International Journal of Allied Practice, Research and Review
Website: www.ijaprr.com (ISSN 2350-1294)

Identifying More Efficient Ways of Load balancing the Web (http) Requests.

Mukesh Negi
Project Manager, TechMahindra, Noida, India
mukesh_negi77@yahoo.com

Abstract— **Load balancing of web requests is one of majorly focused area in the field of Information Technology. As everybody knows, We are moving towards the E-business rapidly and today almost more than 60% of business fully dependent on web and rest 40% is also moving towards the same. it's the best way to cover customers all over the country and nation instead of just running a business within limit of a particular city, state or country. As a client, you can give contract to an IT organization for your ebusiness infrastructure and availability based on some SLA contracts but, For an IT organization it's always been a challenge to make your ebusiness solutions available 24*7 within SLA contracts. Even though they have all necessary solutions, designs and engineers for same but, due to the challenges coming day by day, it's very necessary to review all existing solutions regularly and then explore them further to enhance further. Load balancing is one of the way to keep your web applications highly available in the background without knowing where they are actually running and from where client are accessing it. There are lots of solutions are available in market for load balancing and making your web applications highly available, but as it's always a part of CIP (continuous improvement plans) So, there are further options and areas need to be explorer to find out further efficient ways of load balancing the web requests among different application servers and hosts.**

Keywords*— *Web; Load Balancing; E-business; High Availability; Continuous Improvement Plans

I.INTRODUCTION

Web, or you can refer WWW or W3, is a simple term which now almost everybody knows and would able to define But, Hardly a bunch of peoples know the complexities behind it. In a simple definition, You can say, It's a route to access all of the information's scattered around the world from any cornet of the world where you have a basic internet connectivity. You have to just type and enter the URL of website in the address bar of your browser, Doesn't matter where is the information on web across the world, it will display on your computer screen within a seconds. Yes, it's a magic of Web.

If you are not aware about anything apart from just your browser and web site address, then let me give you an small overview on same. Just like you access some document on your local desktop of laptop, same way the information like texts, images, videos etc are exist over web on some high configuration machines over the world. Web or www is the way to access that information locally from any corner of the world. So, How do you connect to that machine over the web or net when you just only access a web address or URL. Simple, the address you access as a web is called a DNS name. Like, if you access yahoo.com, then yahoo is a DNS name, purchased by organization running yahoo. There is a corresponding I.P address of DNS, and this is the I.P address of machine, also called host is the actual machine of yahoo which holds all of data you access when accessed yahoo.com. There is DNS server over the net, which holds all the DNS names and their corresponding I.P addresses. So, when you access or hit any URL on your browser, it first reach the DNS server, DNS server return corresponding I.P address, and then reaches that I.P address over the net where actual information exist, fetch the same and display on your local desktop or laptop.

Another complications behind that is, How that information parsed and accessed, because you get the information in proper format, indents, line, paragraph, colour and of course within a seconds or fractions of seconds depend on your net speed. So, for all of this you need manpower to design your website, Write code in some specific web supported language, an web server or application server to host you website, a machine to run your web or application server, further to publish it o net, a static I.P, a unique Domain name. Once you have all these things are in place, you would able to host your website.

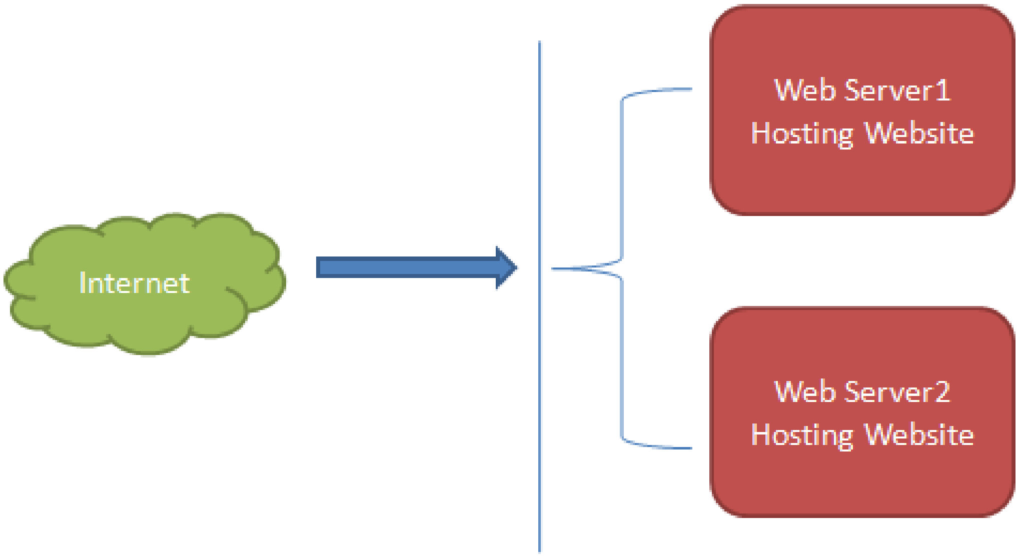
Now, further complications are, How you would make your website highly available. For example, just consider facebook.com. How many users are using it on daily basis every minutes and seconds 24*7. after all, it's some machine which hosting this website, so first point is how efficient is your machine to hold such a large traffic, because each user request will come to your machine which need a fraction of your machine resources like CPU, RAM etc, and second point is even you have good configuration machine, how efficient to handle multiple simultaneous requests. So, this is the actual complicated part which really deals with the high availability. If you don't have good resources at backend to handle loads then it will directly impact your business because as much as your website is slow, as much as users will go away from that.

So, to make your highly available and to handle multiple simultaneous requests, you have to consider the mechanism of load balancing. For example, If you purchased some multiple items in market, then to carry them on hand, you do the proper load balancing of items according to their weights on both hand. If you will not do same and carry all items in a single hand, then you would not able to sustain it for long. Load balancing concept also works on same theory, where you have to use multiple solutions like softwares, hardware's, dbms etc to allow all multiple requests to load balance between all instead of just overloading a particular machine or software.

II. Load Balancing Applications with HLB and Web Servers

As shown in below figure, It’s the easiest and more widely used way of load balancing to make your applications more available. There you can use more than one web servers to host and load balancing the requests. In between you can use a Hardware Load Balancer to divert the requests among different web servers. Here Web Server1 and Web Server2 hosting the same web applications and exist across two different machines. Another benefit of this architecture is your applications will be available in case any of the machines goes down. From requests load balancing perspectives, hardware load balancer will load balance the requests between two web servers.

Only the drawback is, In case of any performance or overload kind of issues with any one of the web server machine, your hardware load balancer will continue to send request unless it went down.



Users will access the URL from internet that request will hit the DNS server, and then to your hardware load balancer and then hardware load balancer will forward or load balance the requests between backend web servers where your web application or website is hosted.

IV. Load Balancing your Application Servers

Another approach is, behind your web servers you can further use application servers like web logic server, web sphere etc to make your application highly available. Web servers you can use as proxy to divert the requests between different application servers.

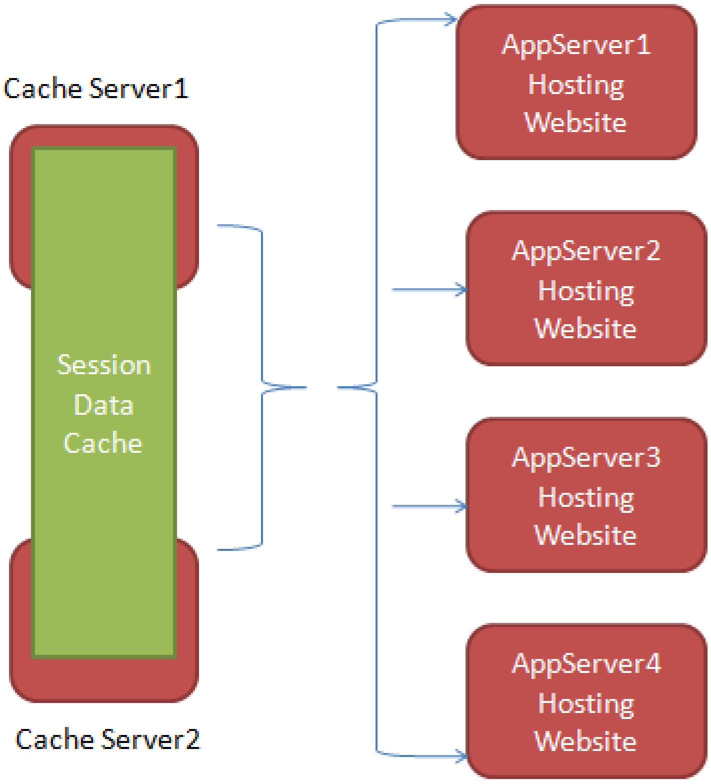
As shown in below architecture, Hardware load balancer will divert the request to backend web servers, and further each web server will divert the requests to backend applications servers.

Limitation is same, In your any of the backend server is under heavy load or under any performance issues, your hardware load balancer or webserver will continue to send request to that server until that got down.

V. Current Solutions

One more enhanced and efficient way to load balancing and making your application highly available is external cache server (For example Oracle Coherence Server) where all web sessions data stored in external JVM which make backend application servers more reliable. In this way, application servers get more memory (heap) for other application operations, as well there is no need to local session replications to reflect on other server in case primary server goes down. Here, all sessions data stored in external JVM, and case of any server goes down, session data will reflect from the common cache. Further, request will load balance from external cache server to backend application or web server to make it highly available.

Limitations would be same, In case any backend server is under high load or performing poorly due to any issues, your balancer will continue to send request to that server until it would go down and impact user expectations and satisfactions.


Cache Server1
Session Data Cache
Cache Server2
AppServer1 Hosting Website
AppServer2 Hosting Website
AppServer3 Hosting Website
AppServer4 Hosting Website

IV.CONCLUSION

As we have seen in complete article that one of the major limitations across all existing solutions is, In case of any of the backend server or servers are in overload condition, or not performing well up to the expectations due to any reasons, your balancer will continue to send request to that/those servers till it would go down. After studying and analyzing all the existing solutions, I have recommended one of the best approaches to deal with the same limitations. Approach is, before sending requests to any of the backend web or application servers, load balancer should consider and check the important performance parameters first, and forward request to that server only if feel server is in good condition to handle the requests after analyzing all of the current health parameters. Gathering all these things at run time would create further performance issues, so another better approach for same is, your load balancer should get status of all important parameters from backend web or application server in a regular interval and store in local cache or file, and then compare against that before sending request. In this way, current valid request will forward only to the available servers and users will not face any performance or availability issues.

VI.REFERENCES

[1] An Approach on Semi-Distributed Load Balancing Algorithm for Cloud Computing System , International Journal of Computer Applications (0975 – 8887) Volume 56– No.12, October 2012

[2] Ant colony Optimization: A Solution of Load balancing in Cloud , International Journal of Web & Semantic Technology (IJWesT) Vol.3, No.2, April 2012

[3] A Cluster Based Replication Architecture for Load Balancing in Peer-to-Peer Content Distribution, International Journal of Computer Networks & Communications (IJCNC) Vol.2, No.5, September 2010

[4] Client-Side Load Balancing and Resource Monitoring in Cloud, Miss.Rudra Koteswaramma / International Journal of Engineering Research and Applications, Vol. 2, Issue 6, November- December 2012, pp.167-171

[5] Co-operative Scheduled Energy Aware Load-Balancing,technique for an Efficient Computational Cloud, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011 ISSN (Online): 1694-0814

[6] Comparison of Load Balancing Strategies on Cluster-based Web Servers, YM Teo and R Ayani, Transactions of the Society for Modeling and Simulation (accepted for publication), 2001.,

[7] Dispatcher Based Dynamic Load Balancing on Web Server System, Harikesh Singh, Dr. Shishir Kumar, nternational Journal of Grid and Distributed Computing Vol. 4, No. 3, September, 2011

[8] International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 6, June 2012

[9] Load Balancing In Public Cloud, Shrikant M. Lanjewar, Susmit S. Surwade,Sachin P. Patil,Pratik S. Ghumatkar, ,Prof Y.B. GURAV, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver. VI (Feb. 2014), PP 82-87

[10] Load Balancing Approaches in Grid Computing Environment, International Journal of Computer Applications Volume 72 - Number 12,Year of Publication: 2013\

[11] Multi-agent Optimized Load Balancing Using Spanning Tree for Mobile Services, ©2010 International Journal of Computer Applications (0975 – 8887) Volume 1 – No. 6

[12] Review on Existing Load Balancing Techniques of Cloud Computing, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 2, February 2014