

Dynamic Load Balancing in Cloud Using Live Migration of Virtual Machine

Sandeep Bhargava, Swati Goyal(Suresh Gyan Vihar University,Jaipur)

Abstract— Cloud computing is promising as a latest feature of widespread distributed computing. Cloud computing is a framework for provide a use of appropriate, on-demand network right to use to a collective pool of computing resources. Load balancing is one of the major and important part in cloud computing which is necessary to share out the dynamic work load across numerous nodes to make sure that no single node is staggered. It helps in most favorable consumption of resources and hence in augmenting the performance of the system. The objective of load balancing is to increase the throughput of the processor. By using virtual machine server balance the load .This determines that when user send request to server, it provide proper throughput to user and zero downtime in order of millisecond. The load balancing criteria have significant influence on the performance, and they can change the behavior of the system. By using this research work basically we can redistribute the work load among different server of the distributed system. .[1]

Index Terms— Cloud Computing, Load Balancing, Virtualization, VMware Workstation.

I. INTRODUCTION

Cloud computing

Cloud computing is Internet based development and use of computer technology. The concept incorporates infrastructure as a service (IaaS), platform as a service (PaaS) and software as a service (SaaS) as well as Web 2.0 and other recent technology style. One of the most significant thoughts behind cloud computing is scalability, and the key method that makes that likely is virtualization. Virtualization allows enhanced use of a server by assuming multiple operating systems and applications on a single shared computer. Virtualization also allow online migration so that if a server overloaded, an example of an operating system (and its applications) can be migrated to a other one, less cluttered server.

Cloud Components:

In Cloud organization included of 3 key modules:

- 1.Clients
- 2.Data center
- 3.Distributed servers.

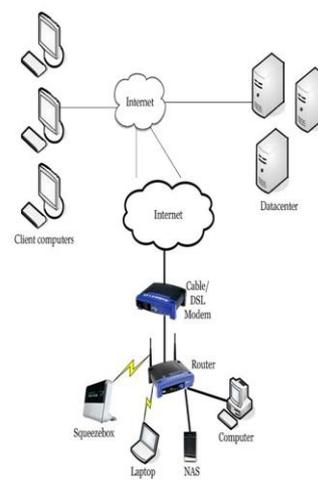


Figure 1: A cloud is used in network to show the Internet

Type of Clouds :

There are three types of cloud:

1. Public
2. Private
3. Hybrid

Services provided by Cloud computing :

Services in a cloud are of 3 kinds:

- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Hardware as a Service (HaaS) or Infrastructure as a Service (IaaS)

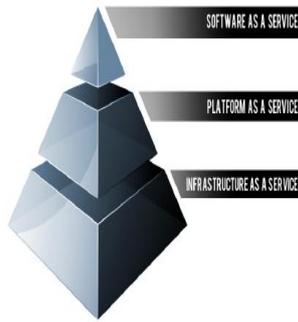


Fig 4 :Services of Cloud computing

Pros and Cons Of Cloud Computing Model:

Pros of cloud computing:

Cost reduction - Cloud computing decreases paperwork, drops transaction costs, and reduces the finance in hardware. Moving your business to ‘the cloud’ and it is also reduces the need for an Information Technology (IT) staff.

It’s scalable - Like electricity and water, some cloud computing facilities permit businesses to only n only pay for what they use for his/her work. And as there business raises, you can accommodate by adding more server space.

Easier collaboration - Since services that are provided in the cloud can be retrieved anytime from any computer, it’s easy to team up with employees in detached locations.

Security: It is provide security at dissimilar level.

Simplicity: It is very easy to use.

Little or no capital investment – It’s have very little cost.

Little or no maintenance cost – maintenance is all from a workstation or configuration screen. We never have to go touch a physical server.

Cons of Cloud Computing :

Little or no capital investment – services do not criticize over years as capital expenses do, so there could be a tax drawback over time.

Monitoring as well as maintenance tackle are not mature yet – visibility into the cloud is inadequate, in spite of current announcements by BMC, CA, Novell and others that they’re altering their data-center management tenders to provide improved control over data in Amazon’s EC2 and other cloud services.

Immature standards – groups like a the dispersed Management Task Force, the Cloud safety connection and the Open Cloud association are mounting standards for data migration, security and other standards, but actual standards at the excellence levels business IT needs are still a pair of years away, most analysts agree.

II LOAD BALANCING

In clouds, **load balancing**, as a method, is functional transversely dissimilar data centers to make sure the network ease of use by minimizing use of computer physical, software failures and mitigating recourse limitations. Load balancing techniques, in the area of cloud computing, reduces costs associated with document management systems and maximizes availability of resources reducing the amount of downtime that affect businesses during outages. ”Cloud computing” is a term, which involves virtualization and distributed computing. It also include networking, software and web services.

Goals of Load balancing :[3]

- To progress the performance significantly
- To have a backup facility it the system fails even partially
- To preserve(maintain) the system constancy
- To provide future alteration(modification) in the system

Type of load balancing algorithm:

- 1) Sender Initiated
- 2) Receiver Initiated
- 3) Symatric

Existing Load Balancing Techniques in Cloud Computing:[3][7][8]

- 1)Honeybee Foraging Algorithm
- 2)Active Clustering
- 3)Biased Random Sampling

Metrics for Load Balancing:

Overhead Associate: It determines the amount of overhead involved while implementing a load-balancing algorithm. It is self-possessed of overhead due to group of tasks, inter- processor and inter-process message. This should be reduced so that a load balancing technique can work professionally.

Throughput: It is used to calculate the no. of tasks whose execution has been completed. It should be high to improve the performance of the system.

Performance: It is used to check the efficiency of the system. It has to be superior at a sensible cost e.g. reduce response time while keeping acceptable delays.

Resource Utilization: It is used to make sure the utilization of resources. It should be optimized for an well-organized load balancing.

Scalability: It is the ability of an algorithm to perform load balancing for a system with any finite number of nodes. This metric should be improved.

Response Time: It is the amount of time taken to respond by a particular load balancing algorithm in a dispersed system. This stricture should be reduced.

Fault Tolerance: It is the ability of an algorithm to perform uniform load balancing in spite of arbitrary node or link failure. The load balancing should be a good fault-tolerant technique.

Migration time: It is the time to migrate the jobs or resources from one node to other. It should be minimized in order to enhance the performance of the system .

III Virtualization

In a computing machine, virtualization means to create a virtual form of a device or resource, like a server, storage device, network or an operating system where the arrangement divides the resource into one or more executing environments.

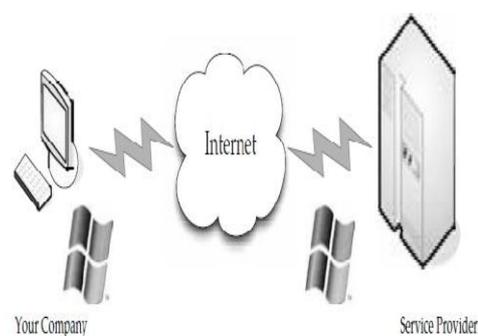
It is a very useful concept in cloud systems. By Virtualizations means that “**something which is imaginary**”, but gives all the services of like a reality. This the software practical implementation of a computer that will be execute unlike programs like a real machine.

Virtualizations is related to cloud computing, because using virtualization the ultimate consumer can use different services of a cloud computing. The remote computer centre will provide different services in a fully or partial virtualized manner.

Or we can say that “Virtualization provides a means to separate the physical hardware (Computer) and the virtual Software (operating system and application) by imitating hardware using software.”

Full Virtualization:

In full virtualization a whole installation of one



machine is

Fig : Full Virtualization

Done on the machine. It will result for a virtual machine which will have all the software that is present in the actual server

2. Paravirtualization

In paravirtualisation, the hardware permits multiple operating systems to run on sole machine .by well-ordered use of system resources such as memory and processor. example VMware software. Rather the full services, here all the services are partially provided.

II. 4 PROPOSED METHODOLOGY & TOOLS

From this paper my main motive is to balance the load among server. I already discuss a lot of about load balancing in previous topic and about cloud computing. Along with cloud computing I already mention what is load balancing and virtual machine. Now I am going to discuss about which hardware tools we needed to do load balancing. For load balancing we will use VMware Workstation. By using it we can create virtual machines. for creating the cloud we have to use VMware Esxi 5.0. In this manner we can create cloud for balancing the load among servers. For balancing the load I am going to use some combination of hardware and software.

For Load Balancing we have to install our service or application onto numerous servers that are configured to share the entire load on server. This kind of pattern is a “*load-balanced cluster*”. Load balancing equilibrium the performance of server-based programs, such as a Web server as well as by scattering client requests crossways numerous servers. Load balancing technologies, usually referred to as *load balancers*, get incoming requests and redirect them to a precise host if necessary. The load-balanced hosts at the same time as respond to different client requests, still numerous requirements from the same to same client. For example, a Web browser may attain the numerous images within a particular Web page from dissimilar hosts in the cluster. This method sprinkled the load, speeds up processing, and minimize the response time to clients.

Load balancers use diverse algorithms to control traffic. The objective of these algorithms is to intelligently distribute load and/or maximize the utilization of all servers within the cluster. Some examples of the techiques include:

In this paper I am going to use Round Robin Algorithm because it is more efficient. Round Robin Load balancing allows us to distribute client requests across multiple servers. Load balancers perk up server fault tolerance as well as end-user response time. Load balancing scattered client requests across several servers to optimize more resource utilization.

Proposed Work:

This is the final structure that I am doing in my research work. In this manner we can balance the load among

server. Client have no knowledge what is going on behind. Till now we discussed about how to Fo

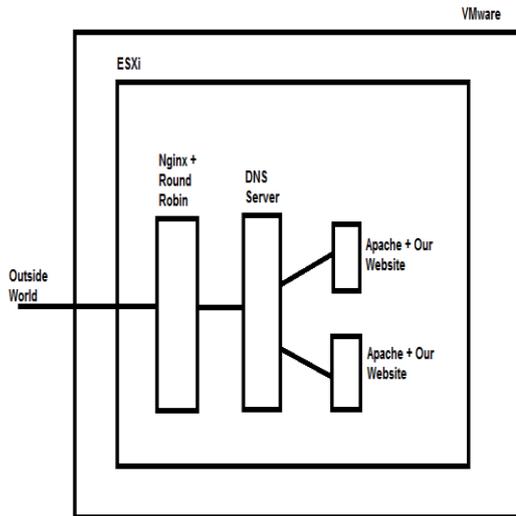


Fig 23: Proposed Work

This is the final structure that I am doing in my research work. In this manner we can balance the load among server. Client have no knowledge what is going on behind. Till now we discussed about how to about how to use VM Ware and VMware ESXi also. Now how to configure DNS for this work.

RESULT ANALYSIS

Accuracy assessment is necessary in case of load balancing for that we have to compare the performance using load balancing and without load balancing. To analysis the performance we are going to use different-different tools.

Tools are:

- (1) GNOME SYSTEM MONITOR
- (2) DARKSTAT
- (3) CACTI Network Monitoring Tool
- (4) Nagios Monitoring Tool

Graph with Load balancing:

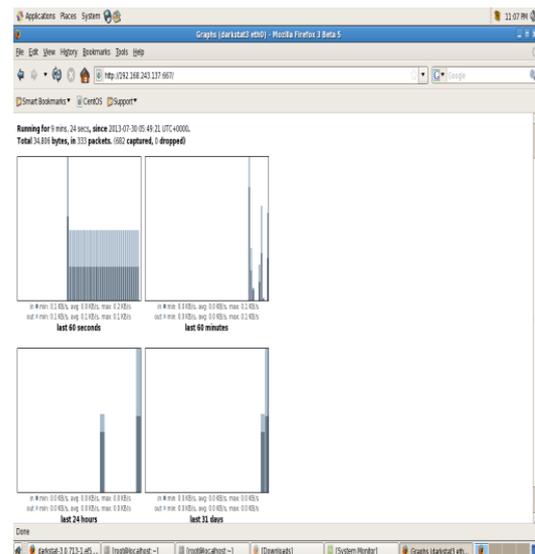


Fig : Snapshot of graph in load balancing

Graph without load Balancing –

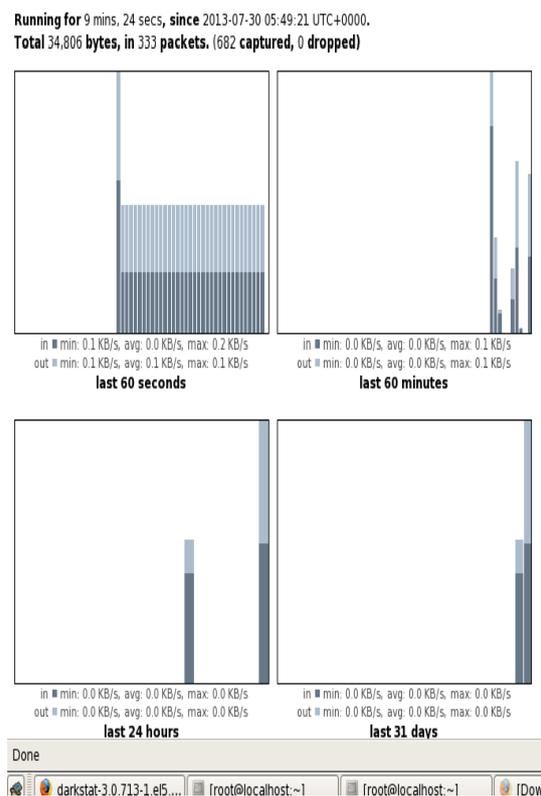


Fig 29: Snapshot of graph without load balancing

Using GNOME SYSTEM MONITOR TOOL : Gnome System Monitor is a too good, that is GNOME process viewer and system monitor with a good easy-to-use interface, It has few good quality , such as a tree view for process dependencies, icons for processes, the ability to bury processes that you don't want to see, graphical time histories of CPU/memory/swap custom, the ability to kill/renice processes needing root access, as well as the standard features that you might expect from a process viewer.

By using this tool we can measure performance CPU process and memory consumptions

CPU and Memory performance

Performance without load Balancing

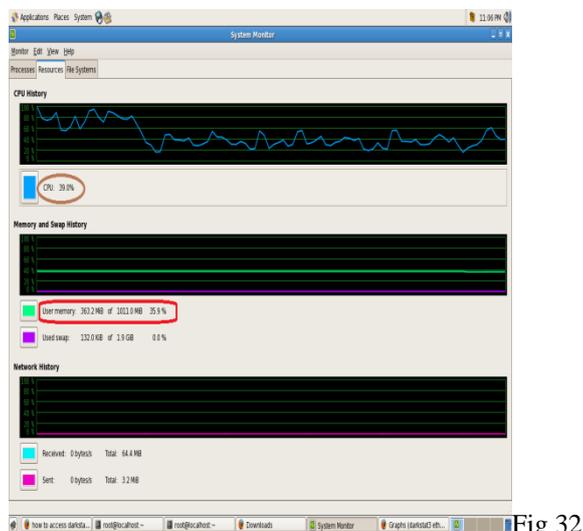


Fig 32

:Snapshot of performance without load balancing

using GNOME SYSTEM MONITOR

Performance without load Balancing

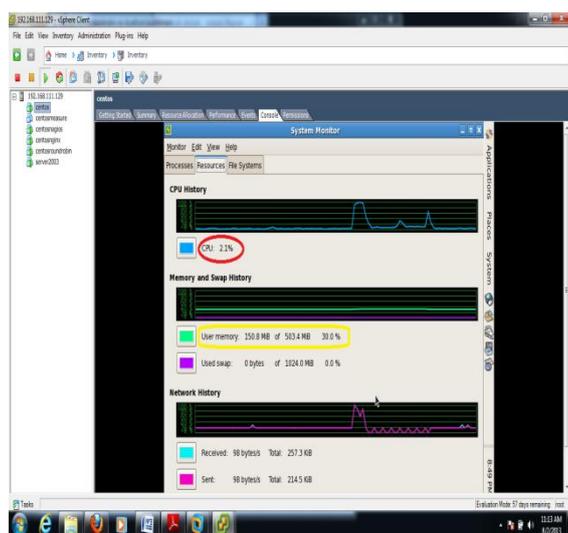


Fig 33: Snapshot of performance with load balancing

using GNOME SYSTEM MONITOR

Without load balancing snapshot we can see that CPU have load 39 % for one request and User Memory consume 35.9%.In second case that is measured with load balancing has CPU load only 2.1% and User memory consume 30%.So we can say that after load balancing CPU have minimum load.

CONCLUSION

Broadly considered, the aim of this paper has been to examine how we can increase performance ratio as well as how we will give proper response of user request.

Without load balancing some processors in the system are idle while other processors have many ready tasks that are waiting for run when the average processors utilizations under 100% .Load balancing function make use of the processor resources more efficiently and produce huge improvement under this condition. When the processor utilization is higher than 100% then user request migrate from one server to another server. In this way user will get response time to time. The load balancing criteria have significant influence on the performance, and they can change the behavior of the system. By using this research work basically we can redistribute the work load among different server of the distributed system. If a server response a request of client and at the same time second request will come, in this case to balance the load second request is transfer to another server and so on. In this manner we can improve both resource utilization and job response time while also avoiding a situation where some nodes are heavily loaded while others are idle or doing little work.

FUTURE DIRECTION:

In this thesis I have been researched how we can increase the performance of server and how to balance the load. Now my studies have undertaken when user sent request to server the request is encrypting means that client request is converted into another language that is not understandable by third person. And the client request is decrypted means converted in original form on server side who response on request. For that I am going to create my own algorithm to provide encryption and decryption. Means that I want to provide more security of user data.

References:

- [1] A. Brian,“Load Balancing in clouds:Tools, tips and techniques”. A technical white paper in solution architect Right Scale.
- [2] Z Chaczko, V Mahadevan, S Aslanzadeh and C Mcdermid,“Availability and Load balancing in Cloud Computing”. International Conference on Computer and Software Modeling, vol 14, Singapore. 2011.
- [3] J. Hu, J. Gu, G. Sun, and T. Zhao, “A Scheduling Strategy onLoad Balancing of Virtual Machine Resources in CloudComputing Environment”, Third International Symposium on Parallel Architectures, Algorithms and Programming(PAAP), 2010, pages 89-96.
- [4] N Kansal, I Chana,“Cloud Load Balancing Techniques: A Step Towards GreenComputing”.IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1, January 2012.
- [5] M. Randles, D. Lamb, and A. Taleb-Bendiab, “AComparative Study into Distributed Load BalancingAlgorithms for Cloud Computing”, Proceedings of 24thIEEE International Conference on Advanced

InformationNetworking and Applications Workshops, Perth, Australia, April 2010, pages 551-556.

[6] Y. Zhao, and W. Huang, “Adaptive Distributed LoadBalancing Algorithm based on Live Migration of VirtualMachines in Cloud”, Proceedings of 5th IEEE InternationalJoint Conference on INC, IMS and IDC, Seoul, Republic ofKorea, August 2009, pages 170-175.

[7] Ali M. Alakeel, “A Guide to Dynamic Load Balancing in Distributed Computer Systems” College of Computing and Information Technology, University of Tabuk, Tabuk, Saudi Arabia, published in the IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.6 June 2010

[8] Sachin Kumar, Niraj Singhal, “A Study on the Assessment of Load Balancing Algorithms in Grid Based Network”, published in the International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012.

[9] Sandeep Sharma, Sarabjit Singh, and Meenakshi Sharma, “Performance Analysis of load Balancing Algorithms”, published in the proceedings of World Academy of Science, Engineering and Technology Volume, 28 April 2008, ISSN 2070-3740.

[10] Zenon Chaczko, Venkatesh Mahadevan, Shahrzad Aslanzadeh and Christopher Mcdermid, “Availability and Load Balancing in Cloud Computing”, IPCSIT vol.14 (2011) © (2011) IACSIT Press, Singapore 2011 International Conference on Computer and Software Modeling.

[11] Eduardo Pinheiro, Ricardo Bianchini, Enrique V. Carrera, and Taliver Heath, “Load Balancing and Unbalancing for Power and Performance in Cluster-Based Systems” Technical Report DCS–TR–440, May 2001

[12] Jiani Guo,” Load Balancing in a Cluster-Based Web Server for Multimedia Applications” IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, IEEE, and Laxmi Narayan Bhuyan, Fellow, IEEE

[13] Z. Zhang, and X. Zhang, “A Load Balancing Mechanism Based on Ant Colony and Complex Network Theory in Open Cloud Computing Federation”, Proceedings of 2ndInternational Conference on Industrial Mechatronics and Automation (ICIMA), Wuhan, China, May 2010, pages 240-243.

[14] H. Mehta, P. Kanungo, and M. Chandwani, “Decentralized content aware load balancing algorithm for distributed computing environments”, Proceedings of the International Conference Workshop on

Emerging Trends in Technology (ICWET), February 2011, pages 370-375.

[15] Y. Lua, Q. Xie, G. Kliot, A. Geller, J. R. Larus, and A. Greenberg, “Join-Idle-Queue: A novel load balancing algorithm for dynamically scalable web services”, An international Journal on Performance evaluation, In Press, Accepted Manuscript, Available online 3 August 2011.