# Chapter 77
# On the Conceptualization of Elastic Service Evaluation in Cloud Computing

**Danqing Feng**
*Harbin Institute of Technology, Harbin, China; & AirForce Communication NCO Academy, Dalian, China*

**Zhibo Wu**
*Harbin Institute of Technology, Harbin, China*

**Zhan Zhang**
*Harbin Institute of Technology, Harbin, China*

**Jinwei Fu**
*Harbin Institute of Technology, Harbin, China*

## ABSTRACT

*Cloud computing is becoming an urgent technology in the enterprises. One key characteristic in the cloud computing is the elasticity. Then, it is urgent for the users how to rank the renting services reasonably. Considering the main features of the elasticity, this article gives classification on resource optimization. However, one of the major challenges is how to optimize resource allocation in an elastic manner. Due to the special pay-as-you-go manner, resource optimizing strategies are associated with the goal of minimizing the costs on the premise of service-level-agreement (SLA). Another challenge of resource optimizing strategies is to how to dynamically respond to the application demands. In this paper, the authors sketch the elastic definition more clearly. Secondly, different dimensions are described on elastic resource allocations. Thirdly, it is important to seek out the proper resource allocation strategy. Finally, the challenges and conclusions are discussed in this article.*

DOI: 10.4018/978-1-7998-5339-8.ch077

## INTRODUCTION

Recently, cloud computing (Rimal et al., 2009) has been developed rapidly in the technical era. The basic principle is that users can manage and store data everywhere at any time, and pay as they have used in the cloud computing, such as electricity and water (Buyya et al., 2009). Then, one of the core features in the cloud computing is to allocate and reconfigure the resources in a dynamic and flexible manner. That is, a reasonable resource-optimized strategy is to implement an efficient elastic resource allocation. However, it is a challenge to exactly identify elastic resource demands. It is prone to two conditions as described. On one hand, the under-provisioned resources are generating the state of insufficient resources. On the other hand, the over-provisioned resources are causing the wasted resource and experiencing the SLA violations (Van et al., 2009). To avoid both improper situations, the efficient way in the resource allocation is to exactly meet the application demands (Buchholz et al., 2012)). However, it is still far from the perfect elasticity (Kuperberg et al., 2011). Nowadays, it is critical and urgent to find a reasonable elastic resource allocation strategy.

In traditional resource deploying manner, the threshold is set for the static configuration, which causes more wasted resource. In contrast, the dynamic resource configuration (Ding, 2015; Chaudhry, 2015; Xiong, 2016) is allocated in a flexible way. In this paper, we present a survey on how to rank the elastic service from the aspect of resource allocation. The goal is to make a further research on an auto-scaling resource allocation. However, for the providers and the users, what they focus is different. The providers focus more on VM start time, available resource, the utilization, the infrastructure and the cost; On the contrary, the users focus on service response time, application demands and the expenditure on leasing resources. And the SLA is focused by both the providers and the users. The involved parameters are listed in Table 1. In fact, for the purpose of overcoming the pros and cons, the researchers also pose cloud brokers to manage and allocate resources to maximize the benefit between the providers and the users. Actually, it is at issue how to rank the elastic service according to the Key Performance Indicators (KPIs).

*Table 1. Input parameters*

| Parameter | Provider | Customer |
|---|---|---|
| Service Level Agreement(SLA) | √ | √ |
| Response Time | | √ |
| Availability | √ | |
| Utilization | √ | |
| Application Demand | | √ |
| Infrastructure | √ | |
| Cost | √ | |
| Expenditure | | √ |

In a word, the contribution of this paper is two-fold: Firstly, this paper gives a classification on the elasticity. From the aspects on resource optimization, we make the further research on elastic resource allocations. We first discuss the definition deeply. Then, considering with resource optimization, we partition resource allocation strategies combining from the angle of the elasticity. Further, the key opti-