A Review : Load Balancing in Cloud

Harneet Kaur[1], Raju Sharma[2]

[1]Assistant Professor, Computer Science and Engineering, Lovely Professional University, Jalandhar, Punjab, India

[2]Computer Science and Engineering, Lovely Professional University, Jalandhar, Punjab, India

ABSTRACT

Presently, Cloud Computing is an internet-based technology which is growing rapidly, and we are adopting it in a fine way where we have more than enough technologies to develop advance communication technologies such as IOT, Big data, Machine learning and AI. Among all these technologies we have load balancing, which is the process of distributing a set of tasks and network traffic across multiple servers. This has increased the availability of applications and websites for end-users. This paper presents a detailed review about the load balancing techniques. It also ensures the advantage and limitations of the existing methods and highlighted them with extreme challenges. This paper also suggests the new insight and best usage towards load balancing in cloud computing.

Keywords : Cloud Computing, Load Balancing, Load Balancing techniques, methods of Load balancing

I. INTRODUCTION

Load balancing is a term which distributes the workload in different kind of nodes in the given specific environment where it ensures that the no node in the system is overloaded.

Load balancing is a concept that refers to the distribution of workload among various types of nodes in a given context, ensuring that no node in the system is overloaded. It's similar to a system that detects overloaded and underloaded nodes and then distributes the actual load evenly among them.

In every area of the IT industry, we still have different kinds of advantages, and the cloud computing and its load balancing.

The most significant benefit of cloud computing is the transformation of a single-user physical computer into a multiuser virtual machine. The Cloud Service Provider plays a critical role in the distribution of services to customers, and it's a difficult task considering the limited virtual resources available. Some VMs will receive a high volume of user tasks when fulfilling user requests, while others will receive a lower volume. We have some clear reasons to research load balancing, as we explored the

benefits of cloud computing. Almost any form of service, such as HTTP, SMTP, DNS, FTP, and POP/IMAP, benefits from load balancing. It also adds redundancy, which improves reliability. The balancing service is provided by a dedicated hardware device or program. Cloud-based servers can attain more small or precisely scalability and availability using server load balancing, and could be best in cloud computing. The load balancing in clouds can be among physical hosts or even VMs and this balancing mechanism distributes the dynamic workload in manner among all the nodes. The load balancing in the cloud is also referred as load balancing as a service
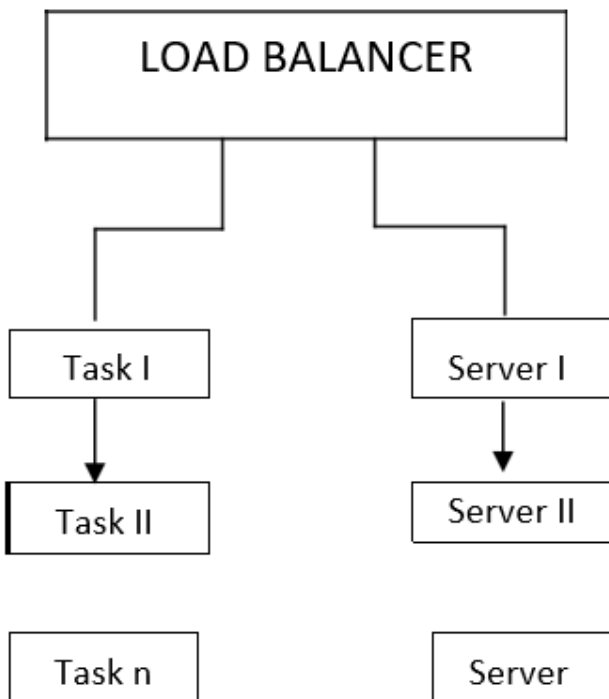

Fig 1. flow chart of load balancer

II. Load Unbalancing

We often go through a process where we encounter an unreliable utility or an unbalanced device. Load unbalancing is an unpredictably occurring occurrence on the CSP (cloud service provider) side that degrades the output and efficiency of computing resources while also compromising guaranteed

Quality of Service. As a result, tasks must be moved to another resource on a different VM. These tasks are handled by three separate units: a load balancer, a resource discovery unit, and a task migration unit, to name a few. So far, we've covered what load balancing and load unbalancing are, as well as how to understand the idea of cloud. investigate the various types of load balancing. Now we will further research about the different types of load balancing.

III. The types of Load balancers

We've already looked at cloud load balancing, and now we'll look at the different types of load balancers. There are three different types of load balancers, as seen in the diagrams below.
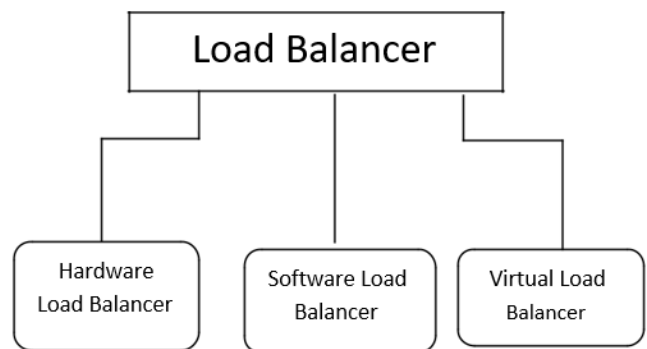

Fig 2. types of load balancing

A. Hardware Load Balancer:

It will depend on hardware or physical, as the name implies. These devices can accommodate a significant amount of traffic, but they come at a price that is difficult to justify due to their limited versatility.

B. Software Load Balancer:

These load balancers are compatible with common hardware such as desktop and laptop computers, as well as common operating systems. This software is available in two versions: commercial and open source, and it's important to note that both must be

installed before you can use it. In general, these are less expensive and easier to manage.

C. Virtual Load Balancer:

This load balancer differs from software load balancers in that it uses a virtual machine to run the software of a hardware load balancing system.

In terms of different cloud-based balancers that are available in the IT, we have more forms of load balancing.

Network load balancing, HTTPs load balancing, and internal load balancing are a few examples.

Network Load balancing as its namesuggests it is like network layer information to decide where to send network traffics and it is considered as the fastest of all the load balancing solutions.

HTTPs Load Balancing is one of the oldestforms of load balancing and it may rely on layer 7 which means that operating in the application layer. This load balancing is most flexible type of load balancing because it allows us to form distribution decisions based on any information that comes with an HTTP address.

Internet Load Balancing is same as oridentical to the network load balancing but can be leveraged to balance internal infrastructure.

So we studied or researched about the different types of the load balancers their merits and demerits that how they are useful for us in cloud communication in advanced IT sector where world is finding every day a new technology on different kind of tools.

IV. Examples of Load Balancers

A. Direct Routing Requesting Dispatching Technique:

In this approach of request dispatching is like to the one implemented in IBM's Net Dispatcher. A real server and load balancer share the virtual IP address. In this, load balancer takes an interface construction with the virtual IP address that accepts request packets and it directly sends the packet to the selected servers and gets the possible outputs.

B. Dispatcher-Based Load Balancing Cluster

A dispatcher does smart load balancing by utilizing server availability, workload, capability and other user-defined criteria to regulate where to send a TCP/IP request. The dispatcher module of a load balancer can split HTTP requests among various nodes in a cluster. The dispatcher splits the load among many servers in a cluster so the services of various nodes seem like a virtual service on an only IP address; consumers interrelate as if it were a single server despite of having information about the final infrastructure.

C. Linux Virtual Load Balancer

It's an open source load balancing solution for creating highly scalable and usable network services like HTTP, POP3, and FTP, as well as SMTP media and caching and Voice over IP. It's a straightforward and effective load balancing and failover solution. The load balancer itself is the primary entry point of server cluster systems and can execute Internet Protocol Virtual Server, which implements transport-layer load balancing in the Linux kernel also known as Layer-4 switching.

V. Weighted Load Balancing

The process of allowing users to set a specific weight for each origin server in a pool is known as weighted load balancing. Because of its potential to rebalance traffic when an origin becomes unhealthily crowded, weighted load balancing should be considered. Traffic will be rebalanced to the remaining open

origins based on their respective weights and the load balancing weight priority.

The nodes, which are just as critical as load balancing in the cloud, are an underappreciated aspect of weighted load balancing. Nodes that restart start with an empty cache, and as the cache is repopulated, the node slows down, slowing down the collection as a whole. This is where heat-weighted load balancing, which aims for low latency, comes into play.

The heat of each node is a factor in enhancing the node selection in the coordinator, so as a node is being rebooted, latency remains at a low level.

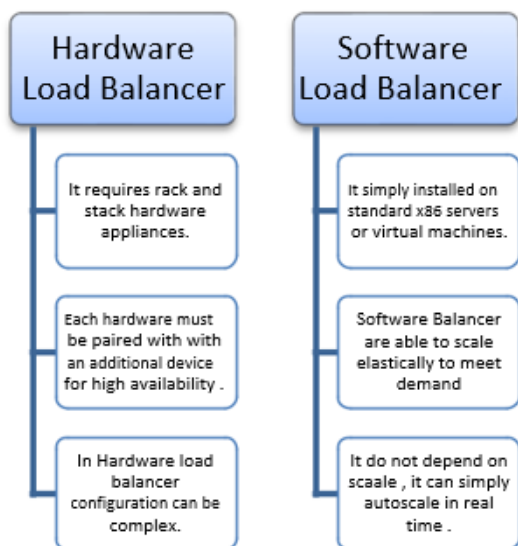## VI. Hardware Load Balancer and Software Load Balancers

FIG. 3 hardware and software load balancer

The above figure shows the aspects of hardware and software load balancer with some differences between them, where we will get to know these terminologies in better way.

VII. Hardware MERIT AND DEMERIT

As we know every coin has its two sides and same for the software and hardware load balancers, we do have the merits and demerits for the same.
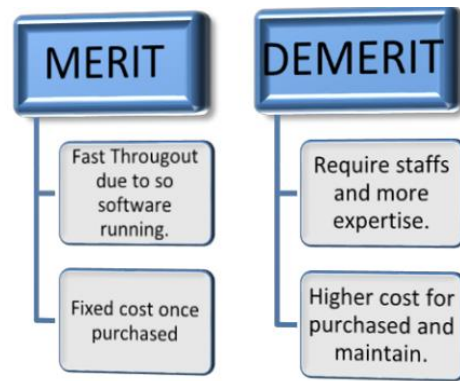

FIG. 4 merit and demerit of hardware

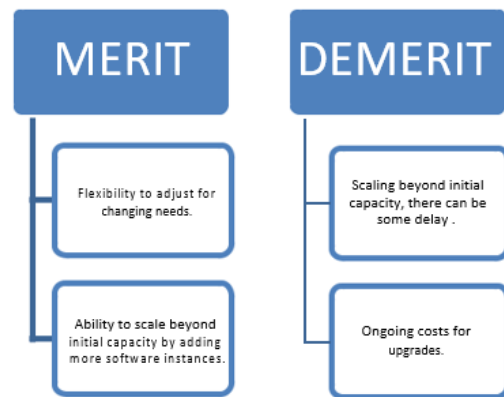## VIII. Software Merit and Demerit

FIG. 5 merit and demerit of software

In the above tables we understood the concepts of hardware load balancer and software load balancer and their pros and cons that how they are beneficial for us to use in different kind of ways, and we studied the demerits for the same.

IX. DNS Load Balancing and Hardware Load Balancing

We've already looked at the majority of load balancing, so now we'll look at another form of load balancing: hardware load balancing, and compare the two in their respective fields. Client requests to a domain within the Domain Name System are spread through various server machines using DNS load balancing, which is a software-defined approach to

load balancing. The DNS system sends a different version of the list of IP addresses each time it responds to a new client request using the round-robin method, therefore distributing the DNS requests evenly to different servers to handle the overall load. DNS load balancing failover safety is then provided by the automated elimination of DNS load balancing. DNS load balancing differs from hardware load balancing in a few ways, but both can be a very efficient solution for distributing traffic. Load balancing at the DNS level has a number of advantage is the scalability and price. A DNS load balancer distributes traffic to several different IP addresses, whereas the hardware solution uses a single IP address and splits traffic leading to it on multiple servers. As for pricing, hardware load balancers require a large upfront cost whereas DNS load balancers can be scaled as needed.

X. Some Load Balancing terms

A. SDN:Load balancing using SDN separatesthe control plane from the data plane for application delivery. This allows the control of multiple load balancing. It also helps the network to function like the virtualized versions of compute and storage. With the centralized control, networking policies and parameters can be programmed directly for more responsive and efficient application services. This is how networks can become more agile.

B. UDP: A UDP load balancer utilizes UDPload balancing is often used for live broadcasts and online games when speed is important and there is little need for error correction. UDP has low latency because it does not provide time-consuming health checks.

C. TCP: A TCP load balancer is a device that distributes traffic across the Internet. TCP load balancing ensures a consistent and error-free flow of

packets to IP addresses that would otherwise be skipped or corrupted.

D. SLB: It uses a system of load balancing algorithms to provide network services and content delivery. It prioritizes responses to unique requests received over the network from clients. Client traffic is distributed to servers using server load balancing.

E. Virtual: Via virtualization, virtual load balancing attempts to emulate software-driven infrastructure. On a virtual machine, it runs the programmed of a physical load balancing appliance. Virtual load balancers, on the other hand, don't get around the architectural issues that conventional hardware appliances have, such as reduced scalability and automation, as well as a lack of central management.

F. Elastic: Elastic Load Balancing adjusts traffic to an application in response to changing demand. It learns the status of application pool members (application servers) using system health checks and routes traffic to available servers, manages fail-over to high availability targets, or automatically spins up additional bandwidth.

G. Geographic: For optimal productivity and security, geographic load balancing redistributes application traffic through data centers in different locations. Local load balancing occurs within a single data center, while regional load balancing utilizes multiple data centers in multiple locations.

H. Multi-site: Multi-site load balancing, also known as global server load balancing (GSLB), distributes traffic among servers in multiple locations around the world. On-premises or in a public or private cloud, the servers may be used. After a disaster in one place makes a server inoperable, multi-site load balancing is critical for fast disaster recovery and business continuity.

I. Load Balancer as a Service:

Load Balancer as a Service uses advances in load balancing technology to meet the agility and application traffic demands of organizations implementing private cloud infrastructure.

XI. CONCLUSION

The paper presents a comparative study on load balancing approaches in different sectors and provide better solutions. The problem of load unbalancing in cloud computing was discussed. An abstracted load balancing model was briefly discussed together with activities involved in load balancing process. The data collected for this study had been gathered from five reputed potential databases that include IEEE Xplore digital library Science Direct, ACM digital library, Springer, and Elsevier. From the review conducted during this research process, it is concluded that there are a lot of issues still open in load balancing process which can be supported in future by applying an efficient and well managed load balancing algorithm most importantly along dimensions of additional QoS metrics and algorithm complexity evaluation.

XII. ABBREVIATIONS

VM: Virtual Machine
CSP: Cloud Service Provider
SDN: Software-Defined Networking
UDP: User Datagram Protocol
TCP: Transmission Control Protocol
SLB: Server Load Balancing
VoIP: Voice over Internet Protocol
HTTP: Hyper Text Transport Protocol

XIII. REFERENCES

[1]. Availability and Load Balancing in Cloud Computing Zenon Chaczko Venkatesh MahadevanShahrzadAslanzadehand ChristopherMcdermid1

[2]. The Gradient Model Load Balancing MethodF.C.H Lin R.M. Keller

[3]. Load Balancing in Cloud Computingusing Modified Throttled AlgorithmShridhar G Domanal G RamMohan Redday

[4]. Load Balancing in Cloud Computing using Stochastic Hill Climbing-A Soft Computing ApproachBrototi Mondal Kaushik Dasgupta Paramartha Dutta

[5]. Cluster based load balancing in cloud computing. Surbhi Kapoor Chetna Dabas

[6]. A Task Scheduling Algorithm Based on Load Balancing in Cloud ComputingYiqiuFangFeiWangJunwei Ge

[7]. https://www.geeksforgeeks.org/load-balancing-in-cloud-computing/

[8]. Load balancing in cloud computing Muhammad Samiullah

[9]. https://cloudcomputingprojects.net/load-balancing-in-cloud-computing-thesis/

[10]. Load balancing in cloud computing: Challenges an issues Shalini Joshi, Uma Kumari

[11]. User priority guided Min-Min scheduling algorithm for load balancing in cloud commuting Huankai Chen, Frank Wang, Na Helian, Gbola Akanu

[12]. Load Balancing in cloud computing Violetta N.Volkova, Liudmila V.Chemenkaya, Elena N.Desyatirikova, Moussa Hajali, Almothana Khodar, Alkaadi Osama

[13]. Load Balancing Cloud Computng: State of art A. Khiyaita, H.El Kakkali, M.Zbakh, Dafir El Kettani