



## A Review on Presented Load Balancing Techniques of Cloud Computing

<sup>1</sup>Mandeep Dhami, <sup>2</sup>Jaspreet Singh<sup>1</sup>M.Tech (IS) Scholar, CEC Landran, Mohali, Punjab, India<sup>2</sup>Assistant Professor, Department of Information Technology, CEC Landran, Mohali, Punjab, India

**Abstract:** Cloud Computing is an innovative development rising in IT situation with enormous necessities of communications and property. "Cloud computing" is an expression, which involve virtualization, dispersed computing, networking, software and web services. A cloud consists of more than a few rudiments such as customers, datacenter and dispersed servers. It includes error acceptance, high ease of use, scalability, elasticity, summary slide for users, condensed cost of possession, on demand service etc. Load Balancing is an important feature of cloud computing atmosphere. A well-organized load balancing plan ensure competent resource operation by provisioning of possessions to cloud user's on-demand basis in pay-as-you-say-manner Central to these issues lays the establishment of an effective load balancing algorithm. The load can be CPU load, recollection ability, delay or network load. Load balancing is the process of distributing the load among various nodes of a distributed system to advance both resource consumption and job response time while also avoid a location where some of the nodes are seriously full while other nodes are unneeded or doing very tiny work. Load balancing ensure that all the mainframe in the system or every node in the network does about the equal amount of work at any immediate of time. This method can be sender initiate, phone initiate or symmetric type (mixture of sender initiate and phone initiate type).

**Keywords:** Load Balancing, cloud computing, distributed system symmetric type.

### I. INTRODUCTION

In the ground of information technology, cloud computing is a current tendency that move computing and data absent from desktop and moveable computers into large data centers. Cloud computing allows everybody to use software and computing services on-demand at anytime, wherever and anywhere through the internet. Cloud computing mostly deals with totaling, software, data right to use and storage services that may not necessitate end-user facts of the physical position and design of the system that is delivers the services [1]. The explanation of cloud computing provide by National Institute of Standards and Technology says that: "Cloud computing is a representation for enable expedient, on-demand network access to a common pool of configurable computing resources (e.g., networks, servers, data storage, software application and other computing services) that can be in alacrity provisioned and unconfined with minimum management effort or service provider interface". By giving out of supply the overall cost reduce [2].

### 1.1 TYPES OF INFRASTRUCTURE IN CLOUD COMPUTING

#### A. CHARACTERSTICS OF CLOUD COMPUTER

Cloud computing is an on-demand, virtualized, cost-effective, expandable, position and device autonomous, and all time obtainable system. Cost is abridged to a significant level as the infrastructure is provide by 3<sup>rd</sup> party and global use of resources which avoid expenditure of resources and computing power. The major goal of cloud computing is to make a better use of dispersed resources by unite them to realize higher throughput [3] and be able to solve large scale totaling problems. Some of the nearly all imperative key individuality is:

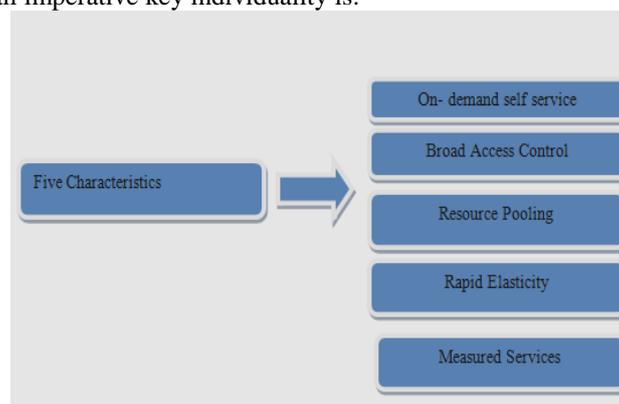


Figure no 1: Characteristics

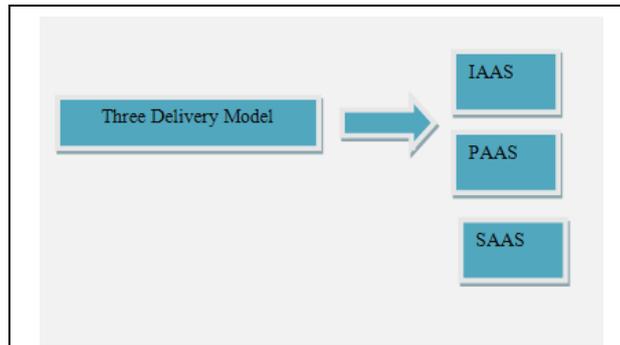


Figure no 2: Cloud Delivery Model

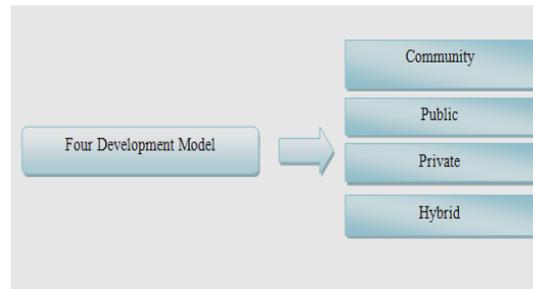


Figure no 3: Development Model

## II. LOAD BALANCING

Load balancing is the process of reassigning the total pack to the entity nodes of the combined system to make effectual resource exploitation and to recover the response time of the jobs, concurrently removing a situation in which some of the nodes are over burdened while some others are below loaded. Load balancing is one of the major [4] issues connected to cloud computing. The consignment can be a recollection, CPU capacity, net or delay load. It is always necessary to share work load amongst the variety of nodes of the disseminated system to get better the reserve utilization and for better presentation of the system. This can help to keep away from the situation where nodes are either greatly loaded or under loaded in the network. Load balancing is the method of ensure the evenly giving out of work load [5] on the pool of system node or supercomputer so that without distressing, the running task is completed. The goals of load balancing are to:

- Advance the performance
- Preserve system stability
- Construct fault tolerance system
- Contain future modification.

Types of load balancing Algorithms:

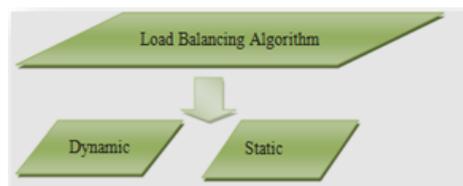


Figure no 4: Types of load balancing algorithm [6]

Table 1: Types of load balancing Algorithm

Load Balancing Algorithm	Description
In static algorithm the transfer is divided equally among the servers. This algorithm requires a preceding knowledge of system capital, so that the choice of irregular of the load does not depend on the existing state of system. Static algorithm is correct in the system which has low disparity in load.	In dynamic algorithm the lightest member of staff serving at table in the whole network or system is searched and preferred for balancing a load. For this real time announcement with network is needed which can increase the transfer in the system. Here present state of the system is used to make decision to straight the load.

## 2.1 CHALLENGES FOR LOAD BALANCING

There are some qualitative metrics that can be improved for better load balancing in cloud computing.

- **Throughput:** It is the total number of responsibilities that have finished execution for a given scale of time. It is necessary to have high through put [7] for better presentation of the system.
- **Associated Overhead:** It describes the amount of overhead during the completion of the load balancing algorithm. It is a work of art of movement of tasks, inter process announcement and inter processor. For load balancing procedure to work properly, bare minimum overhead should be there [8].
- **Fault tolerant:** We can describe it as the ability to carry out load balancing by the suitable algorithm without subjective link or node failure. Every load balancing algorithm should have good fault tolerance come near.
- **Migration time:** It is the amount of time for a development to be transferred from one system node to another node for finishing. For better performance of the system this time should be always less.
- **Response time:** In dispersed system, it is the time taken by a particular load balancing technique to respond. This time should be minimizing for better performance [9].
- **Resource Utilization:** It is the limitation which gives the information within which existing the resource is utilize. For well-organized load balancing in system, optimum resource should be utilized.
- **Scalability:** It is the aptitude of load balancing algorithm for a system with any limited number of processor and machines. This limitation can be improved for better system performance.
- **Performance:** It is the generally efficiency of the system. If all the parameter is enhanced then the overall system presentation can be enhanced.

## III. RELATED WORK

**Tushar Desai, 2013,[8]** Cloud computing is rising knowledge which is a new typical of large scale disseminated computing and parallel compute. It provides shared possessions, information, software correspondence and other possessions as per client necessities at specific time. As cloud computing is increasing rapidly and more users are concerned towards utility computing, better and fast examine needs to be provided. For improved organization of available good load balancing technique are required. So that loads balancing in cloud attractive more involved area of investigate. And from side to side better load balancing in cloud, recital is augmented and user gets better services.

**Rajesh George Rajan, 2013 [9]** Load Balancing is the lone of the majority significant parts of the present virtual surroundings. In the container of cloud computing environment there were various challenge are there in the load balancing technique like security, fault acceptance etc. Many researchers have been planned different technique to perk up the load balancing. This paper describes a survey on load balancing scheme in cloud environments. There were a variety of load balancing technique are used in these identification and their corresponding compensation, disadvantage and presentation metrics are studied in detail.

**Ruhi Gupta, 2014 [11]** Cloud Computing is a rising standard of the computing world which aim to share data over a single platform. Since 2007, cloud computing has emerge as a buzzwords in ICT industry. Two players in cloud computing environments, cloud providers and cloud users, pursue different goals; providers want to maximize revenue by achieving high resource utilization, while users want to minimize expenses while meeting their performance requirements. However, it is difficult to allocate resources in a mutually optimal way due to the lack of information sharing between them. As an increased demand of the resources of cloud computing, load balancing is the usual problem to be faced. Various load balancing algorithms have been designed by various researchers.

## IV. EXISTING LOAD BALANCING TECHNIQUES

- **Honey Bee Foraging Algorithm:** This complete algorithm is based on the process of honeybees finding the food and disturbing others to go and eat the food. First hunter bees go and discover their food. After pending back to their respective beehive, they dance. After seeing the might of their dance, the scout bees go behind the forager bees and get the food. The more vigorous the dance is, the more food existing is [10]. So this whole process is map to congested or under overloaded virtual servers. The server processes the requests of the clients which are related to the food of the bees. As the server gets profound or is stuffed, the bees search for another position i.e. client is encouraged to any other virtual server. In this way, this complete technique works.
- **Task Scheduling Algorithm based on Load Balancing:** It discuss a two level task preparation instrument base on load balancing to meet dynamic supplies of users and get high resource operation. It achieves load balancing by first map tasks to practical machines and then virtual equipment to host possessions.
- **Throttled Load Balancing Algorithm:** This algorithm makes use of individuality of virtual machines. Client requests the ID of virtual machine. Throttled load balancing algorithm income that ID to the user.
- **Ant Colony Optimization Technique:** In this technique, a pheromone table was being calculated which was rationalized by ants as per the resource utilization and node selection formula. Ants move in onward direction in search of the congested or under loaded node. As the overfull node is traverse, then ants move back to fill the recently encounter under loaded node, so a single table is rationalized every time.
- **Role Based Access Control:** RBAC is a practice used to reduce the load of the cloud. In this, a role is assign to each user so that limited application of the cloud can be access by their individual number of users. So by this move toward, the resources are limited to the users.

- **Resource Allocation Scheduling Algorithm:** In this algorithm, practical nodes are shaped first. Then the predictable response time of each virtual node is found. Then according to the least laden node criteria, resourceful virtual node is found [11] and ID of that node is return to the client.

## V. CONCLUSION

In this paper, we have survey various load complementary technique for cloud compute. The main reason of load balancing is to convince the customer obligation by distribute load vigorously among the nodes and to make greatest resource operation by reassigning the total load to person node. These ensure that every supply is dispersed efficiently and consistently. So the recital of the system is augmented. We have also discussed virtualization of cloud and necessary qualitative matrix for load balancing. One of the main issue of cloud computing is system load balancing, since congestion of an exacting node makes it slow down ensuing poor system competence. So there is always an obligation of competent load balancing algorithms for humanizing the operation of computing reserve.

## REFERENCES

- [1] Foster, I., Zhao, Y., Raicu, I. & Lu, S. (2008). Cloud computing and grid computing 360-degree compared. IEEE Grid Computing Environment Workshops.
- [2] Youseff, L. Toward a Unified Ontology of Cloud Computing Retrieved from <http://spsteve.wikispaces.asu.edu/file/view/unified+interface+cloud.pdf>.
- [3] Luo, S., Lin, Z. & Chenm, X. (2011). Virtualization security for cloud computing service. 2011 International Conference on Cloud and Service Computing 978-1-4577-1637-9/11/\$26.00 ©2011 IEEE. Shenzhen, China: ZTE Corporation.
- [4] Sotomayor, B., Montero, R. S., Llorente, I. M. & Foster, I. (2009). Virtual infrastructure management in private and hybrid clouds. IEEE Internet Computing, 13(5), 14-22.
- [5] Radojevic, B. & Zagar, M. (2011). Analysis of issues with load balancing algorithms in hosted (cloud) environments. In proceedings of 34th International Convention on MIPRO, IEEE.
- [6] Lee, R. & Jeng, B. (2011). Load-balancing tactics in cloud. In proc. International Conference on CyberEnabled Distributed Computing and Knowledge Discovery (CyberC), IEEE, (pp. 447-454).
- [7] Dongliang Zhang, Changjun Jiang, Shu Li, "A fast adaptive load balancing method for parallel particle-based simulations", Simulation Modelling Practice and Theory 17 (2009) 1032–1042.4
- [8] 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.
- [9] 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.
- [10] B. Godfrey, K. Lakshminarayanan, S. Surana, R. Karp, I. Stoica, "Load balancing in dynamic structured P2P systems", in: INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 4, IEEE, 2004, pp. 2253–2262.
- [11] 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.