IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol-6, Special Issue-5, Jun 2018 E-ISSN: 2347-2693

Resource Allocation in Cloud Computing: A Review

Vikas Mongia[1], Deepak Kumar[2]

[1]Research Scholar,Guru Kashi University,Talwandi Sabo(Bathinda), India
[2]Guru Kashi University, Talwandi Sabo (Bathinda), India

Available online at: www.ijcseonline.org

***Abstract* -** Cloud computing has quickly appeared as a outstanding standard for contributing IT infrastructure, resources and services on a pay-per-use basis from the last few years. Cloud computing is a promising technology and number of researches has been proposed for solving the issues faced by the cloud. There are number of challenges that a cloud is facing, from which, the main challenge is the resource allocation technique. Cloud permits the provisioning of resource on-demand. This procedure of allocating and re-allocating of resources is the way to accommodate the impulsive demands with an improvement of return on investment by means of infrastructure with the support of cloud. Resource allocation is the method in which the resources are allocated to each cloud user by the providers of cloud services. The varied factors like response time, cost, and dynamic allocation need to be acknowledged while choosing a technique of resource allocation. Though, in spite of the recent growth in cloud computing market, number of problems in the resource allocation remains unaddressed. This source course has introduced the significant concepts and the mechanisms of cloud computing and deliberates some research question on the topic while emphasizing on challenges and state-of-art solutions in the resource allocation. The article will expectantly inspire the future researchers to come up with the optimal and smarter resource allocation algorithms and structures to build up the paradigm of cloud computing.

Keywords: Cloud computing, service models, virtualization, resource allocation

I. Introduction

Cloud computing allows several organizations and customers to use a variety of applications without installing and accessing their personal records on any portable computer with web access [1]. The technology also allows greater proficiency in computing by using centralized data storage, bandwidth, and processing. Cloud computing has become a popular explanation, making cheap and easy access to external information technology an import artificial source. It helps the scientists with a new model for utilizing the computing infrastructure. More and more enterprises (for example, businesses, research centers, etc.) got benefits from cloud computing in the direction of hosting their applications. With virtualization, cloud computing can address the vast client base of heterogeneous computing need with the same type of physical infrastructure. Resources of computer, storage resources and the different applications can dynamically provide charge as per use and later can be released if not needed. Such services are offered with an agreement i.e. SLA, which gives the user the desired Quality of service (QoS). Some enterprises harness the ability of cloud resources with private and public clouds by giving full QoS to users. Cloud computing is flexible and service based infrastructure with the support of multiple programming pattern [2]. In contrast to previous approaches such as clustering and grid computing, cloud computing is not application-oriented, but service-oriented; it also provides on-demand virtualization type resources as a measureable and billing convenience.

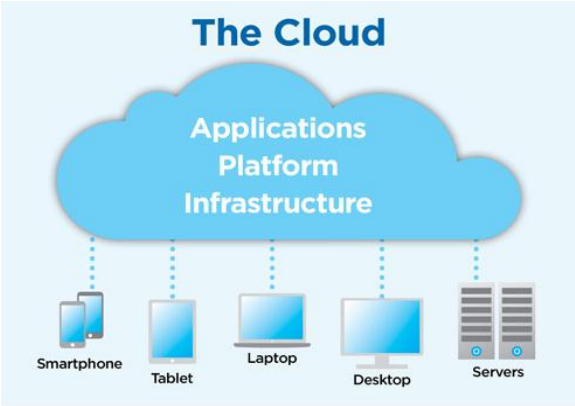

Figure 1: Cloud Architecture

II. Cloud Computing Services

There are three types of service models in cloud computing, namely, IaaS, PaaS and SaaS and the explanation for the same is given below and is shown in below diagram [3]:

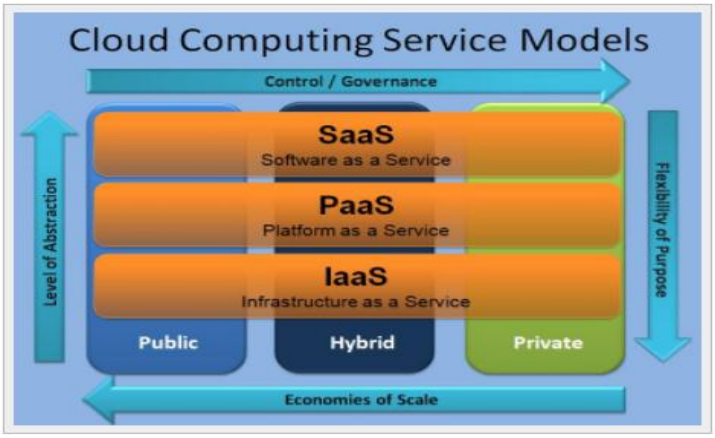

Figure 2: Service models of Cloud computing

Table 1: Cloud Services [4]

Service Models	Description
SaaS (Software as a service)	Software-as-a-service (SaaS) cloud users allow their applications in the main host environment after which it can be taken to use by application users from various clients through the network. It is software that is positioned over the internet in the LAN. It is "Pay-as-you-go" plan and was at beginning deployed for customer relationship Management (CRM) and Sales Force automation. Cloud consumers have no control over cloud infrastructure that often employs multi-tenancy system architecture (software multi-tenancy is a software architecture in which a single object/part of software runs on a server and gives service to multiple tenants (It is group of users who share a common access with given privileges)). SaaS examples include Google Mail, Google Docs.
IaaS (Infrastructure-as-a-service)	Infrastructure as a service gives a platform virtualization scenario as a service. Instead of purchasing servers, data centers pace and network equipment, software clients purchase those resources as a fully outsourced service. Cloud Users straight way takes IT infrastructures provided in IaaS cloud. In order to integrate physical resources in ad-hoc mode to meet the high and low resources demand, virtualization is extensively used. The main plan of virtualization is to set up independent VM's that are way away from hardware and other VMs. Amazon EC2 is an example of IaaS.
PaaS (Platform as a Service)	Platform as a service is a development platform that supports full "Software Lifecycle" allowing cloud consumers to flourish cloud services and applications as SaaS straight on the PaaS cloud. Difference between SaaS and PaaS is that SaaS only hosts completed cloud applications though PaaS gives a development platform that hosts both completed and in-progress cloud applications. So, for processing development infrastructure includes programming environment, tools, configuration management, PaaS in addition requires supporting application hosting environment. An example of PaaS is Google App/Engine.

III. Cloud Computing Deployment Models

The concept of cloud computing is described with number of deployment models, that has the appropriate trade offs for different agencies for the migration of services and for the operations to the cloud dependent environments. Below table defines the comparison of cloud computing deployment models by means of the scope and the security levels [5].

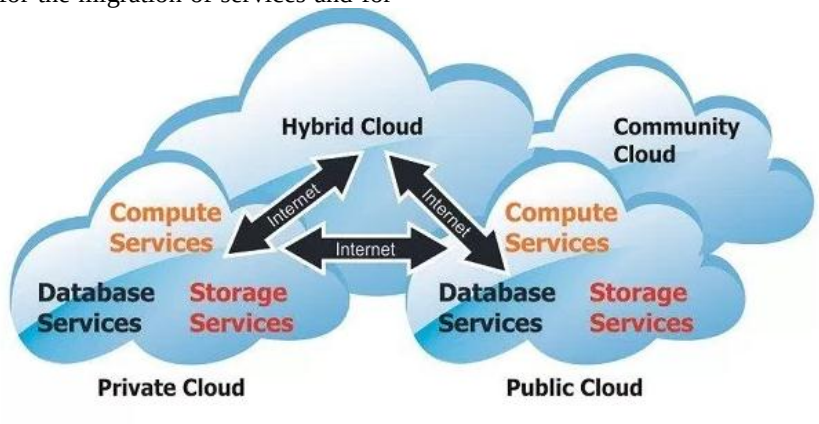

Figure 3: Cloud Deployment models

Table 2: Comparison of cloud deployment models [6]

Deployment Model	Scope of services	Managed by	Security level
Public model	General public and large industry groups	Cloud service provider	Low
Private model	Single organization	Single organization	High
Community model	Organization those share the same policy, mission and same security aspects	Several organization or Cloud service providers	High
Hybrid model	Organization and public	Organization and public	Medium

IV. Virtualization In Cloud

Virtualization technology has affected the modern data centers. As all the applications and operating systems are established as virtual machine images, and are executed by physical servers running on VMM (Virtual machine monitor) or hypervisor computer software that creates and run virtual machine) [7]. In order to fully understand the capabilities of cloud computing, cloud providers need to ensure that they can adapt to their virtual machine (VM) transport to meet different buyer prerequisites while keeping customers away from the basic data center. Virtualization applications dispense many comforts, including consolidation, migration. In this workspace, data centre has becomes a main hub of interchangeable computer resources which are leveraged to run the virtual machine images as needed. A general technique for enhancing data centre energy proficiency is to place VMs by co-coordinating the number of dynamic servers to meet the current needs of VMs and using SLA violations to place the remaining servers in low-control standby mode [8]. Cloud computing allows hosting of multiple services on a globally shared resource pool where resources are allocated to services on demand [4]. It uses virtualized environment for functioning services, because without

virtualization computing is inefficient and not flexible. But, it has some performance degradations of services and also has energy overheads and large amount of power consumption. In past, many researchers have worked on making energy efficient algorithm for reducing energy consumption. Many algorithms were implemented for saving energy of data centers by turning off or by putting idle servers to sleep mode of servers [9]. But these techniques were not so effective because of performance degradations of services and improper resources utilization. Some of the previous works also includes an idea for making energy efficient algorithm for data centers. The researchers has proposed Virtual machine placement algorithm that is minimization of migration (MM), which considers utilization of host CPU as per the list of virtual machines in decreasing order of CPU utilization. The performance of algorithm is better than other placement algorithms but they did not consider SLA parameters while selecting virtual machine for migration, which might be effected by live migration. Most of the violations occur during live migration of virtual machines, migration impacts the parameters of SLA (like availability, response time, throughput, network bandwidth etc.). So, there is a need to develop new approach for SLA aware energy efficient algorithm for resource allocation in data centers [7]. The idea of Virtual Machines (VMs) is connected to diminish energy utilization as it essentially decreases the rate of idle power in the general base [10].

V. Ras (Resource Allocation Strategy) In Cloud

RAS is the combination of cloud providers with the usage and allocation of scarce resources within the range of cloud environment to fulfill the cloud applications. It needs the measures and sorting of the assets needed by each application to complete the user's work. The designation times of the assets are moreover an ideal RAS contribution. A perfect RAS should stay away from the associated criteria that have to be taken after [11]:

As per Cloud supplier's point of view, it is unreasonable that client's anticipation dynamic way, application ways and clients requests are taken place. According to cloud clients, the requirements have to be finished as per the time given and with insignificant cost [12]. The physical assets usually are shared among different figures for provisioning and virtualization. The provisions fulfilled the demands with the mapping of virtualized assets for the physical ones. The programming and the equipments are allocated to cloud applications on the basis of the requests. The enhanced RAS has to shun away the below five situations:

Table 3: A perfect RAS should stay away from the associated criteria

Resource contention situation	Occurs when the applications executes the similar resources at the similar time.
Scarcity of resources	Arises when there are fewer resources.
Resource fragmentation situation	Occurs when there is an isolation of resources.
Over-provisioning of resources	Occurs when the application gets extra resources than needed.
Under-provisioning of resources	Arises when the application is allocated with less number of resources as needed.

Resource allocation is a mechanism that has been implemented in many computing regions, for example data centre management, operating systems, and grid computing [13]. Resource allocation encompasses scheduling of activities as well as allocating the accessible resources in a quite cost-effective way and also it implemented on optimal processes to proficiently assign physical and/or virtual resources to various applications of developers', therefore, by reducing the operational price of the cloud-environment. The consumption of energy associated with the resources allocation should be taken into account. Resource allocation is one of the fundamental technologies of cloud-computing domain, which utilizes the computing resources like bandwidth, energy, and delay and so on, in the network to facilitate the execution of cumbersome tasks that require large-scale computation.

Computers can be made as the device that consumes less energy by using lower power processor, using cooling devices as well as using spinning SSD (Solid State Drive) of small size rather than large size. Intel has developed a process that is going to use less power. It utilizes resources like data centers, heat, light, power etc. in an optimal manner. Indian IT industry has witnessed big growth in the world. However, this huge growth leads to the high consumption of energy as well as power that results in high cost consumption. According to statistics from IBM, it shows that less than 4 percent of the energy going into a data centre to process tasks.

Resource allocation is one of the challenges of cloud computing since end-users could easily access resources from anyplace and at any time. The resources presented in a cloud could not be demanded straightly but it could be opened using SOAP/Restful web APIs which map requests for storage or computations being plotted to virtualized

ICT resources (which are servers, blob storage, elastic IP, and so on.). Since, cloud data-center provides large quArtificiality of resources, the cloud computing model is capable of supporting on-demand allocation of elastic resource. On the other hand, such large quArtificiality also leads towards non-optimal resource allocation. [14]

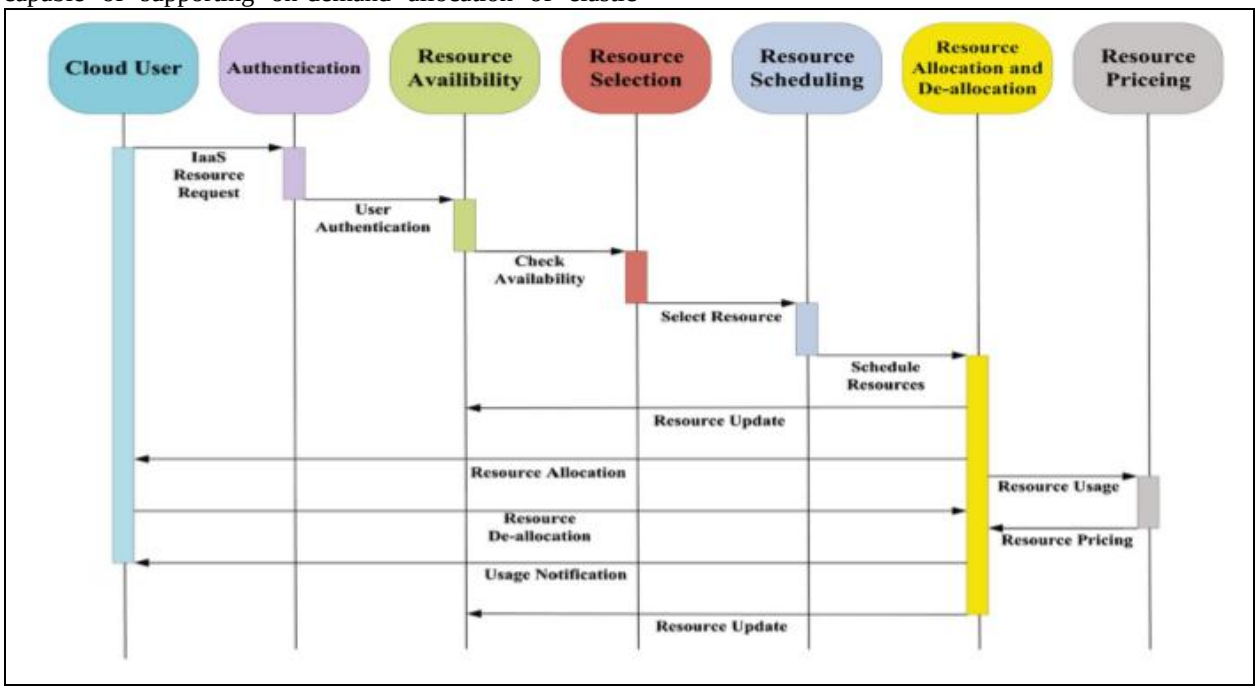

Figure 4: Resource allocation process

In cloud computing paradigm, the main challenge is the allocation of several accessible resources between various end-users having varying requests of resources dependent upon their patterns of application usage. The random as well as varying requests need to run on data-centre resources through Internet. The goal of resource allocation for any specific cloud provider could be either enhances the applications of Quality of Service or increase utilization of resource along with energy proficiency. The key objective is to augment Quality of Service parameters (i.e. response time) which measures the competence of resource-allocation irrespective of the category of ICT resources assigned to any specific end-users. The enhanced Quality of Service parameters could be any measure for example budget, communication delay, space, and time.

Figure 4 shows the process of resource allocation which is dependent on the accessible resources distribution for the necessary cloud application on the internet in the regular way. Furthermore, IaaS layer plays a significant role in the resource allocation with the support of pre-defined policies of resource allocation to the cloud users. Though, if the resources are not being allocated as per cloud user's need than the services would not last for more span of time. The clarification of the issue is to authorize the cloud providers for organizing the each module resources independently. So, the resource allocation is drawn as a segment of resource management and depicts an incredible character in resource allocation effectively and economically. The concept of resource allocation in IaaS is significant issue because of management and resource provisioning.

VI. 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 multi-tenancy therefore, developing a cost- efficient scalable environment. Yumin Wang et al. (2017) has considered

monitoring of metrology flowrate in the industrial piping system for the boiler in the case study and real time execution has been achieved by means of Labview and MyDaQ environment. The information w.r.t DAQ is executed into the cloud network by using subset reduction and regrouping on the basis of features by utilizing Fuzzy C-means clustering algorithm. The outcome has shown fast computation time having less complexity overhead by cloud distribution. Weidong Li et al. (2017) has taken the issue of multi resource fair allocation having bounded number of tasks. The author has proposed LMMNS (Lexicographically max min normalized share) mechanism which is a general DRF generalization and proposed a non-trivial enhanced algorithm for finding LMMNS fair allocation with linear running time by means of number of users. It has been proved that LMMNS has satisfied envy freeness and proof of group strategy and has analyzed the approximation ration of LMMNA by taking few assumptions. Abdul Hameed et al. (2016) ahs identifies the open challenges connected with energy efficient resource allocation. The authors has find out the power and energy inefficiencies in the hardware and software and classified the related methods with the pros and corns which is needed for identifying the general patterns and behaviors.

VII. Conclusion

Cloud computing has now came as a novel technology which has more potentials in market and enterprises. The concept of cloud has make this probable to access the applications and connected data from anywhere. The companies can rent the resources form the cloud storage for some computation purposes so that there should not be reduction of the infrastructure accordingly. The main pitfall in cloud computing is of resource allocation. As of the model uniqueness, the resource allocation is usually executed with an aim of reduction of cost linked with it. The another challenge in resource allocation is to fulfill the user's needs and application requirements. This paper has addressed the concept of cloud following the resource allocation strategy with the challenges.

References

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: 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 energy-efficient resource allocation algorithm based on immune clonal optimization for green cloud computing. *EURASIP Journal on Wireless Communications and Networking*, *2014*(1), 64.
13. Wang, Y., Li, J., & Wang, H. H. (2017). Cluster and cloud computing framework for scientific metrology in flow control. *Cluster Computing*, 1-10.
14. Li, W., Liu, X., Zhang, X., & Zhang, X. (2017, October). Multi-resource fair allocation with bounded number of tasks in cloud computing systems. In *National Conference of Theoretical Computer Science* (pp. 3-17). Springer, Singapore.
15. Hameed, A., Khoshkbarforoushha, A., Ranjan, R., Jayaraman, P. P., Kolodziej, J., Balaji, P., ... & Khan, S. U. (2016). A survey and taxonomy on energy efficient resource allocation techniques for cloud computing systems. *Computing*, *98*(7), 751-774.

