A Review on Load Balancing Techniques in Cloud Computing

Harish Rajpoot[1], Punit Kumar Johari[2]
[1, 2] Department of CSE and IT
[1, 2] MITS, Gwalior, India

Abstract- *Cloud Computing (CC) is a term, which entails virtualization, distributed computing, networking, application and internet services. A cloud consists of a number of elements, namely customers, data centers and servers. It entails fault tolerance, excessive availability, scalability, flexibility, limited overhead for customers. Valuable for those problems lies in the establishment of an efficient load balancing algorithm. The load will also be CPU load, network load etc. Load balancing is a form of distributing. Many nodes of a distributed process for increasing both consumption and response time of job between load. Load balancing ensures that all the processors in the procedure or each node within the network perform about the equal sum of job at any time instant. Optimization Techniques play an important role in load balancing in the cloud as they try to allocate resources in an efficient manner such that access time is decreased and over all throughput is also improved. The study of whole paper goes through optimization techniques and how they useful for solving load balancing problem in the cloud. With the use of optimization approach problem of load balancing in the cloud can be resolved.*

Keywords- Cloud Computing, Load balancing, Optimization, throughput

I. INTRODUCTION

"Cloud computing is a kind of ubiquitous, enabling convenient, on demand network access to a shared pool of configurable computing resources (e.g., storage, applications, networks, servers, and services) that can be quickly released and provisioned with smallest management service provider communication"[1]. In a CC large number of computers are connected through internet. A CC system which does not use load balancing has a number of drawbacks such as there is an uneven distribution of workload which results in server overloading and system may crash, due to this performance is degraded and efficiency is also reduced.

Now a day CC is used for many purposes, for example file storage and sharing, Cloud Database, CRM (customer relation- ship management), Email, Website hosting, E-commerce, File backup etc. CC is used by many organizations like Facebook, google, twitter etc.

II. TYPES OF CLOUDS

They also called as cloud models.their location (location of cloud server) cloud divided into four categories-

- Public cloud
- Private cloud
- Hybrid cloud
- Community cloud

Public cloud:- In public cloud the user doesn't know from where the information or data is updated. There is no any visibility to the client that computer infrastructure is hosting from where. Public cloud is hosted under some organizations.

Private cloud:- In private cloud the computer infrastructure is hosted by a particular user. In this cloud, the user has its own account and password to operate it. For example Facebook, twitter etc.

Hybrid cloud:- Hybrid cloud is grouping of two or more clouds such as private, public, community cloud. Clouds are formed according requirements of the organization or user.

Community cloud:- In communication cloud, the data has been shared between two or more organizations with the help of cloud server, for example: in California, all government organizations to exchange their data by a computer infrastructure to manage the data of citizen and country.

III. SERVICES OF CLOUD COMPUTING

The CC offered mainly three types of services:

1. SAAS (Software as a service):-

The capability furnished to the client is to create to use of the supplier's purposes jogging on a cloud infrastructure. The functions are obtainable from various client contraptions via either a thin purchaser interface, reminiscent of an internet browser (e.g., email based on the web), or a software interface [24].

2. IAAS (Infrastructure as a service):-

The storage, networks, processing is ability provided for clients and other very important computing resources. the place the client is capable to set up and run arbitrary software, which will include running methods and functions [24].

3. PAAS (Platform as a service):-

The skill provided to the patron is to set up onto the cloud infrastructure, customer-created or received application libraries, offering tools supported and created utilizing programming languages via the supplier [25].

IV. RELATED TECHNOLOGIES

The CC has characteristics of these technologies-

A. Virtualization
B. Grid computing
C. Utility computing

A. Virtualization:

In virtualization, a virtual environment is created which is visible to the user. There is a layer present between hardware and software. By this virtualization the user can operate different O.S in one computer.

B. Grid Computing:

In the grid computing the different systems are connected with each other by any hub and topology to do some task in different computers (peer to peer network).

C. Utility computing:

In this, the user has to pay charges as per use for example: billing model of computing resources is similar to how utilities like electricity are traditionally billed.

V. LOAD BALANCING

Load balancing is the procedure of disseminating the load between different resources in any system. Hence load to be disseminated over the resources in a cloud-based architecture so that every resource does practically the equivalent measure of work at any instant of time. Fundamental prerequisite is to give a few procedures to adjust solicitations to give the arrangement of quick reaction for solicitation. Cloud load balancers oversee online activity by appropriating workloads between various servers and assets consequently. They augment throughput, minimize reaction time, and stay away from over-burden [2]. It is the component of disseminating the load among different nodes of a conveyed system to enhance both use of resources and work reaction time. It additionally guarantees that all the processor in the systems or each node in the system does around the equivalent measure of work at any moment of time. Load balancing is finished with the assistance of burden balancers; in burden adjusting load on systems is just as appropriated between individual nodes of a system so that with burden adjusting the measure of work to be done is equivalent on each node.

Cloud load balancers deals with the online workload between various servers. Essential things to be consider while growing such algorithm is: estimation of load, correlation of different loads, dependability of diverse system, execution of framework, association between the nodes, way of a work to be exchanged and so forth. Many times load can be considered as: CPU load, measure of memory utilized, load in network etc.

VI. GOALS OF LOAD BALANCING

The goals of load balancing are :[4]

a) Get better the performance
b) Accommodate future changing
c) Build fault tolerant system
d) Maintain system stability

VII. TECHNIQUES OF LOAD BALANCING

There are numerous sorts of techniques for load balancing which are accessible for CC. These Techniques help to improve the efficiency of the cloud server some of them are :

Geographical distribution:

The geographical distribution of the nodes matters a considerable measure in the aggregate execution of any continuous distributed computing systems, particularly in the event of the huge scaled applications such as hike, Facebook soon. A very much disseminated system of nodes in cloud environment is helpful in taking care of adaptation to internal failure and keeping up the effectiveness of the system. It uses Geological load balancing (GLB) which can be characterized as a progression of choices about relocation of virtual machines (VMs) or computational undertakings to geographically distributed data centers with a specific end goal to meet the service level agreements (SLAs) or service due dates for tasks and to diminish the operational expense of the cloud system [3].

Hierarchical Load Balancing:

Hierarchical load balancing involves unique levels of the cloud in load balancing determination. Such load balancing tactics on the whole function in grasp slave mode. These can also be modeled as tree data constitution where every node within the tree is balanced under the supervision of its parent node. Grasp or supervisor can use the gentle weight agent procedure to get statistics of slave nodes or youngster nodes. Based upon the know-how gathered through the guardian node provisioning or scheduling determination is made. Three-phase hierarchical scheduling proposed, it has a couple of phases of scheduling [19]. Request monitor acts as a head of the community and is dependable for monitoring service manager which in turn display carrier nodes. First stage uses for BTO (Best Task Order) scheduling, second stage uses for EOLB (Enhanced Opportunistic Load Balancing) scheduling and last third stage uses for EMM (Enhanced Min-Min) scheduling.

Static Load Balancing Algorithm:

Static algorithms are capable for system in load with low variations. In the static algorithm the traffic is divided evenly among the servers. This algorithm neede a previous information of system resources the performance of the processors is determined at start of the execution, therefore the decision of shifting of the load does not depend on the current state of the system. However, static load balancing algorithms have a drawback in that the tasks are assigned to the processor or machines only after it is created and that tasks cannot be changed during its execution for load balancing in another machine [23].

Dynamic Load Balancing Algorithm:

Dynamic load balancing algorithms take the different attributes of the host into new account ' like capabilities and network bandwidth. These algorithms trusts on a set of information that depends on past collected information about the hosts and run-time properties collected as the selected hosts process the tasks. These algorithms allocate and reassign tasks to the hosts depend on the attributes gathered and calculated. Similar algorithms need continuous monitoring of the host and job progress and are usually harder to implement. However, they are extra precise and could result in more efficient load balancing. Present condition of the system plays an important role in scheduling of task because most of the decisions are based on it. These algorithms are stronger than static algorithms, can undoubtedly adjust to change and give better results in heterogeneous and dynamic situations [30].

Ant Colony based load balancing:

An ant colony algorithm is a bionic algorithm that comes from a method of ant patrolling the tracks in the nature. Ants could destroy a substance called pheromone in the path during the campaign. The other ants can logic of bio-material, and selected the right path in the cluster. Ant colony was created by the large number of collective behavior and this would show a positive response phenomenon of data: one path that was passed through more ants would later be chosen by the following one with a greater probability. It provides an efficient way of load balancing because ant colony algorithm uses of the positive feedback principle, and can speed up the evolutionary process in a certain extent, and can realize parallel processing [27].

Parallel Genetic Algorithm:

Genetic Algorithms (GA) are most helpful and powerful search techniques that are used to solve complicated problems. They can be very demanding in terms of memory and computation load . Parallel implementations are the kind of Gas that is called Parallel Genetic Algorithm (PGA) that is providing good performance and in terms of scalability .

PGA uses coarse-grained as the model of PGA for scheduling the resources in the cloud. PGA can easily implement on networks of heterogeneous computer [28].

An Energy and Deadline Aware Optimization Framework:

It focuses on minimizing the process cost of a cloud system by maximizing its energy efficiency while ensuring that explain the user deadline in Service Level Agreements are met, that is provides more opportunities for energy and performance optimizations. However, these optimizations require extra supporting efforts for example- virtual machine placement, resource provisioning, and task scheduling. The basic idea is to start with schedules those applications that are deadline oblivious but energy efficient. It will then co-optimize latency for the deadline violating applications and the global energy cost. It is energy awareness but deadline oriented. Once all deadlines have been met, then it will only focus on the energy cost, unless new deadline violations appear [29].

Honey Bee Behavior Inspired Load Balancing Algorithm [6]:

In this, load balance over the virtual machines for amplifying the throughput. The load balancing CC can be accomplished the rummaging conduct of bumble bees. This algorithm is gotten from the conduct of bumble bees that uses the technique to discover and harvest sustenance. In colonies, there is a class of honey bees called the scout honey bees and

the other sort was forager honey bees. The scout honey bee which scrounges for food sources, when they discover the food, they return to the bee sanctuary to promote this news by utilizing a dance called waggle/tremble/vibration dance. Forager honey bees then take after the Scout Bees to the area that they discovered food and afterward start to procure it. After that they come back to the bee sanctuary and do a tremble or a vibrant dance to different honey bees in the hive giving a thought of the amount of food is left. The assignments expelled from the over-burden VMs go about as Honey Bees. Upon accommodation to the under load VM, it will redesign the quantity of different need undertakings and heaps of errands doled out to that VM. This data will be useful for different undertakings. Since all VMs are sorted in an ascending order, the task removed will be submitted to under stacked VMs. Current workload of all accessible VMs can be computed against the data got from the data center. Points of interest are expanding the throughput; holding up time on task is minimum and overhead become minimum. The disadvantage is starvation problem ,i.e., if extra priority based queues are there, then the lower priority load can be stay continuously in the queue.

VIII. CHALLENGES IN LOAD BALANCING

In CC, There are some subjective metrics in better load balancing [4] [5].

a) Throughput:

It is the aggregate number of tasks that have finished execution for a given size of time. It is required to have high throughput for better performance and execution of the system.

b) Associated Overhead:

It depicts the measure of overhead amid the usage of the load balancing algorithm. It is an organization of development of assignments, bury process correspondence and entomb processor. For a load balancing technique to work properly, least overhead ought to be there.

c) Fault tolerant:

We can characterize it as the capacity to perform load balancing by the proper algorithm without discretionary connection or node failure. Each load balancing algorithm ought to have great adaptation to internal failure approach.

d) Migration time:

It is the measure of time for a procedure to be exchanged starting with one system node, then onto the next node for execution. For better execution of the system this time ought to be constantly less.

e) Response time:

In Distributed system, it is the time taken by a specific load balancing system to react. This time ought to be minimized for better execution.

f) Resource Utilization:

It is the parameter which gives the data inside which surviving the resource is used. For effective load balancing in system ideal resource ought to be used.

g) Scalability:

It is the capability of load balancing algorithms for a system with any some number of machines and processes. This parameter can be enhanced for better system execution.

h) Performance:

It is the general proficiency of the system. In the event that every one of the parameters are enhanced than the general system. Execution can be improved.

IX. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization (PSO) is a kind of population and PSO are depend on the stochastic optimization technique make by Dr. Kennedy and Dr. Eberhart in 1995 [26], drived out by social conduct of bird flocking or fish fooling. PSO offers numerous likenesses with transformative algorithm systems, for example, Genetic Algorithms (GA). The framework is introduced populace of irregular arrangements and hunt down optima through overhauling eras. Not at all like GA, PSO has no progress administrators, for instance change and hybrid. In the PSO, the potential arrangements are known as particles, fly through the issue space by taking after the present ideal particles. The point by point data will be given in taking after segments. Contrasted with GA, the benefits of PSO are that PSO is anything but difficult to execute and there are a couple of parameters to modify. PSO effectively connected with numerous regions provides better solutions for example: capacity advancement, artificial neural network training, fuzzy system control, and different regions where GA can be connected.

As stated earlier, PSO simulates the behaviors of bird flocking. Consider the next state of affairs: a group of birds is randomly searching meals in a region. There is just one piece of meals in the region being searched. All the birds have no idea the place the meals. However, they know the way some distance the food is in every generation. So what's the high-quality process to find the meals? The strong one is to follow the bird which is nearest to the food.

PSO realized from this situation and used it to clear up the optimization problems. In PSO, each and every single solution is a "BIRD" in the search area. We describe it "particle". Every particle has health values that are evaluated via the health operate to be optimized, and have velocities which direct the flying of the particles. The particles fly by means of the problem space via following the present most excellent particles.

Collection of random particles (solution) is initialized for PSO and then search for optima with the modifying generations. All particles are updated through using following two "best" values. The primary one is the fine resolution (fitness) it has finished so far (The fitness value can be preserved), this value is called pbest. One more "excellent" value that's tracked with the aid of the particle swarm optimizer is the first-rate price, received to this point with the aid of any particle in the population. This excellent value is a global best and referred to as gbest. When ,take part in a unit of the population (possible solutions) as its the quality value is a nearest, topological neighborst.

Later, discover the two best values, the particle modifies its positions and velocity with following equation (a) and (b)-

v[] = v[] + c1 * rand() * (pbest[] - present[]) + c2 * rand() * (gbest[] - present[]) -(a)

present[] = persent[] + v[] -(b)

v[] is the particle velocity, persent[] is the current particle (solution). gbest[] and pbest[] are explained as stated previous. rand () is a random number between (0,1). c1, c2 are learning factors.

X. RELATED WORKS

Dhinesh et al. [6], devised an algorithm named honey bee conduct influenced load balancing algorithm. In this, load is balanced across the digital machines for maximizing the throughput. This algorithm is derived from the behavior of honey bees that makes use of the system to search out and reap meals. In bee hives, there's a category of bees known as the scout bees and the one other style was forager bees. The scout bee which forage for meals sources, once they in finding the food, they arrive again to the bee hive to promote this information by means of utilizing a technique known as waggle/tremble/vibration technique. The cause of this technique, gives the great plan and/or number of meals and similarly its space from the beehive. Forager bees then comply with the Scout Bees to the place that they found food and then to reap it.

Bin dong et al. [7], present that a dynamic data migration load balancing algorithm founded on selected structure. Considered the enormous file method there have been quite a lot of problems like dynamic file migration, algorithm situated best on centralized process and so on. So these issues are to be evaded by means of the introduction of the algorithm referred to as self acting load balancing algorithm. In the parallel file system the data are transferred between the memory and the storage devices so that the data management is an important part of the parallel file system. There were a variety of challenges that are focused through load balancing in the parallel file system such as- availability and the scalability of the system, network transmission and the load migration. Because, the load on each I/O server are different in dynamic load balancing algorithms because the workload becomes varies always, thus there were various decision making algorithms are needed.

Yunhua et al. [8], proposed an efficient cell resolution scheme and two warmth diffusion established algorithm called international and local diffusion. Considered the dispensed virtual environments there have been more than a few number of users and the weight having access by way of the concurrent customers can intent situation. According to the heat diffusion algorithm, the virtual environment is divided into a large number of square cells and each square cell having objects. The heat diffusion algorithm is in such a way that all nodes in the cell send load to its nearest nodes in alliteration and the move was the difference between the present node to that of nearest node. So it was connected to heat diffusion method,i.e, the movement of heat from high to low object, when they were placed adjacently.

Markus et al. [9], the idea of overlay networks for the interconnections of machines that create the backbone of an online atmosphere. A virtual online world that makes the opportunities to the world for better technological advancements and developments. This system developed Hyper verse architecture, that can be responsible for the proper hosting of the virtual world. There were self organized load balancing method through which the world surface is subdivided in small cells, and this is controlled by a public server. In these

cells various hotspots are present so that the absolute mass of the object in the cell can be calculated by the public server. Hotspot accuracy is better when increasing the network load. This system cannot avoid the overloaded nodes, but find out the number of links that assigned to each node while joining the network. It provides many advantages such as the network becomes reliable, the network becomes resilience, efficient routing, and fault tolerant. The disadvantage is the overload ratio at the beginning is higher so that public servers are initially placed randomly, so some time is used for balancing the load.

Ye et al. [10], concentrated on the migration of more than one virtual machine deliberating different resource reservation methods. In this, on the basis of the experimental results, it finds different ways of optimization which can be completed on the source machine, migrating more than one digital machine parallelly, and workload-aware process for migration choices, to make stronger the migration effectively.

Aloksingh et al. [11], present that Load Balancing Algorithm. In this, the client first requests load and balancer to check the right virtual machine which accessed, easily load and perform operations which is given by user or client. This is completely based on a virtual machine and host. Various performance parameters for throttele are communication cost, network delay, load movement factor. It is dynamic in nature and it has a high load movement factor.

Kumar et al. [12], devised a new algorithm that appears in an under-loaded node whenever there is an overloaded node in the network, and it also assigned precedence to each and every computing node in the approach that's exclusively situated on the computing energy of these nodes.

Ye et al. [13], investigated quite a lot of strategies developed for migrating a whole virtual cluster. This also awarded several challenges in live migration of digital clusters reminis-cent of significant quantity of knowledge, hindrance of network bandwidth, Communications between VMs.

Nuaimi et al. [14], provided the benefits and downsides of various current load balancing algorithms through comparing the prevailing algorithms such as Ant colony, MapReduce, WLC (weighted least connection), DDFTP (twin course downloading algorithm from FTP servers) and VM mapping. It also mentioned the challenges that ought to be addressed to provide most suitable and efficient load balancing algorithms.

Ren et al. [15], Presented a dynamic load balancing algorithm which is completely headquartered on digital computer migration in a cloud environment. Right here on this algorithm a triggering procedure was proposed that was once based on fractal ways. In This algorithm, comparison in different algorithms, resulted in extra balanced and improved resource utilization. The triggering approaches had been founded on a detailed threshold that resulted an instantaneous height load brought about once a virtual computing device used to be migrated in the cloud environment. However, in this, the migration choices were made on the groundwork of the historical past of the burden indicators load know-how to get load worth. When okay load price exceeds the specific detailed price, migration used to be prompted.

Mousumi Paul et al. [16], develop a scheduling method which follows the Lexi - find approach to assign the job to the existing resources. The scheduled task will be maintained by a load balancing algorithm that distribute the pool of task into small partition and then distribute into local middleware. Job scheduling has been resolve as general assignment issues and to search the least cost. In this, probabilistic factor is generated cost matrix that depends on the few most critical state of efficient task scheduling i.e task waiting time, task arrival and the most important task processing time in a resource.

Achar et al. [17], presented algorithm which dynamically allocates assets centered on the necessity and distribute the burden throughout the servers. The experiments were performed on Xen Cloud Platform prove that the proposed algorithm increased the performance of functions running in digital machines by means of using the characteristic scaling and migration. Here a resource allocation algorithm used to be offered to fortify the performance of the functions going for walks in digital computing device in phrases of response time and distribute the load throughout the servers.

Xu et al. [18], introduces a better load balance model for the public cloud headquartered on the cloud partitioning thought with a swap mechanism to prefer distinctive methods for extraordinary occasions. The algorithm applies the game conception to the load balancing technique to reinforce the effectively in the public cloud environment.

Sharma et al. [19], a new enhanced and effective scheduling algorithm is proposed after which applied in CC atmosphere utilizing CloudSim toolkit, in Java language. With the aid of visualizing the referred to parameters in graphs and tables we will simply determine that the overall response time and knowledge center processing time is multiplied as

well as cost is diminished in comparison to the prevailing scheduling parameters.

Nikita et al. [20], load distribution decisions are made a dynamic load balancing algorithm which are depending on the present workload at all hots of the distributed system. As a consequence, this algorithm must furnish a method for gathering and managing method status know-how. The algorithm handles the requests in an expert approach. It begins by using checking the counter variable of each and every server node and knowledge core. After checking, it transfers the weight hence with the aid of deciding upon the minimum value of the counter variable and the request is treated with no trouble and takes a smaller period of time, and offers maximum throughput. The randomly transfer of load can intent some server to closely loaded while different server is evenly loaded. This algorithm no longer most effective balance the weight but in addition it improves the response time for the cloud. While taking into consideration the impact of cost optimization one has to believe in the field of the option to this predicament. This algorithm basically allocates request that is coming from the customer nodes to the flippantly loaded server cluster (knowledge middle) and gives the response in a diminished period of time, by way of doing this, it makes the response to request ratio.

Kousik Das et al. [21], They present a novel load balancing strategy by Genetic Algorithm (GA). The algorithm balance the load of the cloud infrastructure while trying to minimize the create span of a given tasks set. A simple GA is composed of three operations: selection, genetic operation, and replacement. The advantage of this technique is that it can hold a huge find space, applicable to complex goal function and can avoid being trapped in local optimal solution.

Palta et al. [22], presented aspects of CC which entails the challenges, some open issues and load balancing procedures. The concept is to suggest a process that's composed of several replicated machines referred to as " virtual redundant desktop" that may dump the burden once the info centers get overloaded. This may occasionally fulfill our rationale of attaining virtual desktop migration which is without doubt one of the predominant challenges for load balancing mechanism. Better approach administration, better useful resource allocation and higher response time to end customers will also be accomplished then. Even limitless user requests will outcome VRMs handiest. It will result into balance load to begin with at data centers and afterwards crossing a specific threshold, it will migrate load from knowledge facilities to the connected virtual Replicated Machines called as VRMs.

XI. CONCLUSION

Load balancing is a process of re-assigning the full load to the individual nodes of the collective method to create resource utilization robust and to support the response time of the job, at the same time disposing of a condition in which some of the nodes are overloaded while some others are beneath loaded. A load balancing algorithm which is dynamic in nature does now not recollect the earlier state or habits of the procedure, i.e., it is based upon the current habits of the approach. The major things to recollect whilst establishing such algorithm are: estimation of load, assessment of load, steadiness of different procedure, performance of process, communication between the nodes, selecting of nodes, nature of labor to be transferred and plenty of different ones. A more enhanced optimization technique can be utilized for developing a more optimized approach for load balancing.

REFERENCES

[1] Peter mell, Timothy grance "The NIST Definition of Cloud Computing" National Institute of Standards and Technology Special Publication 800-145 September 2011.

[2] Alok singh1 , Vikas Kumar Tiwari2 , Dr. Bhupesh Gour3 "A Survey on Load Balancing in Cloud Computing Using Soft Computing Technique's" International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 9, September 2014 pp.8081-8086.

[3] Shiny "Load Balancing In Cloud Computing:A Review" IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 15, Issue 2 (Nov. - Dec. 2013), PP 22- 29.

[4] Foster, I., Y. Zhao, I. Raicu and S. Lu, "Cloud Computing and Grid Computing 360-degree compared," in proc. Grid omputing Environments Workshop, pp: 99-106, 2008.

[5] Buyya R., R. Ranjan and RN. Calheiros, "InterCloud: Utilityoriented federation of Cloud Computing environments for scaling of application services," in proc. 10th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP), Busan, South Korea, 2010.

[6] Dhinesh Babu L.D, P. VenkataKrishna, "Honey bee behavior inspired load balancing of tasks in Cloud

Computing environments", Applied Soft Computing 13 (2013) 2292–2303.

[7] Bin Dong, Xiuqiao Li, Qimeng Wu, Limin Xiao, Li Ruan, "A dynamic and adaptive load balancing strategy for parallel file system with large-scale I/O servers", J. Parallel Distribution Computing. 72 (2012) 1254–1268.

[8] Yunhua Deng, Rynson W.H. Lau, "Heat diffusion based dynamic load balancing for distributed virtual environments", in: Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology, ACM, 2010, pp. 203–210.

[9] Markus Esch, Eric Tobias, "Decentralized scale-free network construction and load balancing in Massive Multiuser Virtual Environments", in: Collaborative Computing: Networking, Applications and Worksharing, Collaborate Com, 2010, 6th International Conference on, IEEE, 2010, pp. 1–10.

[10] Kejiang Ye et. al. "Live Migration of Multiple Virtual Machines with Resource Reservation in Cloud Computing Environments" IEEE 4th International Conference on Cloud Computing,2011.

[11] Aloksingh et. al. "A Survey on Load Balancing in Cloud Computing Using Soft Computing Technique's" International Journal of Advanced Research in Computer and Communication Engineering, September 2014, ISSN 2319-5940.

[12] Sachin Kumar, Niraj Singhal "A Priority based Dynamic Load Balancing Approach in a Grid based Distributed Computing Network" International Journal of Computer Applications, July 2012.

[13] Kejiang Ye, Xiaohong Jiang, Ran Ma, Fengxi Yan "VC-Migration: Live Migration of Virtual Clusters in the Cloud "ACM/IEEE 13th International Conference on Grid Computing, 2012.

[14] Klaithem Al Nuaimi, Nader Mohamed, Mariam Al Nuaimi and Jameela Al-Jaroodi"A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms" IEEE Second Symposium on Network Cloud Computing and Applications, 2012.

[15] Haozheng Ren, Yihua Lan, Chao Yin "The Load Balancing Algorithm in Cloud Computing Environment" IEEE 2nd International Conference on Computer Science and Network Technology, 2012.

[16] Mousumi Paul Debabrata Samanta Goutam Sanyal"Dynamic job Scheduling in Cloud Computing based on horizontal load Balancing"2011 IJCTA Vol 2 (5), ISSN:2229-6093, pp1552-1556.

[17] Raghavendra Achar et. al. "Load Balancing in Cloud Based on Live Migration of Virtual Machines" IEEE India Conference (INDICON), 2013.

[18] Gaochao Xu, Junjie Pang, And Xiaodong Fu; "A Load Balancing Model Based On Cloud Partitioning For The Public Cloud". Tsinghu A Science And Technology, 2013.

[19] Tejinder Sharma, Vijay Kumar Banga;" Efficient and Enhanced Algorithm in Cloud Computing" International Journal of Soft Computing and Engineering (IJSCE), 2013.

[20] Nikita Haryani, Dhanamma Jagli "Dynamic Method for Load Balancing in CCOSR Journal of Computer Engineering (IOSR-JCE), 2014.

[21] Kousik Dasgupta et. al. "A Genetic Algorithm (GA) based Load Balancing Strategy for Cloud Computing" International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA), 2013.

[22] Rukman Palta, Rubal Jeet; "Load Balancing in the Cloud Computing Using Virtual Machine Migration: A Review". International Journal of Application or Innovation in Engineering & Management, 2014.

[23] Nadeem Shah, Mohammed Farik;" Static Load Balancing Algorithms In Cloud Computing: Challenges & Solution" International Journal Of Scientific & Technology Research Volume 4, Issue 10, October 2015.

[24] LizheWang,Jie Tao, Marcel Kunze"Scientific Cloud Computing: Early Definition and Experience"The 10th IEEE International Confrrence Computing and Communications, 2008.

[25] Software & Information Industry Association, "Softwaor as a Service: Strategic Backgrounder", February 2001.

[26] James Kennedy and Russell Eberhart "Particle Swarm Optimization" IEEE,1995.

[27] Xin Lu, Zilong Gu "A Load-Adapative Cloud Resource Scheduling Mo- del Based On Ant Colony Algorithm" Proceedings of IEEE CCIS2011.

[28] Zhongni et al. “An Approach for Cloud Resource Scheduling Based on Parallel Genetic Algorithm” IEEE 2011.

[29] Yue Gao et al. “An Energy and Deadline Aware Resource Provisioning, Scheduling and Optimization Framework for Cloud Systems” IEEE 2013.

[30] Hareesh M J et al. “A Review on Load Balancing Algorithms in Cloud ” 2011, IJCTA Vol 5(2), ISSN:2229-6093, pp. 640-645.