*Abstract*—
**This work describes a simple cluster management system that operates as a user interface for some user and manager tasks oft performed on a cluster that is used as a parallel machine. The system is implemented with HTML pages and CGI scripts. The use of HTML and CGI has been found adequate for these type of systems.**

*Keywords*— **parallel processing, clusters, cluster management systems.**

## I. Introdução

Considerando o crescimento exponencial na velocidade de execução, verificado nos últimos anos, dos processadores utilizados em computadores pessoais [CUL 98], associado ao significativo crescimento na largura de banda de redes de interconexão comerciais nesse mesmo período [FOS 95], o uso de clusters para execução de tarefas de processamento paralelo tem se tornado cada vez mais freqüente [9]. Clusters aparecem em várias formas: clusters de alto desempenho, cluters de alta disponibilidade, clusters dedicados, clusters não–dedicados, entre outros. É possível "montar" clusters com uma grande variedade de recursos, como supercomputadores, sistemas de armazenamento e pesquisa de dados, e classes especiais de dispositivos geograficamente distribuídos e usá–los como um recurso único.

Também é possível a utilização de Grid Computing [BAK 99], que tem por objetivo o uso cooperativo de recursos distribuidos geograficamente formando um único computador poderoso.

Neste trabalho enfocamos o uso de clusters fechados como sistemas de processamento paralelo, tendo em vista o destino principal do cluster adquirido por nosso grupo de pesquisa.

Uma das principais desvantagens da utilização de clusters para processamento paralelo, tanto para usuários como para programadores, provém do fato de que eles se constituem geralmente de diversos nós que aparecem como recursos isolados, que portanto devem ser tratados individualmente, complicando o processo de desenvolvimento e controle da execução dos programas. Isto é uma decorrência do fato de que os diversos elementos de hardware e software do sistema foram projetados com vista ao uso como sistemas individuais, interligados à Internet.

Uma solução para este problema é o desenvolvimento de um sistema operacional que gerencie as diversas máquinas envolvidas e apresente ao usuário uma visão única do sistema. Esta solução, no entanto, é muito custosa em termos de esforço de desenvolvimento e apresenta a desvantagem adicional de tornar o comportamento do sistema paralelo durante a execução de um programa menos previsível (em termos de desempenho), devido às sobrecargas introduzidas para possibilitar o gerenciamento dos recursos por parte do sistema operacional distribuído.

Um outra solução é o desenvolvimento de um conjunto de ferramentas que sirvam de apoio aos programadores e usuários em suas tarefas. Estas ferramentas poderiam ser um conjunto de programas (ou *scripts*) que se responsabilizam pelo abstração dos detalhes de interface com o sistema operacional e com o sistema de passagem de mensagens, e devem apresentar uma interface única à partir da qual o sistema paralelo possa ser operado para a execução de tarefas e administração, sem necessidade de recurso a *login* explícito em máquinas remotas, anotação de números de identififcação de processos, etc.

Este trabalho descreve uma solução simples usando a segunda forma indicada, ou seja, um conjunto de programas para ajudar os usuários na execução de suas

tarefas em um cluster homogêneo. Com este conjunto de ferramentas o usuário pode submeter seus processos, manter–se informado do andamento dos mesmos ou até finalizar alguma tarefa, caso seja necessário, entre outras opções.

## II. Objetivo

O trabalho aqui descrito se insere no projeto de desenvolvimento de um sistema de processamento paralelo baseado em um *cluster* de computadores pessoais. O objetivo é desenvolver programas de apoio aos usuários de forma a permitir que o *cluster* apareça como um recurso único para várias tarefas de desenvolvimento e execução de programas e gerenciamento do sistema.

O cluster é formado por 16 computadores pessoais, sendo cada um um K6 450 Mhz, com 128 Megabytes de memória RAM, 4 Gigabytes de capacidade de armazenamento em disco, com sistema operacional completamente independente. O sistema operacional utilizado é o Linux Slackware, versão 7.0, e implementa a pilha de comunicações TCP/IP, sobre a qual as rotinas de comunicação do padrão MPI são construídas. Quando um programa paralelo deve ser executado, uma cópia do programa é disparada em cada um dos nós, a partir de um nó inicial, através da execução de comandos remotos (via rsh). Do ponto de vista do sistema operacional, o programa em execução em cada nó é isolado dos outros. A ligação entre os processos é feita puramente pela comunicação entre eles, utilizando os protocolos de rede, e portanto transparente ao sistema.

Associados a esta estruturação estão vários problemas que limitam a utilidade do sistema para as tarefas de execução de programas e gerenciamento:

- Quando um programa vai executar, devem ser fornecidos, além dos parâmetros normais do programa, também parâmetros para o sistema paralelo (indicação das máquinas nas quais o programa deverá ser executado). Normalmente o usuário não possui as informações necessárias (por exemplo, máquinas menos ocupadas) para decidir esses parâmetros de forma a melhorar a eficiência de utilização do sistema paralelo.

- Durante a execução de um programa, não é trivial verificar seu estado, visto que ele consiste em um conjunto de processos distintos em diversas máquinas.

- Durante a operação de um sistema paralelo, é útil em certas situações reservar alguns nós para um usuário ou grupo de usuários específicos. Não existe nenhum mecanismo para garantir essa exclusividade no sistema original.

- Pode ser em certas situações útil estabelecer limites de horário para a execução de certos programas. Também para isso não existe previsão no sistema original.

O sistema desenvolvido pretende prover facilidades:

- para o usuário, facilitando os processos de inicialização do sistema, execução de programas, controle de execução, notificação de problemas, notificação de término, entre outras;

- para o gerente do sistema, o controle do *cluster* em termos de sistema paralelo, como por exemplo alocação de nós a certos úsuários por certos intervalos de tempo.

## III. O Sistema

Para o desenvolvimento da interface com o usuário foi escolhido o uso de HTML com CGI. Isto porque esta ferramenta possiblita maior facilidade de operação remota do sistema. Para o desenvolvimento dos scripts utilizamos Perl [HAR 96], pois Perl é a linguagem de desenvolvimento ideal para trabalhar em servidores Web, por várias razões. Muitos projetos de programação em servidores Web são em alto–nível, o que significa que eles tendem a não envolver manipulação de bits, chamadas diretas ao sistema operacional, ou interação com o hardware do servidor. Pelo contrário, eles focalizam leituras de arquivos, reformulação de saídas, e escritas no browser. As tarefas de alto nível são as melhores executadas pelo Perl. Além desta vantagem, também é uma linguagem flexível e eficiente.

Este trabalho está dividido, inicialmente, em três páginas Web. A primeira, de visualização pública, que possibilita aos usuários verificar os usuários conectados e as máquinas disponíveis por usuário. Além de permitir também ao usuário entrar no sistema. É neste ponto que os usuários são identificados como usuários ou administrador do sistema. Caso seja um usuário ele entrará em uma página para usuários.

A página dos usuários possui opções para:

- submeter processos
- verificar recursos, como espaço em disco e memória em uso
- finalizar processos
- verificar andamento de uma tarefa
- verificar status das máquinas
- gerenciar diretório do usuário
- verificar máquinas disponíveis para execução

A opção submeter processos permite ao usuário submeter os processos instantameamente ou programar, dentro da sua permissão de horário, o horário em que ele deseja que o job comece a trabalhar. Também é possível escolher em quais nós o usuário deseja que os processos sejam executados (dentre os nós habilitados pelo administrador).

Caso o usuário queira saber a ocupação dos discos, deve utilizar o opção "verificar discos", onde o sistema retornará uma página com a demonstração gráfica da utilização de disco de cada nó. Se deseja saber a ocupação de memória, deve usar a opção "verificar memória", que demonstrará graficamente a utilização da memória de cada nó.

Para finalizar processos já inicializados, o usuário utiliza a opção "finalizar processos", onde aparecerá uma lista dos processos inicializados pelo usuário, lista esta utilizada para escolher qual(is) processo(s) deseja finalizar.

A opção "verificar andamento de uma tarefa", exibe para o usuário os problemas, caso estes existam, na execução do sistema e os processos que já finalizaram a execução.

Os processos em execução podem ser verificados utilizando a opção "status das máquinas", que mostra todos os processos em execução submetidos pelo usuário.

Também é possível verificar se os processos a serem inicializados se encontram no diretório do usuário. Para tanto, basta utilizar a opção conteúdo do diretório, que mostrará todos os arquivos armazenados no diretório.

Para saber quais os nós disponíveis para ele, o usuário deve utilizar a opção "máquinas disponíveis".

Se o usuário em questão é o gerente do sistema, será mostrada uma página específica para o gerenciamento do sistema.

As opções desta página são as seguintes:

- incluir usuários;
- remover usuários;
- atribuir permissão de uso das máquinas;
- finalizar processos;
- reboot.

Estas opções serão descritas abaixo.

A opção "incluir usuário", permite ao gerente incluir usuários tanto no sistema operacional como no sistema paralelo.

A opção "remover usuário", ao contrário da opção anteriormente descrita, exclui usuários, tanto do sistema operacional como do sistema paralelo.

Como já descrito acima, uma das particularidades deste sistema é restringir o uso de nós para cada usuário. Cabe ao gerente do sistema esta atribuição. E para tanto, ele deve utilizar a opção "atribuir permissão de uso de máquinas". Ele pode atribuir, bem como restringir a utilização dos mesmos.

Assim como o usuário, o gerente do sistema tem permissão para finalizar um processo, caso seja necessário. A vantagem é que o gerente pode finalizar qualquer processo de qualquer usuário. Para executar esta tarefa, deve ser utilizada a opção "finalizar processos".

Caso seja necessário reinicializar todas as máquinas, somente o gerente tem permissão para realizar esta tarefa. E, para tal, deve utilizar a opção "reboot".

## IV. Trabalhos Relacionados

Vários são os conjuntos de ferramentas, como o acima descrito, já desenvolvidos, entre eles: PBS, DQS, Condor, Codine e outros [CLU 00]. Abaixo segue uma breve descrição dos citados.

O objetivo do projeto Portable Batch System [POR 00], ou simplesmente PBS, foi inicialmente criar um sistema de processamento de batch para satisfazer demandas únicas de uma rede de computadores heterogêneos. O propósito do PBS é oferecer controles adicionais além de inicialização e agendamento de execução de batch processos, e direcionar rotas deses processos entre diferentes hosts. O módulo de agendamento independente do PBS permite ao administrador do sistema definir que tipo de recursos, e quanto de cada recurso poderá ser usado por cada job. O modulo de agendamento tem total conhecimento das filas de job disponíveis, processos trabalhando, e recursos do sistema em uso.

O DQS [DIS 00] (Dynamic Queuing System) é um sistema de fila baseado em UNIX, desenvolvido pelo Supercomputer Computations Research Institute (SCRI) na Florida State University. O DQS é desenvolvido como uma ferramenta de gerenciamento para ajudar na distribuição de recursos computacionais pela rede. DQS oferece arquitetura transparente tanto para usuários como administradores num ambiente heterogêneo.

Condor [CON 00] é um software para execução de batch processos em estações de trabalho que possam estar ociosas. A maioria das ferramentas do Condor são de localização automática e alocação de máquinas ociosas, checkpointing e migração de processos. Todas essas ferramentas são conseguidas sem qualquer modificação nas linhas de kernel do UNIX. Não é preciso modificar o código fonte para executar o Condor, embora os programas devem estar ligados especificamente com bibliotecas do Condor.

Os softwares acima descritos são de pesquisa, sem fins comerciais. Já o Codine [COD 00] é desenvolvido pela Cray, como um pacote comercial. Seu objetivo é ser utilizado em ambientes de redes heterogêneos, em particular em grandes clusters de workstation com

servidores de computadores integrados, como computadores vetoriais e paralelos. Codine oferece um ambiente de enfileiramento de batchs para uma grande variedade de arquiteturas via uma ferramenta de administração baseado em um ambiente gráfico para o usuário (GUI). Ele também oference uma balanceamento de carga dinâmico e estático, checkpointing e suporte batch, processos interativos e paralelos. Por este software ser comercial, ele não foi considerado em nossos estudos, pois não oferecer uma boa flexibilidade e código fonte aberto para possível integração com nosso trabalho.

Todos os softwares mencionados acima, assim como outros existentes no mercado, se preocupam basicamente com tarefas como gerenciamento e escalonamento de processos. Nosso trabalho tem como principal característica a preocupação em simplificar as tarefas mais comuns de administração e uso do cluster. Como proposta para futuros estudos, pensamos em verificar uma possível integração entre a interface por nós desenvolvida e alguns dos softwares desenvolvidos por outras entidades.

## V. Conclusão

Com o objetivo de auxiliar o usuário e o gerente do cluster para processamento paralelo do IFSC em suas tarefas, este trabalho está sendo desenvolvido, possibilitando aos mesmos uma visualização clara e única do cluster. Optamos pelo desenvolvimento de um novo software, apesar da grande variedade de softwares de gerenciamento de cluster disponíveis no mercado, para oferecer uma melhor flexibilidade e adequação às necessidades dos usuários locais.

O trabalho mostra que sistemas desse tipo podem ser adequadamente implementados com o uso de técnicas de interface com o usuário desenvolvidas para a implementação de páginas de rede interativas, e com o uso de softwares amplamente disponíveis.

## Referências

[BAK 99] BAKER, M.; BUYYA, R.; LAFORENZA, D. *The Grid: International Efforts in Global Computing.*

[CLU 00] Cluster Computing Review http://www.npac.syr.edu/techreports/hypertext/sccs-748/cluster-review.html

[COD 00] CODINE http://www.genias.de/genias/english/codine

[CON 00] CONDOR http://www.cs.wisc.edu/condor

[CUL 98] CULLER, D. E.; SINGH, J. P.; GUPTA, A. *Parallel Computer Architecture: a Hardware/Software Aproach,* Morgan Kaufmann, 1998.

[DIS 00] Distributed Queuing System (DQS) http://www.scri.fs.edu/~pasko/dqs.html

[FOS 95] FOSTER, I. *Designing and building parallel programs,* Addison-Wesley, 1995.

[HAR 96] HARLAN, D.; POWERS, S.; DOYLE, P.; FÓGHLÚ, M. Ó. *Using Perl 5 for Web Programming*, QUE Corporation, 1996.

[POR 00] Portable Batch System (PBS) http://pbs.mrj.com

[THE 00] The Berkeley NOW Project http://now.cs.berkeley.edu