

Visualização Simultânea e Multinível de Informações de Monitoramento de *Cluster**

Lucas Mello Schnorr, Benhur Stein, Philippe Navaux

Universidade Federal do Rio Grande do Sul - Instituto de Informática
Av. Bento Gonçalves, 9500 - Bloco IV - CEP 91501-970 Porto Alegre - RS
lmschnorr@inf.ufrgs.br, benhur@inf.ufsm.br, navaux@inf.ufrgs.br

Introdução

O monitoramento de um *cluster* consiste na coleta e análise de informações de estado das máquinas que o compõem. Esse monitoramento pode ser empregado para se observar o comportamento das máquinas e identificar possíveis causas de problemas. Pode ser utilizado também para se acompanhar e otimizar as aplicações executadas nesses *clusters*, especialmente quando associado ao monitoramento de bibliotecas de comunicação.

A utilização de apenas uma fonte de informação pode não fornecer dados suficientes para se entender os motivos de determinado comportamento do *cluster* ou de uma aplicação. Quando isso ocorre, o desenvolvedor pode fazer uso de uma ferramenta complementar de monitoramento para identificar a fonte do problema. A utilização de mais de uma ferramenta de monitoramento e visualização torna a tarefa de monitoramento de um *cluster* mais complicada.

Com base nesse contexto, este trabalho tem como objetivo facilitar a visualização do estado de um *cluster* para tarefas administrativas e de desenvolvimento de aplicações. Para isso, propõe-se a construção de um sistema capaz de coletar e integrar as informações oriundas de diferentes fontes de informação e fornecer uma visualização simultânea e correlata desses dados. Com o monitoramento centralizado, o administrador de um *cluster* ou o desenvolvedor de aplicações paralelas poderá realizar uma análise mais completa das informações coletadas.

A seguir são apresentados detalhes acerca do sistema que está sendo construído.

Descrição do sistema

O sistema que está sendo construído é composto basicamente por duas etapas: a primeira trata da coleta das informações de diferentes fontes enquanto que a segunda faz a tarefa de integrar os dados e permitir a visualização dessas informações.

A etapa da coleta consiste na obtenção de dados registrados pelas diferentes fontes de informação. As fontes de informação que estão sendo utilizadas pelo sistema são as bibliotecas de comunicação DECK [Barreto 1998] e MPI [Gropp 1994], a ferramenta de monitoramento de *cluster* Ganglia [Massie 2003] e dados coletados do sistema operacional. No caso da biblioteca DECK, foi feita uma instrumentação que permite o registro do comportamento de aplicações desenvolvida com essa biblioteca. O MPI, ao

*Financiamento: CNPq através do processo número 380051/03-6

contrário, já possui uma forma de se obter o registro dos eventos bastando construir um coletor dessas informações que entendesse o formato gerado do MPI. A arquitetura do Ganglia foi alterada para se obter informações em intervalos de tempos menores, permitindo uma maior amostragem do que está acontecendo com os dados monitorados. Para se obter as informações do Ganglia, foi construído um coletor independente que obtém as informações do Ganglia ao longo do tempo e as registra em arquivo. Para se obter dados do sistema operacional foi instrumentado o núcleo do Linux para capturar as trocas de contexto durante a execução de um determinado processo e seus fluxos de execução.

A segunda etapa, integração e visualização das informações, se preocupa basicamente na padronização, uniformização e sincronização dos dados além de uma forma de visualizar as informações. A padronização das informações consiste em fazer com que os identificadores dos dados de uma fonte sejam iguais aos de outra fonte de informação. Por exemplo, se a ferramenta de monitoramento **A** identifica as máquinas do *cluster* com números sequenciais e a ferramenta **B** identifica-as através dos nomes, o sistema deve unificar essas duas formas de identificação. A sincronização das datas das informações de diferentes fontes é feita tomando-se por referência o relógio de uma determinada máquina do *cluster*. As datas das demais informações são sincronizadas de acordo com esse relógio de referência. Na uniformização de formato para a visualização, os dados oriundos de diferentes fontes de informação são convertidos para um formato único. No caso, o formato utilizado é o formato de arquivo da ferramenta de visualização Pajé [Stein 2000]. Por fim, a visualização se dá através da utilização da ferramenta Pajé.

Conclusão

Este texto descreve um sistema capaz de integrar e visualizar simultaneamente informações de monitoramento de *cluster*. Com esse sistema, o administrador de um *cluster* ou o desenvolvedor de aplicações paralelas poderá realizar uma análise mais completa das informações de monitoramento das máquinas e das aplicações que nelas são executadas. Como trabalhos futuros, pretende-se aumentar o número de fontes de informação do sistema.

Referências

- [Stein 2000] Stein, B.; Kergommeaux, J. C.; Bernard, P. *Pajé, an interactive visualization tool for tuning multi-threaded parallel applications*. Elsevier. Parallel Computing, 2000.
- [Massie 2003] Massie, M.; Chun, B.; Culler, D. *The Ganglia Distributed Monitoring System: Design, Implementation and Experience*. UCLA, Berkeley Technical Report, 2003.
- [Gropp 1994] Gropp, W.; Lusk, E.; Skjellum, A. *Using MPI: Portable Parallel Programming with the Message-Passing Interface*. MIT Press, Cambridge, MA, October 1994.
- [Barreto 1998] Barreto, M.; Navaux, P.; Rivière, M. *DECK: a new model for a distributed executive kernel integrating communication and multithreading for support of distributed object oriented application with fault tolerance support*. Congreso Argentino de Ciencias de la Computacion. pp 623-637. Neuquém, Argentina. 1998.