



PIBIC/CNPq/UFPG-2012

ESTUDO COMPARATIVO DE SERVIÇOS DE AUTO-SCALING OFERECIDOS NO MERCADO DE COMPUTAÇÃO NA NUVEM

Lília Rodrigues Sampaio¹, Raquel Vigolvinho Lopes²

RESUMO

A Computação na Nuvem se propõe a fornecer serviços em vários níveis, desde o provimento de recursos computacionais sob-demanda até o acesso a um software na Internet. Dentro dessa área, um novo serviço vem aparecendo no mercado, o *auto scaling*. Essa técnica consiste em escalar dinamicamente a quantidade de recursos que executa uma aplicação, baseado em condições definidas pelo usuário considerando métricas coletadas por um monitor ligado a aplicação. Este trabalho foi, então, desenvolvido e organizado em duas partes. A primeira, referente a uma comparação sistemática entre diversos serviços de *auto scaling* oferecidos no mercado, bem como serviços semelhantes de monitoramento e gerenciamento de aplicações, se deu através da definição de critérios relevantes de comparação entre eles, e avaliação da adequação de tais serviços para implementação de um algoritmo tradicional de provisão dinâmica proposto na literatura. Ainda parte desse estudo, estendemos a pesquisa para a identificação de arquiteturas de infraestrutura de *deployment* para tais serviços, possibilitando o entendimento de seu funcionamento em diversos provedores e gerenciadores. A segunda parte incluiu, por fim, a aplicação prática dos conceitos de *auto scaling* estudados, através do suporte ao funcionamento e gerenciamento de uma aplicação web chamada *Forrosquare*, utilizando o provedor Amazon EC2, e uma análise de seu desempenho.

Palavras-chave: computação na nuvem, *auto scaling*, comparativo arquitetural do *auto scaling*.

COMPARATIVE STUDY OF AUTO-SCALING SERVICES OFFERED BY CLOUD COMPUTING MARKET

ABSTRACT

Cloud Computing has the purpose of providing services in different levels, making possible the acquisition of computing resources on demand, as well as the access to a software through the Internet. In this area, a new service has emerged in the market, the *auto scaling*. This technique consists in scheduling dynamically the amount of resources, based on conditions defined by the user considering metrics collected by a monitor connected to the application. This research was, then, developed and organized in two steps. The first one, concerning a comparative study between a diversity of *auto scaling* services offered by the market, as well as similar services of monitoring and management of applications, consisted in the definition of relevant comparative criteria for analyzing them, and an evaluation of the adequacy of such services to the implementation of a traditional dynamic provisioning algorithm proposed by the academia. Still in this study, we extended the research to the identification of possible architectures of deployment infrastructures to such services, making possible the understanding of its behavior on a range of providers and managers. The second part included, finally, the practical application of *auto scaling* concepts studied, towards a support of the functionality and management of one web application called *Forrosquare*, using the Amazon EC2 provider, and an analysis of its performance.

Keywords: cloud computing, *auto scaling*, architectural comparative of *auto scaling*.

¹ Aluna do Curso de Ciência da Computação, Centro de Engenharia Elétrica e Informática, Departamento de Sistemas e Computação, UFPG, Campina Grande, PB, e-mail: lilia.sampaio@ccc.ufcg.edu.br

² Ciência da Computação, Professora Doutora, Centro de Engenharia Elétrica e Informática, Departamento de Sistemas e Computação, UFPG, Campina Grande, PB, e-mail: raquel@dsc.ufcg.edu.br