



A REVIEW ON TASK ALLOCATION PERFORMANCE ISSUES IN CLOUD ENVIRONMENT

Kavita Redishettywar ⁽¹⁾, Prof. Rafik Juber Thekiya ⁽²⁾

⁽¹⁾ Research Scholar, Department of Computer Science and Engineering, Matoshri Pratishthan Group of Institutions, Khupsarwadi, Nanded

kavita.kotalwar@gmail.com

⁽²⁾ Assistant Professor, Department of Computer Science and Engineering, Matoshri Pratishthan Group of Institutions, Khupsarwadi, Nanded

rafikthekiya@gmail.com

ABSTRACT

Cloud computing is an emerging paradigm in the computer industry where the computing is moved to a cloud of computers. It has become one of the buzz words of the industry. The core concept of cloud computing is, quite simply, that the vast computing resources that we need will reside somewhere out there in the cloud of computers and we'll connect to them and use them as and when needed. Computing can be described as any activity of using and/or developing computer hardware and software. It includes everything that sits in the bottom layer, i.e. everything from raw compute power to storage capabilities. Cloud computing ties together all these entities and delivers them as a single integrated entity under its own sophisticated management. Load balancing is a mechanism that distributes the dynamic workload equally across over the nodes or virtual machines within the whole cloud server to avoid a state of conflict wherever some virtual machines are measured as heavily loaded whereas others nodes or hosts are measured as idle or doing very little work. It helps to realize a high client satisfaction and resource utilization magnitude relation, consequently increasing the performance and resource utility of the system. It additionally makes sure that each computing resource in the cloud server is distributed with efficiently and fairly among all the requests of the client. It additionally prevents bottlenecks of the system which can occur because of load imbalance.

Keywords

Cloud Computing, IAAS, SAAS, PAAS, Load Balancing, Virtual Machine.

INTRODUCTION TO CLOUD COMPUTING

Cloud Computing has become one of the most talked about technologies in recent times and has got lots of attention from media as well as analysts because it is offering lots of opportunities. Enterprises have been determined to reduce computing costs and for that reason most of them started using it in IT technology then adapted virtualization technology. For the good of the enterprises it is futuristic to help them in this i.e. Cloud Computing. Cloud Computing has taken the enterprise to new level and allows them to further reduce costs through improved utilization, reduced administration and infrastructure cost and faster deployment cycles. Cloud computing has become an established paradigm for running services on external infrastructure, where virtually unlimited capacity can be dynamically allocated to suit the current needs of customers and where new instances of a service can be deployed within a short time frame. Although the term Cloud computing has come to include several kinds of technologies offering remote execution and service management, it is used in this paper to denote scalable elastic data center infrastructures offering dynamic and cost-efficient service provisioning. There are many different Cloud computing solutions available, such as Amazon Elastic Compute Cloud [1]. However, different Cloud computing solutions are rarely compatible with each other and this creates a kind of vendor lock-in which is not only limiting to the customer, but also limits the potential of Cloud computing as a whole since separate Cloud computing solutions are unable to interoperate. Grid computing can be seen as one of several predecessors to Cloud computing. Grid computing is often about making large computations using large amounts of resources, while as Cloud computing is more about making large amounts of resources available to many different applications over a longer period of time. Clouds leverage modern technologies such as virtualization to provide the infrastructure needed to deploy services as utilities. Still, Cloud computing and Grid computing share a lot of the underlying technology and many concepts from Grid computing can be modified and made suitable for Cloud computing as well. Cloud computing is Internet ("CLOUD") based development and use of computer technology ("COMPUTING"). Cloud computing is a general term for anything that involves delivering hosted services over the Internet. It is used to describe both a platform and type of application. These cloud applications use large data centers and powerful servers that host Web applications and Web services. Anyone with a suitable Internet connection and a standard browser can access a cloud application. [1]

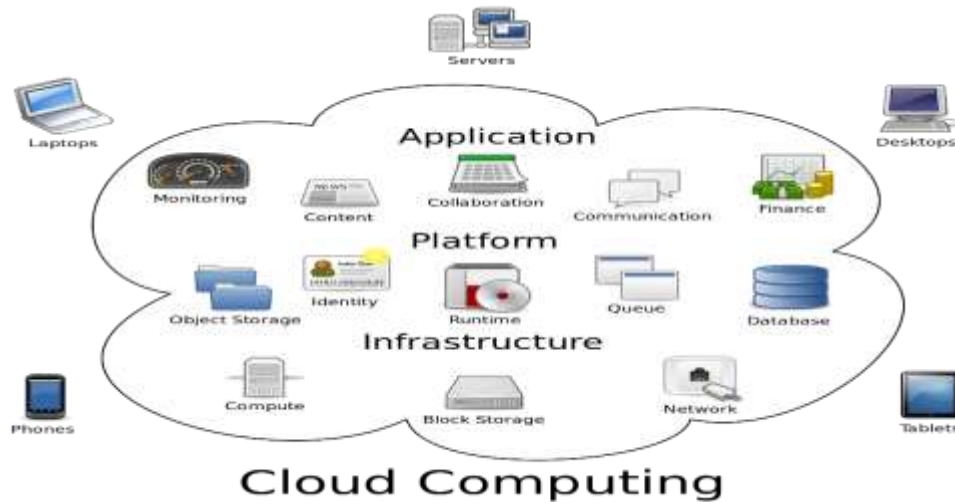


Figure 1. Cloud Computing Model

BENEFITS OF CLOUD COMPUTING

Many of the benefits to be had when using Cloud Computing are the lower costs associated. At the infrastructure level, virtual images can be scaled and contracted with complete disregard for any associated hardware costs such as equipment procurement, storage, maintenance and use. This is all taken care of by the service provider and will be factored into the payment for the service: capital expenditure has been converted into operational expenditure. Resources within the cloud can be treated as a commodity, an 'unlimited' medium. At both the platform and software level similar benefits are seen.

- Cloud technology is paid incrementally, saving organizations money.
- Organizations can store more data than on private computer systems.
- No longer do IT personnel need to worry about keeping software up to date.
- Cloud computing offers much more flexibility than past computing methods.
- Employees can access information wherever they are, rather than having to remain at their desks.
- No longer having to worry about constant server updates and other computing issues, government organizations will be free to concentrate on innovation.
- Decoupling and separation of the business service from the infrastructure needed to run it.
- Flexibility to choose multiple vendors that provide reliable and scalable business services, development environments, and infrastructure that can be leveraged out of the box and billed on a metered basis—with no long term contracts.

SERVICE MODELS

There are different types of services are provides by cloud models like: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) [6] which are deployed as public cloud, private cloud, community cloud and hybrid clouds.

1) Software as a Service (SaaS):- The capability provided to the consumer is to use the some applications which is running on a cloud infrastructure. The applications are accessible from many devices through an interface such as a web browser (e.g., web-based email). The consumer does not control the cloud infrastructure which includes network, and servers, all operating systems, and provides storages. Software-as-a-Service (SaaS) is the broadest market. In this case the provider allows the customer only to use its applications. The software interacts with the user through a user interface. These applications can be anything from web based email, to applications like Twitter or Last.fm. This is the idea that someone can offer you a hosted set of software (running on a platform and infrastructure) that you don't own but pay for some element of utilization - by the user, or some other kind of consumption basis. You don't have to do any development or programming, but you may need to come in and configure the (very flexible, configurable and sometimes customizable) software. You don't have to purchase anything. You just pay for what you use. A SaaS provider typically hosts and manages a given application in their own data center and makes it available to multiple tenants and users over the Web. Some SaaS providers run on another cloud provider's PaaS or IaaS service offerings. Oracle CRM on Demand, Salesforce.com, and Netsuite are some of the well-known SaaS examples.

2) Platform as a Service (PaaS):- Platform-as-a-Service (PaaS) is a set of software and development tools hosted on the provider's servers. Google Apps is one of the most famous Platform-as-a-Service providers. This is the idea that someone can provide the hardware (as in IaaS) plus a certain amount of application software - such as integration into a common set

of programming functions or databases as a foundation upon which you can build your application. Platform as a Service (PaaS) is an application development and deployment platform delivered as a service to developers over the Web. It facilitates development and deployment of applications without the cost and complexity of buying and managing the underlying infrastructure, providing all of the facilities required to support the complete life cycle of building and delivering web applications and services entirely. PaaS [5] provides all the resources that are required for implementation of applications and all services completely from the Internet. In this no downloading or installing is required of any software. The capability provided to the consumer is to deploy onto the cloud infrastructure. Consumer uses all the applications by using different programming languages and tools which are provide by the provider. Any consumer has not any control on cloud infrastructure including all networks, servers and operating systems, but has control over the applications which they deployed.

3) Infrastructure as a Service (IaaS):- Infrastructure-as-a-Service (IaaS) provides virtual servers with unique IP addresses and blocks of storage on demand. Customers can pay for exactly the amount of service they use, like for electricity or water, this service is also called utility computing. The capability provided to the consumer is to access all the processing, storage, networks and other many fundamental computing resources. Consumer [5][6] is able to deploy arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed application, and possibly limited control of select networking components.

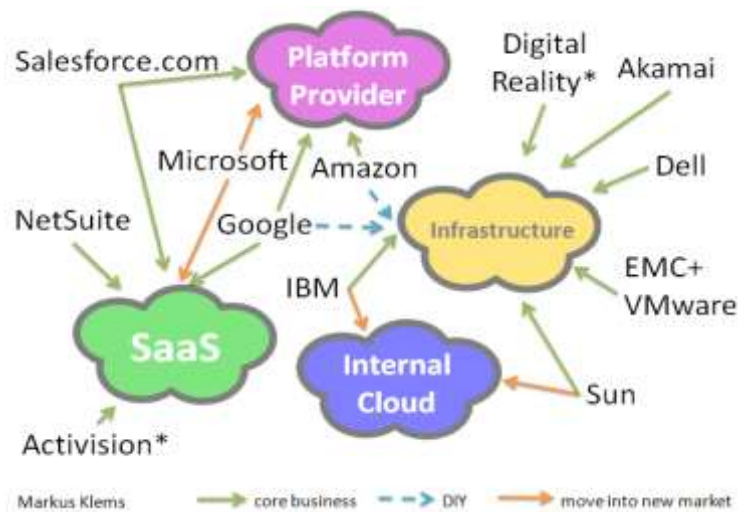


Figure 2. Service Models

DEPLOYMENT MODELS

Depending on infrastructure ownership, there are four deployment models of cloud computing [6].

1) Public Cloud: - Public cloud [9] allows users to access the cloud publicly. It is access by interfaces using internet browsers. Users pay only for that time duration in which they use the service, i.e., pay-per-use. In Public cloud or external cloud resources are dynamically provisioned on a fine-grained, self-service basis over the Internet, via web applications/web services, from an off-site third-party provider who shares resources and bills on a fine-grained utility computing basis.

2) Private Cloud: - A private clouds [10] operation is with in an organization's internal enterprise data center. The main advantage here is that it is very easier to manage security in public cloud. Example of private cloud in our daily life is in tranet. Private cloud and internal cloud products claim to "deliver some benefits of cloud computing without the pitfalls", capitalizing on data security, corporate governance, and reliability concerns. They have been criticized on the basis that users "s till have to buy, build, and manage them and as such do not benefit from lower upfront capital costs and less hands-on management, essentially.

3) Hybrid Cloud: - It is a combination of public cloud [11] and private cloud. .It provide more secure way to control all data and applications .It allows the party to access information over the internet. It allows the organization to serve its needs in the private cloud and if some occasional need occurs it asks the public cloud for some computing resources. A hybrid cloud environment consisting of multiple internal and/or external providers. It can also describe configurations combining virtual and physical, collocated assets

4) Community Cloud: - When cloud infrastructure construct by many organizations jointly, such cloud model is called as a community cloud. The cloud infrastructure could be hosted by a third-party provider or within one of the organizations in the community.

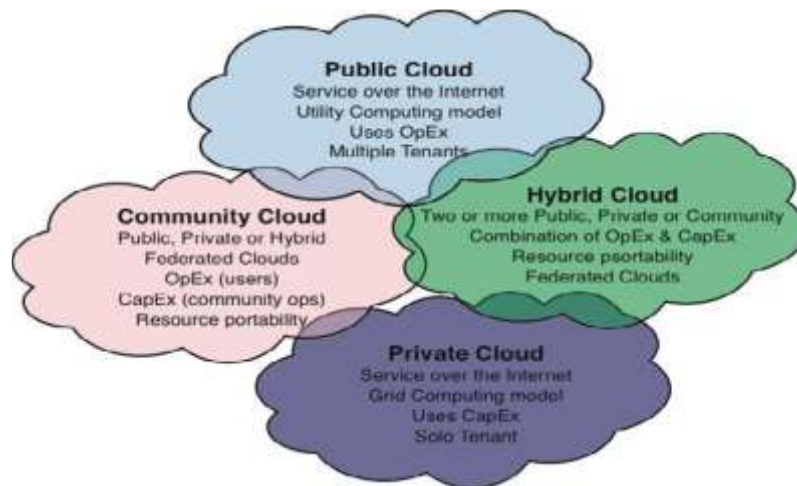


Figure 3. Cloud Service Models and Providers

LOAD BALANCING

It is the mechanism of distributing the load among various nodes of a distributed system to improve both resource utilization and job response time while also avoiding a situation where some of the nodes are heavily loaded while other nodes are idle or doing very little work. It also ensures that all the processor in the system or every node in the network does approximately the equal amount of work at any instant of time. Load Balancing [5] is done with the help of load balancers where each incoming request is redirected and is transparent to client who makes the request. Based on predetermined parameters, such as availability or current load, the load balancer uses various scheduling algorithm to determine which server should handle and forwards the request on to the selected server. To make the final determination, the load balancer retrieves information about the candidate server's health and current workload in order to verify its ability to respond to that request. Load balancing solutions can be divided into software-based load balancers and hardware-based load balancers. Hardware-based load balancers are specialized boxes that include Application Specific Integrated Circuits (ASICs) [32] customized for a specific use. They have the ability to handle the high speed network traffic whereas Software-based load balancers run on standard operating systems and standard hardware components.

LOAD BALANCING ALGORITHMS

In order to balance the requests of the resources it is important to recognize a few major goals of load balancing algorithms:

- Cost effectiveness:** primary aim is to achieve an overall improvement in system performance at a reasonable cost.
- Scalability and flexibility:** the distributed system in which the algorithm is implemented may change in size or topology. So the algorithm must be scalable and flexible enough to allow such changes to be handled easily.
- Priority:** prioritization of the resources or jobs need to be done on beforehand through the algorithm itself for better service to the important or high prioritized jobs in spite of equal service provision for all the jobs regardless of their origin.

Following load balancing algorithms are currently prevalent in clouds:-

OLB: Opportunistic Load Balancing (OLB) assigns each task, in arbitrary order, to the next machine that is expected to be available, regardless of the task's expected execution time on that machine. The intuition behind OLB is to keep all machines as busy as possible [8, 9].

MET: In contrast to OLB, Minimum Execution Time (MET) assigns each task, in arbitrary order, to the machine with the best expected execution time for that task, regardless of that machine's availability. The motivation behind MET is to give each task to its best machine. This can cause a severe load imbalance across machines.

MCT: Minimum Completion Time (MCT) assigns each task, in arbitrary order, to the machine with the minimum expected completion time for that task. This causes some tasks to be assigned to machines that do not have the minimum execution time for them [1].

Min-min: Min-min heuristic uses minimum completion time (MCT) as a metric, meaning that the task which can be completed the earliest is given priority. This heuristic begins with the set U of all unmapped tasks. Then the set of minimum completion times (M), is found.

Max-Min: The Max-min heuristic is very similar to Min-min and its metric is MCT too. It begins with the set U of all unmapped tasks. Then, the set of minimum completion times (M) is found as mentioned in previous section. Next, the task with the overall maximum completion time from M is selected and assigned to the corresponding machine and the workload of the selected machine will be updated. And finally the newly mapped task is removed from U and the process repeats until all tasks are mapped [1, 9].



LJFR-SJFR: LJFR-SJFR heuristic begins with the set U of all unmapped tasks. Then the set of minimum completion times is found the same as Min-min. Next, the task with the overall minimum completion time from M is considered as the shortest job in the fastest resource (SJFR). Also the task with the overall maximum completion time from M is considered as the longest job in the fastest resource (LJFR). At the beginning, this method assigns the m longest tasks to the m available fastest resources (LJFR). Then this method assigns the shortest task to the fastest resource, and the longest task to the fastest resource alternatively [4, 11].

Sufferage: In this heuristic for each task, the minimum and second minimum completion time are found in the first step. The difference between these two values is defined as the sufferage value. In the second step, the task with the maximum sufferage value is assigned to the corresponding machine with minimum completion time [4, 12].

RELATED WORK

Nguyen Khac Chien et al. (2016) has proposed a load balancing algorithm which is used to enhance the performance of the cloud environment based on the method of estimating the end of service time. They have succeeded in enhancing the service time and response time of the user. Ankit Kumar et al (2016) focuses on the load balancing algorithm which distributes the incoming jobs among VMs optimally in cloud data centers. The proposed algorithm in this research work has been implemented using Cloud Analyst simulator and the performance of the proposed algorithm is compared with the three algorithms which are preexists on the basis of response time. In the cloud computing milieu, the cloud data centres and the users of the cloud-computing are globally situated, therefore it is a big challenge for cloud data centres to efficiently handle the requests which are coming from millions of users and service them in an efficient manner. S.Yakhchi et al. (2015) discusses that the energy consumption has become a major challenge in cloud computing infrastructures. They proposed a novel power aware load balancing method, named ICAMMT to manage power consumption in cloud computing data centres. We have exploited the Imperialism Competitive Algorithm (ICA) for detecting over utilized hosts and then we migrate one or several virtual machines of these hosts to the other hosts to decrease their utilization. Finally, we consider other hosts as underutilized host and if it is possible, migrate all of their VMs to the other hosts and switch them to the sleep mode. Surbhi Kapoor et al. (2015) aims at achieving high user satisfaction by minimizing response time of the tasks and improving resource utilization through even and fair allocation of cloud resources. The traditional Throttled load balancing algorithm is a good approach for load balancing in cloud computing as it distributes the incoming jobs evenly among the VMs. But the major drawback is that this algorithm works well for environments with homogeneous VMS, does not considers the resource specific demands of the tasks and has additional overhead of scanning the entire list of VMs every time a task comes. The issues have been addressed by proposing an algorithm Cluster based load balancing which works well in heterogeneous nodes environment, considers resource specific demands of the tasks and reduces scanning overhead by dividing the machines into clusters.

Shikha Garg et al. (2015) aims to distribute workload among multiple cloud systems or nodes to get better resource utilization. It is the prominent means to achieve efficient resource sharing and utilization. Load balancing has become a challenge issue now in cloud computing systems. To meet the user's huge number of demands, there is a need of distributed solution because practically it is not always possible or cost efficient to handle one or more idle services. Servers cannot be assigned to particular clients individually. Cloud Computing comprises of a large network and components that are present throughout a wide area. Hence, there is a need of load balancing on its different servers or virtual machines. They have proposed an algorithm that focuses on load balancing to reduce the situation of overload or under load on virtual machines that leads to improve the performance of cloud substantially. Reena Panwar et al. (2015) describes that the cloud computing has become essential buzzword in the Information Technology and is a next stage the evolution of Internet, The Load balancing problem of cloud computing is an important problem and critical component adequate operations in cloud computing system and it can also prevent the rapid development of cloud computing. Many clients from all around the world are demanding the various services rapid rate in the recent time. Although various load balancing algorithms have been designed that are efficient in request allocation by the selection of correct virtual machines. A dynamic load management algorithm has been proposed for distribution of the entire incoming request among the virtual machines effectively. Mohamed Belkhouraf et al. (2015) aims to deliver different services for users, such as infrastructure, platform or software with a reasonable and more and more decreasing cost for the clients. To achieve those goals, some matters have to be addressed, mainly using the available resources in an effective way in order to improve the overall performance, while taking into consideration the security and the availability sides of the cloud. Hence, one of the most studied aspects by researchers is load balancing in cloud computing especially for the big distributed cloud systems that deal with many clients and big amounts of data and requests. The proposed approach mainly ensures a better overall performance with efficient load balancing, the continuous availability and a security aspect.

Lu Kang et al. (2015) improves the weighted least connections scheduling algorithm, and designs the Adaptive Scheduling Algorithm Based on Minimum Traffic (ASAMT). ASAMT conducts the real-time minimum loads scheduling to the node service requests and configures the available idle resources in advance to ensure the service QoS requirements. Being adopted for simulation of the traffic scheduling algorithm, OPNET is applied to the cloud computing architecture.

Hiren H. Bhatt et al. (2015) presents a Flexible load sharing algorithm (FLS) which introduce the third function. The third function makes partition the system in to domain. This function is helpful for the selection of other nodes which are present in the same domain. By applying the flexible load sharing to the particular domains in to the distribute system, the performance can be improved when any node is in overloaded situation. Shanti Swaroop Moharana et al. (2015) specifies that the Cloud Computing is high utility software having the ability to change the IT software industry and making the software even more attractive. It has also changed the way IT companies used to buy and design hardware. The elasticity of resources without paying a premium for large scale is unprecedented in the history of IT industry. The increase in web traffic and

different services are increasing day by day making load balancing a big research topic. Cloud computing is a new technology which uses virtual machine instead of physical machine to host, store and network the different components. Saraswathi AT et al (2015) focuses on allocation of VM to the user, based on analyzing the characteristics of the job. Main principle of this work is that low priority jobs (deadline of the job is high) should not delay the execution of high priority jobs (deadline of the job is low) and to dynamically allocate VM resources for a user job within deadline. It builds upon advances of virtualization and distributed computing to support cost efficient usage of computing resources, emphasizing on resource scalability and on-demand services. It allows business outcomes to scale up and down their resources based on needs. Managing the customer demand creates the challenges of on demand resource allocation.

Rajwinder Kaur et al. (2013) discusses the main challenges in cloud computing which is required to distribute the dynamic workload across multiple nodes to ensure that no single node is overwhelmed. It helps in optimal utilization of resources and hence in enhancing the performance of the system. A few existing scheduling algorithms can maintain load balancing and provide better strategies through efficient job scheduling and resource allocation techniques as well. In computing, the load may be of various types like memory load, CPU load and network load etc. Load balancing is the process of searching overloaded node and transferring the extra load of the overloaded node to other nodes. Amandeep Kaur Sidhu et al. (2013) aims to share data, calculations, and service transparently over a scalable network of nodes. Since Cloud computing stores the data and disseminated resources in the open environment. So, the amount of data storage increases quickly. In the cloud storage, load balancing is a key issue. It would consume a lot of cost to maintain load information, since the system is too huge to timely disperse load. O. M. Elzeki et al. (2012) proposed a unique modification of Max-min algorithm. The algorithm is built based on comprehensive study of the impact of RASA algorithm in scheduling tasks and the atom concept of Max-min strategy. An Improved version of Max-min algorithm is proposed to outperform scheduling map at least similar to RASA map in total complete time for submitted jobs. Improved Max-min is based on the expected execution time instead of complete time as a selection basis. Kumar Nishant et al. (2012) proposed an algorithm for load distribution of workloads among nodes of a cloud by the use of Ant Colony Optimization (ACO). This is a modified approach of ant colony optimization that has been applied from the perspective of cloud or grid network systems with the main aim of load balancing of nodes. This modified algorithm has an edge over the original approach in which each ant build their own individual result set and it is later on built into a complete solution. The system, which is incurring a cost for the user should function smoothly and should have algorithms that can continue the proper system functioning even at peak usage hours.

Ms. Parin. V. Patel et al. (2012) has discussed load Balancing approach. Many types of load concern with cloud like memory load, CPU load and network load. Load balancing is the process of distributing load over the different nodes which provides good resource utilization when nodes are overloaded with job. Load balancing has to handle the load when one node is overloaded. When node is overloaded at that time load is distributed over the other ideal nodes. Many algorithms are available for load balancing like Static load balancing and Dynamic load balancing. Yang Xu et al. (2011) put forwards a novel model to balance data distribution to improve cloud computing performance in data-intensive applications, such as distributed data mining. By extending the classic Map Reduce model with an agent-aid layer and abstracting working load requests for data blocks as tokens, the agents can reason from previously received tokens about where to send other tokens in order to balance the working tasks and improve system performance. Our key contribution lies in building an efficient token routing algorithm in spite of agents' unknowing to the global state of data distribution in cloud.

RESEARCH GAP

Cloud computing thus involving distributed technologies to satisfy a variety of applications and user needs. Sharing resources, software, information via internet are the main functions of cloud computing with an objective to reduced capital and operational cost, better performance in terms of response time and data processing time, maintain the system stability and to accommodate future modification in the system. So there are various technical challenges that needs to be addressed like Virtual machine migration, server consolidation, fault tolerance, high availability and scalability but central issue is the load balancing, it is the mechanism of distributing the load among various nodes of a distributed system to improve both resource utilization and job response time while also avoiding a situation where some of the nodes are heavily loaded while other nodes are idle or doing very little work. It also ensures that all the processor in the system or every node in the network does approximately the equal amount of work at any instant of time. To make the final determination, the load balancer retrieves information about the candidate server's health and current workload in order to verify its ability to respond to that request. Load balancing solutions can be divided into software-based load balancers and hardware-based load balancers. Hardware-based load balancers are specialized boxes that include Application Specific Integrated Circuits (ASICs) customized for a specific use. They have the ability to handle the high-speed network traffic whereas Software-based load balancers run on standard operating systems and standard hardware components. Various problems have been identified in the current work and are listed below:

- There is no sorting mechanism defined on to clusters.
- Overhead of scanning the VM's in each cluster.
- Cloudlets are assigned one by one to each VM, so lot of time is consumed here.
- There is no priority mechanism defined on the cloudlets.

CLOUD SIM

The CloudSim simulation layer provides support for modeling and simulation of virtualized Cloud-based data center environments including dedicated management interfaces for VMs, memory, storage, and bandwidth. The fundamental

issues, such as provisioning of hosts to VMs, managing application execution, and monitoring dynamic system state, are handled by this layer. A Cloud provider, who wants to study the efficiency of different policies in allocating its hosts to VMs (VM provisioning), would need to implement his strategies at this layer. Such implementation can be done by programmatically extending the core VM provisioning functionality. There is a clear distinction at this layer related to provisioning of hosts to VMs. A Cloud host can be concurrently allocated to a set of VMs that execute applications based on SaaS provider's defined QoS levels. This layer also exposes the functionalities that a Cloud application developer can extend to perform complex workload profiling and application performance study. The top-most layer in the CloudSim stack is the User Code that exposes basic entities for hosts (number of machines, their specification, and so on), applications (number of tasks and their requirements), VMs, number of users and their application types, and broker scheduling policies. By extending the basic entities given at this layer, a Cloud application developer can perform the following activities: (i) generate a mix of workload request distributions, application configurations; (ii) model Cloud availability scenarios and perform robust tests based on the custom configurations; and (iii) implement custom application provisioning techniques for clouds and their federation.

CONCLUSION

This paper is based on cloud computing technology which has a very vast potential and is still unexplored. The capabilities of cloud computing are endless. Cloud computing provides everything to the user as a service which includes platform as a service, application as a service, infrastructure as a service. Load balancing is dividing work load between a set of computers in order to receive the good response time and all the nodes are equally loaded and, in general, all users get served quicker. Load balancing may be enforced with hardware, software, or a mix of each. Typically, load balancing is that the main reason for server's unbalanced response time. Load balancing plans to optimize the usage of resources, maximize overall success ratio, minimize waiting time interval, and evade overloading of the resources. By the utilization of multiple algorithms and mechanisms with load balancing rather than one algorithm might increase reliability and efficiency. Load balancing within the cloud differs from classical thinking on load balancing design and implementation by misuse of data center servers to perform the requests on the basis of first come first serve basis. The older load balancing algorithm allocates the requests according to the incoming requests of the client. Load balancing is one amongst the central problems in cloud computing.

REFERENCES

- [1] S. Yakhchi, S. Ghafari, M. Yakhchi, M. Fazeli and A. Patooghy, "ICA-MMT: A Load Balancing Method in Cloud Computing Environment," IEEE, 2015.
- [2] S. Kapoor and D. C. Dabas, "Cluster Based Load Balancing in Cloud Computing," IEEE, 2015.
- [3] S. Garg, R. Kumar and H. Chauhan, "Efficient Utilization of Virtual Machines in Cloud Computing using Synchronized Throttled Load Balancing," 1st International Conference on Next Generation Computing Technologies (NGCT-2015), pp. 77-80, 2015.
- [4] R. Panwar and D. B. Mallick, "Load Balancing in Cloud Computing Using Dynamic Load Management Algorithm," IEEE, pp. 773-778, 2015.
- [5] M. Belkhouraf, A. Kartit, H. Ouahmane, H. K. Idrissi, Z. Kartit and M.E. Marraki, "A secured load balancing architecture for cloud computing based on multiple clusters," IEEE, 2015.
- [6] L. Kang and X. Ting, "Application of Adaptive Load Balancing Algorithm Based on Minimum Traffic in Cloud Computing Architecture," IEEE, 2015.
- [7] N. K. Chien, N. H. Son and H. D. Loc, "Load Balancing Algorithm Based on Estimating Finish Time of Services in Cloud Computing," ICACT, pp. 228-233, 2016.
- [8] H. H. Bhatt and H. A. Bheda, "Enhance Load Balancing using Flexible Load Sharing in Cloud Computing," IEEE, pp. 72-76, 2015.
- [9] S. S. MOHARANA, R. D. RAMESH and D. POWAR, "ANALYSIS OF LOAD BALANCERS IN CLOUD COMPUTING," International Journal of Computer Science and Engineering (IJCSE), pp. 102-107, 2013.
- [10] M. P. V. Patel, H. D. Patel and P. J. Patel, "A Survey On Load Balancing In Cloud Computing," International Journal of Engineering Research & Technology (IJERT), pp. 1-5, 2012.
- [11] R. Kaur and P. Luthra, "LOAD BALANCING IN CLOUD COMPUTING," Int. J. of Network Security, pp. 1-11, 2013.
- [12] Kumar Nishant, P. Sharma, V. Krishna, Nitin and R. Rastogi, "Load Balancing of Nodes in Cloud Using Ant Colony Optimization," IEEE, pp. 3-9, 2012.
- [13] Y. Xu, L. Wu, L. Guo, Z. Chen, L. Yang and Z. Shi, "An Intelligent Load Balancing Algorithm Towards Efficient Cloud Computing," AI for Data Center Management and Cloud Computing: Papers from the 2011 AAAI Workshop (WS-11-08), pp. 27-32, 2011.
- [14] A. K. Sidhu and S. Kinger, "Analysis of Load Balancing Techniques in Cloud Computing," International Journal of Computers & Technology Volume 4 No. 2, March-April, 2013, ISSN 2277-3061, pp. 737-741, 2013.
- [15] O. M. Elzeki, M. Z. Reshad and M. A. Elsoud, "Improved Max-Min Algorithm in Cloud Computing," International Journal of Computer Applications (0975 – 8887), pp. 22-27, 2012.



- [16] B. Kruekaew and W. Kimpan, "Virtual Machine Scheduling Management on Cloud Computing Using Artificial Bee Colony," Proceedings of the International MultiConference of Engineers and Computer Scientists 2014 Vol I, IMECS 2014, 2014.
- [17] R.-S. Chang, J.-S. Chang and P.-S. Lin, "An ant algorithm for balanced job scheduling in grids," Future Generation Computer Systems 25 (2009) 20–27, pp. 21-27, 2009.
- [18] Z. Chaczko, V. Mahadevan, S. Aslanzadeh and C. Mcdermid, "Availability and Load Balancing in Cloud Computing," International Conference on Computer and Software Modeling IPC SIT vol.14 (2011) © (2011) IACSIT Press, Singapore, pp. 134-140, 2011.
- [19] R. K. S, S. V and V. M, "Enhanced Load Balancing Approach to Avoid Deadlocks in Cloud," Special Issue of International Journal of Computer Applications (0975 – 8887) on Advanced Computing and Communication Technologies for HPC Applications - ACCTHPCA, June 2012, pp. 31-35, 2012.
- [20] Kumar Nishant, P. Sharma, V. Krishna, N. and R. Rastogi, "Load Balancing of Nodes in Cloud Using Ant Colony Optimization," IEEE, pp. 3-9, 2012.
- [21] Ankit Kumar, Mala Kalra, "Load Balancing in Cloud Data Center Using Modified Active Monitoring Load Balancer", IEEE pp. 1-5, 2016.
- [22] Saraswathi AT, Kalaashri.Y.RA, Dr.S. Padmavathi, "Dynamic Resource Allocation Scheme in Cloud Computing", ELSEVIER, pp. 30-36, 2015



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).