

# A Survey of Various Load Balancing Techniques and Proposed Load Balancing Method in Cloud Computing.

Kalyani Ghuge

PG Student, Dept. of Computer Engg.  
G.H. Rasoni College of Engineering & Management,  
Pune, India

Mrs Minaxi Doorwar

Assistant Professor Dept. of Information Technology.  
G.H. Rasoni College of Engineering & Management,  
Pune, India

**Abstract—** The Cloud Computing data Centre has abundant hosts as well as Solicitation requests. A Cloud Computing platform that manoeuvre behind the corporate firewall and within control of an organisation IT department. Cloud deals numerous aids such as self-service, agility, efficiency, security, better resource utilization and faster time to market. To improve the working enactment of the cloud service provider load balancing is one of the indispensable aspects. Load Balancing distribute the dynamic workload across all nodes based on some prediction. In cloud, Load Balancing as a technique is realistic across different data centre to ensure the network convenience by diminishing use of computer hardware, Software catastrophes and proper resource utilization. In this paper, discuss the available techniques of load balancing and comparing all techniques with already available techniques and merits, demerits of algorithm.

**Keywords:** Cloud Computing, Efficiency, Resource Utilization, Utilization, Service Provider, Load Balancing.

## I. INTRODUCTION

Cloud Computing became widespread in last rare year. Cloud provides stretchy software as service (SaaS), Platform as Service (PaaS), Infrastructure as Service (IaaS)[3]. The cloud computing an internet based enlargement in which dynamically accessible and frequently virtualized assets are provided as a service over the internet has become a substantial dispute[3]. Using internet technologies enormously ascendable IT-related competencies are provided to multiple peripheral consumers “as a Service” in the cloud computing [13]. Private and Public are the two types of the cloud. Depending on the application demand cloud computing infrastructure countenance forming a capricious number of virtual machine instances. There are supplementary service that are offered from cloud apart from the basic cloud service[14]. Enterprise communication elucidation to customers are outsource by communication as a service (CaaS) [15]. The CaaS vendors blameable for administration of hardware and software voice over Ip, immidate messaging service, video conferencing, soft phone, multimedia conferencing which mean all activity related to communication. Essential for user sharing of database, remotely hosted database service provides by Database as a Service (DaaS) and make to function as if the database were local. Monitoring as a Service (Maas) provides monitoring for security such as outward thread, susceptibility exposure, monitoring for trouble shooting, monitoring for service level agreement (SLA) compliance and quality of services (QOS) [15]. Have a need of having decidedly available resource to service the request on demand where the consumers pay for the resource utilized promptly, so cloud is a pay-go model [3]. To distribute the large amount of workload over a node using the load balancing technique. In the load balancing contribute the load over the node. To distribute or divide the load among the node considering some prediction. Dynamic resource can be effectively managed on a cloud computing platform using the load balancing concept [1]. In general, load balancing algorithm can be sketchily categorized as Dynamic or Static, Centralized or Decentralized, Periodic or Non-Periodic and those with threshold or without threshold [5]. Load balancing in the cloud computing based on standard load balancing but differs from classical thinking on load-balancing architecture and implementation by using commodity servers to perform the load balancing, which provides for new opportunities and economies of scale as well as presenting its own unique set of challenges.

## II. DEPLOYMENT MODEL OF CLOUD

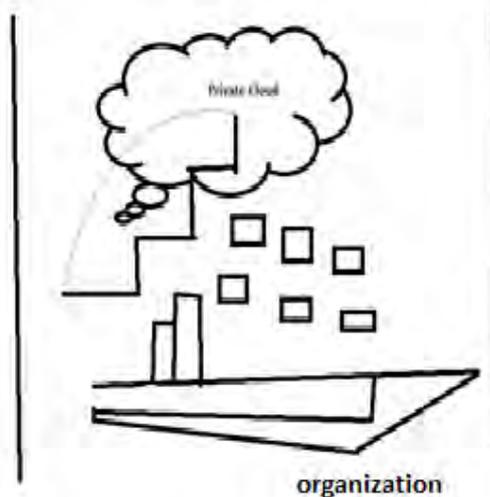


Fig 1. Private Cloud.

**A. Private Cloud:** The cloud infrastructure is wrought exclusively for all organization build his over cloud. Private cloud permits organization to retain total control over their infrastructure, application and data[8].

**B. Public Cloud:** It is the most shared and prevalent form in which cloud infrastructure is made available to the universal public or a huge organization to universal is maintained by an industry trade cloud service which means that pay as per use[8].

**C. Community Cloud:** Community Cloud resources are shared and managed by numerous organizations supports a specific community that has communal apprehensions

**D. Hybrid Cloud:** Hybrid cloud is opus of two or more cloud (public, private, community) that remain unique entities.

In Private cloud there is superior governor over data, while in Public cloud, cloud service provider are responsible for many aspect related to fortifying organisation and data. Public clouds have enhanced resource utilization proportion with private cloud.

## III. LITRATURE SURVEY

In the Load balancing important things are estimation of load, comparison of load, stability of different system, performance of system, interaction between the nodes, nature of work to be transferred, selecting of nodes and many other ones to consider while developing such algorithm [Ali M. Alakeel, "A Guide to Dynamic Load Balancing in Distributed Computer Systems", IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.6, June 2010] [4] . Two main categories of load balancing algorithms are available; they are named as static and dynamic. Static load balancing algorithms needs prior knowledge of the System.

### **Existing Load Balancing Techniques in Cloud Computing:**

**A. Task Scheduling based on LB-** Y. Fang et al. [Fang Y., Wang F. and Ge J.][6] discussed a two-level task scheduling mechanism based on load balanc-ing to meet dynamic requirements of users and obtain high resource utilization. It achieves load balancing by first map-ping tasks to virtual machines and then virtual machines to host resources thereby improving the task response time, resource utilization and overall performance of the cloud com-puting environment.

**B. Honeybee Foraging Behavior-** M. Randles et al. [Randles M., Lamb D. and Taleb-Bendiab A. (2010) 24th Inter-national Conference on Advanced Information Networking and Applications Workshops, 551-556] [4]investi-gated a decentralized honeybee-based load balancing tech-nique that is a nature-inspired algorithm for self-organization. It achieves global load balancing through local server actions. Performance of the system is enhanced with increased sys-tem diversity but throughput is not increased with an increase in system size. It is best suited for the conditions where the diverse population of service types is required.

**C. Biased Random Sampling-** M. Randles et al. [Randles M., Lamb D. and Taleb-Bendiab A. (2010) 24th Inter-national Conference on Advanced Information Networking and Applications Workshops, 551-556][10] investigated a distributed and scalable load balancing approach that uses random sampling of the system domain to achieve self-organization thus balancing the load across all nodes of the system. The performance of the

system is improved with high and similar population of resources thus resulting in an increased throughput by effectively utilizing the increased system resources. It is degraded with an increase in population diversity.

**D. Active Clustering-** M. Randles et al. [Randles M., Lamb D. and Taleb-Bendiab A. (2010) 24th International Conference on Advanced Information Networking and Applications Workshops, 551-556] investigated a self-aggregation load balancing technique that is a self-aggregation algorithm to optimize job assignments by connecting similar services using local re-wiring. The performance of the system is enhanced with high resources thereby increasing the throughput by using these resources effectively. It is degraded with an increase in system diversity.[4]

**E.ACCLB-** Z. Zhang et al. [Zhang Z. and Zhang X.] proposed a load balancing mechanism based on ant colony and complex network theory (ACCLB) in an open cloud computing federation. It uses small-world and scale-free characteristics of a complex network to achieve better load balancing. This technique overcomes heterogeneity, is adaptive to dynamic environments, is excellent in fault tolerance and has good scalability hence helps in improving the performance of the system.[3]

**F. (OLB + LBMM)-** S.-C. Wang et al. [Wang S., Yan K., Liao W. and Wang S. (2010) 3rd International Conference on Computer Science and Information Technology, 108-113.][12] proposed a two-phase scheduling algorithm that combines OLB (Opportunistic Load Balancing) and LBMM (Load Balance Min-Min) scheduling algorithms to utilize better executing efficiency and maintain the load balancing of the system. OLB scheduling algorithm, keeps every node in working state to achieve the goal of load balance and LBMM scheduling algorithm is utilized to minimize the execution time of each task on the node thereby minimizing the overall completion time. This combined approach hence helps in an efficient utilization of resources and enhances the work efficiency.

**G. Decentralized content aware-** H. Mehta et al. [Mehta H., Kanungo P. and Chandwani M. (2011) International Conference Workshop on Emerging Trends in Technology, 370-375.] proposed a new content aware load balancing policy named as work-load and client aware policy (WCAP). It uses a parameter named as USP to specify the unique and special property of the requests as well as computing nodes. USP helps the scheduler to decide the best suitable node for processing therequests. This strategy is implemented in a decentralized manner with low overhead. By using the content information to narrow down the search, it improves the searching performance overall performance of the system. It also helps in reducing the idle time of the computing nodes hence improving their utilization.[11]

**H. Server-based LB for Internet distributed services-** A. M. Nakai et al. [Nakai A.M., Madeira E. and Buzato L.E. (2011) 5th Latin-American Symposium on Dependable Computing, 156-165.] proposed a new server-based load balancing policy for web servers which are distributed all over the world. It helps in reducing the service response times by using a protocol that limits the redirection of requests to the closest remote servers without overloading them. A middleware is described to implement this protocol. It also uses a heuristic to help web servers to endure overloads.

**I. Join-Idle-Queue-** Y. Lua et al. [ Lua Y., Xie Q., Kliot G., Gellerb A., Larusb J. R. and Greenber A. (2011) Int. Journal on Performance evaluation.] proposed a Join-Idle-Queue load balancing algorithm for dynamically scalable web services. This algorithm provides large-scale load balancing with distributed dispatchers by, first load balancing idle processors across dispatchers for the availability of idle processors at each dispatcher and then, assigning jobs to processors to reduce average queue length at each processor. By removing the load balancing work from the critical path of request processing, it effectively reduces the system load, incurs no communication overhead at job arrivals and does not increase actual response time.[6]

#### IV. PROPOSED MODEL

Load Balancing algorithms are designed for all intents and tenacities to equally manage the load on processors and maximize their utilization while minimizing the total task execution time. There is patchy dissemination of marvellous upsurge in workload which domino effect in server overburdening and server may clang. Due to the load, it is decisive to control workloads to improve system performance and maintain stability [18]. The load which is on a cloud is a variable and depends on various factors [18]. To knob this problematic of inequality of load on cloud and to intensification its working efficiency provide the proposed work for that. The proposed system basically divide into three model such as server, submitter (Cloud User) and the processors. This model is basically develop for the balancing the jobs or the load which overcome the problem of the server crash and the inequality of the load.

**General steps which included into proposed model.**

- a) Cloud user which called submitter submit task to the server.
- b) Load balancer is used to balance or divide the load on the basis of the constraints using the liner programming.
- c) Divide the load and submit it to the proper processor.
- d) Processoer consider the quality of the factors such as cost, utilization of resource etc.
- e) Prossessor process the load and submit the result after proper load distribution and task completion to the server.
- f) Server submit result to cloud user with the low response time.

In the proposed System the steps we follows get an proper utilization with the minimum cost of balancing the laod and resource utilization or provisioning. For the balancing the workload consider the quality of factor as a constraints which depends on the task which submitter submit to the server and give the optimal solution to the user.

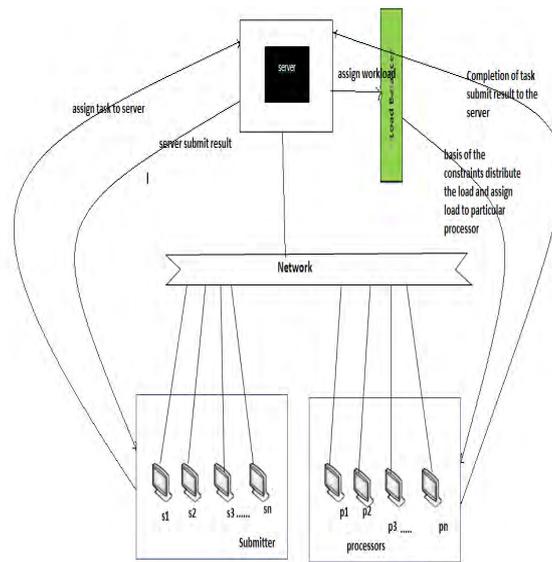


Fig 2 . Proposed Model for Load Balancing

TABLE 1 : SURVEY TABLE

SR.NO	TECHNIQUE	AUTHOR /YEAR	ENVIRONMENT	DESCRIPTION
1.	VectorDot	A. Singh et al. / 2008	Datacenters with integrated server and storage virtualization	1. Uses dot product to distinguish node based on the item requirement
2.	Carton	R. Stanojevic et al. / 2009	Unifying framework for cloud control	1. Uses Load Balancing to minimize the associated cost and uses Distributed Rate Limiting for fair allocation of resources
3.	COMPARE AND BALANCE	Y. Zhao et al. / 2009	Intra-Cloud	1. Based on sampling 2. Uses adaptive live migration of virtual machines
4.	Event-driven	V. Nae et al. / 2010	Massively Multiplayer Online Games	1. Uses complete capacity event as input, analyzes its components and generates the game
5.	Scheduling strategy on LB of VM resources	J. Hu et al. / 2010	Cloud Computing	1. Uses Genetic algorithm, historical data and current state of system to achieve best load balancing and to reduce dynamic migration
6.	LBVS	H. Liu et al. / 2010	Cloud Storage	1. Uses Fair-Share Replication strategy to achieve Replica Load balancing module which in turn controls the access load balancing 2. Uses writing balancing algorithm to control data writing load balancing
7.	Task Scheduling based on LB	Y. Fang et al. / 2010	Cloud Computing	1. First maps tasks to virtual machines and then virtual machines to host resource
8.	Honeybee Foraging Behavior	M. Randles et al. / 2010	Large scale Cloud	1. Achieves global load balancing through local server action
9.	Join-Idle-Queue	Y. Lua et al. / 2011	Cloud data centers	1. First assigns idle processors to dispatchers for the availability of the idle processors at
10.	Lock-free multiprocessing solution for LB	X. Liu et al. /2011	Multi-core	1. Runs multiple load-balancing processes in one load balancer

## V. CONCLUSION AND FUTURE WORK

In this paper, deliberate various algorithm for load balancing for cloud computing. Discuss the algorithm techniques suggested by author and environment which used by them. Design the proposed system which gives the optimal and feasible solution to cloud user. Future work is related to the better utilization of the resources with less amount of time and fast throughput of the system.

## VI. REFERENCES

- [1] Lal Shri Vratt Singh, Jawed Ahmed "A GREEDY ALGORITHM FOR TASK SCHEDULING & RESOURCE ALLOCATION PROBLEMS IN CLOUD COMPUTING" International Journal of Research & Development in Technology and Management Science –Kailash Volume- 21| Issue 1 | ISBN - 978-1-63102-445-0 | March 2014
- [2] Lori M. Kaufman, Bruce Potter "Monitoring Cloud Computing by Layer, Part 1" COPUBLISHED BY THE IEEE COMPUTER AND RELIABILITY SOCIETIES 1540-7993/11/© 2011 IEEE MARCH/APRIL 2011
- [3] Klaitheem Al Nuaimi, Nader Mohamed, Mariam Al Nuaimi and Jameela Al-Jaroodi" A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms" 2012 IEEE Second Symposium on Network Cloud Computing and Applications.
- [4] Venubabu Kunamneni" Dynamic Load Balancing for the Cloud" International Journal of Computer Science and Electrical Engineering (IJCSEE) ISSN No. 2315-4209, Vol-1 Iss-1, 2012
- [5] Ratan Mishra, Anant Jaiswal" Ant colony Optimization: A Solution of Load balancing in Cloud" International Journal of Web & Semantic Technology (IJWesT) Vol.3, No.2, April 2012
- [6] Tushar Desai, Jignesh Prajapati "A Survey Of Various Load Balancing Techniques And Challenges In Cloud Computing" INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 2, ISSUE 11, NOVEMBER 2013
- [7] Mohit Dhingra, J. Lakshmi, S. K. Nandy CAD Lab, Indian Institute of Science Bangalore 560 012 (India)" Resource Usage Monitoring in Clouds"
- [8] Harpreet Kaur, Maninder Singh" A Task Scheduling and Resource Allocation Algorithm for Cloud using Live Migration and Priorities" International Journal of Computer Applications (0975 – 8887) Volume 84 – No 13, December 2013
- [9] Prabavathy, B, Priya, K, Chitra Babu "A Load Balancing Algorithm For Private Cloud Storage" IEEE - 316614th ICCCNT 2013 July 4-6, 2013, Tiruchengode, India
- [10] Gaochao Xu, Junjie Pang, and Xiaodong Fu "A Load Balancing Model Based on Cloud Partitioning for the Public Cloud" IEEE TRANSACTIONS ON CLOUD COMPUTING YEAR 2013
- [11] Martin Randles, David Lamb, A. Taleb-Bendiab "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing" 2010 IEEE 24<sup>th</sup> International Conference on Advanced Information Networking and Applications Workshops
- [12] Albert Y. Zomaya, Senior Member, IEEE, and Yee-Hwei The "Observations on Using Genetic Algorithms for Dynamic Load-Balancing" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 12, NO. 9, SEPTEMBER 2001
- [13] Shu-Ching Wang, Kuo-Qin Yan, Wen-Pin Liao and Shun-Sheng Wang "Towards a Load Balancing in a Three-level Cloud Computing Network" 978-1-4244-5540-9/10/©2010 IEEE
- [14] Jenn-Wei Lin, Chien-Hung Chen, Chi-Yi Lin "Integrating QoS awareness with virtualization in cloud computing systems for delaysensitive applications" © 2014 Elsevier B.V.
- [15] A.Meera, S.Swamynathan "Agent based Resource Monitoring system in IaaS Cloud Environment" International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013
- [16] Radojevic, B., & Zagar, M. (2011). Analysis of issues with load balancing algorithms in hosted (cloud) environments. In MIPRO, 2011 Proceedings of the 34th International Convention, 416-420. IEEE.
- [17] Jorge M. Cortés-Mendoza, Andrei Tchernykh, Johnatan Pecero, Pascal Bouvry, Laura Cruz-Reyes " Distributed VoIP Load Balancing in Cloud Computing"
- [18] Ms. Shilpa D. More, Mrs. Smita Chaudhari "Reviews of Load Balancing Based on Partitioning in Cloud Computing" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3), 2014, 3965-3967