

An optimized Load Balancing Technique for Virtual Machine Migration in Cloud Computing

Deepak Mahawar
M.Tech. Scholar, Dept. of CSE
Arya College of Engineering & IT
Jaipur

Dr. Vishal Shrivastava
Professor, Dept. of CSE
Arya College of Engineering & IT
Jaipur
Vishal500371@yahoo.co.in

Abstract—Cloud computing (CC) is a service that uses subscription storage & computing power. Load balancing in distributed systems is one of the most critical pieces. CC has been a very interesting and important area of research because CC is one of the best systems that stores data with reduced costs and can be viewed over the internet at all times. Load balance facilitates maintaining high user retention & resource utilization by ensuring that each computing resource is correctly and properly distributed. This paper describes cloud-based load balancing systems. CC is virtualization of hardware like storage, computing, and security by virtual machines (VM). The live relocation of these machines provides many advantages, including high availability, hardware repair, fault tolerance, or workload balancing. In addition to various VM migration facilities, during the migration process, it is subject to significant security risks which the industry hesitates to accept. In this paper we have discussed CC besides this we also emphasize various existing load balancing algorithms, advantages & also we describe the PSO optimization technique.

Keywords—Cloud Computing, Load Balancing, Virtualization, Virtual Machines, PSO Mutation Operators.

I. INTRODUCTION

CC is the fastest-growing innovation in information technology (IT) that has seen a strong growth trend in recent years. CC is one of the largest buzz words & the number of CC providers is growing rapidly. Numerous companies like Microsoft & other heavyweight Internet technology businesses, including Amazon and Google, that might be called giant computing industries are developing cloud services. [1,2]. CC offers a brand-new quality service from the beginning, promising low initial investment, anticipated efficiency, limitless scalability, highly available, etc. [3]. The idea of cloud computing & virtualization has grown much in popularity in data technology over the past few years. Several organizations, through increased computer usage, decreased administration time decreased infrastructure charges, have begun to incorporate these emerging technologies further.

CC grew as favorites in recent times. Due to the service's part, this provides flexibility in retrieving data and an easy way for keeping files for making large files and sets of data accessible for numerous consumers to overall world. Managing such kinds of huge sets of data call for many approaches for enhancing and simplifying operations as well as provide perfect efficiency levels for consumers. Load Balancing (LB) in CC is one of the key problems. CPU, memory, delay, or networking load may be load. Load balancing is a technique that distributes workload among several nodes within the presented workspace so that this ensures no nodes within the system are idle or overloaded for every moment.[5].

LB is the mechanism of moving the workload between processors to improve the efficiency of the operation. Working on a computer means the average time it takes for all of the

functions allocated to the machine to be executed. Balancing a uniform load of VMs ensures that others are not idle or loaded moderately when other machines are heavily loaded. LB is one of the main factors in growing the cloud service provider's efficiency levels. The advantages of transferring workload include an increasing usage ratio of services that further increases total efficiency to ensure optimum customer satisfaction. [6].

Virtual machines have the freedom, separation, and encapsulation of hardware. Virtualization advantages include catastrophe management, preparation, product assessments, monitoring, quality compliance, software design, enhanced security, reduced delivery periods, cloud reduction, expanded hardware use, and streamlined administration. [7]. A special middleware, hypervisor, resumes physical hardware, provides so-called virtual machines with their hardware power, operating like actual computers. Live migration defines the mechanism by which a VM is copied to another physical machine while a VM is still powered. The virtualization server has special advantages and is a good method for several situations. Besides that, load balancing methods can be used in cloud computing to divide the quantity with VMM. Cloud information centers are typically adapted to facilitate hundreds of loads that could lead to low resource consumption & energy consumption. Techniques for load balancing including optimization of PSO [8].

PSO is a stochastic, population-based, globally optimized computational intelligence technology proposed by Kennedy and Eberhart in 1995. The social activity of flocking birds in search of food is encouraging. Due to its special search function, basic definition, measurement efficiency, and fast execution, PSO has been commonly used in many fields of

engineering optimization. Every swarm component represents a high-dimensional solution with 4-vector, its current position, the best position to date, its neighborhood best position and speed, and changes its position in the search area based on its perfect place (pbest) as well as its best position (gbest) throughout the search process. [9].

II. CLOUD COMPUTING

The newest word for the long-dreamed view of computer services is cloud computing. This cloud offers a consolidated pool of configurable computer services on-demand access through networks that can be quickly delivered and released with reduced maintenance activity or contact between service providers. Cloud computing now shifts fast the internet service to make it easier for small companies to develop a smartphone application for consumers. The provision of information technologies and software is significantly improved with cloud computing. Cloud computing provides a basis for innovations such as grid computing, including clustering, Server or dynamic application virtualization, and also large-scale automation and SOA shared networks. Cloud computing refers to a network computing paradigm in which the software or app operates instead of on a local computer system on attached servers. To maintain consistency, cloud storage depends on resource sharing. The larger definition of converged networks and pooled resources is this cloud computing [10].

A. Cloud Architecture

Cloud architecture comprises many modules, a front end, a back end, cloud applications e.g.saaS,paas, iaas. The front end includes customers, mobile devices that enable cloud storage system access. The history includes disc servers, virtual machines, and protection frameworks. Both ends are linked via the internet. The system is also in the monitoring and management of customer demand and traffic[11] and is controlled by central servers. Using the "Middleware" program to contact computers linked in the network.

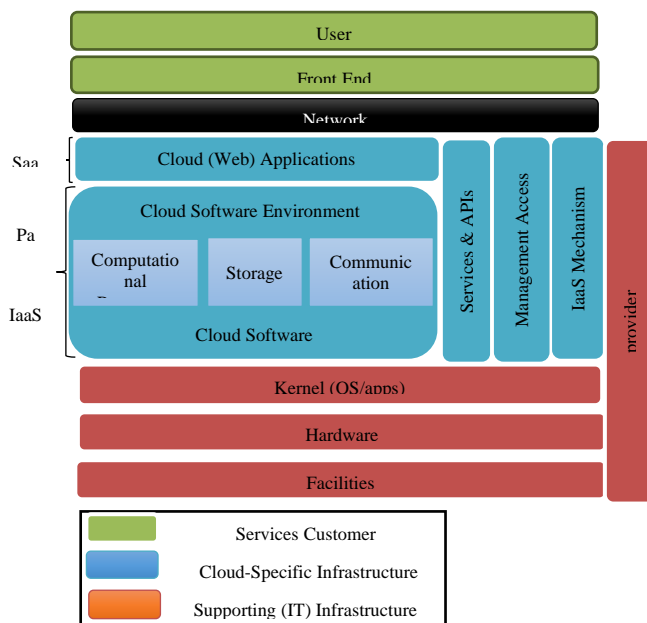


Figure 1: ArchitectureLayers ofCloud Computing

B. Cloud Services Models

1. SaaS

The most common method of cloud computing is likely Software as a Service (SaaS). SaaS users Web to provide software run by a third-party supplier and accessible on the customer's gui. SaaS makes management and services simple for organizations as anything can be handled by vendors: program, runtime, files, middleware, O/S, virtualization, servers, servers & network. Instances of SaaS include Gmail, Google Apps, Microsoft Office 365.

2. PaaS

Platform as a service (PaaS) provides a platform of computer tools. What developers learn from PaaS is to design or customize apps on a platform. PaaS quickly, simply and cost-effectively develops, tests and deploys applications to avoid the need to purchase the underlying hard- & software layers. AWS Elastic Beanstalk, Windows Azure, Heroku, Force.com & Google App Engine are examples of PaaS.

3. IaaS

Infrastructure as a Service (IaaS) provides computing infrastructure, storing, and networking (including the virtualization platform environment). Rather than needing to buy applications, computers, or networks, customers will purchase them as a completely outsourced service which is mostly charged with the value of the services they are using. Essentially, a third party can install a virtual server on its IT infrastructure in preparation for a rental fee. Amazon EC2, Windows Azure, Rack Space, Google Compute Engine are some examples of IaaS. [12].

C. Cloud Deployment Models

1. Public Cloud

The paradigm for public cloud deployment is real cloud hosting. This model provides utilities across a publicly accessible network. The provider here can be delivered services for free by the vendor, or under a pay-per-user license policy. The general public has access to this cloud computing model and is operated by a third-party cloud service provider (CSP). This model is ideally tailored to market needs, using transitional resources for component testing. It lowers the capital time and minimizing operational IT costs.

2. Private Cloud

A single company owns a private cloud deployment model. In this model, cloud computing is run internally or by a third person either by a single entity, and is hosted privately or publicly. Private cloud uses virtualization solutions that focus on strengthening dispersed IT systems often within the company's data centers. In this model, the company maintains complete control of company records, safety rules, and device output.

3. Community

Cloud computing is shared by related organizations in this model. Several organizations or services sharing special needs

including stability, compliance, or competence are collectively procured in this model cloud architecture. It is a common private cloud that a specific business can use a private cloud. The CSP is responsible for group cloud maintenance. Especially in comparison to a private cloud, the model helps minimize expense.

4. Hybrid Cloud

A hybrid cloud is made up of more than two clouds that remain distinct but linked and provide the advantages of multiple implementation models. Hybrid cloud will also mean that dedicated services with cloud infrastructure are linked and administered. Hybrid deployment models are complicated and need careful preparation, in particular, if coordination is required between two cloud deployments [13].

D. Characteristics of Cloud Computing

1. Universal access

Features are accessible over the network and accessed by the normal process, which facilitates the use of thin or thick consumer platforms such as cell phones, computers, laptops...

2. Scalable Services

In terms of nodes and utilities, the cloud architecture is very scalable to extend. Cloud providers can connect additional nodes to the cloud or add services to cloud customers.

3. On-Demand self-services

Cloud allows us the ability to continuously consume computational services e.g. server time, storage, and networking without human intervention.

4. Pay-Per-Use scenario

Cloud services offer services at no cost; users are vital to pay only for resources they are using for access and use.

5. Collaboration

Cloud enables several different organizations to cooperate to solve a challenge or to work together to solve any analysis problem. Cloud reliability, customization, calculated, managed and virtualized services are among other features. [14].

III. LOAD BALANCING

It is fact to disperse load through many resources within each system. In such a manner, the load must be distributed over resources in construction modeling based on cloud, as all resources do around task's identical quantity at each point of time. The elementary need is delivering few approaches for stabilizing demands for giving a choice of application quicker [15]. Method of load balancing which makes every processor busy along with completing works around within same time.

As the users raise the load, the rise in the number of users leads to inadequate resource utilization when the cloud provider is not set up with a strong load balancing system and even cloud servers cannot be appropriately used. The output of the highly loaded node is seized or captured. If a good strategy

for load balancing is used, The load can be similarly separated, and hence we can optimize the resource use (the concept here also describes low loads on heavily loaded nodes, and more loads on nodes with fewer loads now). One of the keytests with cloud computing is the complex division of workload [16].

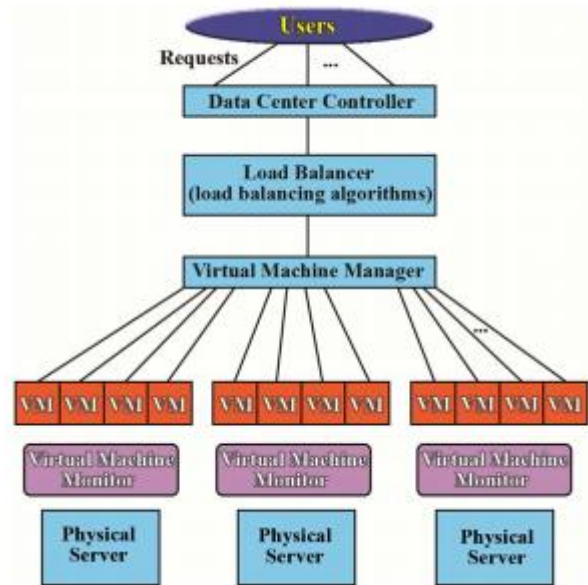


Figure 2: The model of LB

A. Goals of LB

Load balancing's goals are as follows:

- The system's stability remains on track.
- Have capability in altering this as per extend or modification within the setup of the system.
- Promote a system of fault-tolerant concerning stamina, performance under the system's partial failure.
- Achieve huge improvement within the performance.
- Fully utilization of shared resources.
- Increase system's adoptability for adjusting to modifications [17].

B. Types of LB

Load balancing could be classified as the system's current state, like dynamic and static load balancing.

1. Static

Load balancing is knowledge of past related to resources and software of the system. The choice for moving workload should not exactly depend on the system's present status. It relates with load balancing which distributes workload derived which is related to policies' constant set, associated with qualities of workload. It is not defensive. Hence, each system has a minimum of one allotted assignment for itself [18]. The static algorithms don't consider dynamic modifications during

runtime. Few algorithms of static LB are Min-Max, RR & MM algorithms.

2. Dynamic

It doesn't consider the system's prior state and no previous understanding is needed. It depends on the machine's present status. The usual method is permitted by it for relocating from the machines that are heavily loaded dynamically for obtaining quick execution. In such conditions, there is a rise of case communication and turns in more if there is an enhancement in a variety of processors [19]. Few dynamic load balancing is honey-bee foraging, joint-idle queue, biased random sampling, and active clustering.

C. Algorithms of LB

The algorithm for load balancing can be used to allocate charge between multiple device nodes to optimize the utilization of resources as well as the system's response time. The key aim of these algorithms is to solve the condition in which the node is strongly loaded & other nodes are idle, which causes the application to fail [20]. Typically, inside CC was used according to LB algorithms:

1. Round Robin Algorithm (RRA)

It is an algorithm for static LB which uses the fashion of round-robin to allocate jobs. This scheduling is quite efficient and effective time scheduling policy. The algorithm randomly selects nodes for load balancing. Here, an essential role is played by data centers in the handling process of load balancing within cloud computing [21].

2. Min-Min LB Algorithm (MLLB)

This is a fast or simple algorithm that gives improved performance. The algorithm includes a task set. There are not assigned tasks initially to any nodes. Hence minimum time for completion is calculated for every available node within the system. After calculation, a task is chosen which has a minimum time for completion and assign to a separate node. The time that is currently available for execution is uploaded, then there is the removal of tasks from an available set of tasks. The process is performed until every task would be allocated in equivalent machines [22,23].

3. Opportunistic LB Algorithm (OLB)

The algorithm doesn't analyze the virtual machine's current state as this is a static load balancing algorithm. This makes effort in keeping all nodes busy. The algorithm manages unexecuted tasks rapidly to nodes that are available within the system. Every task could be randomly assigned to a node. The algorithm doesn't give load balance good results [24]. Due to such reason, this doesn't calculate the current time of execution of node, hence the task would process slowly in this manner.

4. Max-Min LB Algorithm

The algorithm is the same as an algorithm of Min-Min LB. In beginning, every available task is submitted to the system, and calculation is done for minimum time for completion for every available task. After the calculation, a task is selected which

has maximum time for completion and the task is allocated to the corresponding machine [25]. This algorithm's performance is better when compared with the MinMin algorithm as if the only single large task is there in the task set, then short tasks are run parallel by the Max-Min algorithm with the large task.

5. Active Monitoring LB Algorithm

It is an algorithm of dynamic LB where the load is allocated to VM through finding out the least loaded virtual machine or idle VM in the list. Initially, there is a search for null VM if no null VM is there. Further, the least loaded virtual machine is chosen. Here index table for every request and servers which are assigned to servers currently is maintained with help of a load balancer. When there is a new request, the servers' index table is scanned by the data center which is least loaded or idle. The algorithm uses the concept of first come first serve to assign a load to a server having the least index number for greater than two servers [26].

6. Equally Spread Current Execution Algorithm

It is an algorithm of dynamic LB where effort is made by load balancer for distributing an equal quantity of load between every server which is available in a data center. The processes are assigned a priority at starting of the algorithm, it then checks capacity and size for transferring the load to the server that could handle the load in a smaller period [27]. At such a point, there is a measure of the capacity of a virtual machine and estimation of a load. The load is allocated as per the capacity and size of the matching virtual machine.

7. Active Clustering Algorithm

The algorithm defines virtual machine's clustering to balance load within cloud computing. For the algorithm, clustering is a grouping of the objects together that have similar kind of properties [28]. Hence, virtual machines having similar properties are together grouped in the cluster for handling kind of load.

8. Throttled Load Balancing Algorithm

The algorithm is about VM. Throttled Load Balancer (TLB) maintains every process, as well as monitors, work on servers. Hence, in the algorithm, the best virtual machine is found by load balancer for client requests which could effectively handle the load and quite easily. Different virtual machines have different properties and capacities for handling different loads. Hence, as per load, the right virtual machine should be selected for a load.

D. Activities involved in load balancing

- Identification of user task requirements
- Identification of resource details of a VM.
- Task scheduling
- Resource allocation
- *Migration*

In cloud LB, migration is an essential aspect and the latter is incomplete another without former. Two forms of migration are considered in the cloud based on entities - VM migration or task migration. VM migration (VMM) is the migration of a VM to eliminate the overloading issue from one physical host and is classified in various forms like VM migration or non-living movement. The migration of tasks through VMs is likewise a task and has 2 kinds: migration of intra-VM tasks as well as the migration of inter-VM work. Literature has suggested a wide range of migration methods. An effective migration technique results in a cost-effective LB. From the comprehensive study, it was concluded that the method of task migration is faster and less expensive than VM migration. [29].

IV. VIRTUAL MACHINE ARCHITECTURE

A virtual machine, resource scheduling technology, migration technology, and safety technology play an important role in evaluating the general fairness and efficiency of a virtualized system in a VM system. The virtual machine architecture traditionally has been based on sharing processor resources between domains. Even so, the quality of the system's operation can be bad and/or unpredictable. As hardware & system resources are rapidly expanding, the evaluation system of resource operation in VM is increasingly important. This is why it is a vital factor to increase VM system operation efficiency.

In specific, Figure 3 demonstrates the architecture of the VM method. Several VMs use having similar physical "machine" or host in the design. The host OS kernel or VMM has a resource distribution on VMs at the lowest level right above the hardware layer. For each VM several functions, like file access including network support, are run on top of "Guest" OS which, in turn, offers a customary set of high-level abstractions for VM-running apps. Indeed, VM has the same architecture, and operates an operating system, as a real host machine. A virtual system runs several VMs in a host computer, each of which can execute an operating system. VM allows users to easily manipulate the operation state of the machine, to copy and save (checkpoint) read or modify, share, migrate & rollback. This versatility gives users and administrators great value [30].

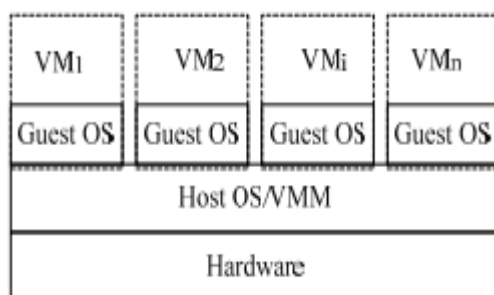


Figure 3: The architecture of virtual machine system

The load balancer generates requests from users as well as executes load-balancing algorithms to exchange requests between VMs. LB determines which VM the next request should be allocated. Task management is carried out by the

data center controller. The load balancer conducts load balancing algorithms for allocating duties to an appropriate VM. The VM Admin is responsible for VMs. Virtualization is a prominent cloud-based application. Virtualization is primarily targeted at complex and expensive sharing of hardware between VMs. VM is a computer software application that runs on operating development and information. VMs manage users' requests. Applications are spread all around the world with unpredictable requests. [31]. Virtualization functionality enables various operating programs and services to be running on identical servers in meantime. virtualization technology is referred to as a virtual system screen or virtual manager that separates computer conditions from the real physical infrastructure. Virtualization allows the physical hardware layer independent of computers, workstations, storage, and various systems [32].

V. PARTICLE SWARM OPTIMIZATION (PSO)

PSO uses the social conduct of the gatherings of the populace in nature, for example, creature groups or flying creatures rushing, or tutoring of fish. PSO comprises a populace called the swarm and every individual from the swarm is known as a particle. The particles search through the global ideal with a set velocity. Since the particles alter and refresh the situation regarding themselves and their neighborhood, it can do both nearby and global search. PSO tests the problem by including several candidate solutions identified as particles in search space that fit into mathematical equations over positions and velocity of the particle [33].

Local best-known positions so even in best-known positions in the search field, they are improved and then identified by other particles. The best particle, called a swarm, is supposed to move to the best solutions [34]. The approach for optimizing nonlinear continuous function was established by James Kennedy or Russell Eberhart in 1995. The authors showed their work using ideas from various fields. Particle swarm optimization (PSO) for a neural network using this method for optimization has been in the center of attraction. PSO excels at global search which is a very common gradient-based method for training connection weights of ANNs, PSO shows faster convergence [35]. The particle swarm optimizer was able to train the network to achieve 92 percent correct [36].

A. PSO Based Mutation Operators

Usage of PSO algorithms by mutation operators. PSO is an effective global convergence algorithm, but can quickly get stuck in local minima. PSO converges very easily, it is widely agreed. The results of an analysis by Vesterstrom etc. (2004), for example, found that PSO often meets at the fastest of tested algos with its results of differential evolution, PSO & evolutionary algorithms. This is a double-edged sword – fast convergence in optimization algo is attractive, but this is feared that the algo may stagnate after a minimum local one has been found. In contrast to a local search for late stages of PSO service there are many methods to slower convergence and hence maximize the time the algorithm uses during the initial exploratory process. One solution is to establish new swarm topologies as it has been proven that a dissimilar topology will impact the swarm operation in convergence to let the trade-off between study and exploitation be modified.

For instance, see Kennedy's (1999) or Kennedy et al. (2002) paper. [37].

VI. LITERATURE REVIEW

S. Garg, et al. (2015)In this study, an algorithm focusing on LB is recommended to minimize the overload or load problem on VMs, which significantly improves the efficiency of the cloud. CloudAnalyst is used to do a comparative study. [38].

P. K. Tiwari & S. Joshi (2016)The presented approach generates performance comparable to Equally Spread Current LB Algorithm (ESCEL) as well as Push-Pull Algorithms for migration time, scalability, efficiency, & availability factor. It depends on other methods of LB& the potential field of study in load management mechanism [39].

A. Kumar & M. Kalra (2016)The suggested CloudAnalyst algorithm was applied and the efficiency of the suggested algorithm is measured with the three algorithms preexisting based on response time. The study suggests that the algorithm suggested operates better than the existing algorithms. [40].

R. K. Devi et al. (2017)Research paper describes a method of LB based on a graph structure known as a weighted graph, complemented by a dominant CDC set definition. The suggested solution increases the use of CDC services by balancing the load and also allows fault tolerances [41].

S. Ghosh & C. Banerjee (2018)Presented a new execution time slice dynamic RR algorithm, that was used to identify a dynamic time quantum (TQ) in each round for specific needs. The Priority Component (PC) of the procedure has been used here to measure. Their test results provide improved efficiency to minimize average waiting times, total context switches as well as the average turnaround time. [42].

K. D. Patel & T. M. Bhalodia (2019)implemented an LB algorithm by combining two cloud-based workload-balancing algorithms. For priority-based tasks & enhanced weighted RR algorithms used for a non-priority-based task, they used enhanced honey bee behavior inspired algorithms. Their study is to increase the efficiency of the scheme, improve the use of resources and minimize completion time. [43].

G. J. Mirobi & L. Arockiam (2019)The cloud providers suggested are classified as overloaded VMs, underloaded VMs, or balanced VMs. If an overcharged VM exists, the suggested Enhanced Throttled Load Balancer will locate the appropriate underloaded VM and will immediately initiate the method of work migration, migrating task from overloaded VM into the underloaded VM, thus balance the load on VMs. This research provides a dynamic solution for load balance using Enhanced TLB that helps reduce the operation time & response time. [44].

A. Mishra and D. Tiwari (2020)Research aims to have used a simulator for the CloudSim cloud platform. Using unit response time as a metric, comprehensively implemented, and operated system. Events show the functionality of the

algorithm implemented in a VM that can be used on a test scale and also improve application performance. [45].

S. Swarnakar et al. (2020)To manage load balancing in a cloud world, suggested present study using genetic-based algorithmic methods. As the suggested algorithm cloudlets take less time to perform & more efficiently LB in cloud environments by using powerful, fittest function to distribute clouds to suitable virtual data center machines in the cloud environments [46].

R. Agarwal et al. (2020)The approach suggested depends on the PSO based on a mutation to LB between the data centers. An effective LB algorithm is built here, which minimizes output parameters such as MakeSpan time or improves fitness function in CC [47].

VII. CONCLUSION

In the past years, CC has been a modern technology. Now in large and small companies, there are a lot of cloud services. How these platforms can be understanding and used is a huge problem. The meaning, types, characters & CC resources were defined by us in this paper. While each CC platform has its strength, it must be noted that there are many outstanding problems regardless of the kind of platform. We discussed multiple algorithms in this paper or discussed the various algorithms for LB in CC& metrics for LB in the cloud. The main problem with CC is LB. LB is important to consistently distribute the additional complex local workload over an entire cloud to achieve high customer satisfaction or use of resources. Load balancing aims to achieve an effective and equal distribution of every computational capacity to ensure high customer satisfaction or use of resources. The new paradigm of computing on demand is cloud computing. In CC-VMMs plays a significant role. In CC, VMM is a big challenge. With the growing popularity of cloud computing services, virtual machines are migrating through data centers or different resource pools. Finally, we concluded that CC is a major area of research. From the perspective of LB, In CC is one of the most discussing topics. This can be done on various optimization techniques.

VIII. REFERENCES

1. L. Qian, Z. Luo, Y. Du, and L. Guo, "Cloud Computing: An Overview," in CloudCom '09: Proceedings of the 1st International Conference on Cloud Computing. Springer-Verlag, 2009, pp. 626–631.
2. G. Lin, D. Fu, J. Zhu, and G. Dasmalchi, "Cloud Computing: IT as a Service," IT Professional, vol. 11, no. 2, pp. 10–13, 2009.
3. Mohamed Magdy Mosbah, "Current Services In Cloud Computing: A Survey", International Journal of Computer Science, Engineering and Information Technology (IJCEIT), Vol.3, No.5, October 2013
4. N. Venkateswara Rao, "A Survey of Cloud Computing: Cloud Computing Concerns And Issues", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 2 Issue 4, April – 2013
5. Ghomi, E. J., Rahmani, A. M., & Qader, N. N. (2017). Loadbalancing algorithms in cloud computing: A survey. Journal of Network and Computer Applications, 88, 50-71.
6. Ali M Alakeel, "A Guide To Dynamic Load Balancing In Distributed Computer Systems", International Journal of

- Computer Science and Network Security, Vol. 10 No. 6, June 2010.
7. D.Kiran Kumar," Review on Virtualization for Cloud Computing", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 8, August 2014
 8. Pradip D. Patel," Live Virtual Machine Migration Techniques in Cloud Computing: A Survey", International Journal of Computer Applications (0975 – 8887), Volume 86 – No 16, January 2014
 9. Binitha S, S Siva Sathya," Survey of Bio-inspired Optimization Algorithms", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-2, May 2012
 10. Harshitha. K. Raj," A Survey on Cloud Computing", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 7, July 2014
 11. Miklau and D. Suci, "Controlling access to published data using cryptography," in Proceedings of the 29th international conference, 2011, page no- 31- 46
 12. E.Gorelik, "Cloud Computing Models", Massachusetts Institute of Technology Cambridge, MA,2013. Available: <http://web.mit.edu/smadnick/www/wp/2013-01.pdf>
 13. Harshitha. K. Raj," A Survey on Cloud Computing", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 7, July 2014
 14. A. Sivakumar," A Survey on Cloud Computing Models and it's Applications", International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 12 | Dec 2019
 15. Joshi, S., & Kumari, U. (2016, December). Load balancing in cloud computing: challenges & issues. In 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I) (pp. 120-125). IEEE.
 16. Foram F Kherani," Load Balancing in cloud computing", 2014 IJEDR, Volume 2, Issue 1
 17. Hota, A., Mohapatra, S., & Mohanty, S. (2019). Survey of different load balancing approach-based algorithms in cloud computing: a comprehensive review. In Computational Intelligence in Data Mining (pp. 99-110). Springer, Singapore
 18. Lakhani, B., & Agrawal, A. (2018). Pros and Cons of Load Balancing Algorithms for Cloud Computing. International Journal of Computer Science Engineering & Technology, 8(4).
 19. Mukati, L., & Upadhyay, A. (2019). A Survey on Static and Dynamic Load Balancing Algorithms in Cloud Computing. Available at SSRN 3365568.
 20. Punit Gupta," Trust and Reliability based Load Balancing Algorithm for Cloud IaaS", 978-1-4673-4529-3/12@ 2012 IEEE
 21. Shakir, M. S., & Razzaque, A. (2017, October). Performance comparison of load balancing algorithms using cloud analyst in cloud computing. In 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON) (pp. 509-513). IEEE.
 22. Jain, S., & Saxena, A. K. (2016, November). A survey of load balancing challenges in cloud environment. In 2016 International Conference System Modeling & Advancement in Research Trends (SMART) (pp. 291-293). IEEE.
 23. Xu, M., Tian, W., & Buyya, R. (2017). A survey on load balancing algorithms for virtual machines placement in cloud computing. Concurrency and Computation: Practice and Experience, 29(12), e4123.
 24. Mazher, N., Riaz, F., Tariq, M., Ishaque, F., & Shakir, H. S. (2018). Performance Comparison of Load Balancing Algorithms using Cloud Analyst in Cloud Computing.
 25. Kumar, P., & Kumar, R. (2019). Issues and challenges of load balancing techniques in cloud computing: a survey. ACM Computing Surveys (CSUR), 51(6), 1-35.
 26. Babu, K. R., & Samuel, P. (2016). Enhanced bee colony algorithm for efficient load balancing and scheduling in cloud. In Innovations in bio-inspired computing and applications (pp. 67-78). Springer, Cham.
 27. Milani, A. S., & Navimipour, N. J. (2016). Load balancing mechanisms and techniques in the cloud environments: Systematic literature review and future trends. Journal of Network and Computer Applications, 71, 86-98.
 28. Adhikari, M., & Amgoth, T. (2018). Heuristic-based loadbalancing algorithm for IaaS cloud. Future Generation Computer Systems, 81, 156-165. [19] Khan, R. Z., & Ahmad, M. O. (2017). A survey on load balancing algorithms in cloud computing. International Journal of Autonomic Computing, 2(4), 366-383.
 29. <https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-019-0146-7>
 30. Yunfa Li, Wanqing Li, Congfeng Jiang," A Survey of Virtual Machine System: Current Technology and Future Trends", 2010 Third International Symposium on Electronic Commerce and Security
 31. EinollahJafarnejadGhomi," Load-balancing algorithms in cloud computing: A survey", Journal of Network and Computer Applications 88 (2017) 50–71
 32. Preeti Thakur," Virtualization in Cloud Computing", International Journal of Recent Trends in Engineering & Research (IJRTER), Volume 02, Issue 0X; Month - 2016 [ISSN: 2455-1457]
 33. Gudise, V. G., and Venayagamoorthy, G. K. (2003), Comparison of particle swarm optimization and backpropagation as training algorithms for neural networks, in Swarm Intelligence Symposium, Proceedings of the IEEE, pp. 110-117.
 34. Singh, H., & Gangwar, R. C. (2014). Comparative Study of Load Balancing Algorithms in Cloud Environment. International Journal on Recent and Innovation Trends in Computing and Communication, Vol: 2, 3195- 3199.
 35. Hashem, W., Nashaat, H., & Rizk, R. (2017). Honey Bee Based Load Balancing in Cloud Computing. KSII Transactions on Internet and Information Systems (TIIS), 11(12), 5694-5711.
 36. Nema, L., Sharma, A., & Jain, S. (2016). Load Balancing Algorithm in Cloud Computing: An Extensive Survey. International Journal of Engineering Science and Computing, 7463-7468.
 37. VytautasJancauskas," Empirical Study of Particle Swarm Optimization Mutation Operators", Baltic J. Modern Computing, Vol. 2 (2014), No. 4, pp. 199–21
 38. S. Garg, R. K. Dwivedi and H. Chauhan, "Efficient utilization of virtual machines in cloud computing using Synchronized Throttled Load Balancing," 2015 1st International Conference on Next Generation Computing Technologies (NGCT), Dehradun, India, 2015, pp. 77-80, doi: 10.1109/NGCT.2015.7375086.
 39. P. K. Tiwari and S. Joshi, "Dynamic weighted virtual machine live migration mechanism to manages load balancing in cloud computing," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016, pp. 1-5, doi: 10.1109/ICCIC.2016.7919581.

40. A. Kumar and M. Kalra, "Load balancing in cloud data center using modified active monitoring load balancer," 2016 International Conference on Advances in Computing, Communication, & Automation (ICACCA) (Spring), Dehradun, India, 2016, pp. 1-5, doi: 10.1109/ICACCA.2016.7578903.
41. R. K. Devi, G. Murugaboopathi and P. Vijayakumar, "A Graph-Based Mathematical Model for an Efficient Load Balancing and Fault Tolerance in Cloud Computing," 2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM), Tindivanam, 2017, pp. 136-140, doi: 10.1109/ICRTCCM.2017.25.
42. S. Ghosh and C. Banerjee, "Dynamic Time Quantum Priority Based Round Robin for Load Balancing In Cloud Environment," 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, India, 2018, pp. 33-37, doi: 10.1109/ICRCICN.2018.8718694.
43. K. D. Patel and T. M. Bhalodia, "An Efficient Dynamic Load Balancing Algorithm for Virtual Machine in Cloud Computing," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 145-150, doi: 10.1109/ICCS45141.2019.9065292.
44. G. J. Mirobi and L. Arockiam, "Dynamic Load Balancing Approach for Minimizing the Response Time Using An Enhanced Throttled Load Balancer in Cloud Computing," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2019, pp. 570-575, doi: 10.1109/ICSSIT46314.2019.8987845.
45. A. Mishra and D. Tiwari, "A Proficient Load Balancing Using Priority Algorithm In Cloud Computing," 2020 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), Hyderabad, India, 2020, pp. 1-6, doi: 10.1109/ICMLANT50963.2020.9355972.
46. S. Swarnakar, N. Kumar, A. Kumar and C. Banerjee, "Modified Genetic Based Algorithm for Load Balancing in Cloud Computing," 2020 IEEE 1st International Conference for Convergence in Engineering (ICCE), Kolkata, India, 2020, pp. 255-259, doi: 10.1109/ICCE50343.2020.9290563.
47. R. Agarwal, N. Baghel and M. A. Khan, "Load Balancing in Cloud Computing using Mutation Based Particle Swarm Optimization," 2020 International Conference on Contemporary Computing and Applications (IC3A), Lucknow, India, 2020, pp. 191-195, doi: 10.1109/IC3A48958.2020.233295.