International Journal of Engineering & Technology
CENTRE OF PROFESSIONAL RESEARCH PUBLICATIONS
ISSN 2049-3444

International Journal of Engineering and Technology Volume 2 No. 12, December, 2012

ResourceGate: A New Solution for Cloud Computing Resource Allocation

Abdullah A. Sheikh
Department of Computer Science and Software Engineering
The University of Melbourne, Australia

ABSTRACT

Cloud computing has taken place to be focused by educational and business communities. These concerns include their needs to improve the Quality of Services (QoS) provided, also services such as reliability, performance and reducing costs. Cloud computing provides many benefits in terms of low cost and accessibility of data. Ensuring these benefits is considered to be the major factor in the cloud computing environment. This paper surveys recent research related to cloud computing resource allocation and addresses possible solutions. It presents some contributions of resource allocation that have been introduced in these fields. The proposed solution aims to develop resource management system for PaaS in cloud Environment. It provides reducing IT costs and enhancing performance of resource allocation in PaaS. In addition, our solution allows users of sharing software's licenses. It is found that there is a much room of research in reducing cost and sharing application licences in cloud computing environment.

Keywords: *Cloud Computing, Resource Allocation, Platform as a Service*

1. INTRODUCTION

The use of cloud computing has increased rapidly in many organizations Cloud computing is increasingly attracting both academic and industrial communities concerns [17]. These concerns include their needs to improve the Quality of Services (QoS) provided, also services reliability, performance and reducing costs[3]. Cloud computing provides many benefits in terms of low cost and accessibility of data. Subashini and Kavitha [10] argue that small and medium companies use cloud computing services for various reasons, including because these services provide fast access to their applications and reduce their infrastructure costs.

This paper focuses on reviewing recent research related to cloud computing resource allocation and addresses possible solutions. The proposed solution is about reducing IT costs and enhancing performance of resource allocation in PaaS. In addition, our solution allows users of sharing software's licenses. Also, it can help managing the services by giving the user new type of services as service on demand.

This paper is organized as follows. Section 2 describes the beginning of cloud computing and its components. Section 3 discusses and analyses an example of current resource allocation system, as well as examining their limitations. Section 4 presents suggestions for a possible solution. Section 5 will conclude the paper with the suggestion of future work.

2. BACKGROUND

NIST [1] gives a basic definition of cloud computing as "a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".

2.1 Cloud Computing Architecture

NIST defines cloud computing model into five characteristics, three delivery models, and four deployment models [1]. The five key characteristics of cloud computing are: location-independent resource pooling, on-demand self-service, rapid elasticity, broad network access, and measured service [11]. These five characteristics represent the first layer in the cloud environment architecture (see Figure1).

Service models mean the level of services provided in cloud computing [7-9]. The cloud community has used three types of service models such as Infrastructure (IaaS), Platform (PaaS) and Software (SaaS) [3, 6]. It explained as following:

Infrastructure as a Service (IaaS) means that delivering the computing resources is in virtual machine form (VM) [10] [4],which provides to the user a view of the identified server. The user is able to manage the system within a

virtual machine and deploy the required software. Amazon EC2 and Google Compute Engine are examples of IaaS.

Software as a Service (SaaS) this application level presents the software services provided to the user that runs over the internet [10] which also running on the cloud infrastructure [4]. Application from Amazon Web Services Marketplace and Salesforce.com are good examples of SaaS.

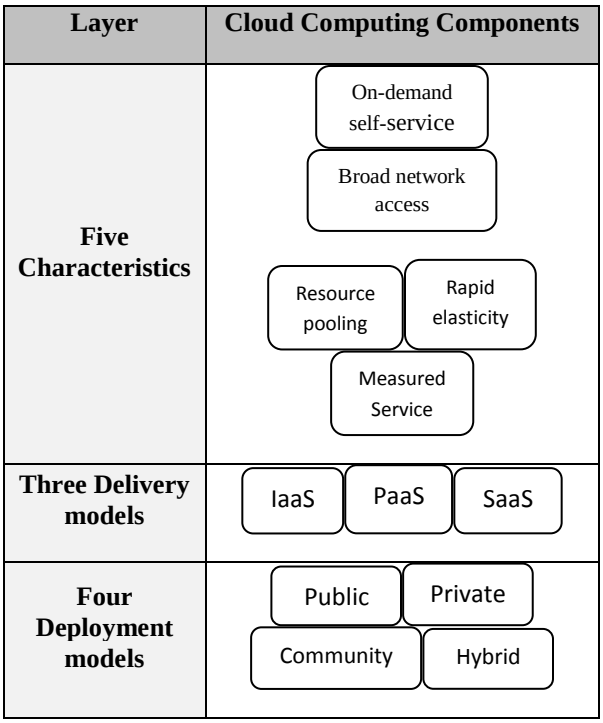

Figure 1: Cloud Environment Architecture [2].

Platform as a Service (PaaS) this models means the access to resources which provided as an Application Programming Interface (API) form. That used for development and deployment the applications [10]. In this model, there is no direct access to the application by users and the resource allocation is managed automatically by the platform[4]. Google App Engine and Microsoft Azure are examples of PaaS. Cloud deployment models include public, private, community, and hybrid clouds. A cloud environment that is accessible for multi-tenants and is available to the public is called a public cloud. A private cloud is available for a particular group, while a community cloud is modified for a specific group of customers. Hybrid cloud infrastructure is a composition of two or more clouds (private, community, or public cloud) [11]. This model represents the third layer in the cloud environment architecture.

3. RESOURCE ALLOCATION

3.1 Resource Allocation: Preliminary

Resource allocation is about sharing softwares and applications through the cloud platforms. Moreover, it helps to reduce costs of technologies that have been used, and to reduce the time of the services that provided loading or preloading [4, 14]. For example, sharing software licences in one lab and sharing applications accessible will be more convenient [15]. Resource allocation has a big implication on cloud computing in many different prospective such as, accessing resources and bandwidth allocation [13].

3.2 Related works of Resource Allocation

Qiang et al. [9] offer a solution to improve the performance by making resources virtual based on virtual machine (VM) which makes all hardware resources in public as shared space. In addition, they aim to regulate many resources utilisation of service level objective of applications SLOs. Jiayin et al. [6] propose an algorithm that adjust resource allocation based on updating the actual task executions which helps to recalculate the finishing time that assigned to the cloud.

Walsh et al. [12] propose a solution about dividing architecture to two-layers (local and global). The local layer is responsible for calculating the utilities. Where, the global layer computes the near optimal configuration for resources based on result that provided by local layers. This solution is considered as statical way of resource allocation and it implemented to fix the load balancing

with the server cluster which also helps applications scalability [8].

Yazir et al. [18] present a new approach that has been introduced for dynamic autonomous resource management in cloud computing which consists of a distributed architecture of NAs. In addition, it performs resource configurations using Multiple Criteria Decision Analyse MCDA with the PROMETHEE method. This approach is more practicable with large amount of data centre [8].

According to Goudarzi et al. [5], there is a problem considered of resource allocation, which is the optimising the total profits that gained from multi dimensional for multi tire application. Their aimed is to apply resource consolidation techniques to consolidate resources determining the active servers.

3.3 Analysis of Current Solutions

As mentioned before, according to Goudarzi et al. [5], there is a problem has been considered of resource allocation and that is to optimise the total profits gained from multi dimensional for multi tire application. Their aimed is to apply resource consolidation techniques to consolidate resources determining the active servers.

There are some approaches that based on statical solutions of resource allocation and some are based on dynamics. Statical solutions help in some issues like applications saleability and loading balance of the web server. Otherwise, the dynamical solutions have not discussed these issues because they need a minimum response time and high level of reliability from web applications [8]. In addition, dynamical solutions have discussed the cloud virtualisation and reducing the time cost [8].

Comparing all contributions in these fields presents the reason why the large number of cloud resources need to be efficiently managed for allocation, reallocation, and balancing resource access. Desirable resource allocation efficiency is facing some inherent challenges such as unpredictable demands, on-demand resource provisioning, and dynamic availability considering time-variant and high energy costs of data centres. In order for improving resources allocation within clouds, such stated challenges and considerations should be addressed, towards achieving an optimised resource allocation model. Improving resource allocation should meet the changes in policies of resource management and works through the current network traffic situation. Hence, there is a need to pay more attention and spend extra efforts against the resource allocation challenges.

3.4 Example of Resource Allocation Framework

This section presents an example of resource allocation framework. In addition, an overview of how the example works with the discussion of its limitations will be provided.

Tian [16] provides details and technical work about the architecture of the virtual cloud computing lab design and how each side in the environment works (see Figure 2).

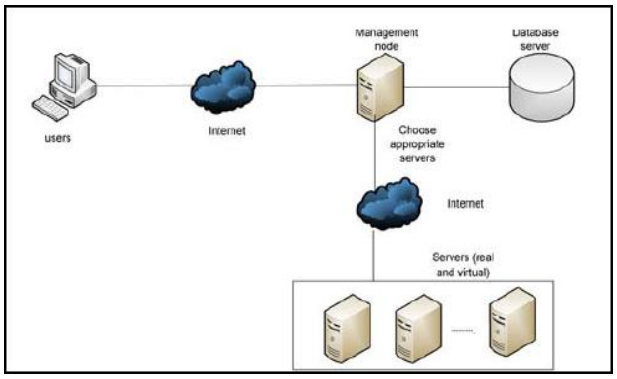

Figure 2 Architecture of cloud computing lab [16]

Figure 2 shows an overview of the Tian [16] cloud computing lab environment. In general, a user sends a request to the computing platform the Management node which makes the verification processes. Then management node finds the available resources from the available virtual server. The Management system will allocate resources to user for a specific period of time, then the user needs to renew the resources again for further use [16].

According to Tian [16], the management system of his framework includes Users Management, Allocation Management, and Connection Management (see Figure 3). For example, Users Management includes basic and login information for users. In addition, Allocation Management includes all information about resources and the status for each resource. The connection manager takes the responsibility for allocating resources to users meeting the authentication criteria.

Limitation that have been found in Tian's system [16] as following:

First, the variables affect the systems are users' authentication and resources status due to time wast. For example, the time needed to make the authentication process affects the performance and the load time. In addition, renewing the use of the allocated resources will affect the system performance as well. Second, the system confounds between users data and connection managers

and that can affect the performance and the cloud virtualisation. Finally, a user for each time needs to renew his request to use the allocated resources. Then the process will go over again which is time cost.

The propose solution covers the limitations in static system and it will be combined between static and dynamic resource allocation types with virtualisation and resource management. It would be better that the system includes the ability to reduce the traffic load because it does not show wither the resources size have been considered. That could affect the cloud computing for users who intend to use resources from outside.

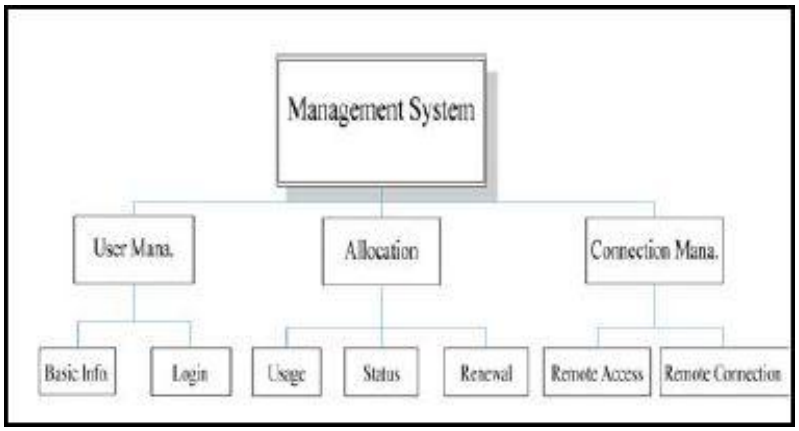

Figure 3 Management System [16]

4. PROPOSED SOLUTION

This section discusses the proposed solution and gives an over view of the system environment. Also, it explains what advantages it could be solved comparing with current works. Then it gives a brief risks and pay-off that could face the approach implementation.

4.1 The Approach Overview

As mentioned before statical approaches of resource allocation are different than dynamics. Statical solutions can help in issues like applications saleability and web server's loading balance. On the other hand, dynamical solutions have not been discussed these issues because it needs a minimum response time and high level of reliability from web applications. Additionally, dynamical solutions have discussed the cloud virtualisation and reducing the time cost. Moreover, most approaches are auto resource selections; users do not play any role of choosing the better resources for them. The following Figure (see Figure 4) presents the overall system environment called ResourceGate.

The basic operation of the system that the client makes the request through the application which consider as a controller then via the network connection which may be internet or intranet the request will be processed. If the user has given the authentication the system will response with the proper virtual server to load the resources to the client including his account profile and a virtual image of his resources and system.

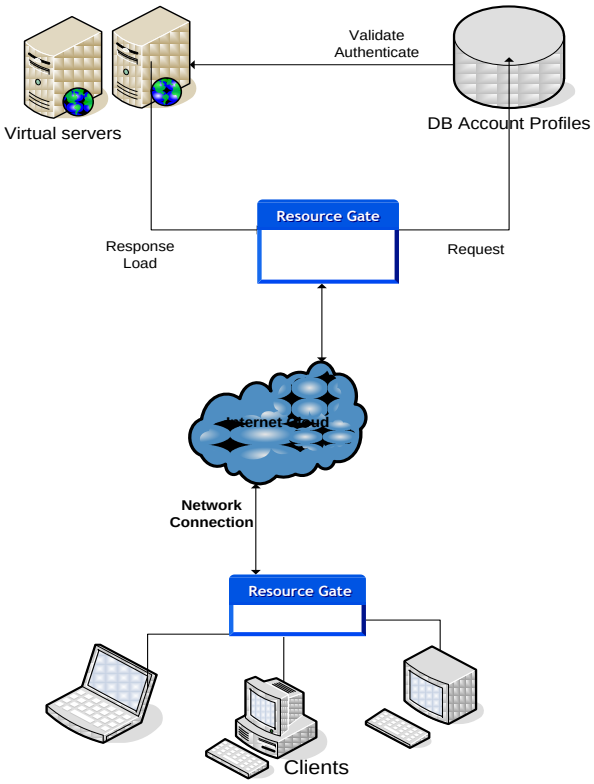

Figure 4: System overview of ResourceGate

From software engineering prospective, time of using the service will be controlled via sessions, so it will be dynamically renovate while the user is online. Using

server session will be more suitable to force any issues related such as bandwidth issues, security and integrity, responsiveness, cancelling the work and crashes, which may occur if the session implemented on the client side.

4.2 Benefits of the Proposed Solution

ResourceGate will make cloud computing resource allocation more efficient, faster and with less cost, also it will make application more available on demands. For example, it will be done by allowing users the chance to modify their setting to choose the desire resources in any time they want. Most users are concern about what type of services they get, and involving making a choice in their services can gain more trust and more efforts. Give the users more experience by reducing the time that cost when they need to load the services or waiting response time.

Renew the use of the allocated resources will be dynamically using session state functions. For example, the allocated resources will remain in use as the user still online through the session technique which will be automatically renewable.

However, this solution will help Cloud computing providers are the target of the research's outcomes. They need to improve the services and get the users trust. The proposed framework will help to improve the quality of services provided, and services reliability, flexible services , performance and reducing costs technology and data transmission. Resource allocation load level will be decreased to be better rather than load all services in one time but make it on demand services.

4.3 Risks and Pay-off

There are many barriers that facing the proposed solution some are technical and some are in business side. Technically, there is some implementation complexity and the assumption should be demonstrated theoretically before coding stage. In business side, there are some boundaries and role that control the services providers before evolving users to choose what they desire in the cloud computing such security and regulating concerns.

5. CONCLUSION AND FUTURE WORK

This paper has briefly introduced the cloud computing and its components. Also, it has given a brief over view of cloud computing resource allocation presenting the current works related. Furthermore, it analysed the current situation of the current works providing the most recent work as an example. In addition, it showed what limitations existed of some works such as time and IT cost. Finally, it proposed an approach that solved some constraints from the current solutions and presented what are the benefits and risks of the new approach. For future work, we are going to test a simulation of the new solution on a cloud computing lab.

ACKNOWLEDGMENT

Abdullah is supported by grant from Taif University in Kingdom of Saudi Arabia.

REFERENCES

[1] (NIST), *http://www.nist.gov/itl/cloud/*, Accessed in May-2011.

[2] M. A. AlZain, E. Pardede, B. Soh and J. A. Thom, *Cloud Computing Security: From Single to Multi-clouds*, *System Science (HICSS), 2012 45th Hawaii International Conference on*, pp. 5490-5499.

[3] A. Dastjerdi and R. Buyya, *A Taxonomy of QoS Management and Service Selection Methodologies for Cloud Computing*, 6 (2011), pp. 109-131.

[4] T. Dillon, C. Wu and E. Chang, *Cloud computing: Issues and challenges*, Ieee, pp. 27-33.

[5] H. Goudarzi and M. Pedram, *Multi-dimensional sla-based resource allocation for multi-tier cloud computing systems*, IEEE, pp. 324-331.

[6] L. Jiayin, Q. Meikang, N. Jian-Wei, C. Yu and M. Zhong, *Adaptive resource allocation for preemptable jobs in cloud systems*, *Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on*, pp. 31-36.

[7] S. Kamara and K. Lauter, *Cryptographic cloud storage*, Financial Cryptography and Data Security, 6054 (2010), pp. 136-149.

[8] C. S. Pawar and R. Wagh, *A review of resource allocation policies in cloud computing*, World Journal of Science and Technology, 2.

[9] L. Qiang, H. Qinfen, X. Limin and L. Zhoujun, *Adaptive Management of Virtualized Resources in Cloud Computing Using Feedback Control*, *Information Science and Engineering (ICISE), 2009 1st International Conference on*, 2009, pp. 99-102.

[10] S. Subashini and V. Kavitha, *A survey on security issues in service delivery models of cloud computing*, Journal of Network and Computer Applications (2011), pp. 1-11.

[11] H. Takabi, J. B. D. Joshi and G. Ahn, *Security and Privacy Challenges in Cloud Computing Environments*, Security & Privacy, IEEE, 8 (2010), pp. 24-31.

[12] W. E. Walsh, G. Tesauro, J. O. Kephart and R. Das, *Utility functions in autonomic systems*, IEEE, 2004, pp. 70-77.

[13] J. Weinman, *Network implications of cloud computing*, *Telecom World (ITU WT), 2011 Technical Symposium at ITU*, IEEE, pp. 75-81.

[14] T. Wenhong, *Three Ways to Improve the Efficiency of Virtual/Cloud Computing Lab*, *Apperceiving Computing and Intelligence Analysis, 2008. ICACIA 2008. International Conference on*, 2008, pp. 160-163.

[15] T. Wenhong, S. Sheng and L. Guoming, *A Framework for Implementing and Managing Platform as a Service in a Virtual Cloud Computing Lab*, *Education Technology and Computer Science (ETCS), 2010 Second International Workshop on*, pp. 273-276.

[16] T. Wenhong, S. Sheng and L. Guoming, *A Framework for Implementing and Managing Platform as a Service in a Virtual Cloud Computing Lab*, *2010 Second International Workshop on Education Technology and Computer Science (ETCS)*, 2010, pp. 273-276.

[17] Z. Yang, *1Look Beyond Utility Modelâ€"Disruptive Cloud Computing For Future ICT4D Research*.

[18] Y. O. Yazir, C. Matthews, R. Farahbod, S. Neville, A. Guitouni, S. Ganti and Y. Coady, *Dynamic resource allocation in computing clouds using distributed multiple criteria decision analysis*, Ieee, pp. 91-98.

