

Avaliando a Disponibilidade de Recursos em um *Cluster Virtual*

G. P. Pezzi, N. Maillard, C. A. F. De Rose, K. Saikoski

CPAD – PUCRS/HP

Av. Ipiranga, 6681 - Faculdade de Informática

pezzi@cpad.pucrs.br, {nmaillard, derose, saikoski}@inf.pucrs.br

Introdução

O ambiente VCLUSTER (*cluster virtual*) é um projeto entre a PUCRS e HP Brasil que visa agregar os recursos ociosos dos laboratórios da FACIN para fornecer um ambiente de alto desempenho cujos recursos se alteram dinamicamente. Torna-se portanto imprescindível ter uma previsão dos recursos disponíveis para fornecer ao usuário.

Este trabalho apresenta o estudo relativo à monitoração do VCLUSTER, realizado durante dois meses. Foi implementado um sistema de troca de informações no qual os nós mandam mensagens ao sair do modo *cluster* para informar sua atividade. Assim, foi possível ter uma imagem média da disponibilidade dos nós do VCLUSTER.

Existem diversos projetos que focam o uso de ciclos ociosos: o projeto Condor [LIV 97], VCLUSTER [DER 2002], NWS, etc. O VCLUSTER é um ambiente semelhante ao dos *clusters* dedicados que ao mesmo tempo protege os dados do usuário usando um sistema de *reboot* do nó [RIC 2002] e torna possível a execução de aplicações paralelas em máquinas de usuário. No entanto, a dinamicidade não nos permite afirmar a disponibilidade dos nós em um dado momento, de modo que é necessária uma previsão da disponibilidade das máquinas baseada nos dados de monitoração recolhidos.

Para saber quando um nó está disponível, o parâmetro utilizado é a troca de contexto da máquina e o *boot* do novo sistema operacional. Cada vez que um nó volta para o sistema do usuário, ele envia um *e-mail* contendo informações sobre seu uso.

Ambientes Dinâmicos e Monitoração

Esta seção apresenta as soluções classicamente encontradas para monitorar ambientes dinâmicos: uma delas (Condor) por ser usada em um ambiente próximo ao VCLUSTER. A outra (NWS) por permitir a monitoração e a previsão da disponibilidade dos recursos.

O Projeto VCLUSTER: O CPAD desenvolve desde 2001, em colaboração com a HP Brasil, o ambiente VCLUSTER [DER 2002], que agrega as estações ociosas de um laboratório em um *cluster virtual* de forma que um usuário possa alocar nós e disparar aplicações usando os mesmos *scripts* disponíveis para os outros *clusters*. Para detectar quais nós estão entrando e saindo do VCLUSTER foi usado um módulo de troca de contexto (Mode-Switch). O Mode-Switch detecta recursos ociosos e inicializa um novo sistema

operacional que é executado em uma outra partição do disco. Se o usuário local precisar da máquina novamente o nó volta para o sistema operacional original sem qualquer perda de dados¹. O ambiente VCLUSTER encontra-se instalado em 25 máquinas do laboratório da Faculdade de Informática.

Condor: é um projeto que começou há 15 anos e foca no uso de recursos ociosos. O objetivo inicial era usar o sistema de chamadas de procedimentos remotas junto com *checkpointing* para executar tarefas assíncronas em processadores ociosos. No que diz respeito à monitoração, Condor usa o módulo Hawkeye que tem um monitor central e um agente por nó usado pelo Condor. Cada agente fica responsável por transmitir ao monitor um objeto “ClassAd” descrevendo o atual estado do nó. Um “ClassAd” é uma classe padrão definida pelo Condor para a descrição de recursos computacionais. Os agentes mandam regularmente os classAds para o monitor e o não recebimento é visto como um desligamento. O monitor é também responsável por fornecer as estatísticas (geradas através de uma base de dados) e pode gerar saídas em HTML.

Network Weather Service: O ex-projeto AppLeS (agora nomeado Grail [CAS 2002]) foca o escalonamento de tarefas em grades de processadores. O módulo NWS [WOL 99] está especificamente encarregado de monitorar e prever o desempenho da rede e dos diversos recursos usados. Assim como Condor/Hawkeye, NWS usa um “sensor” local a cada nó, o qual pode monitorar o desempenho da rede, bem como o da CPU ou da memória. Além dos sensores existentes, NWS permite de forma simples a programação de novos sensores que um usuário queira implementar. Além dos sensores e do módulo centralizando os dados, NWS fornece ferramentas matemáticas de tratamento de séries temporais, a fim de prever a evolução futura dos recursos monitorados (daí o nome).

Monitoração do VCLUSTER

Devido à dinamicidade de um *cluster virtual*, surgiu a necessidade de avaliar o comportamento em termos da quantidade de recursos disponível a cada hora do dia. De fato, é possível contabilizar os dados do passado para que se possa fazer uma previsão para o futuro.

Escolha de um método para aquisição e formato dos dados: A abordagem escolhida foi o envio de *e-mails*, por não interferir no desempenho do nó e de não ser necessária a realização de grandes alterações na imagem do VCLUSTER. A mensagem deve conter nome da máquina, data incluindo dia da semana, hora de *login* e hora de saída do modo *cluster*.

Análise estatística dos dados: Foram abordadas duas formas de representação: o número médio de nós disponíveis por hora do dia e a disponibilidade média de cada nó separadamente. Este dado permite identificar, caso existam, diferenças e semelhanças no comportamento das máquinas, a fim de separá-las em grupos que terão tratamento distinto.

¹O Mode-Switch se encarrega do salvamento do contexto

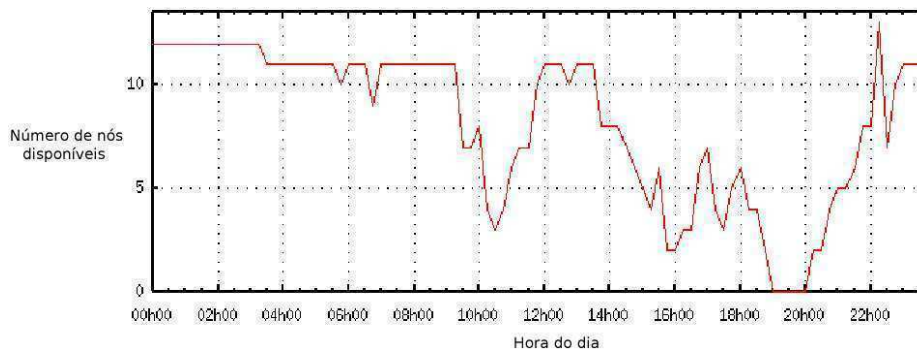


Figura 1: Resultado de um dia de medição no vCLUSTER.

Resultados Experimentais

Com o *script* de monitoração instalado em algumas máquinas do vCLUSTER e a ferramenta de análise dos dados pronta já foi possível ter uma idéia da utilização das máquinas. A Figura 1 mostra os horários de pico de utilização e ociosidade em um dia normal de aula.

Escolha de um intervalo de amostragem apropriado para testar a disponibilidade dos nós: Para isso foram feitos gráficos com intervalos de amostragem diferentes, de 15, 30 e 60 minutos. Os gráficos gerados representam a média e o intervalo de confiança dos nós disponíveis em um certo período de dias úteis. A Figura 2 apresenta uma sobreposição dos gráficos de 15 e 60 minutos.

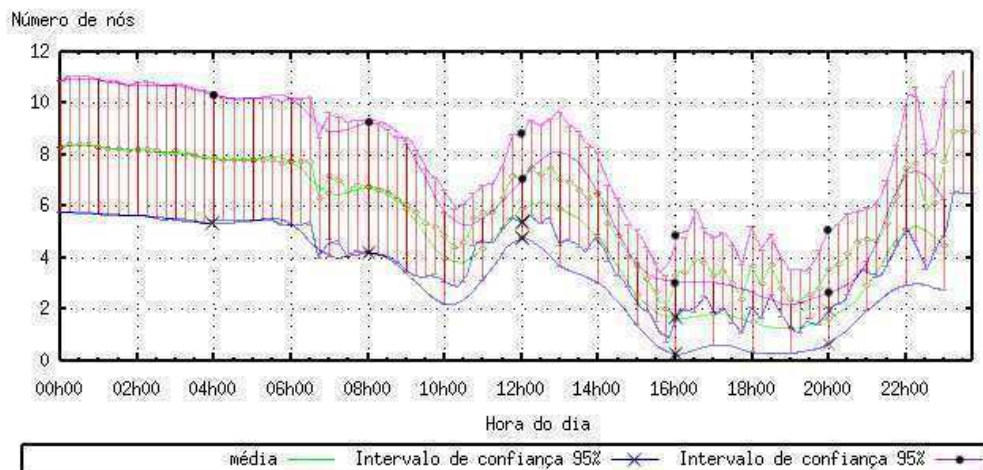


Figura 2: Número médio de nós disponíveis no vCLUSTER.

As linhas mais suaves (e com valores mais baixos) representam as medições com

60 minutos. Por haver uma diferença de comportamento considerável (de aproximadamente 25% em alguns pontos), optamos pelo intervalo de 15 minutos que é mais preciso.

Considerações sobre a análise dos nós separadamente: A disponibilidade média é uma porcentagem das vezes que o nó foi encontrado disponível no intervalo medido. O objetivo de analisar a disponibilidade de cada nó separadamente é encontrar grupos de máquinas que tenham comportamentos distintos das demais máquinas. Isto possibilita que, no futuro, esses grupos sejam identificados e tratados diferenciadamente.

Considerações Finais

Com esse estudo foi possível visualizar os horários em que se tem maior e menor utilização do laboratório. Esta informação já permite que os usuários tenham uma idéia de qual o melhor horário para executar as suas aplicações no VCLUSTER. Para aproximar a estrutura fornecida pelo VCLUSTER ainda mais à de um *cluster* dedicado, será necessário integrar a ferramenta de alocação com um módulo estatístico, de forma que a ferramenta de alocação faça automaticamente a previsão da disponibilidade para o usuário. A ferramenta de alocação pode mostrar, por exemplo, o número de máquinas previstas para estarem no modo usuário como já alocadas.

Cabe ainda fazer algumas variações nos gráficos, para tentar encontrar informações que possam ser importantes. Outras alternativas de análise dos dados podem ser utilizadas de acordo com as especificidades de cada ambiente, como por exemplo, a frequência de troca de contexto de cada nó. Para o laboratório em estudo neste trabalho, um outro gráfico que poderia interessar seria uma média da disponibilidade por dia da semana, uma vez que a utilização do laboratório depende do esquema de aulas que é semanal.

Referências

- [CAS 2002] CASANOVA, H.; HAYES, J.; YANG, Y. **Algorithms and software to schedule and deploy independent tasks in grid environments**. Aussois, France: [s.n.], 2002. invited paper to the Workshop on Distributed Computing, Metacomputing, and Resource Globalization.
- [DER 2002] DE ROSE, C. A. F. et al. The virtual cluster: a dynamic environment for exploitation of idle network resources. In: SYMPOSIUM ON COMPUTER ARCHITECTURE AND HIGH PERFORMANCE COMPUTING (SBAC-PAD'2002), 14., 2002, Vitória, ES, Brasil. **Proceedings...** [S.l.: s.n.], 2002. p.141–148.
- [LIV 97] LIVNY, M. et al. Mechanisms for high throughput computing. **Speedup Journal**, v.11, n.1, June 1997.
- [RIC 2002] RICHARD, B. I-cluster: the execution sandbox. In: IEEE INTERNATIONAL CONFERENCE ON CLUSTER COMPUTING 2002, 2002. **Proceedings...** [S.l.: s.n.], 2002.
- [WOL 99] WOLSKI, R.; SPRING, N.; HAYES, J. The network weather service: a distributed resource performance forecasting service for metacomputing. **Journal of Future Generation Computing Systems**, v.15, n.5-6, p.757–768, October 1999.