Vishakha et al. International Journal of Institutional & Industrial Research ISSN: 2456-1274, Vol. 1, Issue 1, Jan-April 2016, pp.1-5

Performance Analysis of Cloud Load Balancing Algorithms

Vishakha, Surjeet Dalal

Department of CSE, SRM university, Haryana, India

***Abstract*- Cloud computing is the new word that describes an internet based computing technology which enables the users to access information and use various resources from the clouds from any location. This technology is evolving and developing with the increasing demands in the IT Sector and business environments. However out of the various issues that surround it, load balancing is one such important issue which aims for the even distribution of workload in the system for enhancing performance. This paper presents a review on various load balancing techniques and compares various parameters that are important for performance.**

Keywords— Cloud computing; resources; load balancing, stochastic hill algorithm; clustering

1. Introduction

Cloud computing is a term that has gained immense popularity over the years in the field of computing. It is a computing environment in which all the application services such as storage, networking, processing and other capabilities are provided in a shared and distributed manner. With internet as its main underlying medium, resources can be granted and released in an on request and pay according to the usage fashion. As the demand for computing services in the IT sector is elevating there is a requirement to create a hybrid set of infrastructure which offers productive, reliable and expandable set of services. The IT industry is obtaining the services from cloud service provider as it involves a list of appealing characteristics. Absolutely no beforehand cost-Users can rent the resources for as long as they require and pay only according to their usage. Maintenance Cost-It eliminates the need to maintain the infrastructure thereby reducing the maintenance cost for the companies outsourcing the services. Cloud computing also offers scalability in which the service providers can easily accommodate the rise in resource requests by expanding their services [1].

Cloud computing offers an easy and reliable way to rent hardware and software services through devices equipped with internet facilities fulfilling the demands for computing services.

There are basically three types of service models is cloud computing

- **Infrastructure as a Service**: This service type stands at the bottom of service architecture. It administers a set of hardware, networking & processing components such as servers, storage devices &other infrastructure components. Users can rent the infrastructure according to their requirement.
- **Platform as a Service**: In this service model, users can gain access to the various tools needed for the development of applications without having to worry about managing the infrastructure/hardware.
- **Software as a Service***: In this service model users can gain direct access to applications through the medium of an interface*

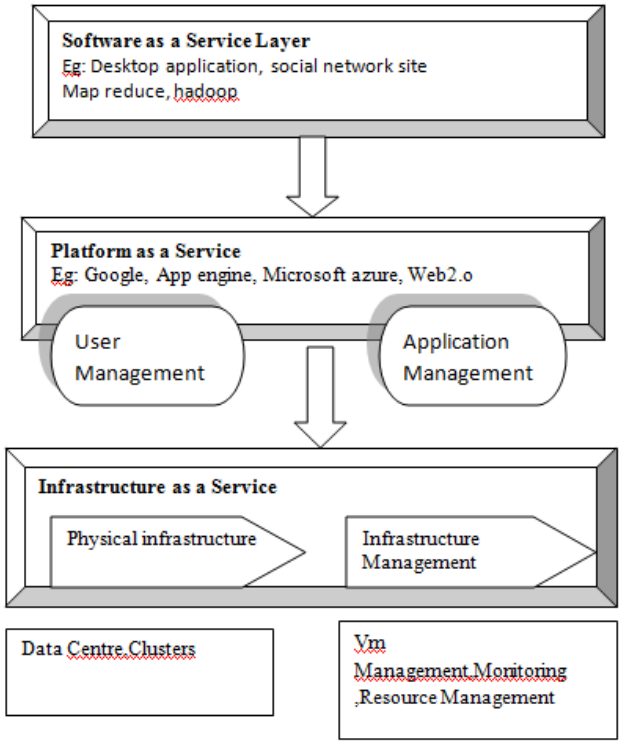

Figure 1 Cloud Environment

Clouds can also be divided into three basic types

- *Public Cloud*: In this type of cloud services are open to common people via internet on a pay per usage manner. Although Public clouds can be exposed to security issues & attacks due to a large set of general users involved.
- *Private Cloud*: Private cloud offers services to a particular organisation. All the members of the organisation can have access to cloud services rented by the organisation through any service provider.
- *Hybrid Cloud*: This type of cloud combines the features of both public and private cloud. Private cloud can gain

services from public cloud whenever the need of scalability arises.[2][4]

1.1. Some issues in cloud computing-:

Even though the cloud computing technology has evolved in the recent years there are still a no. of issues surrounding in its way of development. Security is posed to be the most important issue surrounding cloud as there exists a no. of users in the cloud subjecting a threat to the integrity of data. The data is stored on remote servers which is why it remains under the constant threat of loss and attacks. Several studies have been conducted to address this issue and till now the scope of bringing new changes to resolve this problem exists.

Load balancing is another important issue which surrounds the cloud environment. As the no. of users are growing, there is a need to develop a mechanism to balance the load of resource requests so that there exists proper even distribution of load amongst all the nodes. It facilitates efficient use of all resources which ultimately enhances system performance. This issue plays a very important role in the performance of the system. The main objective of the service provider is to guarantee quality of service with high user satisfaction. Several approaches have been presented to find a solution to this problem but the scope of building better solutions with improved results exists.

2. Load balancing

2.1. Introduction

Load balancing is the technique to divide and allot the job requests evenly amongst all the processing nodes. It plays a very important role in the performance of the system. When the load is allocated appropriately in the whole system, there would be an increase in user satisfaction as job execution would be performed with less processing time and quality of service is provided. If the load is not distributed properly, there often exists a situation where job requests are executed with more time, queued or even rejected in the worst of scenarios. It also affects the utilization factor of all the resources. Two of the important situations arising in the process of load balancing is overloaded and under loaded nodes. The main task is to ensure that no processing node is loaded excessively while other nodes are left either with fewer loads or even a case with no load. Load balancing involves balancing and transfer of workload from overloaded nodes to under loaded nodes to facilitate effective utilization of resources and thereby improving the performance of the whole system. A lot of studies and new methods were formulated to make an improved system with even distribution of load amongst various processing nodes:

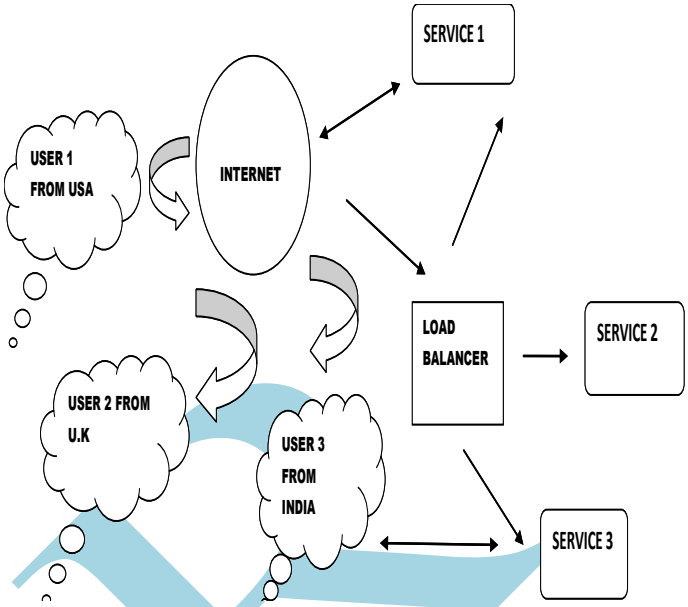

Figure 2 Cloud load manager

2.2. Classification of load balancing algorithm

There are two types of algorithm for load balancing:[5]

1. Static Algorithm: In this type of algorithm information regarding the job request and the resources available needs to be known beforehand prior to the job execution. The transfer of workload is independent of existing state of system and does not occur during the processing of a request.
   Example: Round robin, throttled algorithm, Ant colony algorithm.
2. Dynamic Algorithm: In dynamic algorithm, no information is required before hand prior to job execution .The transfer of load depends upon the existing current state of system and migration of workload from heavily loaded to lightly loaded nodes occurs.

2.3. A brief discussion of some algorithms:

Round Robin: In round robin, the jobs are allocated to the nodes in a rounded fashion. It allocates a job upon arrival to a node chosen randomly. Further requests are allocated in a rounded manner to all other nodes.

Throttled Algorithm: Throttled algorithm works on the principle that for a particular job requests, appropriate node is chosen according to the requirement of the job. The information regarding the virtual machine is mentioned by the job manager. If the appropriate node appears unavailable, the job is queued to wait for the availability of the nodes.

Ant-colony algorithm :This algorithm is inspired by the technique used by ants in finding the best route leading to their location for food .The ants lay a substance called pheromones denoting the rout for their destination .The amount of pheromones denote the quality of food ,the status of the route. Once the job request is received, the ants allocate to a node and head in a forward direction. The movement involves tracing of nodes whether they are

overloaded or not. The information updated /collected is conveyed.

Honey bee: This algorithm is based on the method used by honey bees in finding themselves food. There exist a class of honey bees which is responsible for discovering food. Once the food is found it gives indication about the quality and location of food through a waggle dance. Based on this behaviour the resource allocation takes place. The server accepts requests according to the status of virtual machines. The transfer of workload from overloaded to under loaded VM is done to execute a high priority task which is queued on a machine.

Table 1: Static Vs Dynamic Algorithms

Algorithm	Type	Advantages	Disadvantages
1.Round-robin	Static	Considers whole system equivalent	No information is known beforehand
2.Throttled	Static	Considers status of VM before placing any task	Works best only when identical hardware configuration of VM's exists
3. Ant Colony	Static	Feedback concept generates better solutions	Analyzing theoretical study is difficult
4. Active clustering	Dynamic	Grouping of identical nodes yield better results	Results deteriorate with addition of more nodes
5. Virtual-machine Assign	Dynamic	Considers least loaded VM not used in last iteration	-
6. Honey bee	Dynamic	Improved results in terms of response time	Throughput parameter is ignored

Active Clustering: It is a dynamic based approach which introduces a new technique of clustering. It is based on the concept of combining identical nodes together. A match make node is selected which is used for connecting to a node. This node disconnects itself to ensure even balancing of workload

VM assign Algorithm: This algorithm is another dynamic approach which focuses on assigning jobs to virtual machines by communicating information regarding available VM's between data centre controller & load balancer. It considers least loaded VM which is not used in the last iteration for allocating a resource request. This results in effective utilization of resources

2.4. Parameters For Load Balancing

There are certain parameters/ metrics in load balancing which are very essential for analysing and calculating balancing of load in the system. Some of them which are found to be very influential are:

Response Time: It is defined as that amount of time which is taken by any load balancing mechanism to respond to a resource request. The unit for measuring response time in seconds. It has a very strong impact in the performance of the system.

Resource Utilization: It is another parameter which is a metric to discover how many and to what amount are the resources of a system utilized. It is a key performance indicator in evaluating the performance of the system. It is measured in terms of percentage as determining what percentage of CPU, Memory or Disk memory is utilized.

Migration Time: This parameter defines the amount of time taken in shifting workload for one virtual machine to another in the event of overload or under load. It is measured in seconds and contributes towards the evaluation of performance.

Availability: Availability is found to be an important parameter for load balancing. It determines the availability of a system for execution of job request.

Throughput: It defines the amount of requests that are executed in a given amount of time. It takes the amount of data transferred from system to user and is measured in bytes/sec.

2.5. Mathematical modeling

A mathematical model of a load balancing system consists of a list of input variables and output values. The input variables are the parameters which are essential for optimization. Based on the input variables a function is defined which is maximized or minimized according to the output.

As per the study,[13]the author has taken the two attributes remaining CPU amount and memory amount of the Physical hosts.

Input variables: H(j),C(j),M(j)

Where H(j)=C(j)+M(j)

Requested resource amount is represented with another variable say S(j)

It aims to allocated tasks to optimal hosts as per their required resource amount and by considering the amount left in the hosts.

A standard deviation formula is proposed which is to minimized to obtain the objectives.

3. Existing Solution

S. Yak chi [6] employed imperialism competitive algorithm which detected over utilized and underutilised nodes followed by necessary migration of VM from one to other hosts. This approach considered SLA violation by taking three parameters (SLATAH), PDM (Performance Degradation due to migration) and SLAV. Results showed that to five different policies, ICAMMT had the least power consumption with reduced SLA parameter followed by significant reduction is number of VM migration.

Vishwas Bagwaiya [7] presented a hybrid approach combining throttled and ESCE algorithm. Algorithm works by maintaining a list of allocated VM and allocation count of requests. If VM was assigned else requests was queued. Simulation results illustrated that algorithm had better response time and improved processing along with reduced costs when compared to other methods.

Youssef Fahim [8] proposed another hybrid algorithm considering the state where a few tasks assigned to a virtual machine get blocked due to the availability to detect blocked tasks and executes them. The algorithm was able to reallocate the blocked tasks to another available virtual machine. However no simulation results were illustrated to support the increased performance or availability of the virtual machine.

Mohammad Reza Masbahi[9] presented a unique cloud light weight solution for balancing load in cloud computing . The algorithm works by considering the attribute of VM's following an event-driven architecture. Enterprise Services Bus (ESB) is responsible for managing events. Two types of CLW architecture was proposed: sender and receiver initiated CLW. Simulation results showed better finish time, response time and average CPU utilization than Round-robin. It balanced load by shifting load from overload VM's to under loaded VM's until all VM's are able to obtain a normal status.

Sanjay.k.Dhurander[10] proposed a cluster based load balancing algorithm . It presented a decentralised architecture where network is divided into various clusters. There existed the Master slave concept where the master is involved in broadcasting the tasks to particular slave node of a cluster. Simulation was performed evaluating three performance parameter -load per node, through put and processing wire.

MD. S.Q Zulkar Nine [11] explained an approach based on fuzzy logic. A fuzzy inference system with over 75 rules is created. Fuzzy decision system considered three constraints: memory, bandwidth and disk space requirements. Based on the value of these constraints, decision regarding the assignment of tasks to the virtual machine is made. Simulation results illustrate that algorithm delivered better response time and processing time among various other methods.

Brototi Mondal [12] the author used a derived form of hill-climbing approach stochastic hill climbing. This algorithm works by selecting at random a value from the various uphill values. Selection of an uphill move is dependent on the steepness of uphill value. This algorithm is used for organising job requests to various virtual machine .Simulation is performed using cloud analyst simulator .Performance is evaluated in terms of average response time which outperformed the results of round –robin and FCFS.

Jia,Zhae [13] the authors performed an approach based on Bayes theorem and clustering algorithm considering the three attributes of host CPU ,Memory and posterior probability, a set of physical host capable of performing computation according to requests is created. Based on the combination of both Bayes and clustering strategy, optical hosts are chosen. Simulation is performed through cloud sim simulator and results were compared taking throughput, failure number, make span parameters. Compared with existing work UB-BC presented better improved results.

Saraswathi [14] author explained a dynamic resource allocation scheme based on the attributes of a task. The jobs are considered on the bases of priority. High priority jobs are given preference over low priority jobs. The proposed algorithm selects hosts for the execution of high priority jobs even when all the resources are allocated by putting the low priority the job on pause. The author analysed the performance of the algorithm with cloud sim simulator.

Gao Cho Xu [15] author performed a load partitioning concept. Load balancing is performed by creating sub-areas of clouds which are known as partition at different geographical locations. There exists one main controller which manages the different partition of clouds. A load balancer exists for every partition maintaining a status about its load. No simulation is performed to analyse the performance of this model.

Shridhar[16] author performed a modified version of throttled algorithm. This algorithm differs from the original algorithm in the sense that whenever the data centre controller queries that load balances for availability parsing begins form the virtual machine next to the already assigned virtual machine unlike the parsing from the beginning every time. The author performed the simulation and analysed the performance of algorithm with response time. The results were better when compared to the previous algorithms and author methods.

Yatendra Sahu[17] the author explained load balancing with a dynamic compare and balance algorithm. This algorithm presents a combination of compare and balance algorithm and server consolidation. By assigning threshold value to the host machine, decision regarding migration of virtual machine from over loaded host is made. Simulation results illustrated better performance on load balancing with reduced cost of services offered in cloud.

Kousik Das Gupta [18] performed a genetic algorithm based strategy for load balancing. The algorithm focused on providing effective resource utilization. The algorithm aimed to find the best processing node for execution and thereby balancing the load in the whole system. The simulation results show that proposed algorithm outperformed among other compared methods quantitatively.

Chao Yin [19] explained a load balancing algorithm based on trigger strategy. Virtual machine migration occurs whenever overloading is triggered. Threshold value is considered regarding the decision of migration by evaluating load of a node in terms of CPU load, Memory, Bandwidth. Performance Analysis presented improved performance and better balancing of load.

Aarti Singh [20] author performed an autonomous agent based load balancing algorithm .The algorithm tried to balance the load and allocate the resources though the co-ordination of three agents –load agent ,migration agent and channels agent simulation results were performed using JAVA by taking a number of parameters. Performance was analysed through response time which had better outcomes when exposed to difficult cases.

Table 2: Performance analysis of various algorithms

Algorithm	Resource Utilization	Performance	Availability	Response time	Migration time
ICA-MMT	More	More	No impact	No impact	Less
Hybrid(Throttled&ESCE)	More	More	No impact	Less	No impact
Hybrid algorithm	-	-	-	-	-
Cloud light weight	More	More	More	No impact	Less
Cluster based	No impact	More	More	less	No impact
Fuzzy logic based	No impact	More	No impact	less	No impact
Stochastic Hill	more	More	No impact	less	No impact
Bayes & clustering	No impact	More	More	No impact	Less
Dynamic Resource Allocation	More	More	Priority based	Less	No impact
Load partitioning	-	-	-	-	-
Modified throttled	more	More	No impact	Less	No impact
Dynamic compare and balance	More	More	No impact	No impact	No impact
Genetic Algorithm	No impact	More	No impact	Less	No impact
Autonomous agent based	More	More	More	Less	No impact

4. Conclusion

Cloud Computing is the new emerging technology in the field of computing services. It has the ability to fulfil all business/IT Sector requirements. Due to the advantageous features it offers, there is an immediate need to address the issues surrounding it and come up with different improved solutions. Although many contributions have been made to suggest suitable proposals to the problem of load balancing, there are certain parameters which are collectively important and every solution is not capable of addressing them together. Therefore scope for generating more proposals with effective and improved results still exists.

References

[1]. Qi Zhang , Lu Chang (2010) " Cloud Computing : State-Of-the-Art And Research Challenges" (2010) 1:7-18

[2]. Rajkumar Buyya "Master Cloud Computing " Copyright 2013 by Mc Grawhill Education

[3]. Kuyoro S.O., Ibikunle F.(2011) " Cloud Computing Securitiy Issues And Challenges" (IJCN) Vol 3 Issue 5 2011

[4]. "Cloud Computing" Nick Antonopoulos Springer International Edition 2010

[5]. D.Saranay, L. Sankara Maheswari (July 2015) "Load Balancing Algorithms In Cloud Computing : A Review" Vol 5 issue 7

[6]. S.Yakhchi, M.Fazeli (2015) "ICA-MMT:A load Balancing Method In Cloud Computing Environment" IEEE Conference 2015

[7]. Vishwas Bagwaiya, Sandeep K. Raghuwanshi "Hybrid Approach Using Throttled And ECSE Load Balancing Algorithms In cloud Computing

[8]. Youssef FAHIM, Elhabib BEN LAHMAR (2014) "The load Balancing Improvement Of A Data Centre By A Hybrid In Cloud Computing"IEEE Conference 2014

[9]. Mohmmadreza Mesbahi, Amir Masoud Rahmani (2014) "Cloud Light Weight: A New Solution For Loud Balancing In Cloud Computing" IEEE 2014 International Conference on Data Science & Engineering(ICDSE)

[10]. Sanjay K.Dhurandher, Mohammad S. Obaidat(2014)"A Cluster-Based Load Balancing Algorithm In Cloud Computing" IEEE ICC 2014-Mobile and Networking Symposium

[11]. Md.S.Q.Zulkar Nine, Md.Abul Kalam Azad "Fuzzy Logic Based Dynamic Load Balancing In Virtualized Data Centres"

[12]. Brototi Mondal, Kousik Dasgupta(2012)"Load Balancing In Cloud Computing Using Stochastic Hill Climbing – A Soft Computing Approach" by Elsevier Ltd Procedia Technology

[13]. Jia Zhao , Kun Yang (February 2016) "A Heuristic Clustering – Based Task Deployment Approach For Load Balancing Using Bayes Theorem In Cloud Environment" IEEE Transactions on Parallel and Distributed Systems Vol 27 No. 2

[14]. Sarawathi AT a, Kalaashri. Y.RA b (2015)"Dynamic Resource Allocation Scheme In Cloud Computing " by Elsevier Procedia Computer Science 47(2015) 30-36

[15]. Gaochao Xu , Junjie Pang (February 2013) "A Load Balancing Model Based On Clouding Partitioning For The Public Cloud " Tsinghua Science And Technology ISSN Vol18 No.1

[16]. Shridhar G. Domanal , G. Ram Mohana Reddy "Load Balancing in Cloud Computing Using Modified Throttled Algorithm"

[17]. Yatendra Sahu , R.K. Pateriya (2013) "Cloud Server Optimization With Load Balancing And Computing Techniques Using Dynamic Compare And Balance Algorithm" 2013 5th International Conference on Computational Intelligence And Communication Networks

[18]. Kousik Dasgupta , Brototi Mandal (2013) "A Genetic Algorithm (GA) Based Load Balancing Strategy For Cloud Computing " Elsevier Ltd Procedia Technology 10(2013)340-370

[19]. Haozheng Ren , Yihua Lan (2012) " The Load Balancing Algorithm In Cloud Computing Environment " 2012 2nd International Conference on Computer Science and Network Technology

[20]. Aarti Singh , Dimple Juneja (2015) "Autonomous Agent Based Load Balancing Algorithm In Cloud Computing" Elsevier International Conference on Advance Computing Technologies And Applications (ICACTA-2015)

.

