Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 110 (2017) 94–101

Procedia Computer Science

www.elsevier.com/locate/procedia

The 14th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2017)

A Resource Provision Strategy for Software-as-a-Service in Cloud Computing

Wen-Hwa Liao*, Peng-Wen Chen, Ssu-Chi Kuai

Abstract

The business model of Software-as-a-Service (SaaS) is gradually becoming more refined. Reducing the operation cost of cloud service providers while maintaining service efficiency has become an important topic. The present study focuses on suggesting a resource provisioning strategy for load changes during the operations of an SaaS cloud service provider (CSP). The strategy is to suggest corresponding provision portfolio configurations that can reduce operation cost and maintain efficiency of applications based on surplus resources after reducing the use of virtual machines rented for the SaaS CSP. The underlying concept is identical to memory paging management in operating systems: minimizing fragmented space to utilize resources effectively. Testing results confirm that the strategy suggested in this study can effectively configure a provision portfolio with reduced operation cost. Compared with methods suggested by IaaS CSPs in the mainstream market, the method proposed in the present study can reduce the hourly rental of virtual machines by 40% and enable SaaS CSPs to realize considerable profit by reducing cost.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Cloud computing, Resource provision, Software-as-a-Service (SaaS)

1. Introduction

Undoubtedly, the most significant advantage introduced by cloud computing is the changes it has made to previous information system establishment procedures and methods. This includes the ability to access computing resources according to requirements[2] similar to daily life resources such as water, electricity, gas, and telecommunication[4]. With cloud computing, implementing information systems no longer requires tedious procedures such as equipment purchases and provisioning. Concerns for resource wastage or missing opportunities

* Corresponding author. Tel.: +886-2-2182-2928 ext. 6793; fax: +886-2-25853966.
E-mail address: whliao@ttu.edu.tw

1877-0509
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.06.123

caused by inaccurate usage estimation can also be eliminated[2]. The most advantageous convenience provided by cloud computing is its rapid elasticity[10]. In common internet application systems, tenants can increase the number of servicing virtual machines to maintain system performance when there is an increase in system load[8]. During off-peak periods and when the system is placed under an idle state, tenants can also reduce the number of virtual machines to reduce servicing cost[2]. Infrastructure-as-a-Service cloud service providers (IaaS CSPs) such as Amazon Web Services (AWS) by Amazon, Google Cloud Platform by Google, and Microsoft Azure by Microsoft are operating based on this concept to allow tenants to adjust their computing resources themselves. These IaaS CSP resource portfolios include CPU computation power, memory, saving space, load balancing, firewall, and internet bandwidth.

However, to lower cost practically, system administrators must perform resource management and deployment based on the loading conditions; unfortunately, these tasks are not without effort. In addition to the observation of system operation states, assessments and analyses on growth and/or operating trends of the system are also necessary. From the point of view of system function assembly, there are differing levels of demand for resources by different system functions. This significantly increases the difficulty of creating a deployment environment through resource management. An example would be the usage of simple text information as opposed to multimedia type data such as graphics, audios, and videos. The process of transmission, displaying, and processing these types of data require varying amounts of CPU computing power, memory, hard disk space, and internet bandwidth. Furthermore, from the perspective of operations, the life cycle of the system and system functions also require constant improvements and adjustments based on demand. Even if the system was configured to manage peak loads during the initial stages, it may not be suitable once conditions change.

As mentioned previously, the main role of the IaaS CSP is to provide resources and accessories of different levels for tenants to choose to assemble their own information system environment. Regarding how to manage expansion and reduction of the configuration system effectively based on data usage, and how to satisfy users' demands while saving resources, which is most important to firms, tenants must develop their own plans. Tenants that we mention in this study refer to users of SaaS CSPs. During the operation of SaaS CSPs, the compulsory cost for renting virtual machines involves operating cost; this situation leads to a typical resource management problem. In this study, the environment provided by the IaaS CSPs is utilized and a set of application service provision strategies for SaaS CSP operation are suggested based on the application service resource demand of the SaaS CSPs to meet the goal of satisfying application service performance and reducing cost according to operation conditions.

The remainder of this paper is organized as follows. Section 2 discusses related work in resource provision. The proposed resource provision strategy for Software-as-a-Service in cloud computing is presented in Section 3. Section 4 presents the results of the performance evaluation. Section 5 concludes this paper.

2. Related Work

Regarding system load capacity evaluation, methods suggested by researches[6,8] can be referenced. All these studies employ the queuing theory model to estimate assignment waiting time at different points in the system and further estimate an acceptable system processing power when the system is in an idle state. Nevertheless, reference data required by these formulas require recording and observation under actual system operation. Another method that deviates from the methods mentioned above proposes the waiving of the evaluation process, employing a black box method directly, utilizing the performance response by monitoring the operating condition as reference data and using these data as feedback to adjust the provisioning[3,9]. Given that IaaS CSPs provides this environment, we are able to manage resources flexibly. The method discussed in this paper is a combination of the two important foundations mentioned above. This study discusses the provision configuration strategies of SaaS CSP application services and IaaS CSP virtual machines after completing assessments of the system load capacity or confirmation of the magnitude of expansion or reduction that occurs after the expansion or reduction demand has been calculated by the monitoring operations.

For the expansion method during system operation, when the system operation encounters a bottleneck, adding resources is necessary to guarantee service quality. However, what kind of resources should be added to overcome the bottleneck? For example, if the bottleneck is caused by excessive requests by dynamic multimedia documents, blindly adding resources to the application host computer could result in a waste of resources. Research[6] mentioned