

CPAD–PUCRS/HP : GNU/Linux como Plataforma para Pesquisa em Alto Desempenho

Tiago Coelho Ferreto – ferreto@cpad.pucrs.br
Fausto Richetti Blanco – blanco@cpad.pucrs.br
Marco Aurélio Stelmar Netto – stelmar@cpad.pucrs.br
César Augusto FonticIELha De Rose – derose@inf.pucrs.br

Pontifícia Universidade Católica do Rio Grande do Sul
CPAD–PUCRS/HP – Centro de Pesquisa em Alto Desempenho
CEP 90619–900 – Porto Alegre, RS, Brasil
Tel: (051) 320–3500 Ramal: 4463

Resumo

A área de processamento de alto desempenho está fortemente relacionada ao software livre, devido ao fato de que a maioria dos sistemas e ferramentas utilizadas na área são distribuídos como tal. O sistema operacional GNU/Linux é um exemplo disto. Ele é utilizado atualmente em diversos clusters devido principalmente à sua eficiência e estabilidade. Este artigo apresenta a experiência do Centro de Pesquisa em Alto Desempenho (CPAD – PUCRS/HP) na utilização do GNU/Linux em clusters. Todos os sistemas e ferramentas desenvolvidos no centro são distribuídos como software livre, com o intuito de promover a área e compartilhar as experiências adquiridas com outros centros.

Abstract

The High Performance Computing Area is closely related with free software, because mostly of the systems and tools used in the area are distributed as free software. The operating system GNU/Linux is an example of this. It is used nowadays in many clusters mainly for its efficiency and stability. This paper presents the Research Center of High Performance Computing (CPAD – PUCRS/HP) experience using GNU/Linux on clusters. All the systems and tools developed in the center are distributed as free software, to promote the area and share the obtained experiences with other centers.

1. Introdução

O processamento de alto desempenho é considerado uma ferramenta fundamental para as áreas de ciência e tecnologia [1]. Sua importância estratégica é demonstrada pela quantidade de iniciativas em pesquisa e desenvolvimento nestas áreas, financiadas por organizações do mundo todo. Um exemplo é a definição nos Estados Unidos de um conjunto de aplicações prioritárias (*Grand Challenges* [2], ou grandes desafios) que estão exigindo alto desempenho (e investimentos do governo para resolvê-las). O processamento de alto desempenho, contudo, depende por sua vez de técnicas do processamento paralelo, capazes de prover o desempenho necessário àquelas aplicações [2].

Existem diversas arquiteturas para a realização de processamento de alto desempenho [3]. Atualmente a tendência são as arquiteturas baseadas em agregados, ou simplesmente clusters. Cluster é uma coleção de computadores [4] (estações de trabalho, máquinas pessoais ou servidores), chamados de nodos, os quais são utilizados exclusivamente para obtenção de alto desempenho. Estas máquinas são fisicamente interconectadas por uma rede local comum ou de alta performance. As principais vantagens desta classe [3] são:

- Comportam-se como redes de estações porque os nodos da rede são servidores ou PCs e podem ser usados em aplicações convencionais;
- Com o uso de placas de comunicação de alta velocidade (SCI [5], Myrinet [6]), o sistema de comunicação tem um desempenho que se aproxima das MPPs (*Massively Parallel Processors*),

onde a vazão alcança a ordem de centenas de Mbytes/s e sua latência alguns poucos microsegundos;

- Estações e placas são produzidas em grande escala, o que resulta em um custo total bem menor que um MPP tanto na compra quanto na manutenção.

Como dito anteriormente, este ramo da informática é utilizado por diversas áreas da ciência. Normalmente, existem em grandes universidades centros de pesquisa que disponibilizam este tipo de arquitetura. O Centro de Pesquisa em Alto Desempenho (CPAD) é um destes centros. O projeto CPAD [7] é o resultado de uma parceria entre a Pontifícia Universidade Católica do Rio Grande do Sul e a Hewlett Packard do Brasil. Os principais objetivos do projeto são:

- Estimular a pesquisa e capacitação nas áreas de arquitetura de software e hardware para processamento paralelo e distribuído;
- Desenvolver camadas de software destinadas a otimizar a utilização do sistema operacional GNU/Linux em processamento paralelo e distribuído, colocando-as à disposição da comunidade científica na forma de código aberto;
- Estreitar o relacionamento entre os setores científico e industrial;
- Implantar um centro de pesquisa que forneça a infra-estrutura necessária à pesquisa na área de processamento paralelo e distribuído.

O CPAD conta hoje com dois clusters. Um deles, chamado de cluster principal, é composto por 16 servidores HP/E60, cada um contendo 2 processadores Pentium III de 550MHz e 128 MBytes de memória principal. Estes são interligados por duas redes, uma Fast Ethernet, utilizada para disparar processos e gerenciar o cluster, e uma Myrinet [6], utilizada para troca de mensagens entre os nodos. O outro cluster é chamado de cluster experimental e é composto por 4 PCs HP/Vectra VE. Cada nó possui 1 processador Pentium III de 500MHz e 128 MBytes de memória principal. Estes nodos são interligados por duas redes, uma Fast Ethernet, utilizada para disparar processos e gerenciar o cluster, e uma rede SCI [5], utilizada para comunicação entre os nodos através de segmentos de memória compartilhados. Para obter maiores informações é possível acessar o *site* do CPAD pelo endereço <http://www.cpad.pucrs.br>.

2. A Função do Sistema Operacional em Clusters

O sistema operacional em clusters possui os mesmos objetivos que em uma estação de trabalho [8]. Estes objetivos são: escalonar os múltiplos processos dos usuários num conjunto único de componentes de hardware (gerenciamento de recursos, tendo como ênfase os múltiplos processadores e memórias), e prover abstrações para o software de alto-nível. Algumas destas abstrações incluem a proteção de acesso do usuário a áreas restritas, sincronização de processos, threads e comunicação, assim como o gerenciamento de dispositivos.

Existem diversas características desejáveis em um sistema operacional de clusters. Entretanto, nota-se que estas características possuem em certas ocasiões uma exclusividade mútua, onde consegue-se uma melhoria em um certo ponto decorrente da negatividade de outro fator. As principais características [8] são:

- Gerenciamento: A administração deve ser realizada de forma remota e intuitiva. O gerenciamento possui relacionamento direto com o conceito de SSI (*Single System Image* ou Imagem Única do Sistema), onde cria-se a visão de uma única máquina (transparência), ao invés de diversas máquinas interconectadas.
- Estabilidade: As aplicações executadas nos clusters, pela sua complexidade, costumam durar muito tempo. Logo, é necessário que o sistema possua solidez contra processos com execução anormal, recuperação após erro com reconfiguração dinâmica, e usabilidade sobre carga pesada.
- Performance: A performance é essencial para um sistema operacional de cluster, pois os clusters são construídos para a obtenção de alto desempenho.

- Extensibilidade: O sistema deve possuir mecanismos de integração com novos dispositivos (eg. um novo tipo de rede de alta performance), através de drivers ou módulos que devem ser incorporados ao sistema operacional.
- Escalabilidade: A escalabilidade se refere a capacidade do cluster ser aumentado (inclusão de nodos na máquina), e suporte do sistema neste aumento sem perder eficiência ou apresentar problemas de funcionamento.

Para obter as características descritas, e tendo em vista a complexidade da criação de um sistema operacional para clusters, a solução mais usada é a utilização de um sistema operacional convencional juntamente com uma camada intermediária (*middleware*). Esta camada deve desenvolver serviços em nível de usuário capazes de satisfazer as principais necessidades de uma aplicação ou sistema paralelo/distribuído.

3. GNU/Linux como Sistema Operacional para Clusters

O sistema operacional GNU/Linux [9] é atualmente a plataforma mais utilizada em clusters, pois atende muito bem aos requisitos descritos em 2. Para o CPAD, a escolha da plataforma a ser utilizada nas suas máquinas deveria, além de possuir os principais requisitos de um sistema operacional para clusters, atender às configurações dos clusters, isto é, possuir suporte às máquinas (nodos) e tecnologias de rede (SCI [5], Myrinet [6]) utilizadas na arquitetura. O GNU/Linux foi escolhido por satisfazer a estes requisitos. As principais vantagens encontradas no sistema GNU/Linux foram:

- Estabilidade: O GNU/Linux é baseado no sistema operacional Unix, o qual é muito utilizado por administradores de sistemas devido à sua estabilidade.
- Eficiência: A eficiência do Linux é decorrente do enfoque adotado no seu desenvolvimento, o qual visa um sistema enxuto e com a utilização de todos os recursos oferecidos pela arquitetura.
- Sistema de rede: O Linux foi criado como um sistema de rede. Já que um cluster é uma rede de máquinas, sua utilização é direta.
- SMP: O Linux possui suporte a SMP (*Symmetric Multi-Processing*), o qual habilita o escalonamento de processos e *kernel-threads* entre os processadores da máquina [11]. Isto é fundamental, pois a utilização de máquinas SMP é uma tendência na construção de clusters.
- Código aberto: O código-fonte do Linux, por ser aberto, possibilita que partes de sua estrutura sejam modificados, visando, neste caso, uma melhor adaptação às características do cluster no qual é utilizado.
- Custo: O Linux é gratuito. Ele pode ser adquirido pela Internet sem custo algum. Isto estimula a utilização e distribuição de software livre, e evita também problemas com licenças de software (eg. Um cluster com 1000 nodos deveria possuir as 1000 licenças do sistema operacional utilizado).

O principal problema encontrado no GNU/Linux são as suas diversas distribuições, tornando algumas distribuições incompatíveis com outras. Isto ocorre pois ainda não existe um padrão na criação de distribuições GNU/Linux. Este problema logo deve ser resolvido devido a criação do LSB (*Linux Standard Base*) [12], o qual visa criar uma plataforma padrão a ser seguida por todos os distribuidores.

A distribuição escolhida pelo CPAD foi a Slackware [10]. Isto ocorreu devido às suas principais características [13]:

- É uma das distribuições mais antigas de GNU/Linux, mas está ainda em evidência.
- Possui *scripts* de inicialização e configuração bastante compreensíveis.
- É uma distribuição que fornece bastante liberdade ao administrador.

O Slackware tem como problema principal, o fato de possuir pouca ou nenhuma organização no que diz respeito ao gerenciamento de dependência de pacotes, ou seja, é possível remover um

pacote de software do qual outros pacotes necessitam para funcionar sem sequer ser avisado disto. Um outro problema encontrado é relacionado ao *upgrade* da distribuição, o qual é muito complicado, sendo mais simples fazer backup de dados de usuários e configurações e instalar o novo Slackware a partir do início.

Para atuar como parte da camada intermediária (middleware), a ser utilizada juntamente com o GNU/Linux no gerenciamento do cluster, o CPAD desenvolveu um mecanismo de controle de acesso ao cluster baseado em fila de espera, onde os usuários fazem os pedidos de alocação dos nodos e o sistema vai atendendo pela ordem de chegada. Este mecanismo possibilita também que sejam alocadas partes do cluster, tendo em vista que mais de um usuário possa utilizar os nodos ao mesmo tempo sem que interfira na execução dos processos dos demais usuários. Além disso, o CPAD desenvolveu um sistema de reservas para execução de programas que possuam um tempo de processamento maior que o permitido pelo sistema baseado em fila de espera.

4. Conclusão

O GNU/Linux possui as principais características necessárias a um sistema operacional de clusters. Ele possui eficiência, estabilidade e flexibilidade em relação a mudanças no seu código interno. Desta forma a sua utilização na área de alto desempenho cresce cada vez mais. O CPAD desenvolve sistemas e ferramentas utilizando o GNU/Linux, e os distribui como código aberto, facilitando a troca de experiências entre as diversas instituições. O desenvolvimento de software livre utilizando o sistema operacional GNU/Linux garante a sua continuidade e sobrevivência na pesquisa em alto desempenho.

Referências Bibliográficas

- [1] DE ROSE, César A.F., ZORZO, F., COSTA, Celso M. da, DOTTI, Fernando L.. **Centro de Pesquisa em Alto Desempenho – CPAD–PUCRS/HP**. 1a. ERAD – Escola Regional de Alto Desempenho. Gramado, 2001.
- [2] LEWIS, Ted G., EL–REWINI, H.. **Introduction to Parallel Computing**.
- [3] MAI, G. DE ROSE, C.. **Arquiteturas paralelas versáteis e de baixo custo para a pesquisa e o ensino na área de processamento paralelo e distribuído**. Conferência Latino Americana de Informática, Assunção, Paraguai. 1999.
- [4] HWANG, Kai, XU, Zhiwei. **Scalable Parallel Computing**. McGraw–Hill. 1998.
- [5] IEEE; IEEE Standart for Scalable Coherent Interface (SCI). **IEEE standart**, 1596–1992. New York, 1993.
- [6] SEITZ, C. L. et. al. Myrinet – A Gigabit–per–Second Local–Area Network. **IEEE Micro**, vol. 15, n. 1, Fevereiro 1995, pp. 29–36.
- [7] CPAD – Centro de Pesquisa em Alto Desempenho. <http://www.cpad.pucrs.br>. 2001.
- [8] CHAPIN, Steve, WORRINGEN, Joachim. **Operating Systems**. Cluster Computing White Paper. Versão 2.0. Dezembro de 2000.
- [9] Linux – How Good Is It? Executive Summary, D.H. Brown Associates Inc., <http://www.dhbrown.com>, 1999.
- [10] The Slackware Linux Project, <http://www.slackware.com>.
- [11] MENTRÉ, David. Linux SMP–HOWTO. **Linux HowTos**. <http://www.irisa.fr/prive/mentre/smp-howto>.
- [12] Linux Standard Base, <http://www.linuxbase.org>.
- [13] MACAN, Eduardo M.. **As distribuições Linux**. Revista de Informação e Tecnologia CCUEC Unicamp.