International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 6 ǁ 2022 ǁ PP. 1798-1803

RecentTechniques Resource Allocationin Cloud Computing: A Review

GUGULOTH LACHIRAM[1], BANOTHU SUNEETHA[2]
[1]*Assistant Professor,SBIT,Khammam,Telangana,India.*
[2]*Assistant Professor,Andhra Loyola Institute of Engineering & Technology, Vijayawada, India.*

Abstract
When it comes to the development and deployment of software, cloud computing has the potential to revolutionize several sectors. The dynamic nature of cloud computing's workload presents various issues for on-demand resource allocation and provisioning. Resources are allocated and scheduled cost-effectively in cloud computing using resource allocation. The following reasons contribute to the difficulty of resource management: Large, contemporary data centres with a variety of resource kinds and a fluctuating, unpredictable workload. This is the hallmark of the modern data centre. These services need a lot of resources from cloud service providers, and it's difficult for them to provide all of them to their customers. As a result, the purpose of this work is to describe the theoretical foundations that underlie the various resource allocation approaches that have previously been created.

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

Date of Submission: 14-06-2022 Date of acceptance: 29-06-2022

I. Introduction

Multiple companies and consumers may now utilize several apps without having to install and access their data on any portable device that has an online connection [1]. Centralized data storage, bandwidth, and processing are all made possible by this new technology, which also enhances computer skills. An artificial source of inexpensive and quick access to external information technologies has been made popular by cloud computing. It provides the researchers with a fresh perspective on how to make use of the computational resources available to them. More and more organizations (such as companies and research institutes) are reaping the advantages of cloud computing by putting their apps on the platform. Using virtualization, a cloud computing service may serve a wide range of heterogeneous computing needs using the same physical infrastructure. Computer resources, storage resources, and the resources of many programmers may be changed dynamically and freed later if no longer required. SLAs, or service level agreements, are used to guarantee that the quality of service provided to customers is up to snuff (QoS).

Certain companies can employ both private and public cloud resources to provide consumers with complete Quality of Service (QoS). Multiple programming patterns may be supported by cloud computing's flexible and service-based architecture [2]. Instead of focusing on a specific application like clustering or grid computing, cloud computing focuses instead on a service-oriented approach and on-demand virtualization type resources.

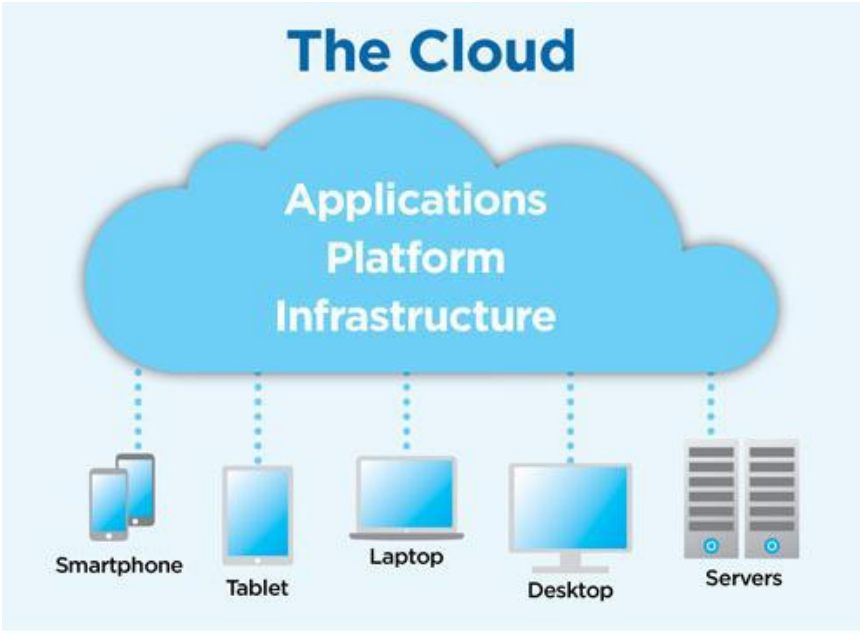

Figure 1: Cloud Architecture

II. Cloud Computing Services

Cloud computing has three service models: IaaS, PaaS, and SaaS. The rationale for each is provided below and shown in the graphic below [3]:

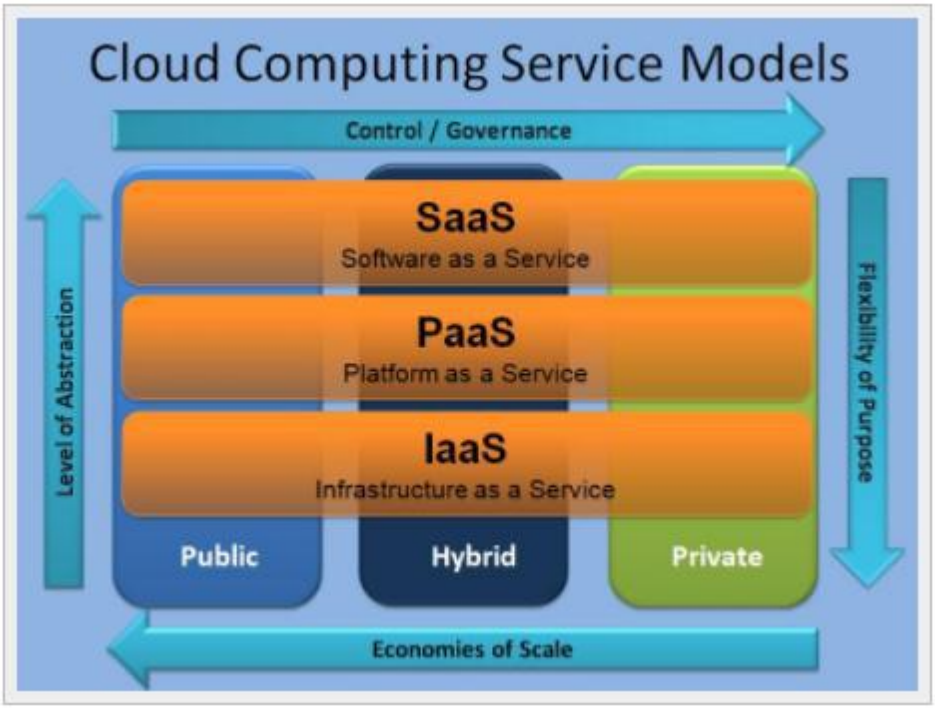

Figure 2: Service models of Cloud computing

2,1. Cloud Computing Deployment Models

According to the definition of cloud computing, there are a variety of deployment methods that may be used to migrate services and operations to the cloud-based environment [4]. Below is a comparison of the breadth and security levels of several cloud computing deployment methodologies [5].

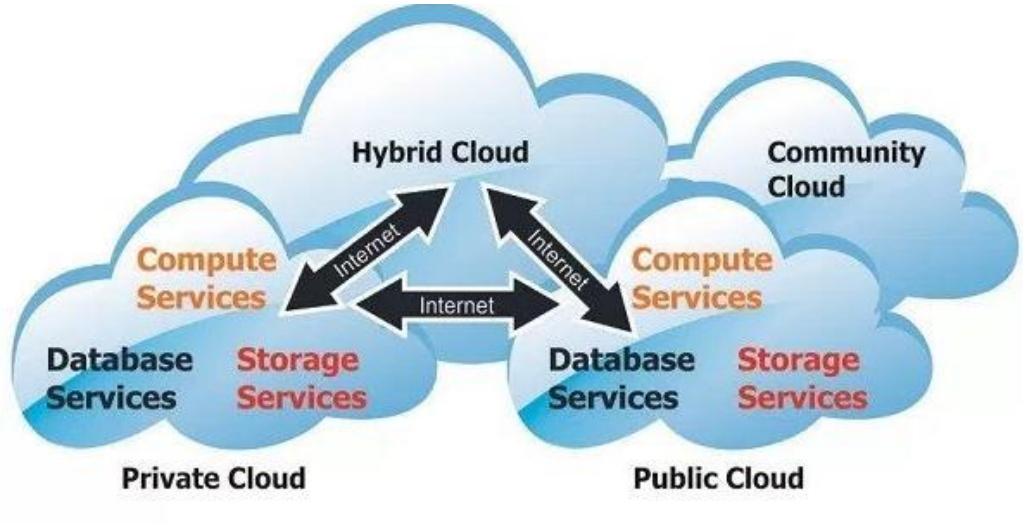

Figure 3: Cloud Deployment models

Table 1: 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	Many organizations or Cloud service providers	High
Hybrid model	Organization and public	Organization and public	Medium

III. Virtualization in Cloud

Today's data centres are being impacted by virtualization technologies. Virtual machine monitors (VMMs) and hypervisors (the software that generates and runs virtual machines) are used to run all of the software and operating systems on physical servers. Cloud computing suppliers must guarantee that they can change their virtual machine (VM) transport to satisfy diverse buyer requirements while keeping clients away from the fundamental data centre to fully appreciate the potential of cloud computing. Consolidation and migration are two of the numerous benefits that virtualization program provide. To execute the virtual machines, this workspace relies on a central hub of interchangeable computing resources, the data centre. VMs may be placed in low-control standby mode by leveraging SLA violations and a coordinated number of dynamic servers to suit current VM demands [8]. This approach can also be used to improve data centre energy efficiency. Cloud computing enables numerous services to be hosted on a globally shared resource pool, where resources are distributed to services as needed.

Because without virtualization, computing is wasteful and inflexible, it employs a virtualized environment to run services. Despite these drawbacks, it is nevertheless a viable option for a huge number of businesses. The development of an energy-efficient algorithm has been the focus of numerous academics in the past. It is possible to reduce the power consumption of data centres by shutting down or sleeping unused servers [9]. These methods, however, were shown to be ineffective due to service degradations and inefficient use of available resources. An energy-efficient algorithm for data centres has also been discussed in the past. Minimization of migration (MM) is a virtual machine placement strategy that takes into account host CPU use based on a list of virtual machines ordered by decreasing CPU usage. When picking virtual machines for migration, they did not take SLA settings into account, which might have an impact on the live migration process. More than a third of SLA breaches are caused by live migrations of virtual machines (like availability, response time, throughput, network bandwidth etc.). A novel technique for SLA aware and energy-efficient resource allocation in data centres must be developed. The concept of virtual machines (VMs) is linked to reduced energy consumption since it reduces the general base's rate of idle power. [10].

3.1. RAS (Resource Allocation Strategy) In Cloud

To meet the needs of cloud applications, RAS brings together cloud service providers and makes use of and distributes limited resources across a wide variety of cloud environments. Each application's assets must be measured and sorted to accomplish the user's task. A further RAS contributor is the asset designation times. A perfect RAS should not violate any of the following requirements [11]:

From the perspective of the Cloud provider, it is illogical to take into account client expectations, application methods, and requirements. Cloud users expect their requests to be completed on time and at a minimal cost [12]. Typically, the physical assets are shared across many entities for provisioning and virtualization purposes. With the mapping of virtualized assets to physical ones, the provisions met the needs. Requests determine how much code and hardware are allotted to certain cloud apps. The following five circumstances must be avoided by the improved RAS.

The kind and quantity of resources required by each program to fulfil a user task are input into the Resource Allocation Strategy (RAS). When it comes to allocating resources, there are two main groups to consider.

1. Static Resource Allocation Strategies
2. Dynamic Resource Allocation Strategies.

Instead of allocating resources in real-time, static resource allocation solutions do it when jobs are still offline. The user's allocation of resources remains constant throughout the application. A predetermined collection of applications, machines, and machine characteristics are sent into the system as an input source.

At runtime, scheduling and resource allocation choices are based on the platform's state, the length of various tasks on various resources available, and the machine's state in dynamic resource allocation algorithms. Resource allocation based on the quality of service (QoS) considers factors such as customer happiness, cost-effectiveness, and the most efficient use of available resources. Due to the inclusion of numerous QoS criteria such as CPU memory, speed, and stability, Cloud computing's resource allocation differs from the conventional distributed computing environment. Energy efficiency is an important consideration when allocating resources. Increased energy savings are achieved in the internal cloud data centre using rescheduled operations. An application's resource allocation is determined by how much work it does and how many times it rearranges the resources of old and new servers it uses.

Table 2 describes the present cloud computing resource allocation approaches, their benefits, and the future work that will be done with these techniques.

Table 2: Comparison of the existing Resource Allocation Techniques in Cloud Computing

Author Name	Technique Used	Parameters Used	Advantages	Future Work
Javier Espadas, Arturo Molina, Guillermo Jiménez, Martín Molina, Raúl Ramírez, David Concha	Tenant-based [13]	Number of VM instances; Workload pikes (increment and peak based); VM capacity	Average Underutilized is statistically improved	Can be performed on other different scenarios such as HPC
Yanbing Liu , Shasha Yang, Qingguo Lin, Gyoung-Bae Kim	Loyaltybased [14]	Transaction rate	Improve the successful transaction rate of the system under the environment of cloud computing	Can be performed taking other scenarios
Gihun Jung, Kwang Mong Sim	Agent-based [15]	The average number of visiting locations for a request; average geographical distance after allocation; average allocation time for a request; success rate; the average number of request denial for the success rate	It has high performance in terms of average allocation time and geographical distance. The model has a fast allocation time.	Work can be done on different sizes of memory spaces, live migration and dynamic workloads.
Hemant Kumar Mehta and Eshan Gupta	Economy-based [16]	Profit, amount of accepted lease, execution cost of the lease	Reduces the cost of execution of the consumer's lease and increases the profit of the provider to a considerable extent	Can be performed considering other scenarios
Rajkamal Kaur Grewal & Pushpendra Kumar Pateriya	Rule-based [17]	Resource utilization	Effective increase in resource utilization and implemented in IaaS	This technique can be used on other services
Dorian Minarolli and Bernd Freisleben	Utility-based [18]	Demand for the webserver in VM; the number of active nodes; VM web server response time; global system utility	Effective cost reduction while improving global system utility using VM migration with better performance	Reliability cost can be included and inclusion of memory, disk and network in a distributed environment

Anna Schwanengel, Alexander Schönl, and Claudia Linnhoff-Popien	Locationbased [19]	Response time, Transfer Time, Processing Time	Reduction of the latency with reducing cost and high user satisfaction	Including the designed control entity for adding and releasing instances in the cloud
Gunho Lee, Niraj Tolia	Topologybased [20]	Job completion t efficiency	Reduce Job completion time up to 59% than simple techniques	Increase the objective functions by involving infrastructure costs, power and reliability
Rerngvit Yanggratoke, Fetahi Wuhib and Rolf Stadler,	Gossip-based [21]	Scalability	Minimizing power consumption through server consolidation when the system is in underload and fair resource allocation in case of overload	Determination of the convergence rate of GRMP-Q and its dependence on CPU and memory demands and to implement in heterogeneous environment making protocol robust from machine failures.
Narander Kumar, Swati Saxena	Preference-based [22]	Actual payment of a winner with his bestprice payment, variation in utilities earned by winners, Revenues Earned	High-Performance benefit in revenues to the service provider and payments of cloud users besides ensuring optimum resources use.	Including an energy-efficient scheduling strategy to allot auction winners' tasks to suitable VMs

IV. Conclusion

This article analyses and discusses the benefits and potential future work of the different resource allocation approaches that are currently available. Resource allocation in cloud computing is becoming more crucial due to a growth in the number of requests from cloud users, and as a result, an effective technique is needed to satisfy those customers. Cloud service companies must have a Resource Allocation Strategy in place if they want to satisfy their customers and make a profit. A customer's application, services, and infrastructure influence the resource allocation method. Cloud computing resource allocation approaches, their benefits, and what's to come are summarised here.

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 ofcloud 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). Modelling 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]. J. Espadas, A. Molina, G. Jiménez, M. Molina,R. Ramírez, D. Concha, "A tenant-basedresource allocation model for scaling Software-as-a-Service applications over cloud computinginfrastructures", Future Generation ComputerSystems, Elsevier, Vol. 29, pp. 273–286, 2013.

[14]. Y. Liu, S. Yang, Qingguo Lin, Gyoung-BaeKim, "Loyalty-Based Resource AllocationMechanism in Cloud Computing", RecentAdvances in Computer Science and InformationEngineering, Vol. **2**, pp. 233-238, 2012.

[15]. G. Jung, K. Mong Sim, "Agent-based AdaptiveResource Allocation on the Cloud ComputingEnvironment", in the Proceedings ofInternational Conference on Parallel ProcessingWorkshops, IEEE, pp. 345-351, 2011.

[16]. H. Kumar Mehta and E. Gupta, "EconomyBased Resource Allocation in IaaS Cloud",International Journal of Cloud Applications andComputing (IJCAC), Vol. **3** (2), pp. 1-11, 2013.

[17]. R. Kaur Grewal & P. Kumar Pateriya, "A Rule-basedApproach for Effective ResourceProvisioning in Hybrid Cloud Environment",International Journal of Computer Science andInformatics, Vol. **1** (4), pp. 101-106, 2012.

[18]. D. Minarolli and B. Freisleben, "Utility-basedResource Allocation for Virtual Machines inCloud Computing", IEEE Symposium onComputers and Communications (ISCC), pp.410-417, 2011.

[19]. A. Schwanengel, A. Schönl, and C. Linnhoff-Popien, "Location-Based Cloud ResourceAllocation Based on Information of the SocialWeb", International Journal of Advanced CloudComputing and Applied Research, Vol. **1** (1),pp. 1-12, 2013.

[20]. G. Lee, N. Tolia, "Topology-Aware ResourceAllocation for Data-Intensive Workloads", inthe Proceedings of the First ACM Asia-pacificworkshop on Workshop on Systems (APSys '10),ACM, pp. 1-6, 2010.
[21]. R. Yanggratoke, F. Wuhib and R. Stadler,"Gossip-based Resource Allocation for GreenComputing in Large Clouds ", in theProceedings of 7th International Conference onNetwork and Service Management, pp. 1-9,2011.