A Survey On Load Balancing In Cloud Computing

Ms. Parin. V. Patel [*1], *Mr. Hitesh. D. Patel** [*2], *Asst. Prof. Pinal. J. Patel* [*3]

** C.S.E. Department, Government College of Engineering, Gandhinagar*
Gujarat Technology University, Gujarat, India.

** C.S.E. Department, Saffrony Engineering collage, Mahesana*
Gujarat Technology University, Gujarat, India.

** C.S.E. Department, Government College of Engineering, Gandhinagar*
Gujarat Technology University, Gujarat, India

Abstract

In present 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 it has more critical issue like security, load management and fault tolerance. In this paper we are discussing Load Balancing approach. Many types of load concern with cloud like memory load, CPU load and network load. 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 has to handle the load when one node is overloaded. When node is overloaded at that time load is distributed over the other ideal nodes. Many algorithms are available for load balancing like Static load balancing and Dynamic load balancing.

Keywords– Cloud Computing, Load balancing, virtualization

1. Introduction

A) What is cloud computing?

The term "cloud" originates from the world of telecommunications when providers began using virtual private network (VPN) services for data communications [1]. The definition of cloud computing provided by National Institute of Standards and Technology (NIST) says that: "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. [2]".So through this cloud computing there is no need to store the data on desktops, portables etc. You can store the data on servers and you can access the data through internet.

Cloud computing provides better utilization of distributed resources over a large data and they can access remotely through the internet.

B) History

The underlying concept of cloud computing was introduced way back in 1960s by John McCarthy. His opinion was that "computation may someday be organized as a public utility [3]". Also the characteristics of cloud computing were explored for the first time in 1966 by Douglas Parkhill in his book, The Challenge of the Computer Utility[3].

The history of the term cloud is from the telecommunications world, where telecom companies started offering Virtual Private Network (VPN) services with comparable quality of service at a much lower cost. Initially before VPN, they provided dedicated point-to-point data circuits which were wastage of bandwidth. But by using VPN services, they can switch traffic to balance utilization of the overall network. Cloud computing now extends this to cover servers and network infrastructure [4].

Many players in the industry have jumped into cloud computing and implemented it. Amazon has played a key role and launched the Amazon Web

Service (AWS) in 2006. Also Google and IBM have started research projects in cloud computing. Eucalyptus became the first open source platform for deploying private clouds [4].

C) Cloud Architecture

Cloud computing system is divided into two sections: the front end and the back end. Front end through which user can interact with the server and backend is the server which provides data to the client. Between server and client network is working as middleware.

Layers and Services of Cloud Computing Architecture:

The below diagram shows the different layers of cloud Computing architecture [3].

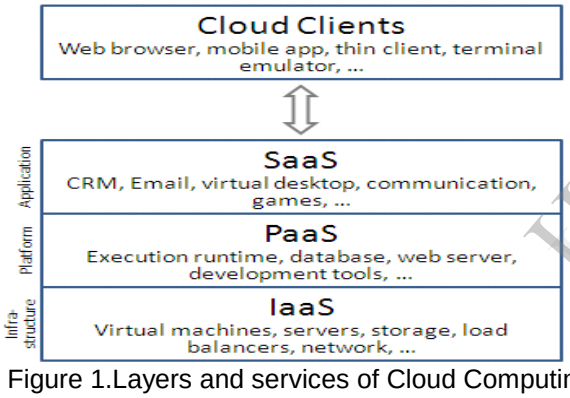


Figure 1.Layers and services of Cloud Computing

A cloud client consists of computer hardware and/or computer software which relies on cloud computing for application delivery, or that is specifically designed for delivery of cloud services [9].

(Cloud) Infrastructure as a Service (IaaS) is also referred as Resource Code, provide (managed and scalable) resources as services to the user- in other words, they basically provide enhanced virtualization capabilities. Accordingly, different resources may be provided via a service interface [5].

(Cloud) Platform as a Service (PaaS) is provides computational resources via a platform upon which applications and services can be developed and hosted. Example: Google Docs, SAP business by design [5].

(Clouds) Software as a Service (SaaS) is also sometimes referred to as Service or application clouds. These clouds are offering implementation of specific business functions and business processes that are provided with specific cloud capabilities, i.e. they provide applications/services using a cloud infrastructure or platform, rather than providing cloud features themselves [5].

Deployment of Cloud Computing Service

For deploying a cloud computing solution, the major task is to decide on the type of cloud to be implemented. Presently three types of cloud deployment takes place - public cloud, private cloud and hybrid cloud Figure below shows the overview of the deployment of these three clouds [6]:

Public cloud allows the user to access cloud via network. This cloud is publicly available on internet so security is the big problem. In this cloud upgradation and maintenance is difficult. This cloud is on "Pay and Use" basis. You need to pay only the time duration that you have use.

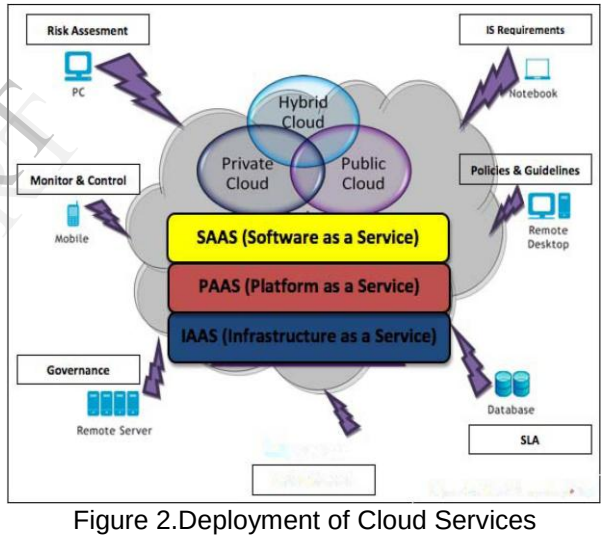


Figure 2.Deployment of Cloud Services

Private Cloud is within an organization. This stores the internal data of organization. It is more secure and maintenance is also easy. Only the internal user can access that data.

The Hybrid Cloud is a combination of any two (or all) of the three models discussed above. Standardization of APIs has led to easier distribution of applications across different cloud models. This enables newer models such as "Surge Computing" in which workload spikes from the private cloud is offset to the public cloud [7].

Community cloud is constructed by many organizations according to their requirements. Cloud infrastructure is managed by third party or one of the organizations.

D) Advantages and Disadvantages of Cloud Computing:

Advantages:

- Easy to Maintain.
- Less cost.
- We can use on pay and use basis.
- Personalized Backup and recovery.
- Remote access.
- Green computing.

Disadvantages:

- Security and privacy.
- Higher operational cost.

2. Virtualization

Virtualization (or virtualization) is the creation of a virtual (rather than actual) version of something, such as a hardware platform, operating system (OS), storage device, or network resources. A virtual machine is subjectively a complete machine (or very close), but objectively merely a set of files and running programs on an actual, physical machine [8].

Types of Virtualization:

1) Full virtualization (Native virtualization):
In this virtualization complete installation is done on machine.so it contains all hardware and software as actual server. This is used by many packages like virtual server, virtual pc and VMware.
2) Para virtualization:
In this virtualization multiple operating systems can run on single machine which use resources like memory and processor. Examples are Microsoft Hyper-V and VMware ESX Server.

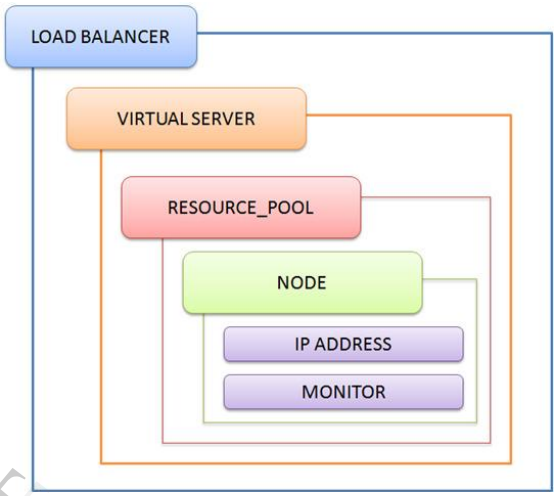
3. Load Balancing

A) What is Load Balancing?

Load Balancing is a technique in which the workload on the resources of a node is shifts to respective resources on the other node in a network without disturbing the running task. A standard way to scale web applications is by using a hardware-based load balancer [9].

The load balancer assumes the IP address of the web application, so all communication with the web application hits the load balancer first. The load balancer is connected to one or more identical web servers in the back-end. Depending on the user session and the load on each web server, the load balancer forwards packets to different web servers for processing. The hardware-based load balancer is designed to handle high-level of load, so it can easily scale [10].

However, a hardware-based load balancer uses application specific hardware-based components, thus it is typically expensive. Because of cloud's commodity business model, a hardware-based load balancer is rarely occurred by cloud providers as a service. Instead, one has to use a software based load balancer running on a generic server [10].



B) Goals of Load balancing

As given in [12], the goals of load balancing are:

- To improve the performance substantially.
- To have a backup plan in case the system fails even partially.
- To maintain the system stability.
- To accommodate future modification in the system.

C) Types of Load Balancing Algorithms:

1) Static Algorithms:
Static algorithms divide the traffic equivalently between servers. By this approach the traffic on the servers will be disdained easily and consequently it will make the situation more imperfectly. This algorithm, which divides the traffic equally, is announced as round robin algorithm. However, there were lots of problems appeared in this algorithm. Therefore, weighted round robin was defined to improve the critical challenges associated with round robin. In this algorithm each servers have been assigned a weight and according to the highest weight they received more connections. In the situation that all the weights are equal, servers will receive balanced traffic [11].
2) Dynamic Algorithms:
Dynamic algorithms designated proper weights

on servers and by searching in whole network a lightest server preferred to balance the traffic. However, selecting an appropriate server needed real time communication with the networks, which will lead to extra traffic added on system. In comparison between these two algorithms, although round robin algorithms based on simple rule, more loads conceived on servers and thus imbalanced traffic discovered as a result [11].

D) Existing Load Balancing Techniques in Clouds

Following load balancing techniques are currently prevalent in clouds.

1) VectorDot- A. Singh et al. [13] proposed a novel load balancing algorithm called VectorDot. It handles the hierarchical complexity of the data-center and multidimensionality of resource loads across servers, network switches, and storage in an agile data center that has integrated server and storage virtualization technologies.

2) CARTON- R. Stanojevic et al. [14] proposed a mechanism CARTON for cloud control that unifies the use of LB and DRL. LB (Load Balancing) is used to equally distribute the jobs to different servers so that the associated costs can be mini-mized and DRL (Distributed Rate Limiting) is used to make sure that the resources are distributed in a way to keep a fair resource allocation.

3) Compare and Balance- Y. Zhao et al. [15] addressed the problem of intra-cloud load balancing amongst physical hosts by adaptive live migration of virtual machines. A 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, etc. and to keep virtual machines' zero-downtime in the process.

4) Event-driven- V. Nae et al. [16] presented an event driven load balancing algorithm for real-time Massively Multiplayer Online Games (MMOG). This algorithm after receiving capacity events as input, analysis its components in context of the resources and the global state of the game session, thereby generating the game session load balancing actions.

5) Scheduling strategy on LB of VM resources – J. Hu et al. [17] proposed a scheduling strategy on load balancing of VM resources that uses historical data and current state of the system. This strategy achieves the best load balancing and reduced dynamic migration by using a genetic algorithm.

6) CLBVM- A. Bhadani et al. [18] proposed a Central Load Balancing Policy for Virtual Machines (CLBVM) that balances the load evenly in a distributed virtual machine/cloud computing environment.

7) LBVS- H. Liu et al. [19] 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. 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.

8) Task Scheduling based on LB- Y. Fang et al. [20] discussed a two-level task scheduling mechanism based on load balancing to meet dynamic requirements of users and obtain high resource utilization. It achieves load balancing by first mapping tasks to virtual machines and then virtual machines to host resources thereby improving the task response time, resource utilization and overall performance of the cloud computing environment.

9) Honeybee Foraging Behavior- M. Randles et al. [21] investigated a decentralized honeybee-based load balancing technique that is a nature-inspired algorithm for self-organization. It 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. It is best suited for the conditions where the diverse population of service types is required.

10) Biased Random Sampling- M. Randles et al. [21] 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.

11) Active Clustering- M. Randles et al. [21] 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. The performance of the system is enhanced with high resources thereby in-creasing the throughput by using these resources effectively. It is degraded with an increase in system diversity.

12) ACCLB- Z. Zhang et al. [22] proposed a load balancing mechanism based on ant colony and complex network theory (ACCLB) in an open cloud computing federation. It uses small-world and scale-free characteristics of a complex network to achieve better load balancing. This technique overcomes heterogeneity, is adaptive to dynamic environments, is excel-lent in fault tolerance and

has good scalability hence helps in improving the performance of the system.

13) (OLB + LBMM)- S.-C. Wang et al. [23] 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. OLB scheduling algorithm, keeps every node in working state to achieve the goal of load balance and LBMM scheduling algorithm is utilized to mini-mize the execution time of each task on the node thereby minimizing the overall completion time.

14) Decentralized content aware- H. Mehta et al. [24] proposed a new content aware load balancing policy named as work-load and client aware policy (WCAP). It uses a parameter named as USP to specify the unique and special property of the requests as well as computing nodes. USP helps the scheduler to decide the best suitable node for processing the requests.

15) Join-Idle-Queue- Y. Lua et al. [25] proposed a Join-Idle-Queue load balancing algorithm for dynamically scalable web services. This algorithm 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.

4. Conclusion

In this paper we have survey load balancing. In cloud computing load balancing is the main issue. Because when client is requesting for service it should be available to the client. When node is overloaded with job at that time load balancer has to set that load on another free node.so, to balance the load is necessary in cloud computing.so in our paper we have disscuss all the existing techniques for Load balancing. And we have also disscuss the virtualization and cloud computing.

5. References

[I] 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] http://en.wikipedia.org/wiki/Cloud_computing.
[4] Yashpalsinh Jadeja and Kirit Modi, "Cloud Computing - Concepts, Architecture and Challenges", International Conference on Computing, Electronics and Electrical Technologies [ICCEET], IEEE-2012.
[5] Samerjeet kaur, "Cryptography and Encryption in Cloud Computing", VSRD International Journal of Computer Science and Information Technology, VSRD-IJCSIT, Vol. 2 (3), 2012.
[6] Ramgovind S, Eloff MM, Smith E, 'The management of security in cloud computing", IEEE – 2010.
[7] Aderemi A. Atayero and Oluwaseyi Feyisetan," Security Issues in Cloud Computing: The Potentials of Homomorphic Encryption" Journal of Emerging Trends in Computing and Information Sciences, VOL. 2, NO. 10, October 2011.
[8] Turban, E; King, D; Lee, J; Viehland," Chapter 19: Building E-Commerce Applications and Infrastructure". Electronic Commerce A Managerial Perspective. pp. 27, 2008.
[9] J. Kruskall and M. Liberman."The Symmetric Time Warping Problem: From Continuous to Discrete. In Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison", pp. 125-161, Addison-Wesley Publishing Co., 1983.
[10] Mr. Nitin S. More, Mrs. Swapnaja R. Hiray and Mrs. Smita Shukla Patel," Load Balancing and Resource Monitoring in Cloud", International Journal of Advances in Computing and Information Researches ISSN: 2277-4068, Volume 1– No.2, April 2012.
[11] R. X. T. and X. F. Z,"A Load Balancing Strategy Based on the Combination of Static and Dynamic, in Database Technology and Applications (DBTA)",2nd International Workshop,2010.
[12] David Escalante and Andrew J. Korty, "Cloud Services: Policy and Assessment", EDUCAUSE Review, vol. 46, no. 4 (July/August 2011).
[13] Singh A., Korupolu M. and Mohapatra D., ACM/IEEE conference on Supercomputing, 2008.
[14] Stanojevic R. and Shorten R., IEEE ICC, 1-6, 2009.
[15] Zhao Y. and Huang W., 5th International Joint Confer-ence on INC, IMS and IDC, 170-175, 2009.
[16] Nae V., Prodan R. and Fahringer T., 11th IEEE/ACM International Conference on Grid Computing (Grid), 9-17,2010.
[17] Hu J., Gu J., Sun G. and Zhao T., 3rd International Symposium on Parallel Architectures, Algorithms and Programming, 89-96, 2010.
[18] Bhadani A. and Chaudhary S., 3rd Annual ACM Bangalore Conference, 2010.
[19] Liu H., Liu S., Meng X., Yang C. and Zhang Y.,International Conference on Service Sciences (ICSS), 257-262,2010.
[20] Fang Y., Wang F. and Ge J.,Lecture Notes in Computer Science, 6318, 271-277,2010.
[21] Randles M., Lamb D. and Taleb-Bendiab A., 24th International Conference on Advanced Information Networking and Applications Workshops, 551-556,2010.
[22] Zhang Z. and Zhang X, 2nd International Conference on Industrial Mechatronics and Automation, 240-243, 2011.
[23] Wang S., Yan K., Liao W. and Wang S, 3rd International Conference on Computer Science and Information Technol-ogy, 108-113, 2010.
[24] Mehta H., Kanungo P. and Chandwani M., International Conference Workshop on Emerging Trends in Technology, 370-375, 2011.
[25] Lua Y., Xiea Q., Kliotb G., Gellerb A., Larusb J. R. and Green-ber A,"Int. Journal on Performance evaluation",2011.