Improving Efficiency of Private Cloud Using a Load Balancing Model: A Survey

Richa Smita Nayak[1] , Prof Dr. J. A. Laxminarayana[2]

[1](Computer Department, Goa College Of Engineering/ Goa University, India)
[2](Computer Department, Goa College Of Engineering/ Goa University, India)
Corresponding Auther: Richa Smita Nayak

Abstract*: Load balancing in cloud computing has played a pioneer role in improving the efficiency. Over a period of time it has changed in a better manner and has had a significant impact on various system parameters. The most important among them being the performance. Load balancing makes efficient use of resources and thus helps in achieving higher user satisfaction. We present a survey of the research contributions regarding a load balancing model which will attempt to disseminate the load evenly across all the nodes via virtualization and thereby promote dynamism and flexible scaling.*
Keywords: *dynamism , efficiency, load balancing, scaling, virtualization*

I.   Introduction

Load balancing in cloud computing is a boon to the field of computer science and technology. Taking all major aspects and metrics into account it has proven its worth in the IT industry. It has become one of the means in which various highly virtualized resources can be managed effectively. Virtualization is the creation of a virtual version of a device or resource, such as server, storage device, network or even an operating system where the framework divides the resource into one or more execution environments. It is capable of handling a huge amount of data and thus caters to the customers need by making services accessible to them with flexible scaling. It also provides a smart framework that encourages dynamism of resources making it more versatile. When we talk about cloud storage it usually refers to the distribution of storage resources and services over the internet. Private cloud storage deals with storage resources that is confined to a particular organization or a company. Therefore the technical risk factors that may arise is very low and security is enhanced as a result of it. Hence, the work in private cloud is facilitated by exploiting the commodity machines and the information available within the organization. As the usage of such private storage and services increase, the demand for storage also intensifies. It leads to the expansion of the cloud storage with additional storage nodes. As there is an expansion of cloud storage nodes, proper balancing of storage nodes in terms of load has to be carried out . In order to maintain the load evenly across all the nodes, the data needs to be disseminated across the nodes. This migration of data may result in the consumption of more network bandwidth.

A lot of research has been done in cloud computing field related to this domain. Good load balancing makes cloud computing more efficient and also improves user satisfaction. There have been many studies of load balancing for the cloud environment. Numerous techniques have been proposed by researchers in the field of cloud computing. It is not feasible to describe all. The major sources of literature are Google Scholar, IEEE Xplore, Science Direct and various other journals.

A comprehensive survey regarding load balancing issues is conducted and presented in this paper. In what follows, section II explains various cloud partitioning techniques. This will be followed by section III and section IV which will explain the various load balancing algorithms and the applications associated to it. Relevant studies of which are briefly mentioned here.

II.   Cloud Partitioning Techniques

There have been many studies of load balancing for the cloud environment. [1] and [3] presented various techniques on cloud partitioning. [1] has a system having main controller, balancers, servers and a client is implemented. It introduces a switch mechanism to choose different strategies for different situations. This paper divides the cloud into cloud partitions and applies different strategies to balance the load on cloud. The load balancing is taken care by the main controller and the balancers. The main controller takes the initiative of which partitioning should get the job and then the balancers decide upon job assignment strategy to the respective nodes. This gives a better idea for balancing the load on clouds. It thus helps to avoid overloading of servers and improve response times. The proposed method in [3] determines the partitioning of the cloud based

on the load status of the system. If the load status of the system is idle then a simple load balancing strategy is adopted wherein a simple round robin algorithm based on load degree evaluation is used. Here the nodes in the load balancing table based on load degree is put in the ascending order . A circular queue is built by the system and it walks through it. The jobs are allocated to the nodes having low load degrees. On refreshment of the load by the balancer the node order will be changed.

III. Load Balancing Algorithms

There are many load balancing algorithms, such as Round Robin, Equally Spread Current Execution Algorithm, and Ant Colony algorithm. In many research works like [2],[7],[9],and [10] an insight to many load balancing algorithms have been provided.

A noble way of identifying various resources and metrics of load balancing is presented in [2]. It also discusses about the various types of load balancing algorithms. Some of them include Carton, Compare and Balance, Event-driven, Biased random sampling, Active clustering and many more. Carton deals with distributing the jobs in such a manner so that the cost is minimized. This is effective because there is less computation the computation overhead is very less. Compare and Balance addresses the problem encountered in intra-cloud load balancing among the physical host by employing the concept of adaptive live migration of virtualized machines. Event-driven considers capacity events as input and analyses its components accordingly to carry out load balancing actions. The Biased random sampling employs the scalable and distributed load balancing approach by performing random sampling of the system domain. The performance of the system is improved ensuring increased throughout. However, load balancing in the cloud is still a new problem that needs new architectures to adapt to many changes.

Load balancing in cloud computing was described in a white paper written by Adler[7] who introduced the tools and techniques commonly used for load balancing in the cloud. There are many load balancing techniques which are commonly available. Nishant et al. The work presented in [9] used the ant colony optimization method in nodes load balancing. It provides an insight to ant colonies that work in collaboration in foraging behavior. This idea explains how ants work collaboratively to search the new food sources and also use the food sources that are existing to shift the food back to the nest. The traversal (forward or backward) of the ants is also tracked as they leave a trail upon moving from one node to the other. This becomes really effective in detecting the overloaded and underloaded nodes and performing operations accordingly on the identified nodes. The traversal as mentioned earlier is of two types. The forward movement focuses on ants movement for food extraction or searching for the food sources. So the ants here move in forward direction continuously in cloud encountering over loaded and under loaded node. Whereas in backward movement the ants move back to the nest for storing food after picking up food from the food sources. So if an ant encounters an overloaded node during its movement when it has previously encountered an under loaded node then it goes back to the under loaded node to check if the node is still under loaded or not and if it finds that it is still under loaded then it redistributes the work to the under loaded node. The vice-versa is also feasible and possible.

Randles et al. [10] gave a compared analysis of some algorithms in cloud computing by checking the performance time and cost. They concluded that the ESCE algorithm and throttled algorithm are better than the Round Robin algorithm. Some of the classical load balancing methods are similar to the allocation method in the operating system, for example, the Round Robin algorithm and the First Come First Served (FCFS) rules. The Round Robin algorithm is used here because it is fairly simple.

IV. Applications

Chaczko et al. [8] described the role that load balancing plays in improving the performance and maintaining stability. It discusses about the load balancing in distributed systems and Message Oriented Model. An application of Hospital Data Management (HDM) is also discussed. It aims to enhance the efficiency of the data retrieval from multiple databases with respect to cloud infrastructure. So the HDM aims to provides the following functionalities:

- It collects usage patterns from the past data
- Forecasts usage patterns which can be used in future.
- Scales database requirements proactively.
- Compresses the data before sending/storing.
- Tracks data transfer.

Proactive scaling and compression has an impact on HDM and effectively help in load balancing of the databases.

V. Comparison

On the basis of comprehensive analysis certain comparison can be made. Some of them have been listed below :

- Biased Random Sampling and Active clustering perform really work well when the number of nodes is increased or in other words when the number of resources is high. It thereby ensures effective utilization of resources. It offers throughput , performance and scalability.
  However, they don't perform well as the system diversity increases thereby adding to its demerit.
- Carton can be implemented easily if the communication overhead is less and also when there is low computation. This way it ensures fair allocation of resources and the cost is also reduced as a result of it. It offers better performance, more resource utilization and low overheads.
- Event driven helps in establishing game sessions. Based on the load of the user it can scale the resources. So the Quality of Service is not ensured here as the quality breaches . However, it offers high resource utilization.
- Round robin ensures fair allocation of resources by making sure that all processes are given same priority. However, the improved version of Round Robin eliminates overheads and thus is more consistent.
- Message Oriented Model plays an important role in real time communication. It is an open technology and improves load balancing in distributed networks. It offers good response time and better performance.

This comparison made evidently gives an insight to choose the algorithms effectively depending on the scenario taken into consideration. It can be used accordingly to load balance the systems actively.
The aggregation of these algorithms depending on the compatibility can also offer better performance. However the simulation of it can be a tedious process.

VI. Conclusion

One of the important issues addressed in private cloud computing is load balancing. It helps in effective utilization of resources and ensure better performance. A thorough study on techniques of cloud partitioning was done and a study on various resources, factors and metrics was made. Also, a thorough analysis on various effective algorithms was done. It gave an insight to the techniques on how balancing of load can be carried out successfully and comparison of the algorithms was made. It provides better strategies which incorporate better scheduling and allocation of resources.

References

[1]. Gaochao Xu, Junjie Pang, and Xiaodong Fu **"**A Load Balancing Model Based on Cloud Partitioning for the Public Cloud".
[2]. Zeinab Goudarzi, Ahmad Faraahi, "Effective Load Balancing in Cloud Computing. International Journal of Intelligent Information Systems", Special Issue: Research and Practices in Information Systems and Technologies in Developing Countries. Vol. 3, No. 6-1, 2014, pp. 1-9.
[3]. CH. VISHNU PRIYA, J. PRAVEEN KUMAR, "Analysis of the Public Cloud Based on Cloud Partitioning using Load Balancing Model", ISSN 2348–2370 Vol.06,Issue.06, September-2014,Pages:510-515
[4]. GoogleTrends,Cloudcomputing,http://www.google.com/trends/explore#q=cloud%20computing,2012
[5]. N. G. Shivaratri, P. Krueger, and M. Singhal, Load distributing for locally distributed systems, Computer, vol. 25, no. 12, pp. 33-44, Dec. 1992.
[6]. P. Mell and T. Grance, The NIST definition of cloud computing, http://csrc.nist.gov/ publications/nistpubs/800145/SP800-145.pdf, 2012.
[7]. B. Adler, Load balancing in the cloud: Tools, tips and techniques, http://www.rightscale.com/infocenter/whitepapers/Load-Balancing-in-the-Cloud.pdf, 2012
[8]. Z. Chaczko, V. Mahadevan, S.Aslanzadeh, and C Mcdermid, Availability and load balancing in cloud computing,presentedatthe2011InternationalConference on Computer and Software Modeling, Singapore, 2011.
[9]. K. Nishant, P. Sharma, V. Krishna, C. Gupta, K. P. Singh, N. Nitin, and R. Rastogi, Load balancing of nodes in cloud using ant colony optimization, in Proc. 14th International Conference on Computer Modelling and Simulation (UKSim), Cambridgeshire, United Kingdom, Mar. 2012, pp. 28-30.
[10]. M. Randles, D. Lamb, and A. Taleb-Bendiab, A comparative study into distributed load balancing algorithms for cloud computing, in Proc. IEEE 24th International Conference on Advanced Information Networking and Applications, Perth, Australia, 2010, pp. 551-556.