

## **Analisando Padrões de Uso para o Gerenciamento Automático de Recursos em Cloud Computing**

**Júlio Cezar Santos Pires, Cristiano André da Costa, Rodrigo da Rosa Righi**

<sup>1</sup>Programa de Pós-Graduação em Computação Aplicada - PIPCA

Universidade do Vale do Rio do Sinos - UNISINOS

Av. Unisinos, 950 - Bairro Cristo Rei - CEP 93.022-000 - São Leopoldo – RS – Brasil

juliocspires@gmail.com, {cac,rrrighi}@unisinos.br

### **1. Introdução**

Cloud Computing, ou Computação em Nuvem, é um modelo para permitir o acesso sob demanda para um conjunto de recursos de computação configurável. Redes, servidores, sistemas de armazenamento, aplicativos e serviços podem ser rapidamente fornecidos e liberados com o mínimo esforço (em teoria) na interação entre usuários e provedores de serviço [Vázquez et al. 2011]. Para viabilizar essa dinâmica e oferecer controle para usuários e administradores da Nuvem, a tarefa de monitoramento é primordial. Elasticidade, dinamicidade da infraestrutura e das aplicações, tolerância a falhas, balanceamento de carga e acordos de nível de serviço (SLA) são características relacionadas a Nuvem e que necessitam de um sistema de monitoramento para controlá-las e efetivá-las [De Chaves et al. 2011]. Nesse contexto, esse artigo aborda a área de monitoramento que trata da coleta e análise de padrões de utilização de serviços em Nuvem. A verificação deles é pertinente para esse ambiente, uma vez que podem antecipar decisões de gerenciamento, reduzir consumo energético e minimizar custos financeiros. Esse artigo descreve as motivações e as vantagens de trabalhar com padrões em Nuvem, bem como trabalhos relacionados e os primeiros passos da pesquisa.

### **2. Motivação e Passos Iniciais da Pesquisa**

Elasticidade é uma característica chave da Nuvem. Uma das empresas mais tradicionais na área, a Amazon, disponibiliza essa propriedade através da cooperação entre dois de seus sistemas: Cloud Watch e Elastic Compute Cloud (EC2). O primeiro atua como um sistema de monitoramento, no qual são definidos alarmes e métricas para coleta de dados. O segundo permite definir regras que quando acionadas podem redimensionar as instâncias de máquinas virtuais em execução. Ambos possuem uma interface em linha de comando na qual possibilita fazer as requisições. Cada novo conjunto de serviços a serem executados necessitam da criação de novas regras específicas para atender a escalabilidade. Além disso, recursos com diferentes características podem precisar da reformulação das regras da mesma maneira.

Analisando essa questão, esse trabalho aborda maneiras de oferecer elasticidade sem onerar o usuário com dados específicos de recursos ou de sua aplicação. Nesse sentido, serão estudados padrões de uso como objeto de pesquisa para se atingir esse objetivo. Além do controle automático da elasticidade, ações baseadas em padrões podem ser úteis nas seguintes situações: (i) tratamento da energia consumida pela Nuvem; (ii) migração de máquinas virtuais; (iii) auxílio na administração dos recursos para o provedor da nuvem; (iv) possibilidade de estimar o tempo para executar um serviço e; (v) gerenciamento do SLA entre consumidores e o provedor de serviço.

A economia de energia elétrica pode ser dada através da desalocação, ou consolidação, de recursos. Um padrão pode ser composto pela análise do histórico de uma ou mais métricas, tais como variação nas taxas de utilização de CPU, disco e memória RAM. Assim, é possível estabelecer curvas de comportamento e tomar ações proativas de elasticidade que trazem, além de benefícios no âmbito de energia, vantagens para ambos, usuário e administrador da Nuvem. O primeiro executa a sua aplicação com um conjunto de recursos ideal ou próximo disso para garantir o SLA. O segundo consegue planejar o uso dos recursos de sua Nuvem e assim pré alocá-los a requisições que entram em vigor no futuro. Contrariamente à ideia largamente difundida, uma Nuvem tem recursos limitados e a administração deles perante as requisições de usuários é pertinente para a adoção do próprio sistema de Nuvem. Em adição, uma vez que recursos podem ser desalocados, além da diminuição da energia com o desligamento das máquinas, é possível adaptar o sistema de resfriamento para consumir menos uma vez que os alvos foram reduzidos.

Um padrão é caracterizado por uma ação cíclica por parte da aplicação, no qual interfere no uso dos recursos. A verificação dele passa por um estudo de similaridades. Elas estão baseadas no fato que o estado de utilização atual do recurso possui uma probabilidade de já ter ocorrido no passado. A análise pode envolver predições que envolvem um grande volume de dados, tornando-a muitas vezes custosa. Visando tratar de forma eficiente este problema, será estudado inicialmente o sistema Apache Hadoop. Ele apresenta uma plataforma escalável que permite verificar e processar grandes quantidades de dados distribuídos entre diferentes nós. Além do Hadoop, a funcionalidade de monitoramento será estudada nos seguintes sistemas: OpenNebula, Eucalyptus, Nagios, Microsoft Azure e Ganglia. Em especial, OpenNebula trabalha com o OpenNebula Watch para capturar informações contábeis e se destaca em relação a concorrentes por oferecer escalonamento dinâmico para o tratamento da escalabilidade de máquinas virtuais [Peng et al. 2009].

### 3. Conclusões e Trabalhos Futuros

Os desafios na área de padrões de uso podem ser sintetizados nas seguintes subáreas: (i) mensurar o grau de acerto nas decisões; (ii) baixa intrusividade sobre a execução da aplicação e; (iii) observar quais métricas são pertinentes para cada tipo de serviço. Os próximos passos compreendem o estudo de métodos computacionais para a interpretação de padrões, bem como a execução de aplicações paralelas que possuem comportamento cíclico juntamente com o Apache Hadoop.

### Referências

- De Chaves, S., Uriarte, R., and Westphall, C. (2011). Toward an architecture for monitoring private clouds. *Communications Magazine, IEEE*, 49(12):130–137.
- Peng, J., Zhang, X., Lei, Z., Zhang, B., Zhang, W., and Li, Q. (2009). Comparison of several cloud computing platforms. In *Information Science and Engineering (ISISE), 2009 Second International Symposium on*, pages 23–27.
- Vázquez, C., Huedo, E., Montero, R. S., and Llorente, I. M. (2011). On the use of clouds for grid resource provisioning. *Future Gener. Comput. Syst.*, 27:600–605.