Quantificação de nodos em diferentes estados de consumo de energia para ambiente de nuvem

Fábio Diniz Rossi, César A. F. De Rose

[1]Programa de Pós-Graduação em Ciência da Computação
Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)
Av. Ipiranga, 6681 – Porto Alegre – RS – Brazil

fabio.diniz@acad.pucrs.br, cesar.derose@pucrs.br

Resumo. *Ambientes de computação de nuvem fornecem uma grande quantidade de nodos ociosos, que poderiam ser desligados visando economia de energia. O tempo no processo de desligar/religar nodos infuencia em métricas de desempenho das aplicações. Este artigo apresenta um modelo que calcula a quantidade de nodos que devem ser mantidos em um estado de suspensão, possibilitando melhorar o trade-off entre economia de energia e desempenho.*

1. Introdução

Com o objetivo de economizar energia sobre nodos em ambientes de nuvem, pesquisadores têm respondido ao desafio de projetar soluções para o gerenciamento do consumo de energia através da utilização de várias técnicas e mecanismos suportados por sistemas de computação modernos [Maillard et al. 2010]. Para oferecer uma relação aceitável entre economia de energia e desempenho dos aplicativos, muitas abordagens são propostas na literatura. A Figura 1 apresenta uma arquitetura proposta em [Rossi and De Rose 2014], voltada para a gestão de técnicas de economia de energia, a fim de alcançar um melhor *trade-off* entre economia de energia e desempenho. Na Figura 1 existem nodos em execução (*Running Nodes*), e nodos desligados (*Powered off Nodes*). Intermediariamente, apresentamos nodos em um estado de suspensão (*Sleep State Nodes*), com o objetivo de responder de forma mais rápida a uma nova demanda computacional do que nodos desligados, ao passo que possibilita economizar mais energia quando comparados a ambientes tradicionais que mantêm nodos em estado ocioso.

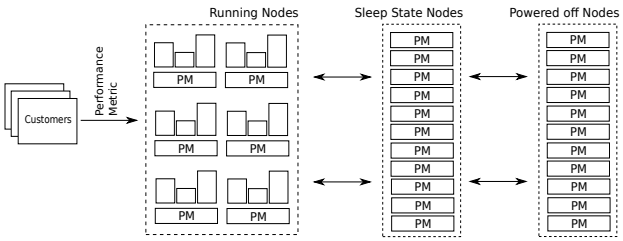

Figura 1. Arquitetura de Nuvem Energeticamente Eficiente

Dependendo da carga imposta pelas aplicações, e para manter o tempo de resposta acordado, o balanceador de carga do ambiente de nuvem deve instanciar novas máquinas virtuais sobre novos nodos físicos (PM). Em caso de subutilização, máquinas virtuais são desalocadas. Com base na dinamicidade desse ambiente, a arquitetura apresentada na Figura 1 possibilita gerenciar nodos em diferentes estados: executando, em suspensão, desligados. A grande questão nessa arquitetura é decidir a quantidade de nodos necessários

para esse estado intermediário de suspensão (*standby*), de maneira que possa suprir às novas demandas das aplicações, ao passo que também permita melhor ajuste de nodos em estado de desligado.

2. Proposta

Assim, nesse artigo propomos um modelo preliminar de ajuste sobre a quantidade de nodos em um estado intermediário $|N_{(sleep)}|$, baseando-se na quantidade de nodos em execução e na variabilidade da carga das aplicações, como mostrado na Equação 1,

$$N_{(sleep)} = N_{(run)}\ x\ (1 - \frac{(\frac{RT_{(SLA)}}{RT_{(now)}})}{\alpha}) \tag{1}$$

onde

$N_{(sleep)}$: consiste na quantidade de nodos em *standby*;
$N_{(run)}$: consiste na quantidade de nodos em execução;
$RT_{(SLA)}$: representa o tempo de resposta acordado em SLA;
$RT_{(now)}$: representa o tempo de respota atual;
α: apresenta a agressividade do modelo (α = 2, nodos em standby contabilizam 50% dos nodos em execução).

Portanto, esse modelo primeiramente faz uma comparação entre o tempo de resposta acordado e o tempo de resposta atual da aplicação. Com essa informação, pode-se visualizar se novas máquinas virtuais serão instanciadas/desalocadas sobre/de novos nodos físicos em um futuro próximo, visando suportar uma crescente demanda da aplicação. Somado a isso, a variabilidade entre os tempos de resposta é armazenada na variável α, que ajusta a quantidade de nodos que devem se manter em estado intermediário, em uma fatia maior ou menor (agressividade), dependendo da necessidade de alocação de novas máquinas virtuais.

3. Perspectivas

A maioria dos trabalhos apresentam estratégias que utilizam diversas técnicas que possibilitam economia de energia em ambientes de nuvem. Porém, todas as estratégias propostas são bastante impactantes no desempenho das aplicações. Acreditamos que nosso modelo preliminar possa ser refinado, possibilitando o ajuste necessário para a gerência de um ambiente de nuvem, com foco na economia de energia, levando em consideração o desempenho das aplicações. Avaliações do modelo proposto com objetivo de reduzir esse *trade-off* ainda estão sendo realizadas.

Referências

Maillard, N., Navaux, P., and De Rose, C. (2010). Energy-aware scheduling of parallel programs. In *Conferencia Latino Americana de Computación de Alto Rendimiento*, CLCAR, pages 95–101.

Rossi, F. D. and De Rose, C. A. F. (2014). Estratégia para economizar energia em nuvens computacionais. In *14ª Escola Regional de Alto Desempenho*, ERAD 2014.