# Demand allocation policies for cloud service brokerage

## <sup>1</sup>Mihir Khandelwal, <sup>2</sup>Lokesh Lamba, <sup>3</sup>Dr N Bharathi Gopalasamy

<sup>1,2</sup>Student, <sup>3</sup>Professor SRM Institute of science and technology

*Abstract*: Cloud computing is a style of computing in which dynamically scalable and other virtualized resources are provided as a service over the Internet. The energy consumption and makespan associated with the resources allocated should be taken into account. This paper proposes an improved clonal selection algorithm based on time cost and energy consumption models in cloud computing environment. We have analyzed the performance of our approach using the CloudSim toolkit. The experimental results show that our approach has immense potential as it offers significant improvement in the aspects of response time and makespan, demonstrates high potential for the improvement in energy efficiency of the data center, and can effectively meet the service level agreement requested by the users.

## Introduction:

Cloud computing is a hot topic of the computer field as an emerging new computing model [1]. It is a style of computing in which dynamically scalable and other virtualized resources are provided as a service over the Internet [2]. It is the traditional computer and network technology including distributed computing, parallel computing, utility computing, network storage technologies, virtualization, load balance, etc. combined with other products [3]. Cloud computing is a model for enabling ubiquitous, on-demand network access to a shared pool of configurable computing resources by setting up basic hardware and software infrastructures in a data center. The aim of green cloud computing is to design a high-performance, low-power computing infrastructure while meeting an energy-efficient and safe service mode.

Resource allocation is the key technology of cloud computing, which utilizes the computing resources in the network to facilitate the execution of complicated tasks that require large-scale computation [4]. Resource allocation needs to consider many factors, such as load balancing, makespan, and energy consumption. Selecting favorable resource nodes to execute a task in cloud computing must be considered, and they have to be properly selected according to the properties of the task [5]. In particular, cloud resources need to be allocated not only to satisfy quality of service (QoS) requirements specified by users via service level agreements (SLAs) but also to reduce energy consumption [6,  $\underline{7}$ ].

## **Cloud computing services:**

There are two types of cloud models: the cloud service model and the cloud deployment model.

### **A. Cloud Computing Services**

The cloud offers its services in the form of three types:

1) Software-as-services (SaaS): The cloud provider offers the user SaaS through the Internet to use when he wants without having to install it on his PC This removes the work of updating and maintenance from the end user, cutting the price of buying the software and removing the need for inconvenient software licenses. The provider and the user will pay-as-used to gain access from any terminal with an operating system and web browser without physical restrictions do this automatically.

2) Platform-as-a services(PaaS): With PaaS, the cloud provides the customer a complex IT environment for the development of applications the consumer uses without having control of the network or server. These services provide the company with a development platform to create their own applications (e.g., Google App Engine).

3) Infrastructure-as-a-services (IaaS): With IaaS, the cloud provides the end user with the storage memory, server, and network on a pay-per-use basis, which reduces the cost and is advantageous to the business. These services provide the company with the ability to add or delete services easily, such as Amazon S3.

### **B.** Cloud Deployment Model

There are various cloud models and classification systems that determine how the services are provided to the end user:

1) **Public cloud:** This service is provided to public users and is open to the end user, but the end user cannot control the infrastructure. Examples of public clouds are IBMs' Blue Cloud, Sun Cloud, and Google App Engine.

2) **Private cloud:** This is any service that is provided exclusively to the end user or to a single organization. This offers more security than the public cloud and is more expensive.

3) Hybrid cloud: This cloud is composed of two or more different clouds which may be public or private. More important data can be kept in the private cloud, and the other types of data can be kept in the public cloud. This hybrid arrangement is more secure and less expensive than the purely private cloud.

4) **Community cloud:** In this cloud, the services are provided exclusively to a group of persons or companies who share the same interests.

#### **C. Resource Allocation:**

In Cloud Computing In cloud computing, resource allocation (RA) is a field that is taken into account in many computing areas such as datacenter management, operating systems, and grid computing. RA deals with the division of available resources between cloud users and applications in an economic and effective way. It is one of the challenging tasks in cloud computing based on the IaaS. Furthermore, RA for IaaS in cloud computing provides several benefits: it is cost effective because users do not need to install and update hardware or software to access the applications, its flexibility allows access applications and data on any system in the world, and there are no limitations of the medium or usage site.

In addition, there are two major processes of RA via cloud computing

### **D. Static Allocation**

Static Allocation schemes: assign fixed resources to the cloud user or application. In this case, the cloud user should know the number of resource instances needed for the application and what resources are requested and should aim to confirm the application's peak load requests. But the limitation for static allocation is usually affected by the over-utilization or under-utilization of computing resources based on the normal workload of the application. This is not cost-effective and is related to insufficient use of the resource during off-peak periods.

#### **E. Dynamic Allocation**

Dynamic Allocation schemes provide cloud resources on the fly when the cloud user or application is requested, specifically to avoid over-utilization and under-utilization of resources. A possible drawback when needed resources are requested on the fly is that they might not be accessible. Thus, the service supplier must allocate resources from different participating cloud data centers

Resource allocation strategy (RAS) is related to combining cloud provider functions for utilizing and assigning scarce resources within the boundaries of the cloud system in order to suit the demand of the cloud application.

As cloud computing has its characteristics, the RAS should avoid the following situations as much as possible:

1) Resource contention: This situation occurs when multiple users and applications attempt to allocate the same resource simultaneously.

2) Resource fragmentation: This occurs when applications cannot allocate resources due to isolated resources being small items.

3) Scarcity: This occurs when multiple applications' requirements for the resources are high and there are limited resources, for example, requests for memory, I/O devices, CPUs, and the techniques that should serve that demand.

4) Over provisioning: This occurs when the users and applications obtain more resources than those that are requested to fit the quality of service (QoS) requirements.

5) Under provisioning: This occurs when the users and applications obtain fewer resources than those requested to fit the QoS requirements.

From the perspective of cloud users, RA should be achieved at a lower cost and in as little time as possible. However, for the cloud provider, it is impractical to predict the dynamic of user demands, nature of users, and application demands. Therefore, resource diversity, limited resources, locality restrictions, dynamic nature of resource requests, and environmental necessities require an efficient and dynamic RAS that is suitable for cloud environments. Since the dynamic and uncertainty of resource demand and supply are unpredictable, different strategies for dynamic resource allocation are suggested. This research presents different RAS that are utilized in cloud environments.



### **Related work:**

Number of researchers has contributed their work in this area of cloud computing with a concern of resource allocation. Few of the work being drawn are defined below: Jinn-TsongTsai et al. (2013) has proposed Optimized time scheduling and resource allocation for cloud computing by utilizing IDEA (Improved Differential evolution algorithm) on the basis of time an cost. The proposed algorithm has depicted the effectiveness for optimizing the resource allocation and task scheduling. The decision makers can adopt the GANTT charts for the task scheduling by means of cost, makespan and the hybridization of both can be used for making the decision when the contradictory objectives are there. JavierEspadas et al. (2014) has attempted to implement the formal calculations for under and over positioning of resources virtualization in cloud computing in the cloud infrastructure, significantly for SaaS service and presented a resource allocation model for deploying SaaS application by means of cloud platforms by considering the multitenancy therefore, developing a cost- efficient scalable environment.

#### Conclusion:

We have presented the design, implementation, and evaluation virtualized cloud environment for dynamic resource allocation. This system multiplex the virtual machines based on the demands of the end users. We use SPAR method to provide memory efficiency in the servers. Finally this system achieves green computing and load balancing between heterogeneous end users.

### **Reference:**

1. Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., ... & Zaharia, M. (2010). A view of cloud computing. Communications of the ACM, 53(4), 50-58.

2. Qian, L., Luo, Z., Du, Y., & Guo, L. (2009). Cloud computing: An overview. Cloud computing, 626-631.

3. Dillon, T., Wu, C., & Chang, E. (2010, April). Cloud computing: issues and challenges. In Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on (pp. 27-33). Ieee.

4. Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., ... & Zaharia, M. (2010). A view of cloud computing. Communications of the ACM, 53(4), 50-58.

5. Krutz, R. L., & Vines, R. D. (2010). Cloud security: A comprehensive guide to secure cloud computing. Wiley Publishing.

6. Buyya, R., Ranjan, R., & Calheiros, R. N. (2009, June). Modeling and simulation of scalable Cloud computing environments and the CloudSim toolkit: Challenges and opportunities. In High Performance Computing & Simulation, 2009. HPCS'09. International Conference on (pp. 1-11). IEEE.

7. Xing, Y., & Zhan, Y. (2012). Virtualization and cloud computing. Future Wireless Networks and Information Systems, 305-312.

8. Swathi, T., Srikanth, K., & Reddy, S. R. (2014). Virtualization in cloud computing. International Journal of Computer Science and Mobile Computing, 3(5), 540-546.

9. Zhang, Y. Virtualization and Cloud Computing. Network Function Virtualization: Concepts and Applicability in 5G Networks, 13- 36.

10. Sharma, G. P., Singh, S., Singh, A., & Kaur, R. (2016). Virtualization in Cloud Computing.

11. Tsai, J. T., Fang, J. C., & Chou, J. H. (2013). Optimized task scheduling and resource allocation on cloud computing environment using improved differential evolution algorithm. Computers & Operations Research, 40(12), 3045-3055.

12. Shu, W., Wang, W., & Wang, Y. (2014). A novel energyefficient resource allocation algorithm based on immune clonal optimization for green cloud computing. EURASIP Journal on Wireless Communications and Networking, 2014(1), 64.

