ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

PROPOSED ALGORITHM FOR LOAD BALANCING ON CLOUD DATA CENTER

Pranav Bhokardankar*, Harshal Sharma

[*] Computer science & engineering, indore institute of science & technology, indore, (m.p.), india
Information technology, acropolis institute of technology & research, indore (m.p.), india

DOI: 10.5281/zenodo.218895

ABSTRACT

Cloud servers are basically developed for supporting the efficient communication and the computation remotely. Thus a number of different time zone machines are able to find the services of resources and the computation. Thus each and every time these machines are working in order to resolve the end client request. But sometime these machines are loaded more then their capacity and does not perform as desired and some of the available processing units are considered to be free. This results in the performance loss in the computational servers. In order to find the optimum performance there is need to implement some technique for load balancing. Thus in this presented work the key focus is placed on the investigation of load balancing approaches and the algorithm. In order to find the optimum technique of load balancing, various different algorithms are studied. Among them four most promising techniques are selected which are promising for load balancing in previous studies. These techniques are genetic algorithm which provides the optimization technique for finding most appropriate solution among available solutions, ABC (artificial honey bee colony) algorithm which is also an optimization technique in the similar ways, the round robin which is frequently used for process allocation during CPU scheduling and finally the self-organizing map which is an unsupervised class of machine learning. These methods are implemented and compared on the basis of their performance parameters. The implementation of the proposed comparative study is performed on the JAVA technology and using the Cloud Sim simulator. After simulation of the presented comparative study the performance of the SOM algorithm provides the optimum results as compared to the Genetic algorithm, Honey bee colony and RR (round robin) algorithms.

INTRODUCTION

The cloud infrastructure is invented to deal with the huge number of request and provide efficient computational environment for different aspects of user applications. Thus a significant amount of request is made over the cloud servers and most of the time that busy in resolving the request of users of different time zones. This presented work is an investigation and demonstration of the workload arising on the server and at the same time the different load balancing techniques that utilized to overcome conflicts on server workloads. The given chapter provides the overview of the proposed system and objects which are required to accomplish.

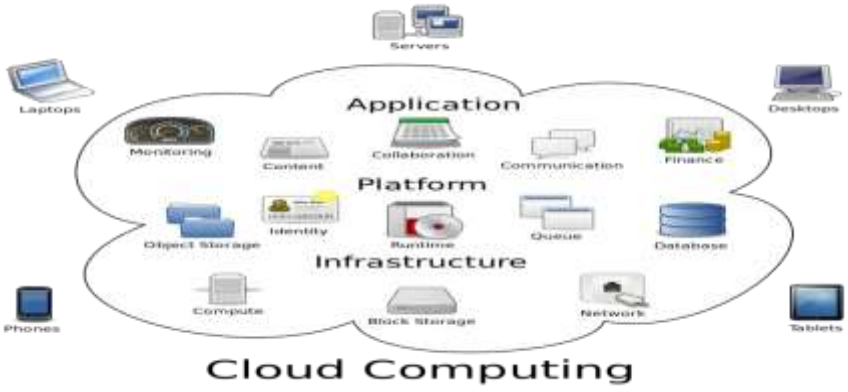

Figure 1.1 cloud computing

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

LOAD BALANCING IN CLOUD

Load balancing is a computer networking method for distributing workloads across multiple computing resources, such as computers, a computer cluster, network links, central processing units or disk drives. Availability is a reoccurring and a growing concern in software intensive systems. Fundamentally, its role is to determine the time that the system is up and running correctly; the length of time between failures and the length of time needed to resume operation after a failure. Availability needs to be analyzed through the use of presence information, forecasting usage patterns and dynamic resource scaling. Load balancing is usually provided by dedicated software or hardware, such as a multilayer switch or a Domain Name System server process.

Load balancing Parameters

Load balancing aims to optimize resource use, maximize throughput, and minimize response time estimation of load, comparison of load, stability of different system, performance of system, interaction between the nodes, nature of work to be transferred and selecting of nodes. A basic example of load balancing in our daily life can be related to websites. Without load balancing, users could experience delays, timeouts and possible long system responses. Load balancing solutions usually apply redundant servers which help a better distribution of the communication traffic so that the website availability is conclusively settled. The parameters are briefly described as follows:

1. Throughput: The total number of tasks that have completed execution is called throughput. A high throughput is required for better performance of the system.
2. Associated Overhead: The amount of overhead that is produced by the execution of the load balancing algorithm. Minimum overhead is expected for successful implementation of the algorithm.
3. Fault tolerant: It is the ability of the algorithm to perform correctly and uniformly even in conditions of failure at any arbitrary node in the system.
4. Migration time: The time taken in migration or transfer of a task from one machine to any other machine in the system. This time should be minimum for improving the performance of the system.
5. Response time: It is the minimum time that a distributed system executing a specific load balancing algorithm takes to respond.
6. Resource Utilization: It is the degree to which the resources of the system are utilized. A good load balancing algorithm provides maximum resource utilization.
7. Scalability: It determines the ability of the system to accomplish load balancing algorithm with a restricted number of processors or machines.
8. Performance: It represents the effectiveness of the system after performing load balancing. If all the above parameters are satisfied optimally then it will highly improve the performance of the system.

Load balancing need

The random arrival of load in cloud environment can cause some server to be heavily loaded while other server is idle or only lightly loaded. Equally load distributing improves performance by transferring load from heavily loaded server. Efficient scheduling and resource allocation is a critical characteristic of cloud computing based on which the performance of the system is estimated. In clouds, load balancing, as a method, is applied across different data centres to ensure the network availability by minimizing use of computer hardware, software failures and mitigating recourse limitations. Cloud vendors are based on automatic load balancing services, which allowed entities to increase the number of CPUs or memories for their resources to scale with the increased demands. This service is optional and depends on the entity's business needs. Therefore load balancers served two important needs, primarily to promote availability of cloud resources and secondarily to promote performance. Goals of load balancing include:

- Substantial improvement in performance
- Stability maintenance of the system
- Increase flexibility of the system so as to adapt to the modifications.
- Build a fault tolerant system by creating backups.

CHARACTERISTICS OF CLOUD

Cloud computing has a variety of characteristics, with the main ones being: [4]

- Shared Infrastructure — Uses a virtualized software model, enabling the sharing of physical services, storage, and networking capabilities. The cloud infrastructure, regardless of deployment model, seeks to make the most of the available infrastructure across a number of users.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

- Dynamic Provisioning — Allows for the provision of services based on current demand requirements. This is done automatically using software automation, enabling the expansion and contraction of service capability, as needed. This dynamic scaling needs to be done while maintaining high levels of reliability and security.
- Network Access — needs to be accessed across the internet from a broad range of devices such as PCs, laptops, and mobile devices, using standards-based APIs (for example, ones based on HTTP). Deployments of services in the cloud include everything from using business applications to the latest application on the newest smartphones.
- Managed Metering — uses metering for managing and optimizing the service and to provide reporting and billing information. In this way, consumers are billed for services according to how much they have actually used during the billing period.

In short, cloud computing allows for the sharing and scalable deployment of services, as needed, from almost any location, and for which the customer can be billed based on actual usage.

PROBLEM DOMAIN

The use of internet and the internet based applications are increases rapidly in last ten years. Individual hands are fully mounted with the smart devices and these devices are frequently usages the services and applications. In order to keep in track the proper functioning of these applications efficient and high performance computing devices are required. Thus cloud computing is providing a way to deliver high performance remote computing technology. But due to this a significant amount of request are arises in each fraction of seconds thus most of the time the systems are becomes busy and experience high computational load. These loads can be responsible for various kinds of network related or process related faults.

PROPOSED SOLUTION

Thus in order to overcome these situations such as fault avoidance, or deadlocks the load balancing or load optimization techniques is required. These load balancing techniques are provides the efficient way of scheduling to allocate a job to the processors in such manner by which the load affect can be minimizing and maximizing the performance of computational servers. Therefore this presented work is intended to find most optimum technique for load distribution on computing units. Thus a number of techniques and algorithms are studied and four most promising techniques are selected namely genetic algorithm, artificial honey bee colony algorithm, round robin algorithm and the SOM (self-organizing map) based method for demonstration of load balancing in cloud servers.

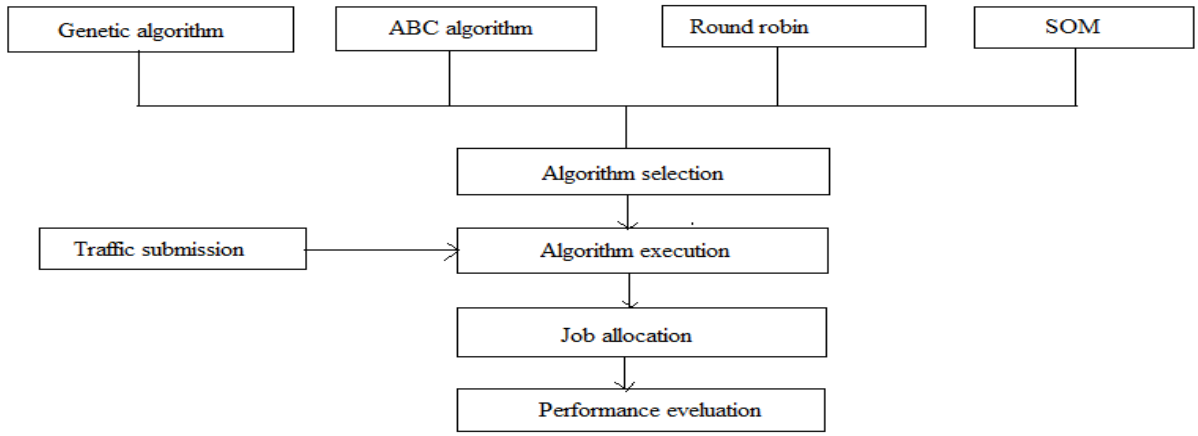

Figure 1.2 proposed simulation architecture

CONCLUSION

Cloud computing is a new generation technology which provide ease in efficient computing, managing the resources and easy pay per usage basis interface of different software and hardware resources. Therefore the demand of this infrastructure is rapidly increases among the various organizations. Due to this traffic or workload on servers are arises frequently and the servers are most of the time are overloaded. These loads on server results the software and hardware label conflicts and issues. Thus in order to overcome these issues scheduling or load balancing techniques are utilized which are help to find the less loaded processing units among the available

computational elements and by allocating the processes into these elements they frequently able to resolve the load on server machines.

Therefore in this presented work the different available techniques of load balancing is investigated and for finding the most optimum technique for load balancing the comparative study is performed.

REFERENCES

[1] Akshay Daryapurkar, Mrs. V.M. Deshmukh, "Efficient Load Balancing Algorithm in Cloud Environment", International Journal Of Computer Science And Applications Vol. 6, No.2, Apr 2013
[2] Argha Roy, Diptam Dutta, "Dynamic Load Balancing: Improve Efficiency in Cloud Computing", International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-2, Issue-4), April 2013
[3] Alexa Huth and James Cebula, "The Basics of Cloud Computing", © 2011 Carnegie Mellon University, Produced for US-CERT
[4] Introduction to Cloud Computing, White Paper, Dialogic, 2013
[5] Nariman Mirzaei, Cloud Computing, Fall 2008,Community Grids Lab, Indiana University Pervasive Technology Institute
[6] Mike Ricciuti, "Stallman: Cloud computing is 'stupidity'", http://news.cnet.com/8301-1001_3-10054253-92.html
[7] Understanding The Cloud Computing Stack SaaS, Paas, IaaS, © Diversity Limited, 2011 Non-commercial reuse with attribution permitted, http://broadcast.rackspace.com/hosting_knowledge/whitepapers/Understanding-the-Cloud-Computing-Stack.pdf
[8] Maimit A. Patel, Asst. Prof. Rutvik Mehta, "A Comparative Study of Heuristic Load Balancing in Cloud Enviournment", International Journal of Advance Engineering and Research Development Volume 2, Issue 1, January -2015
[9] Pradeep Naik, Surbhi Agrawal, Srikanta Murthy, "A survey on various task scheduling algorithms toward load balancing in public cloud", American Journal of Applied Mathematics 2015; 3(1-2): 14-17 Published online December 30, 2014
[10] Siva Theja Maguluri, R. Srikant, Lei Ying, "Stochastic Models of Load Balancing and Scheduling in Cloud Computing Clusters", 2012 Proceedings IEEE Infocom
[11] Gaochao Xu, Junjie Pang, and Xiaodong Fu, "A Load Balancing Model Based on Cloud Partitioning for the Public Cloud", IEEE Transactions on Cloud Computing Year 2013