International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 5 Issue 2, February-2016, Impact Factor: 1.544

A Literature Review of Various Load Balancing techniques in Cloud Computing Environment

Amit Garg[1]
Prof. Kailash Patidar (HOD)[2], Prof. Gaurav Kumar Saxena[3], Prof. Megha Jain[4]

ABSTRACT

In modern days cloud computing is one of the greatest platform which provides storage of data in very lower cost and available for all time over the internet. But the cloud computing has more critical issue like security, load balancing and fault tolerance ability. In this paper we are focusing on Load Balancing approach. The Load balancing is the process of distributing load over the different nodes which provides good resource utilization when nodes are overloaded with job. Load balancing is required to handle the load when one node is overloaded. When the node is overloaded at that time load is distributed over the other ideal nodes. Many load balancing algorithms are available for load balancing like Static load balancing and Dynamic load balancing. The survey of modern load balancing algorithm is presented in this paper. The Load balancing is the process of distributing load over the different nodes which provides good resource utilization when nodes are overloaded with job. Load balancing is required to handle the load when one node is overloaded. When the node is overloaded at that time load is distributed over the other ideal nodes. Many load balancing algorithms are available for load balancing like Static load balancing and Dynamic load balancing.

Keywords: Cloud Computing, virtualization, Load balancing.

INTRODUCTION TO CLOUD COMPUTING

CLOUD COMPUTING [1,2] is one of the most emerging and new way of computer science engineering, where flexible environment gives number of user can get access desired services as per their requirement where any information may available at anytime and anywhere in world. So there is lots of possibilities arisen to access public and private information by using internet. User needs to be begin to use computing services at remote location store the information in private cloud for confidentiality and share in public cloud. It is refers to distributed architecture that provides computing resources over the internet. Than name of the cloud computing services given because cloud is metaphor for internet, where user can see the cloud but doesn't know what inside it. This provides services pay per consumption basis, reduce the cost of operating system and networks. No need to purchase hardware and software licenses and other benefits to unlimited processing power and storage capacity, high efficiency. While the cloud virtually enables network access and services that combines various distributed resources all over the world, security is an important issue to be dealt in order to protect outsourced data accessed via third-party clouds from network intruders. Hence, the main focus of this paper is to develop a model to securely transfer data with all the three main cloud service layers such as Infrastructure as a service (Iaas), Platform as a service (Paas) and Software as a service (Saas).

Figure 1.1 Cloud Computing Scenario

When any enterprise hires a cloud, security, trust and privacy is the main concerns in public cloud because the data is no longer in their hands. There is no firewalls mechanism, perimeter anymore, IDS/IPS at the internal gateway stopping dishonest user from attacks. Virtual private cloud (VPC) inside public cloud using virtualization have made safe surface from attackers. Virtual private cloud looking for ways to compromise sensitive data or processes exists inside public cloud using specific addressing.

LOAD BALANCING IN CLOUD COMPUTING

The load balancing is the process of distributing the load among various resources in any system. Therefore load need to be distributed over the resources in cloud-based architecture so that each resources does approximately the equal amount of task at any point of time. The basic need is to provide some techniques to balance requests to provide the solution of the application faster. All cloud vendors are based on automatic load balancing services, it allows clients to increase the number of CPUs or memories for their resources to scale with increased demands. These services are optional and depend on the clients business needs. So the load balancing serves two important needs, firstly to promote availability of Cloud resources and secondarily to promote performance.

In order to balance the resources it is important to recognize a few major goals of load balancing algorithms:

a) Cost effectiveness: first aim is to achieve an overall improvement in system performance at a reasonable cost.
b) Scalability and flexibility: distributed system in which the algorithm is implemented may change in size or topology So the algorithm must be scalable and flexible enough to allow such changes to be handled easily.
c) Priority: scheduling of the resources or jobs need to be done on before hand through the algorithm itself for better service to the important or high prioritized jobs in spite of equal service provision for all the jobs regardless of their origin.

LITERATURE SURVEY

The work done by A. Singh et al. [3] proposed a novel load balancing algorithm called VectorDot. This algorithm handles the hierarchical complexity of the datacenter and multidimensionality of resource loads across servers network switches and storage in an agile data center that has integrated server and storage virtualization technologies.

The work done by Stanojevic et al. [4]proposed a mechanism CARTON for cloud control that unifies the use of LB and DRL. The LB (Load Balancing) is used to equally distribute the jobs to different servers so that the associated costs can be minimized and DRL (Distributed Rate Limiting) is used to make sure that the resources are distributed in a way to keep a fair resource allocation.

Author Y. Zhao et al. [5] addressed the problem of intra-cloud load balancing amongst physical hosts by adaptive live migration of virtual machines. The load balancing model is designed and implemented to reduce virtual machines migration time by shared storage to balance load amongst servers according to their processor or IO usage.

Work done by V. Nae et al. [6] presented an event driven load balancing algorithm for real-time Massively Multiplayer Online Games (MMOG). The algorithm after receiving capacity events as input, also analysis its components in context of the resources and the global state of the game session, then generating the game session load balancing actions.

The J. Hu et al. [7] proposed a scheduling strategy on load balancing of VM resources that uses historical data and current state of the system. Proposed strategy achieves the best load balancing and reduced dynamic migration by using a genetic algorithm.

The A. Bhadani et al. [8] proposed a Central Load Balancing Policy for Virtual Machines (CLBVM) that balances the load evenly in a distributed virtual machine/cloud computing environment.

The LBVS H. Liu et al. [9] proposed a load balancing virtual storage strategy (LBVS) that provides a large scale net data storage model and Storage as a Service model based on Cloud Storage. The Storage virtualization is achieved using an architecture that is three-layered and load balancing is achieved using two load balancing modules. It helps in improving the efficiency.

The Y. Fang et al. [10] discussed a two-level task scheduling mechanism based on load balancing to meet dynamic requirements of users and obtain high resource utilization. Algorithm achieves load balancing by first mapping tasks to

virtual machines and then virtual machines to host resources thereby improving the task response time, and resource utilization also overall performance of the cloud computing environment.

Author M. Randles et al. [11] investigated a decentralized honey bee based load balancing technique that is a nature inspired algorithm for self-organization. Algorithm achieves global load balancing through local server actions. Performance of the system is enhanced with increased system diversity but throughput is not increased with an increase in system size. This is best suited for the conditions where the diverse population of service types is required.

The work done by M. Randles et al. [11] investigated a distributed and scalable load balancing approach that uses random sampling of the system domain to achieve self-organization thus balancing the load across all nodes of the system.

Author M. Randles et al. [11] investigated a self-aggregation load balancing technique that is a self-aggregation algorithm to optimize job assignments by connecting similar services using local re-wiring. Overall performance of the system is enhanced with high resources thereby in-creasing the throughput by using these resources effectively.

The authors in [12, 19,20,21] proposed a load balancing mechanism based on ant colony and complex network theory (ACCLB) in an open cloud computing federation. Proposed algorithm uses small-world and scale-free characteristics of a complex network to achieve better load balancing. Proposed technique overcomes heterogeneity is adaptive to dynamic environments and has good scalability hence helps in improving the performance of the system.

Authors [13,17,18] proposed a two-phase scheduling algorithm that combines OLB (Opportunistic Load Balancing) and LBMM (Load Balance Min-Min) scheduling algorithms to utilize better executing efficiency and maintain the load balancing of the system. This OLB scheduling algorithm keeps every node in working state to achieve the goal of load balance and LBMM scheduling algorithm is utilized to minimize the execution time of each task on the node thereby minimizing the overall completion time.

Author [14,18,19] Proposed a new content aware load balancing policy named as work-load and client aware policy (WCAP). Proposed work uses a parameter named as USP to specify the unique and special property of the requests as well as computing nodes. The USP helps the scheduler to decide the best suitable node for processing the requests.

Author s [15,16,17] proposed a Join-Idle-Queue load balancing algorithm for dynamically scalable web services. Work provides large-scale load balancing with distributed dispatchers by, first load balancing idle processors across dispatchers for the availability of idle processors at each dispatcher and then, assigning jobs to processors to reduce average queue length at each processor.

CONCLUSION

In this paper we have proposed a survey of load balancing methods. In cloud computing load balancing is one of the main issue. When client is requesting for service it should be available to the client. When any node is overloaded with job at that time load balancer has to set that load on another free node. Therefore load balancing is necessary in cloud computing. so in this thesis we have discussed all the existing techniques for Load balancing. Load balancing is to increase client satisfaction and maximize resource utilization and substantially increase the performance of the cloud system and minimizing the response time and reducing the number of job rejection.

REFERENCES

[1] John Harauz, Lorti M. Kaufinan. Bruce Potter, "Data Security in the World of Cloud Computing", IEEE Security & Privacy, Copublished by the IEEE Computer and Reliability Societies, July/August 2009.
[2] National Institute of Standards and Technology- Computer Security Resource Center -www.csrc.nist.gov
[3] Singh A., Korupolu M. and Mohapatra D., ACM/IEEE conference on Supercomputing, 2008.
[4] Stanojevic R. and Shorten R., IEEE ICC, 1-6, 2009.
[5] Zhao Y. and Huang W., 5th International Joint Confer-ence on INC, IMS and IDC, 170-175, 2009.
[6] Nae V., Prodan R. and Fahringer T., 11th IEEE/ACM International Conference on Grid Computing (Grid), 9-17,2010.
[7] Hu J., Gu J., Sun G. and Zhao T., 3rd International Symposium on Parallel Architectures, Algorithms and Programming, 89-96, 2010.
[8] Bhadani A. and Chaudhary S., 3rd Annual ACM Bangalore Conference, 2010. [9] Liu H., Liu S., Meng X., Yang C. and Zhang Y.,International Conference on Service Sciences (ICSS), 257-262,2010.
[9] Fang Y., Wang F. and Ge J.,Lecture Notes in Computer Science, 6318, 271-277,2010.
[10] Randles M., Lamb D. and Taleb-Bendiab A., 24th International Conference on Advanced Information Networking and Applications Workshops, 551-556,2010.

[11] Zhang Z. and Zhang X, 2nd International Conference on Industrial Mechatronics and Automation, 240-243, 2011.
[12] Wang S., Yan K., Liao W. and Wang S, 3rd International Conference on Computer Science and Information Technol-ogy, 108-113, 2010.
[13] Mehta H., Kanungo P. and Chandwani M., International Conference Workshop on Emerging Trends in Technology, 370-375, 2011.
[14] Lua Y., Xiea Q., Kliotb G., Gellerb A., Larusb J. R. and Green-ber A,"Int. Journal on Performance evaluation",2011.
[15] Yashpalsinh Jadeja and Kirit Modi, "Cloud Computing - Concepts, Architecture and Challenges", International Conference on Computing, Electronics and Electrical Technologies [ICCEET], IEEE-2012.
[16] Samerjeet kaur, "Cryptography and Encryption in Cloud Computing", VSRD International Journal of Computer Science and Information Technology, VSRDIJCSIT, Vol. 2 (3), 2012.
[17] Chung-Cheng Li and Kuochen Wang, An SLA aware load balancing scheme for cloud datacenters, Information Networking (ICOIN), 2014 International Conference, Feb 2014, Pages:58-63.
[18] Gulshan Soni, Mala Kalra, A Novel Approach for Load Balancing in Cloud Data Center, Advance Computing Conference (IACC), 2014 IEEE International , Feb 2014, Pages:807-812.
[19] Cristian Klein, Alessandro Vittorio Papadopoulos, Manfred Dellkrantz, Jonas Durango, Martina Maggio, KarlErik Arzen, Francisco Hernandez-Rodriguez, Erik Elmroth, Improving Cloud Service Resilience using Brownout-Aware Load-Balancing, Reliable Distributed Systems (SRDS), 2014 IEEE 33rd International Symposium, Oct 2014, Pages:31-40.
[20] Tao Wang, Xin Lv, Fang Yang, Wenhuan Zhou, Rongzhi Qi, HuaiZhi Su, A load balancing scheme for distributed key-value caching system in cloud environment, Distributed Computing and Applications to Business, Engineering and Science (DCABES), 2014 13th International Symposium, Nov 2014, Pages:63-67.