Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 6, June 2014, pg.523 – 527

SURVEY ARTICLE

Survey: Advanced Load Balancing Algorithms in Cloud Computing Environment

Er. Amit Batra[1], Priya[2]

[1]Assistant Professor, Department of Computer Science & Engineering, HCTM Technical Campus, Kaithal, India

Email-Id: amitbatra2011@gmail.com

[2] Research Scholar M.Tech., Department of Computer Science & Engineering, HCTM Technical Campus, Kaithal, India

Email-Id: priya17122014@gmail.com

ABSTRACT

Cloud computing has become very popular in last few years. As a part of its services, it provides a flexible and easy way to keep and retrieve data and files. In cloud computing load balancing is a key issue. It would consume a lot of cost to maintain load information. Good load balancing makes cloud computing more efficient and improves user satisfaction. Many algorithms were suggested to provide efficient mechanisms and algorithms for assigning the client's requests to available Cloud nodes. These approaches aim is to enhance the overall performance of the Cloud. In this paper we present the different load balancing algorithms. We discuss these algorithms to provide an overview of the latest approaches in the field.

INTRODUCTION

Cloud computing is a recent technology that concern with online distribution of computing resources and services. In cloud computing, end-user knowledge about the configuration of service delivering system may not be required because client just use services on pay per model

where all system configuration and resource management is taken care by cloud system automatically [13].
One important issue associated with this field is dynamic load balancing or task scheduling. Load balancing algorithms were investigated heavily in various environments. Cloud Computing the main concerns involve efficiently assigning tasks to the Cloud nodes such that the effort and request processing is done as efficiently as possible, while being able to tolerate the various affecting constraints such as heterogeneity and high communication delays [9].

Cloud computing has emerged as a buzzword in the commercial and academic world, for its great potential to fulfill the envisioned blueprint that customers can enjoy computing infrastructure and services in a pay-as-you-go manner[7]. Generally clouds give customers three levels of access: Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS)[5]. Cloud computing is efficient and scalable but maintaining the stability of processing so many jobs in the cloud computing environment is a very complex problem with load balancing receiving much attention for researchers [16]. Cloud computing is an emerging computing model based on the development of distributed computing, parallel processing and grid computing [11].

Cloud computing system comprises of several servers, virtual machines, data centers, storage devices etc which are interconnected in an efficient way. Nowadays, computing systems heavily rely on Virtualization technology and thus makes the servers feasible for independent applications. Further, virtualization process improves the power efficiency of the datacenters (consolidation of servers) and thereby enabling the assignment of multiple virtual machines (VMs) to a single physical server [14].

In cloud system, virtualization plays a very important role by allowing online sharing of computing resources [13]. Clouds use virtualization technology in distributed data centers to allocate resources to customers as the need them [5]. Virtualization support in cloud allows better flexibility and customization to specific application, software, and programming environment needs of HPC users [12]. Virtualization technologies enable application computation and data to be hosted inside virtual containers (e.g. virtual disks) which are decoupled from the underlying physical resources. These virtualization-based clouds provide a way to build a large computing infrastructure by assessing remote computational, storage and network resources [4].

Load balancing is the process of improving the performance of a parallel and distributed system through a redistribution of load among the processors or nodes.
Load balancing is described in as follows "In a distributed network of computing hosts, the performance of the system can depend crucially on dividing up work effectively across the participating nodes" [3]. Load balancing is the task to be distributed among multiple computers, processes, disk, or other resources in order to get optimal resource utilization and to reduce the computation time.
Load balancing is an important means to achieve effective resource sharing and utilization. It has been the hot issue of distributed computing, grid computing and cloud computing research. Load balancing has two meanings: first, it puts a large number of concurrent accesses or data traffic to multiple nodes respectively to reduce the time users waiting for response; second, it put the

calculation from a single heavy load to the multiple nodes to improve the resource utilization of each node [11].

Load balancing algorithm is used to distribute the load among various nodes in the distributed system to improve the resource utilization and request response time of the system. These algorithms are mainly used to overcome the situation
where a node is heavily loaded and other nodes are idle and because of which the request fails[10].

LITERATURE SURVEY

A Scheduling Strategy on Load Balancing of Virtual Machine Resources in Cloud Computing Environment: Jinhua Hu, Jianhua Gu, Guofei Sun, Tianhai Zhao proposed a scheduling strategy on load balancing of VM resources based on genetic algorithm. According to historical data and current state of the system and through genetic algorithm, this strategy computes ahead the influence it will have on the system after the deployment of the needed VM resources and then chooses the least-affective solution, through which it achieves the best load balancing and reduces or avoids dynamic migration. This strategy solves the problem of load imbalance and high migration cost by traditional algorithms after scheduling. Experimental results prove that this method is able to realize load balancing and reasonable resources utilization both when system load is stable and variant.

Towards a Load Balancing in a Three-level Cloud Computing Network: Shu-Ching Wang performed a work that used low-power hosts to achieve high reliability. This approach was about to utilize the computing resources on the network to facilitate the execution of complicated tasks that require large-scale computation. The proposed scheduling algorithm combines 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.

Load Balancing in Cloud Computing Using Modified Throttled Algorithm: Shridhar G.Domanal and G.Ram Mohana Reddy proposed an efficient approach to handle the load at servers by considering both availability of VMs for a given request and uniform load sharing among the VMs for the number of requests served. The work aimed at efficient method for load balancing, depicted from its two different objectives. One being the response time required to serve the requests and other being the distribution of load among the existing VMs. When compared to existing Round-Robin and Throttled algorithms, the response time for proposed algorithm has improved considerably, evident from the results presented in section V. Distribution of load among the virtual machines in Round-Robin algorithm was nearly uniform, but was found less efficient considering response time. Throttled algorithm with better response time than Round-Robin failed to distribute load uniformly, overloading initial VMs and leaving others underutilized. Proposed algorithm distributes load nearly uniform among VMs, with improved response time compared to existing algorithms. Simulation results have demonstrated that the proposed algorithm has distributed the load uniformly among virtual machines.

Cloud Server Optimization with Load Balancing and Green Computing Techniques Using Dynamic Compare and Balance Algorithm: Yatendra Sahu, R.K. Pateriya, Rajeev Kumar Gupta proposed a
threshold based Dynamic compare and balance algorithm (DCABA) for cloud server optimization. Unlike the traditional server optimization strategies which consider only load balancing and scheduling of resources based on the usage of CPU, RAM and BW in physical servers, DCABA also minimizes the number of host machines to be powered on, for reducing the cost of cloud services. Our approach can serve the purpose of service cost reduction in cloud industry with effective utilization of available resources.

Trust and Reliability based Load balancing Algorithm for Cloud IaaS: Punit Gupta, Mayank Kumar Goyal, Prakash Kumar proposed a suitable trust model based on the existing model that is suitable for trust value management for the cloud IaaS parameters. Based on the above achieved trust values, a suitable load balancing algorithm is proposed for better distribution of load which further enhance the QOS of services being provided to the users. Other algorithms do not consider the property of VMM but it has not taken into consideration the properties of a VMM in a datacenter. So we propose a trust management model to overcome this problem, by taking into consideration VMM characteristics which vary from datacenter to datacenter. Then these trust value are been used by load balancing algorithm proposed to improve the QOS provided to the user and better utilization of resources.

The Load Balancing Algorithm in Cloud Computing Environment Haozheng Ren performed a dynamic load balancing algorithm based on virtual machine migration under cloud computing environment. The algorithm proposed the trigger strategy based on the fractal methods. The strategy determines the timing of the virtual machine migration through forecasting the timing to determine the timing of the virtual machine migration.

AN ONLINE LOAD BALANCING SCHEDULING ALGORITHM FOR CLOUD DATA CENTERS CONSIDERING REAL-TIME MULTI-DIMENSIONAL RESOURCE: Minxian Xu performed an online load balancing resource scheduling algorithm (OLRSA) for Cloud data centers considering real-time and multi-dimensional resources. Author develops and applies integrated measurement for each server and a Cloud data center.

A Hybrid Dynamic Load Balancing Approach for Cloud Storage Yilin Lu performed a hybrid control strategy for load balancing. On the one hand the storage node cluster redistributes the load in its local range. Author also present a dynamic migrating strategy caused by two reasons: overload access or long response time in some storage nodes.

CONCLUSION

This paper is based on cloud computing technology which has a very vast potential and is still unexplored. One of the major issues of cloud computing is load balancing because overloading of a system may lead to poor performance which can make the technology unsuccessful. So there is always a requirement of efficient load balancing algorithm for efficient utilization of resources. Our paper focuses on the various load balancing algorithms and their applicability in cloud computing environment.

REFERENCES

[1] Shu-Ching Wang," Towards a Load Balancing in a Three-level Cloud Computing Network", 978-1-4244-5540-9/10 ©2010 IEEE

[2] Jinhua Hu," A Scheduling Strategy on Load Balancing of Virtual Machine Resources in Cloud Computing Environment", 3rd International Symposium on Parallel Architectures, Algorithms and Programming 978-0-7695-4312-3/10 © 2010 IEEE

[3] Branko Radojevic," Analysis of Issues with Load Balancing Algorithms in Hosted (Cloud) Environments", MIPRO 2011

[4] Rui Wang," Design and Implementation of an Efficient Load-Balancing Method for Virtual Machine Cluster Based on Cloud Service".

[5] Jeffrey Galloway," An Empirical Study of Power Aware Load Balancing in Local Cloud Architectures", 2012 Ninth International Conference on Information Technology- New Generations 978-0-7695-4654-4/12 © 2012 IEEE

[6] Yilin Lu," A Hybrid Dynamic Load Balancing Approach for Cloud Storage", 2012 International Conference on Industrial Control and Electronics Engineering 978-0-7695-4792-3/12 © 2012 IEEE

[7] Zheng Hu,"An Utility- Based Job Scheduling Algorithm For Current Computing Cloud Considering Reliability Factor".

[8] Minxian Xu," AN ONLINE LOAD BALANCING SCHEDULING ALGORITHM FOR CLOUD DATA CENTERS CONSIDERING REAL-TIME MULTI-DIMENSIONAL RESOURCE", IEEE CCIS2012 978-1-4673-1857-0/12 ©2012 IEEE

[9] Klaithem Al Nuaimi," A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms", 2012 IEEE Second Symposium on Network Cloud Computing and Applications 978-0-7695-4943-9/12 © 2012 IEEE

[10] Punit Gupta," Trust and Reliability based Load Balancing Algorithm for Cloud IaaS", 978-1-4673-4529-3/12@ 2012 IEEE

[11] Haozheng Ren," The Load Balancing Algorithm in Cloud Computing Environment",2012 2nd International Conference on Computer Science and Network Technology 978-1-4673-2964-4/12 ©2012 IEEE

[12] Abhishek Gupta," Improving HPC Application Performance in Cloud through Dynamic Load Balancing", 2013 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing 978-0-7695-4996-5/13© 2013 IEEE

[13] Yatendra Sahu," Cloud Server Optimization with Load Balancing and Green Computing Techniques Using Dynamic Compare and Balance Algorithm", 2013 5th International Conference on Computational Intelligence and Communication Networks 978-0-7695-5069-5/13 © 2013 IEEE

[14] Shridhar G.Domanal," Load Balancing in Cloud Computing Using Modified Throttled Algorithm".

[15] Zhenzhang,"A Model Based Load Balancing Method in Iaas Cloud".

[16] Gaochao Xu," A Load Balancing Model Based on Cloud Partitioning for the Public Cloud", TSINGHUA SCIENCE AND TECHNOLOGY ISSNl l1007-0214, 2013

