International Journal of Advanced Technology & Engineering Research (IJATER)
www.ijater.com

COMPARATIVE ANALYSIS OF LOAD BALANCING ALGORITHMS IN CLOUD COMPUTING

1Amritpal Singh, 2Rajeev Kumar Bedi, 3Sunil Kumar Gupta
1Assistant Professor, Global Institute of Management, Amritsar, 2Assistant Professor, BCET, Gurdaspur, 3 Associate Professor, BCET, Gurdaspur

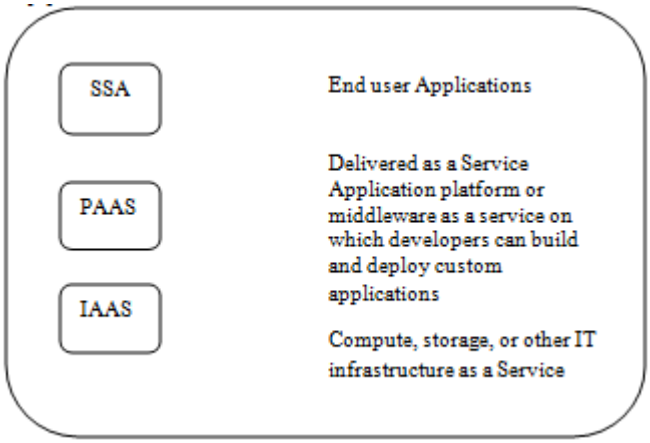
Abstract

The rapid growth of internet usage has brought huge increase in traffic on servers. Due to this reason, users often face number of problems as poor response time or denial of service. It is the responsibility of the service provider to provide best service to the end users. For this load balancing algorithms are used, so that load can be distributed to number of servers and end user can get quick response from the server. There are number of load balancing algorithms that are used as to overcome poor response time and denial of service attacks. In this research, three load balancing algorithms: Round Robin, Equally Spread Current Execution Load and Least Loaded Server Load Balancing algorithm, are simulated and their performances are compared. The simulation is done by using Cloud Analyst simulator. Load indices used for these algorithms are Data Center, VMs, Image Size, Memory, and Bandwidth. The simulation results show that, the choosing of load balancing algorithms have significant impact on the performance of servers.[3] Based on Overall Response Time and Data Center Processing Time, Equally Spread Current Execution Load and Least Loaded Server Load Balancing algorithm have equal performance on these two parameters but if we consider Data Center Request Servicing Time as a parameter then Least Loaded Server Load Balancing algorithm have better performance as compare to other two load balancing algorithms.

Key Terms: - Load Balancing, Internet, Simulator, VMs, Memory, Bandwidth.

Introduction

Cloud Computing is always in demand, when you think about what IT always needs: provide an infrastructure without much investment, training new personnel, or licensing new software. Cloud Computing is a concept where a number of computers are connected through a network. Cloud Computing works on per-use module, it means you will pay to service provider as per your usage. Now these days, Industry and Academia basically totally based on Cloud Computing. In Cloud Computing, one important application is workload shift. In this application local computers have no work to do, when an application is running. The computers that are connected in the network handle all the workload [2].



Load balancing is the process of improving the performance of distributed and parallel computing with the help of distribution of load among the processors or nodes. As the use of the web increasing day by day, with this there is need to increase the requirement for load balancing [5]. The introduction of E-Commerce has lead many businesses to carry out the most of their day-to-day business online. As a result of the increase in demand of the web, web sites providers want to ensure the availability of access for their users and make sure that their requests are processed as quickly as possible [3].

Issues with load balancing methods:

As traffic is increasing due to highly demand of services, so incoming network traffic is distributed on network level by using network load balancing algorithms (like: random allocation, round-robin allocation, etc.). These algorithms use network based parameters of incoming traffic to decide where to forward traffic, without any information from other components of computer system, like current load of application or database servers [4].

The first issues are reasonably easy to correct if they exist. There are some issues that are related to the configuration and features of the load balancer itself, as:

- the priority of the target servers
- the load balancing algorithm used
- the health monitors/probes

Comparison of Algorithms

Round Robin

In this, a number of requests are assigned by datacenter to a list of VMs on a rotating basis. The first request is assigned to a VM- selected randomly from the list of VMs and then the DataCenter controller assigns the particular requests in a circular order. Once the VM is assigned the request, the VM id is moved to the end of the list. In this manner, Round Robin Load Balancer works [10].

Algorithm:-

- RECEIVE_REQUEST
- SERVER_ID=(LAST_SERVER_ID + 1) MOD N
- FORWARD_REQUEST_TO(SERVER_ID)

Equally Spread Current Execution

In AMLB, information about each VMs and the number of requests currently allocated to which VM is maintained. When a request for the allocation of a new VM arrives, it identifies the VM that is least loaded [6]. If there are more than one VMs free, then the VM that comes first is selected. ActiveVmLoadBalancer returns the respective VM id to the Data Center Controller. The data Center Controller sends the corresponding request to the VM referenced by that id. DataCenterController notifies the ActiveVmLoadBalancer of the new allocation and cloudlet is sent to it [10].

Algorithm:-

- Find the NEXT_AVAILABLE_VM
- Check for all current allocation count is less than MAX_LENGTH_VM list allocate the VM
- If AVAILABLE_VM is not allocated CREATE_NEW
- Count the ACTIVE_LOAD_VM
- RETURN_ID_VM which is having least load

Least Connection Scheduling Algorithm

The least-connection scheduling algorithm [12] directs network connections to the server with the least number of established connections. This is one of the dynamic scheduling algorithms; because it needs to count live connections for each server dynamically. For a virtual server that is managing a collection of servers with similar performance, least-connection scheduling is good to smooth distribution when the load of requests vary a lot [12].

Algorithm:-

- RECEIVE_REQUEST
- SERVER_ID=0
- FOR I=1 TO N
  - IF(SERVER(I).CONNECTIONS<SERVER (SERVER_ID).CONNECTIONS)
  - SERVER_ID=I
- FORWARD_REQUEST_TO(SERVER_ID)

Simulation Setup

For the comparison of these three load balancing algorithms, cloud analyst simulator is used.

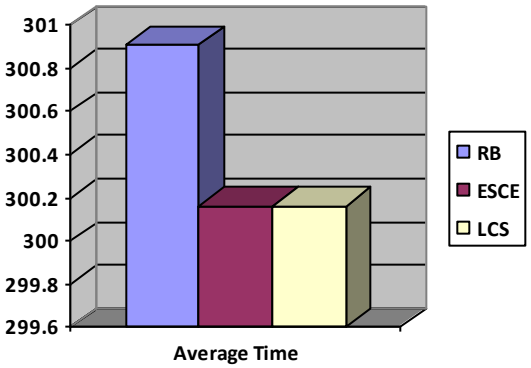
Indices used for the comparison are:- Data Center, VMs, Image Size, Memory, and Bandwidth.

1. **Data Center**
   Data Center provides the specified requirements to the end users. The number of virtual machines, memory required, and bandwidth required etc. There are number of Data Centers in a specific region [7]. As the distance between end user and Data Center increases, it effects on the throughput. So the nearest Data Center always preferred by the end user to avail all the services provided by the Data Center. Three Data Centers are used for the performance analysis of different algorithms.
2. **VMs**
   VM(Virtual Machines) executes the requests that are send from the end users. If a VM have a higher configuration, its response will be quick. So, configuration of VMs play an important role in the execution of requests. As defined above, three Data Centers are used so in each Data Center Ten (10) VMs are used.
3. **Memory**
   Memory denotes the size of RAM in a machine. As size of RAM increases, speed of processor also increases [1]. In case of Data Center Id number 1, 10 VMs are used, each of configuration 1024 Bytes, Data Center Id 2, 10 VMs, each of 2048 Bytes, Data Center Id 3, 10 VMS, each of 4096 Bytes.
4. **Bandwidth**
   The speed of transfer of data from the end user to the specific machine or the time availed by the packet from end user to machine, depends on the Bandwidth provided to the machine [1]. If Bandwidth of the machine is higher, it takes less time to reach its

destination. Each Data Center have same Bandwidth 1000 Bytes.

Results

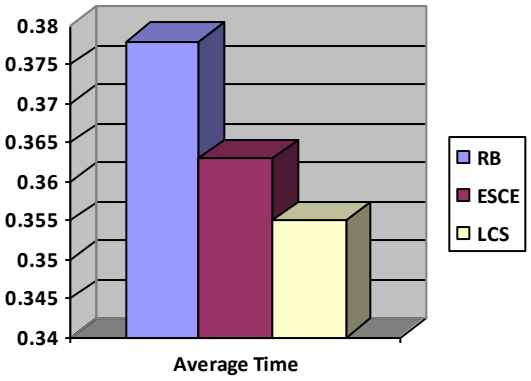
1. Overall Response Time Summary



Graph 1

RB: Round Robin
ESAE:- Equally Spread Current Execution
LCS:- Least Connection Scheduling

2. Data Center Request Servicing Time



Graph 2

RB: Round Robin
ESAE:- Equally Spread Current Execution
LCS:- Least Connection Scheduling

Performance Analysis

Algorithm Used	Overall Response Time in (ms)	Data Center request Servicing Time in (ms)		
		DC1	DC2	DC3
Round Robin	300.91	0.378	0.469	0.491
Equally Spread Current Execution	300.16	0.363	0.479	0.499
Least Connection Scheduling	300.16	0.355	0.486	0.496

Table 1:-Performance Analysis

Conclusion

Time and cost are the two main factors that can enhance the performance of the system. Current strategies analyzed in this paper lacking at many situations, increase in the number of Data Centers lead to increase in the time as well as the cost of the operations. This paper aims towards the development of heterogeneous kind of load balancing algorithms that can easily allocate the load between the servers and give effective response in severe conditions also. Least Connection Scheduling Algorithm give effective results, still there is need to enhance the performance of Load Balancing Algorithms.

References

[1] A. Iyengar, E. Nahum, A. Shaikh, and R. Tewari. "Improving web site performance", in *Practical Handbook of Internet Computing*, vol. 2, M. P. Sing, Ed. CRC Press, 2004.

[2] T. Schroeder, S. Goddard, and B. Ramamurthy. "Scalable web server clustering technologies." *IEEE Network*, vol. 14, no. 3, pp. 38-45, 2000.

[3] D. Andresen, T. Yang, and O. H. Ibarra. "Towards a scalable distributed WWW server on networked workstations." *J. of Parallel and Distributed Computing*, vol. 42, no. 1, pp. 91-100, 1997.

[4] V. Cardellini, E. Casalicchio, M. Colajanni and P. S. Yu. "The State of the Art in Locally Distributed Web-Server Systems." *ACM Computing Surveys*, vol. 34, no. 2, pp. 1-49, June 2002.

[5] Q. Zhang, A. Riska, A. W. Sun, E. Smirni and G. Ciardo. "Workload-Aware Load Balancing for Clustered Web

Servers." *IEEE Transactions on Parallel And Distributed Systems*, vol. 16, no. 3, pp. 219-233, 2005.

[6] V. Carddellini, M. Coljanni and P. Yu. "Dynamic Load Balancing on Web Server Systems." *IEEE Internet Computing*, vol. 3, no. 3, pp. 28-39, 1999.

[7] G. D. Hunt, G. S. Goldszmidt, R. P. King and R. Mukherjee. "Network dispatcher: a connection router for scalable Internet services." *Computer Networks and ISDN Systems*, vol. 30, no. 1-7, pp. 343-347, 1998.

[8] "Cisco Local Director." Internet: http://www.cisco.com/ warp/public/cc/pd/cxsr/400 /index.html.

[9] "Linux Virtual Server project." Internet: http://www.linuxvirtualserver.org/ scheduling.html.

[10] B. Haakon, E. Kloving and Ø. Kure. "A Comparison of Load Balancing Techniques for Scalable Web Servers." *IEEE Network*, vol. 14, no. 4**,** pp. 58-64, 2000.

[11] M. Harchol-Balter, M. E. Crovella and C. D. Murtaz. "On Choosing a Task Assignment Policy for a Distributed Web-Sever System." *J.Parallel Distrib. Computing*, vol. 59, no. 2, pp. 204-228, 1999.

[12] Least Connection Scheduling, http://kb.linuxvirtualserver/wiki/leastconnection- scheduling.

[13] K. H. Yeung, K. W. Suen and Y. K. Wong. "Least Load Dispatching Algorithm for Parallel We

b Server Nodes." *IEE Proceedings on Communications*, vol. 149, no.4, pp. 223- 226, 2002.

[14] P. Barford and M. Crovella. "Generating Representative Web Workloads for Network and Server Performance Evaluation,"*Proceedings of ACM SIGMETRICS*, pp. 151-160, 1998.

Biographies

AMRITPAL SINGH is a student of M.Tech (CSE) at BCET, Gurdaspur, did his graduation in Information technology from BCET, Gurdaspur. His current research area of interest is Cloud Computing. He may be reached at amritpal_bcet@yahoo.co.in.

RAJEEV KUMAR BEDI is an Assistant Professor in the Department of Computer Science & Engineering at BCET, Gurdaspur. His current research area of interest is cloud computing. Mr. Rajeev Kumar Bedi may be reached at rajeevbedi@rediffmail.com.

SUNIL KUMAR GUPTA is an Associate Professor in the Department of Computer Science & Engineering at BCET, Gurdaspur. His current research area of interest is distributed computing. Mr. Sunil kumar Gupta may be reached at skgbcet@gmail.com.