Impact Factor (SJIF): 3.632

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444

(Special Issue for ITECE 2016)

Appraisal of Load Balancing in Hierarchical Cloud Environment

Hemanshi V Dobaria[1], Priyanka Raval[2]

[1]*Computer& Science Engineering, B.H.Gardi Collage of Engineering and Technology*
[2] *Computer& Science Engineering, B.H.Gardi Collage of Engineering and Technology*

Abstract *— Cloud computing is a buzzword in the IT world. There lies a true picture of the future of computing for both in technical perspective and social perspective behind this elaborated flowing phrase. Computers are used everywhere, may be for work, research or in any such field. Computers are used in our day-to-day life with the increasing computing resources for the society. A cloud consists of several elements such as clients, datacenter and distributed servers. It includes fault tolerance, high availability, scalability, flexibility, reduced overhead for users, reduced cost of ownership, on demand services etc. Central to these issues lies the establishment of an effective load balancing algorithm. The load can be of different types such as CPU load, memory capacity, delay or network load. The term Load balancing may be defined as distribution of the load among various nodes of a distributed system to improve both resource utilization and job response time while also avoiding a situation where some of the nodes are heavily loaded while other nodes are idle or doing very little work. All the processor in the system or every node in the network does approximately the equal amount of work at any instant of time is been ensured by load balancing itself.*

Keywords-Load balancing, MinSd , Static Load balancing , Dynamic Load balancing,parameters.

I. INTRODUCTION

Cloud computing is emerging technology. In this Section we give the introduction about cloud computing, the cloud is a set of hardware, networks, storage, services, and interfaces that enable the delivery of computing as a service. Cloud services include the delivery of software, infrastructure, and storage over the Internet (either as separate components or a complete platform) based on user demand [2], U.S. *NIST (National Institute of Standards and Technology)* defines it as that Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.[1] As it were we can say, Cloud Computing is a virtual pool of processing assets which are given through web. In addition, incorporated distributed computing is an entire element processing structure. It gives a required application program environment. It can send, assign or reallocate registering asset powerfully and screen the utilization of assets at all times. As a rule distributed computing has a circulated establishment foundation, and screen the disseminated framework, to accomplish the reason for productive utilization of the framework. [3] There are so many characteristics, but basically 5 essential characteristics, 4 Deployment models and 3 Delivery models which all are discuss in [4, 5, 6].

There are so many challenges like security, efficient load balancing, Performance Monitoring, Resource Scheduling, Required a fast speed internet connection, Scale and QoS management more detail are given in [7] .

Load balancing is one of the major issues in cloud computing. It is a mechanism which distributes the dynamic local workload evenly across all the nodes in the whole cloud. This will avoid the situation where some nodes are heavily loaded while others are idle or doing little work. It helps to achieve a high user satisfaction and resource utilization ratio. Hence, this will improve the overall performance and resource utility of the system. It also ensures that every computing resource is distributed efficiently and fairly [8].

II. LOAD BALANCING

The performance of the framework can be enhanced by load balancing by moving of workload among the processors. The term Workload of a machine can be explained as the total processing time required to execute all the tasks assigned to the machine [8].the reason behind using Load balancing is that to check whether each virtual machine in the cloud system does the same measure of work all through accordingly increasing the throughput and minimizing the response time. One of the important factors to heighten the working performance of the cloud service provider is also load balancing itself .if anyone of the available machine is not idle or partially loaded while others are heavily loaded this means balancing the load of virtual machines is uniform. One of the complex issue of cloud computing is dividing the workload dynamically. Increased resource utilization ratio which further leads to enhancing the overall performance

thereby achieving maximum client satisfaction Utilizing different segments with load balancing is the benefit of distributing the workload. A multilayer key or a Domain Name System server process is the example of Load balancing for the most part which includes committed software or hardware.

The hierarchy of the processing units in a cloud can go up to any number of levels, but the concept of processing the 'Big Data' in small datasets and the re-computing their results remains the same. Load balancing is the procedure of shifting so as to enhance the presentation of the framework of workload among the processors.

2.1 Types of Load balancing

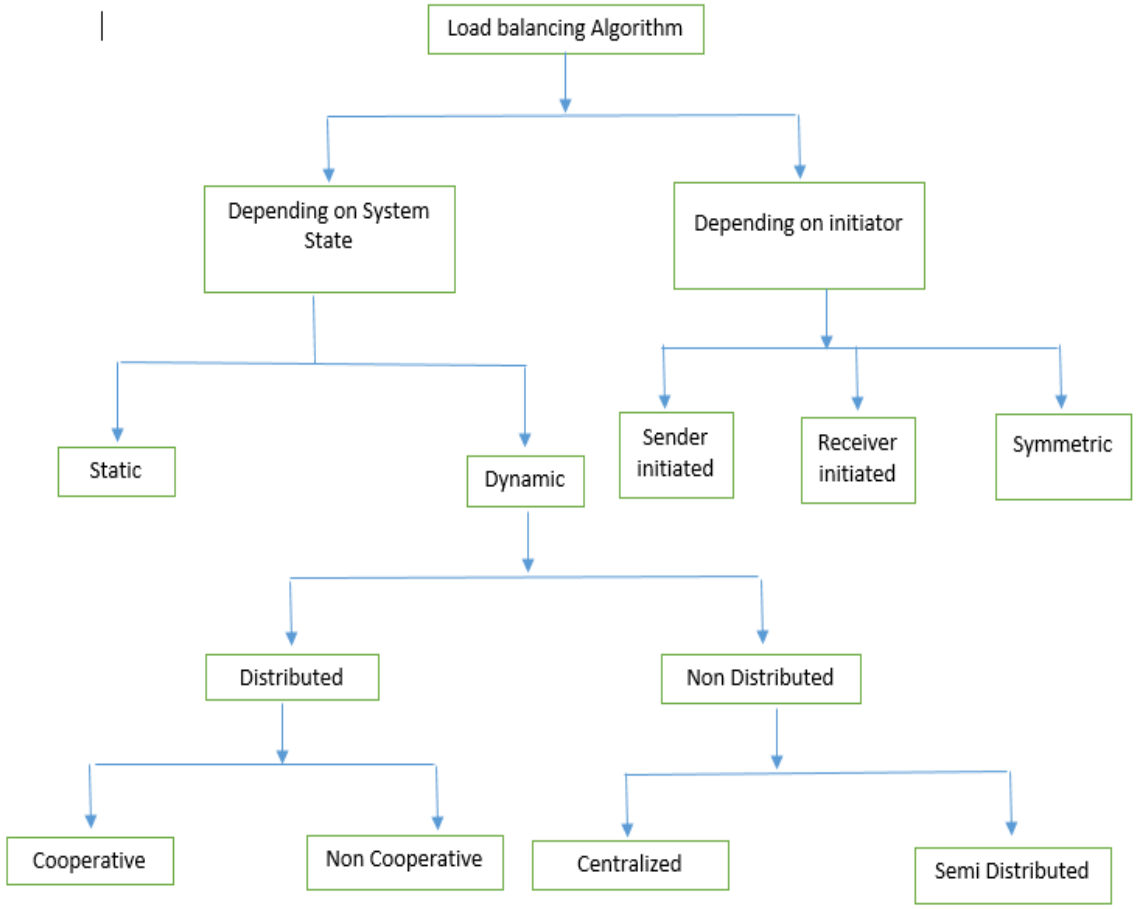

Figure 1 Flow chart of Type of Load Balancing

Sender Initiated: In this type algorithm is initialized by the sender, in this sender send the request message until it finds the receiver to transfer the load

Receiver Initiated: In this type algorithm is initialized by receiver, in this type receiver send request until it get the sender for transmit the load.

Symmetric: It is combination of sender initiated and receiver initiated**.**

Static load balancing: In this homogeneous resources are been mounted by cloud provider. here when environment is static at that time resources are not flexible so user necessity are not subjected to any change at run-time because cloud requires previous knowledge of nodes capacity, processing power, memory, performance and statistics of user requirement. It is easier to simulate but it is not well suited for heterogeneous cloud environment.

Round Robin

This algorithm distributes jobs evenly to all slave processors. All jobs are assigned to slave processors based on Round Robin order, processor choosing is performed in series and will be back to the first processor if the last processor has been reached. Processors choosing are performed locally on each processor, independent of allocations of other processors. [15]

Randomized Algorithm
Random numbers are been chosen by the randomized algorithm to choose slave processors. The slave processors are chosen randomly by the following random numbers generated based on a statistic distribution. Randomized algorithm can attain the best performance among all load balancing algorithms for particular special purpose applications.[16]

Threshold
Processor choosing in this Algorithm is performed based on two threshold values, t_upper and t_lower, that represent upper and lower threshold respectively. Both of these threshold values are used to characterize states of a slave processor State that describe in Table 1. [15]

Processor State	Threshold
Upper Loaded	Load < t_upper
Medium load	t_upper ≤ Load ≤ t_lower
Overloaded	Load > t_lower

Table 1 Characterization of processor's state based on threshold values

Dynamic load balancing: In dynamic environment the cloud provider installs differential property resources. In this condition cloud can't rely on the previous knowledge whereas it takes into account runtime statistics. The resource are flexible and so it can adapt to run time changes in load very easily. Dynamic environment is difficult to be simulated but is highly compatible with cloud computing environment.

Token Routing:
The main objective of the algorithm is to minimize the system cost by moving the tokens around the system. But in a scalable cloud system agents cannot have the enough information of distributing the work load due to communication bottleneck. So the workload distribution among the agents is not fixed. The drawback of the token routing algorithm can be removed with the help of heuristic approach of token based load balancing. This algorithm provides the fast and efficient routing decision. In this algorithm agent does not need to have an idea of the complete knowledge of their global state and neighbors working load. To make their decision where to pass the token they actually build their own knowledge base. This knowledge base is actually derived from the previously received tokens. So in this approach no communication overhead is generated. [17]

Central queuing: This algorithm works on the principal of dynamic distribution. Each new activity arriving at the queue manager is inserted into the queue. When request for an activity is received by the queue manager it removes the first activity from the queue and sends it to the requester. If no ready activity is present in the queue the request is buffered, until a new activity is available. But in case new activity comes to the queue while there are unanswered requests in the queue the first such request is removed from the queue and new activity is assigned to it. When a processor load falls under the threshold then the local load manager sends a request for the new activity to the central load manager. Central manager then answers the request if ready activity is found otherwise queues the request until new activity arrives. [6]

2.2 Major Goals of Load balancing

In load balancing the request of the resource it is important to recognize a few goals of load balancing algorithm**:** [7]

Cost Effectiveness: essential point is to achieve an overall improvement in system performance at a reasonable cost.

Scalability and Flexibility: the system for which load balancing algorithms are implemented may be change in size after some time. So the algorithm must handle these types' situations. So algorithm must be flexible and scalable.

Priority: Prioritization of the resources or jobs needs to be done so higher priority jobs get better chance to execute.

III. LITERATURE REVIEW

The author k. Srivastava and A. kumar discus about cloud basics like its services like IAAS, PAAS, and SAAS with example. In this paper author's discus about cloud services like google services, amazon services, Microsoft service, Salesforce.com, VMware etc. Cloud Architectures and Infrastructure, its services.[9]

The authors S. Zhang et. Discuss about Cloud services as like XaaS where X is Infrastructure, platform, Software, business, Data etc. they also discuss about Cloud computing characteristics like on demand network access , Virtualization , reliability Versatility, High Extendibility . In this paper also mentation about cloud services examples like amazon S3. They also discuss about threats like security, data privacy and its solution. [10]

A.Benletaifa et.al discuss about Network and services , they discuss about IP multimedia Subsystem(IMS) , architecture of IMS, basic of cloud like services , deployment models and they deeply discuss about examples of or compilation of platform as a Services examples like Eucalyptus, open Nebula , Nimbus, Windows Azure, Google app Engine , they also mentation about virtualization and live migration of Vms.[11]

The authors S.Wang et at. Discuss about that cloud computing was utilized by computing resources for perform a function in decentralized manner. The node selection for executing a task in cloud is exploit the effectiveness of the resources, they have to properly select according to the properties of the task. a two-phase scheduling algorithm under a three-level cloud computing network is advanced, they also mentation about proposed algorithm of OLB (Opportunistic Load Balancing) and LBMM (Load Balance Min-Min) scheduling algorithms that can utilize more better executing efficiency and maintain the load balancing of system in three level.[12]

H.Mao et. al. discuss about load balancing different approaches on Bayesian method on different loads as like cpu, memory, network , training they also discuss about System architecture and its characteristics as like load balancing, fault tolerance, Heterogeneity , and its processing elements.[13]

Y. Hao et.al. Discuss about load balancing in three different layers. They also discuss about MinSd algorithm on three levels like PEs, Hosts, Data Centers.in this paper they also discuss simulations on cloudsim to check its performance and its influence on makespan, communication overhead and . And reducing makespan and communication overhead and enhancing throughput of whole system. [14]

IV. Qualitative Parameters

Nature: This relates determination of behavior of load balancing algorithm.
Overload rejection: This is alternative for load balancing. After termination of over loaded situation, then the overload rejection occurs.
Reliability: This relates in case of machine failure, the reliability of algorithms.
Stability: If there is any delay, in transfer of information between load balancing algorithm and processors then the stability comes in picture.
Fault Tolerant: This term relates continuous operation even due to some failure.
Resource Utilization: Automatic load balancing is utilized by resources and also algorithm is able to utilize resource.
Response Time: The time taken to respond distributed system using particular load balancing algorithm.
Waiting time: It is the summation of total time spent waiting in the ready queue.
Turnaround Time: The time period between summation of process and time of completion.
Through put: The data transferred from one place to another.
Makespan: The total application Execution time is known as makespan.
Communication Overhead : The term Communication overhead is related with number of messages over processing elements

V. Related work

Load balancing attitudes traditional distributed system and also various algorithms of load balancing have been projected and executed. In hierarchical cloud environment the hierarchy of levels that are as like PEs, Hosts, Datacenters, Datacenter Brokers. The Physical Structure of cloud is shown in below figure 1.here the phenomena which is followed is at first if at all there is any load on the PE then at first with the help of threshold which lies in the host it checks the sibling PE for its overload to pass the load if not overloaded. if at all it is overloaded then the load is transferred to host and the similar procedure is been followed for data center if host is also over loaded.

At the first level of scheduling the approach is from user's application to virtual machine while the second lies from VM to host resources. Here the first scheduler crafts the task description of VM while the second in host resource discovers the compatible resources for VM.

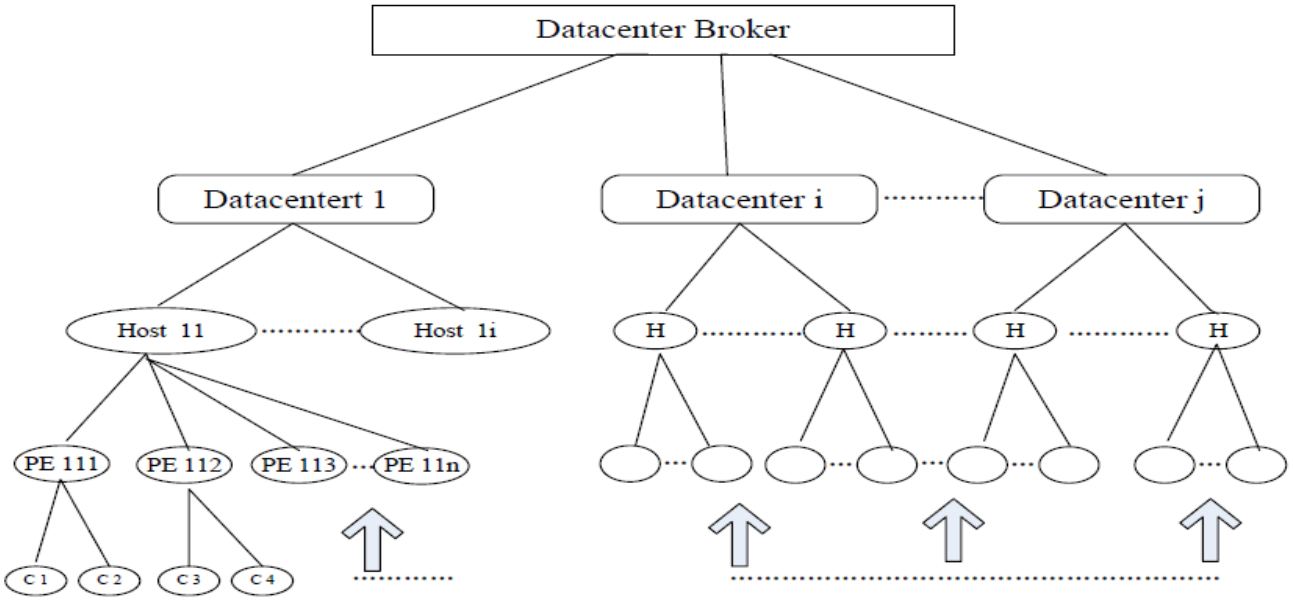

Figure 2. Physical Structure of Cloud [14]

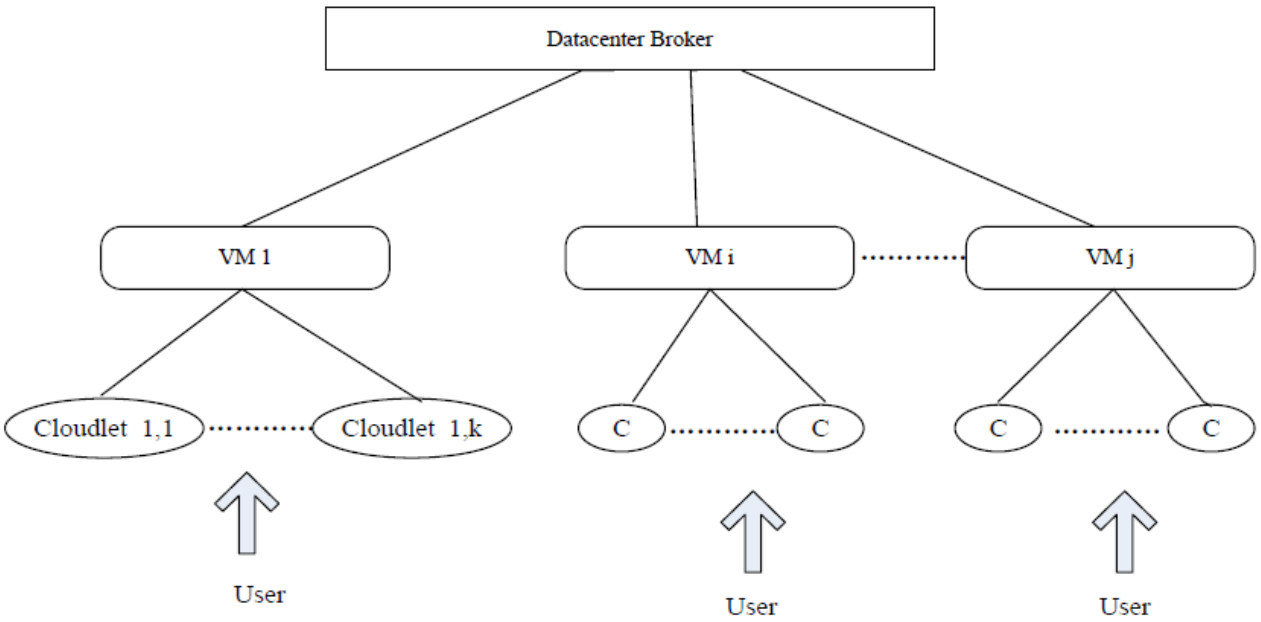

Figure 3 User Views [14]

The anticipated load balancing system works in three levels namely Datacenter, Host and PE. At the time of new Cloudlet which came from a user, it is yield to a VM that is under loaded. We can find from figure 3 that Cloudlet has been assigned to PEs ultimately. And thus load balancing is controlled by the physical resource entity along with Datacenters, Hosts and PEs as shown in figure 2. After a specific interval of time, the resource entity checks its load and classifies them into three categories. Normal load, under load and overload, the standard deviation of load balancing is calculated by below equation. [14]

$$\sigma_r = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(l_i - \bar{l})^2}$$

From this equation σ is the standard deviation of load (%), l_i load of each PE, n total number of load, and Average load of each PE. From this if we fine the scheduling that make the standard deviation value of method is small that means that each load is small.

VI. CONCULSION

This paper provides basic information regarding cloud computing, and its various issues such as efficient load balancing. And it also gives us in detail the discussion about its type and goals. This paper gives information regarding various parameters and it also concludes that MinSd is better than others in hierarchical cloud environment by referring more papers on load balancing and its traditional techniques like round robin, randomized, token ring etc. I also concluded that in future we can work on min Standard Deviation algorithm to improve the efficiency of its parameters like makespan, communication Overhead times and throughput in hierarchical cloud environment.

REFERENCES

[1] P. Mell and T. Grance, "Draft: NIST working definition of cloud computing", Special publication 800-145 21. September 2011

[2] J.Hurwitz, R.Bloor, M. Kaufman, F.Halper, "Cloud Computing for Dummies", Published by Wiley Publishing, Inc, 2010.

[3] S.Zhang, S. Zhan, X.Chen, X.Huo, "Cloud Computing Research and Development Trend", Second International Conference on Future Networks, IEEE2010.

[4] J.SRINIVAS, K.REDDY, Dr.A.QYSER, "Cloud Computing Basics ", International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, July 2012

[5] G.Gill,A.Wadhwa,A.Jatain , "Cloud Computing: Anew Age Of Computing", Fourth International Conference on Advanced Computing & Communication Technologies,IEEE2014.

[6] S.Ray, A.Sarkar, "Execution Analysis of Load Balancing Algorithms In Cloud Computing Environment", International Journal on Cloud Computing: Services and Architecture (IJCCSA), Vol.2, October 2012

[7] R.Kaur ,P.Luthra ,"Load Balancing in Cloud Computing", Association of Computer Electronics and Electrical Engineers ACEEE, 2014

[8] J.Laha, R. Satpathy, k.Dev ," Load Balancing Techniques : Major Challenges in Cloud Computing - A Systematic Review", International Journal of Computer Science and Network, Volume 3, February 2014

[9] K.Srivastava, A.Kumar ,"A New Approach of Cloud: Computing Infrastructure on Demand", Trends in Information Management , July-Dec 2011

[10] S.Zhang,S.Zhang,X.Chen,X.Huo,"Cloud Computing Research and Development Trend", Second International Conference on Future Networks IEEE-2010

[11] A.Benletaifa,A.Haji,M.Jebalia,S.Tabbane," State of the Art and Research Challenges of new services architecture technologies: Virtualization, SOA and Cloud Computing", International Journal of Grid and Distributed Computing, Vol. 3,Dec-2010

[12] S.Wang,K.Yan,W.Liao,S.Wang, "Towards a Load Balancing in a Three-Level Cloud Computing Network", IEEE 2010.

[13] H.Mao, LYuan, Z.Qi," A load Balancing and Overload Controlling Architecture in Clouding Computing", International Conference on Computational Science and Engineering IEEE 2014

[14] Y.Hao,G.Liu,J.Lu,"Three Levels Load Balancing on Clousim", International Journal of Grid Distribution Computing ,Vol.7 IJGDC 2014

[15] H.Rahmawan.Y.Gondokaryono."The Simulation of Static Load Balancing Algorithm", International Conference on Electrical Engineering and Informatics. pp 640-645, Aug 2009

[16] A.Rajguru, S.Apte ," A Comparative Performance Analysis of Load Balancing Algorithms in Distributed System using Qualitative Parameters", International Journal of Recent Technology and Engineering (IJRTE), Volume-1, August 2012

[17] G.Sureshbabu , S.Srivatsa, "A Review Load Balancing Algorithm for Cloud Computing", International Journal Of Engineering And Computer Science , Volume - 3 Issue -9 September, 2014 Page No. 8297-8302