International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Load Balancing in Cloud Computing

Samarsinh Prakash Jadhav[1], Priya R. Deshpande[2]

[1]ME IT Student, MITCOE, Kothrud, Pune University, Maharashtra, India

[2]Asst. Professor, IT Dept., MITCOE, Kothrud, Pune University, Maharashtra, India

Abstract: ***Cloud computing is the next generation of computation. Possibly people can have everything they need on the cloud. Cloud computing provides resources to client on demand. The resources may be software resources or hardware resources. Cloud computing architectures are distributed, parallel and serve the needs of multiple clients in different scenarios. This distributed architecture deploys resources distributive to deliver services efficiently to users in different geographical channels. Clients in a distributed environment generate request randomly in any processor. So the major drawback of this randomness is associated with task assignment. The unequal task assignment to the processor creates imbalance i.e., some of the processors are overloaded and some of them are under loaded. The objective of load balancing is to transfer the load from overloaded process to under loaded process transparently. Load balancing is one of the central issues in cloud computing. To achieve high performance, minimum response time and high resource utilization ratio we need to transfer the tasks between nodes in cloud network. Load balancing technique is used to distribute tasks from over loaded nodes to under loaded or idle nodes. In following sections we are discuss about cloud computing, load balancing techniques and the proposed work of our load balancing system.***

Keywords: Cloud Computing, Load Balancing, Load Balancing Algorithms, IaaS

1. Cloud Computing

There is no proper definition for cloud computing, we can say that cloud computing is collection of distributed servers which provides services on demand [8]. The services may be software or hardware resources as client need. Basically cloud computing have three major components [9]. First is client, the end user interacts with client to avail the services of cloud. The client may be mobile devices, thin clients or thick clients. Second component is data centre; this is collection of servers hosting different applications. This may exist at a large distance from the clients. Now days a concept called virtualization [6] [7] is used to install software that allows multiple instances of virtual server applications. The third component of cloud is distributed servers; these are the parts of a cloud which are present throughout the Internet hosting different applications. But while using the application from the cloud, the user will feel that he is using this application from its own machine. Cloud computing provides three types [5] of services as Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). SaaS provides software to client which need not to install on clients machine. PaaS provides platform to build an applications like database. IaaS provides computational power to user to execute task from another node.

2. Load Balancing

In cloud system it is possible that some nodes to be heavily loaded and other is lightly loaded [9]. This situation can lead to poor performance. The goal of load balancing is distribute the load among nodes in cloud environment. Load balancing is one of the central issues in cloud computing [6].

For better resource utilization, it is desirable for the load in the cloud system to be balanced [9] evenly. Thus, a load balancing algorithm [1][2][4] tries to balance the total system load by transparently transferring the workload from heavily loaded nodes to lightly loaded nodes in an attempt to ensure good overall performance relative to some specific metric of system performance. When considering performance from the point of view, the metric involved is often the response time of the processes. However, when performance is considered from the resource point of view, the metric involved is total system throughput [3]. I n contrast to response time [2], throughput is concerned with seeing that all users are treated fairly and that all are making progress.

To improve the performance of the system and high resource allocation ratio we need load balancing mechanism in cloud. The characteristics of load balancing are [1] [5]:

- Distribute load evenly across all the nodes.
- To achieve a high user satisfaction.
- Improving the overall performance of the system.
- To reduce response time.
- To achieve resource utilization ratio.

Let us take an example for above sited characteristics:

Suppose we have developed one application and deploy it on cloud. Mean while this application is very popular. Thousands of people are using our application. Suppose hundreds of users using this application at the same time from single machine and we did not apply load balancing approach to our application. This time the particular server is very busy to execute the user's tasks and other servers are lightly loaded or idle. The users did not satisfy because of low response and performance of the system. If we apply load balancing on our application, we can distribute some user's tasks to other nodes and we will get the high performance and faster response time. In this way we can achieve above characteristics of load balancing.

2.1 Taxonomy of Load-Balancing Algorithms

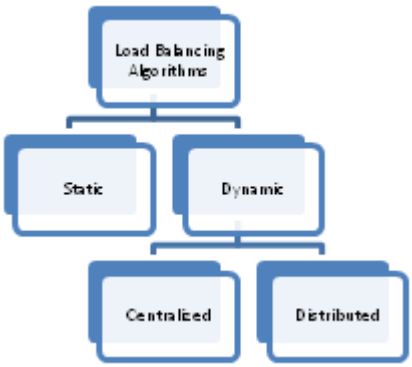

Figure 1: Taxonomy of Load balancing algorithm

There are main two categories of load balancing [3][4]. They are i) Static load balancing and ii) Dynamic load balancing. Static algorithms works statically and do not consider the current state of nodes. Dynamic algorithms [4] work on current state of node and distributes load among the nodes. Static algorithms use only information about the average behaviour of the system, ignoring the current state of system. On the other hand, dynamic algorithms react to the system state that changes dynamically.

Static load balancing [4] algorithms are simpler because there is no need to maintain and process system state information. However, the potential of static algorithm is limited by the fact that they do not react to the current system state. The attraction of dynamic algorithms that they do respond to system state so are better able to avoid those states with unnecessarily poor performance. Owing to this reason, dynamic policies have significantly greater performance benefits than static policies. However, since dynamic algorithms [5] must collect and react to system state information, they are necessarily more complex than static algorithms.

3. Literature Survey

There are many researchers have proposed the work on load balancing in cloud computing, some of them are listed below.

3.1 A Genetic Algorithm [1]

A genetic algorithm approach for optimizing the CMSdynMLB was proposed and implemented. The main difference in this model from previous models is that they considered a practical multiservice dynamic scenario in which at different time steps, clients can change their locations, and each server cluster only handled a specific type of multimedia task so that two performance objectives were optimized at the same time. The main features of this paper included not only the proposal of a mathematical formulation of the CMS-dynMLB problem but also a theoretical analysis for the algorithm convergence.

3.2 Delay Adjustment for Dynamic Load Balancing [2]

The authors are proposed the delay problem on dynamic load balancing for Distributed Virtual Environments (DVEs). Due to communication delays among servers, the load balancing process may be utilizing outdated load information from local servers to compute the balancing flows, while the local servers may be utilizing outdated balancing flows to conduct load migration. This would significantly affect the performance of the load balancing algorithm. To address this problem, authors presented two methods here: uniform adjustment scheme and adaptive adjustment scheme. The first method performs a uniform distribution of the load variation among the neighbor servers, which is a coarse approximation but is very simple to implement. The second method performs limited degree of user tracking but without the need to communicate with neighbor servers.

3.3 Honey Bee Foraging Technique [4]

In this paper, authors have proposed a load balancing technique for cloud computing environments predicated on deportment of honey bee foraging strategy. This algorithm not only balances the load, but withal takes into consideration the priorities of tasks that have been abstracted from heavily loaded Virtual Machines. The tasks abstracted from these VMs are treated as honey bees, which are the information updates globally. This algorithm withal considers the priorities of the tasks. Honey bee deportment inspired load balancing amends the overall throughput of processing and priority predicated balancing fixates on reducing the duration a task has to wait on a queue of the VM. Thus, it reduces the replication of time of VMs. We have compared our proposed algorithm with other subsisting techniques. This load balancing technique works well for heterogeneous cloud computing systems and is for balancing non-preemptive independent tasks. In future, authors orchestrate to elongate this kind of load balancing for workflows with dependent tasks. This algorithm considers priority as the main QoS parameter. In future, they orchestrate to ameliorate this algorithm by considering other QoS factors also.

4. Proposed Work

In today's competitive market, measuring application success as "user interface" alone is no longer enough. Poor availability costs revenue, loyalty and brand image. Application leaders are shifting business-centric metrics to service level management (SLM) [8] to bring IT closer to business. Our aim is to develop a scalable CLOUD solution [6] which is capable of delivering needs of Stock Broking firm without compromising on performance, scalability and cost.

4.1 Features

We will be showing load balancing using following features;

1. User Level Load Balancing on stock application
2. Cloud setup and application deployment [8]

3. Getting Cloud statistics and performance evaluation of each node
4. Resource Monitoring [5] of Cloud Nodes
5. Deploying an application war file on cloud nodes considering their CPU,RAM Usage using cloud controller

4.2 Architecture

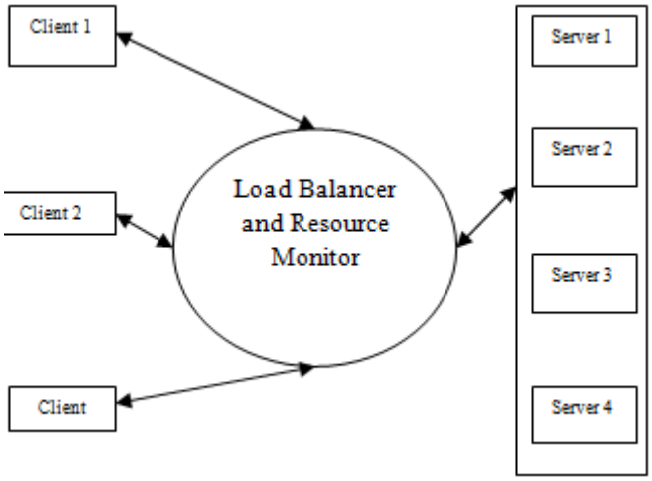

Figure 2: Proposed Architecture

4.3 Scenario of Proposed Algorithm

The VM load balancing algorithm is used to balance the load in the cloud pool. This algorithm check the CPU utilization depends upon the request.

The scenario of proposed algorithm is given below:

1. Get request from client
2. Calculate execution time of each request on each node n1, n2....
3. For each incoming request check resource usage threshold
4. If it goes beyond threshold check resource usage on another node
5. Migrate the request to the node whose resource usage below threshold value and execution time is less.

5. Results

For this project there is need of load testing tool to measure performance of user's request. This project is based on cloud environment we need cloud load testing tool. There are many tools available online to measure load on cloud nodes. For this project we are using Load Storm cloud load testing tool. Load Storm is online testing tool. The summery of load testing result is given below.

Table 1: Summery of the result

	Requests	Response (average s)	Response (max s)	RPS (average)	Throughput (average)	Total Transfer
HTML	288	0.6	1.01	0.24	23.05 kB/s	0.03 GB
Other *	1531	0.22	0.72	1.28	10.31 kB/s	0.01 GB

*Other includes javascript, css, images, pdf, task migration etc. (any content type except html and xml)

The result is shown below images.

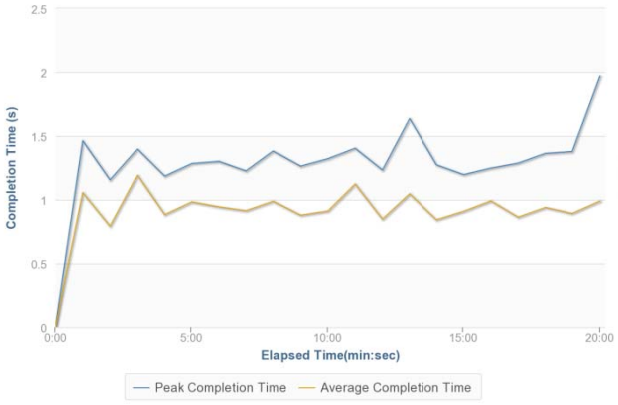

Figure 3: All pages completion time

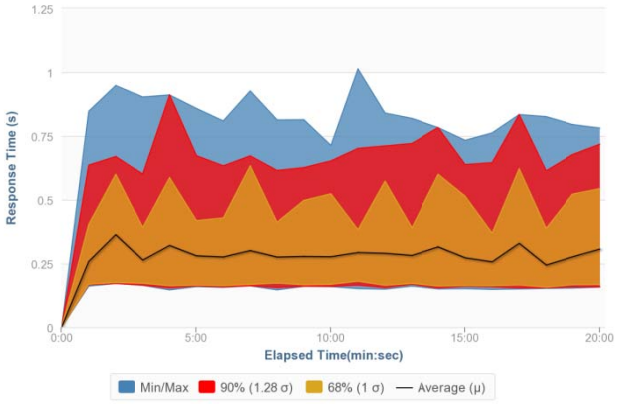

Figure 4: Overall response time

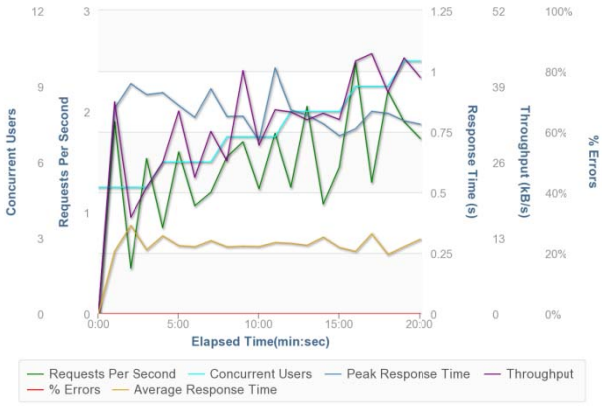

Figure 5: Summery of load testing

6. Conclusion

Cloud Computing has widely been adopted by the industry, though there are many subsisting issues like Server Consolidation, Load Balancing, Energy Management, Virtual Machine Migration, etc. which have not been plenary addressed. Central to these issues is the issue of load balancing, that is required to distribute the excess dynamic local workload evenly to all the nodes in the whole Cloud to achieve a high utilized gratification and resource utilization ratio. It withal ascertains that every computing resource is distributed efficiently and fairly. Subsisting Load Balancing techniques that have been studied mainly fixate on reducing overhead,

accommodation replication time and ameliorating performance etc., but none of the techniques has considered the execution time of any task at the run time. Therefore, there is a desideratum to develop such load balancing technique that can ameliorate the performance of cloud computing along with maximum resource utilization.

7. Future Scope

Future plan is to offer security algorithm with load balancing.

References

[1] Chun-Cheng Lin, Hui-Hsin Chin, Der-Jiunn Deng, "Dynamic Multiservice Load Balancing in Cloud-Based Multimedia System", 1932-8184/$31.00 © 2013 IEEE, DOI 10.1109/JSYST.2013.2256320.
[2] Yinchuan Deng, Rynson W.H. Lau, "On Delay Adjustment for Dynamic Load Balancing in Distributed Virtual Environments", IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 18, NO. 4, APRIL 2012
[3] Daniel Warneke, Odej Kao, "Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 22, NO. 6, JUNE 2011
[4] L.D. Dhinesh Babua, P. Venkata Krishna, "Honey bee behavior inspired load balancing of tasks in cloud computing environments", SciVerse ScienceDirect's Applied Soft Computing, ASOC 1894 1–12, © 2013 Elsevier B.V.
[5] Giuseppe Aceto, Alessio Botta, Walter de Donato, Antonio Pescapè, "Cloud monitoring: A survey", SciVerse ScienceDirect's Computer Networks 57, PP-2093–2115, ASOC 1894 1–12, © 2013 Elsevier B.V.
[6] Mladen A. Vouk, "Cloud Computing – Issues, Research and Implementations", Proceedings of the ITI 2008 30th Int. Conf. on Information Technology Interfaces, June 23-26, 2008, Cavtat, Croatia
[7] J. Sahoo, S. Mohapatra and R. lath "Virtualization: A survey on concepts, taxonomy and associated security issues" computer and network technology (ICCNT), IEEE, pp. 222-226. April 2010.
[8] G. Pallis, "Cloud Computing: The New Frontier of Internet Computing", IEEE Journal of Internet Computing, Vol. 14, No. 5, September/October 2010, pages 70-73.
[9] A. Khiyati, M. Zbakh, H. El Bakkali, D. El Kettani "Load Balancing Cloud Computing: State Of Art", IEEE, 2012.

Author Profile

Samarsinh Prakash Jadhav pursuing Masters of Engineering in Information Technology from MITCOE, Pune University, India. He has near about 4 years of IT industry experience.

Prof. Priya R. Deshpande received the Masters of Engineering in Information Technology degree and having 7.6 years of experience. She is working as Assistant Professor in MITCOE, Pune University, India. Her specialization is in Service Oriented Architecture, Database Management Systems, Software Architecture, and Distributed Systems.