BNU

Beaconhouse National University

Research paper

Comparative Run-Time Analysis of Web Servers on The Basis of Resource Deviation

Submitted to

Sir Saad Saleem Malik

Submitted by

Omer Mustafa – F2016-472
Abdullah Waqar – F2016-064
Mubashar Hussain – F2016-404

Dated

December 29th, 2019

Abstract

With the diversity in available web resources, it is important to streamline important variables in order to effectively highlight the factors that make the biggest impact in determining efficiency for digital infrastructures. Web servers lie at the core of most major digital web infrastructure powering the major avenues of internet usage. In order to effectively simplify decision-making, it is important to have quantified data on the performance deviation of servers on the basis of variability in the allocated resources. This study aims to perform a quantitative analysis on the impact of resource deviation by effective comparison of two diverse web environments and subsequent resource variation to form a verdict about the factors that make the biggest impact.

Keywords: web server, cloud computers, OpenLiteSpeed, Nginx, web server load test

1. Introduction

Internet has become cardinal in our lives. 4.48 billion people around the world actively use internet services (J., 2019) as per a recent survey done by DataReportal in October, 2019. It goes a lot in an effort to serve this unimaginable amount of web traffic. From the core data centers to the network of world wide web, the cost of serving data across multiple points is relative to not only the amount of web traffic but also the infrastructure of the specific applications. The infrastructure that is backing all this up can be quite complex but can be broken down into small systems running services and providing a certain functionality. The web applications we see and interact runs on top of these services, taking advantage of multiple services that an infrastructure offers. An application poorly structured over the internet in terms of software and hardware can be a bottleneck to all the performance issues faced by its users (Shuler, 2019). One of the major root cause reasons for slow application performance is the slow web servers. These web servers act as host attendant between external requests by the internet users and the physical hardware computer connected to the network. In order for web servers to serve data including texts, images, videos, or data of any sort, it has to be stored in the physical storage with access and respective permission for the purpose. As technology transcend leaps and bounds, different types of web servers have been created with consideration of improvements in mind. Each having their advantages and disadvantages. New web servers have been introduced into the space with better performance and speed.

This paper reviews performance of different available web servers along with their cost-benefit analysis, determining options which can benefits internet applications by reducing cost and improving overall performance. We shall review primarily two web servers Nginx (pronounced as ‘engine-x’) and OpenLiteSpeed (open source edition of LiteSpeed web server).

1.1 Overview of web servers

Using an appropriate web server can make or break the quality of a web application that would be served, the decision of what web server should be used for an application depends on the level of sophistication of the service itself which would be provided. “Web servers” can refer to software or hardware or even both of them simultaneously. If it is referred as a hardware, a web server is a powerful computer that stores web server software and the component files of a website and it has to be physically connected to the internet. On the other hand, a web server software includes different parts that control how web users access files hosted on the physical hardware. This piece of software understands web URL addresses, and HTTP (Fielding and Kaiser, 1997). Whenever the required information is needed the request is generated using a piece of software called a web browser, this request reaches the correct web server (physical hardware) and the HTTP web server gets the requests, find the requested document and sends the file back to the browser, considering if the file is available on the server otherwise 404 response is returned.

2. Runtime Analysis of Web Server

It is important to have a sound analysis model to conclude usable results. There are multiple ways we can run the test over different environments including local host to web hosting companies. Performing the test on a local web server can be cost effective but it does not create an environment similar to the web applications running online. On the other hand, dedicated or virtual private servers might also not be the best options to explore the web servers. Cloud computing is the most adopted web hosting solution on the internet and in a survey done by (Flexera, 2019), 94% of the people directly or indirectly use cloud computing and these numbers are bound to increase in year 2020 and onwards.

2.1 Cloud Computing Overview

The term “cloud” was derived from the world of telecommunication. In 1960s John McCarthy gave his opinion that "computation may someday be organized as a public utility”. He was first person to give the underlying concept of cloud computing (Ramanathan, 2014). Cloud computing is the pool of configurable computing resources (e.g. networks, servers, storage applications and services).

Cloud computing in its simplest term means storing and accessing data and programs over the internet instead of your computer's hard drive. The cloud is just a metaphor for the Internet when in fact is coined to refer to the huge and gigantic server-farm infrastructure where resources of multiple servers are pooled together to act as a single unit and has some common characteristics such as accessibility, scalability, flexibility, performance, etc. and it is due to these characteristics that user can compute or assess the data or application or any other service with simply the help of a browser. Providing control and performance on the go.

2.2 Virtualization

Virtualization is a process in computing where multiple servers’ resources are pooled together to act and function as a single unit. Actual resources and the complications of their implementation from end users, for various benefits such as security, optimization, consolidation etc. In computing, there are primarily 2 main kinds of Virtualization. Hardware virtualization and software virtualization (L and D, 2014). In hardware virtualization, the resource is your typical Processor Cores, Memory and may be storage. Cores can be divided by a process called hardware partitioning, which essentially mean that a single processor comprising of 4 cores can be divided to act as 2 processors or machine with 2 cores each and so goes for all the other resources such as RAM, Storage, etc. Software virtualization on the other hand refers to virtualizing the software layers and is achieved through the use of software such as VMware, VirtualBox and others.

This hardware and software complexity create a room for improvement and opportunity. Output resource optimization using hardware is limited to multiple factors such as investment in better hardware and infrastructure, we want to better understand how we could leverage this opportunity using software and generate better outputs.

2.3 Performance Analysis

In order to examine the performance of different popular web servers (Nginx, OpenLiteSpeed and Apache), we load or stress tested their capabilities to determine how the web servers behave when a lot of web users access it simultaneously. Load test (also known as a stress test) is a type of test where we send huge traffic within a short period of time. This test helps to determine how a web server responds on high traffic or when one of your blog posts goes viral and there is a sudden flood of visitors (Varghese, 2019).

2.4 Experimental set-up

In order to implement and simulate an application which is production grade, deployed over internet, we chose the cloud environments for the experiment. To determine the best available web server that can scale with your application without an overhead cost, we performed two experiments;

1. To analyze performance under normal circumstances.
2. To analyze performance under stress or unusual amount of web traffic.

Furthermore, we shall increase the cloud resources to see the impact on the performance up until the performance of both the applications match each other, while performing the load test.

2.5 Test Configurations

- Cloud compute instance/server with 1-GB RAM and 1 virtual CPU (Vultr, 2019)
- Server Location: Frankfurt, Germany
- A simple WordPress website with Twenty Twenty theme installed and activated
- Nginx using Cockpit, a control panel
- OpenLiteSpeed using LiteSpeed Web-Admin panel
- http://nginx.mande.website sub-domain is set-up for Nginx testing
- http://openlitespeed.mande.website sub-domain is set-up for OpenLiteSpeed testing
- No CDN, Redis or Varnish caching
- No Cloudflare proxy (DNS alone)
- https://loader.io, a simple cloud-based load testing tool (SendGrid, 2019)
- MySQL database

Both the cloud servers with equal resource are set-up, one with Nginx (Liu, 2016) and the other with OpenLiteSpeed (OpenLiteSpeed, 2019).

2.6 Load Tests and Results

In order to do performance analysis, three comparison tables have been made in this research work. Table-1 shows the load test performed without cache enabled and on the contrary Table-2 with the cache enabled. Whereas, Table-3 represents primarily the load test of Nginx server, after 4 times increment in its resources to match the performance of the OpenLiteSpeed Server.

Test # 1 – Clients per second and cache disabled.

To evaluate how a server performs when 1000 users connect every second over the period of 1 minute.

TABLE 1

Web Server	No. of clients	Duration (in seconds)	Average Response Time (in milliseconds)	No. of successful requests	Error rate *
Nginx	1000	60	4946	1924	73.6%
OpenLiteSpeed	1000	60	19709	2295	0%

*Errors are HTTP response codes from 400 and up, grouped together in a percentage (SendGrid, 2019).

Result simulation and embedded chart.

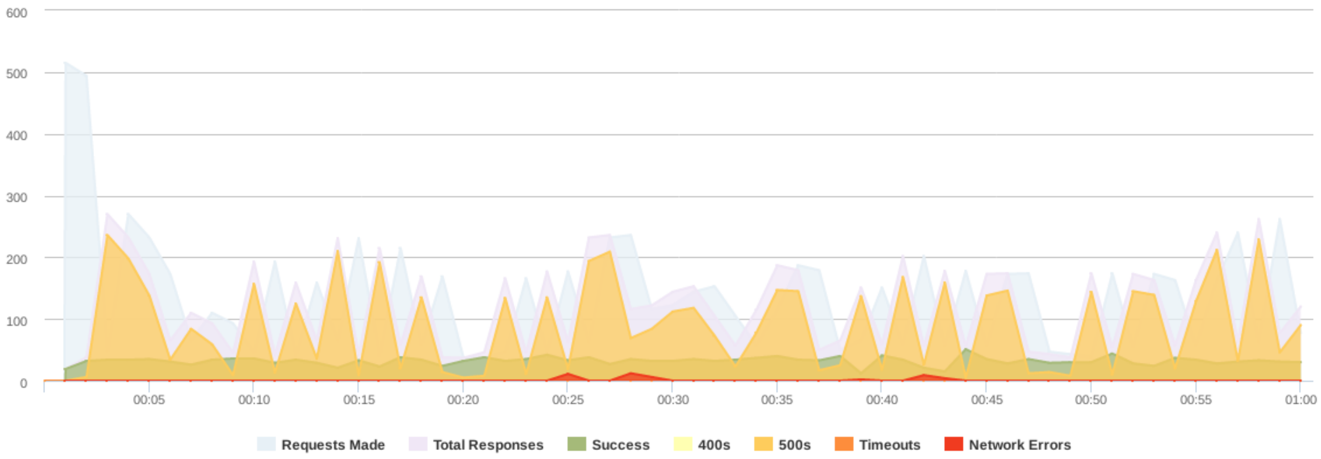

Fig 1. Nginx Stress Test Result, http://bit.ly/2QrNFFd

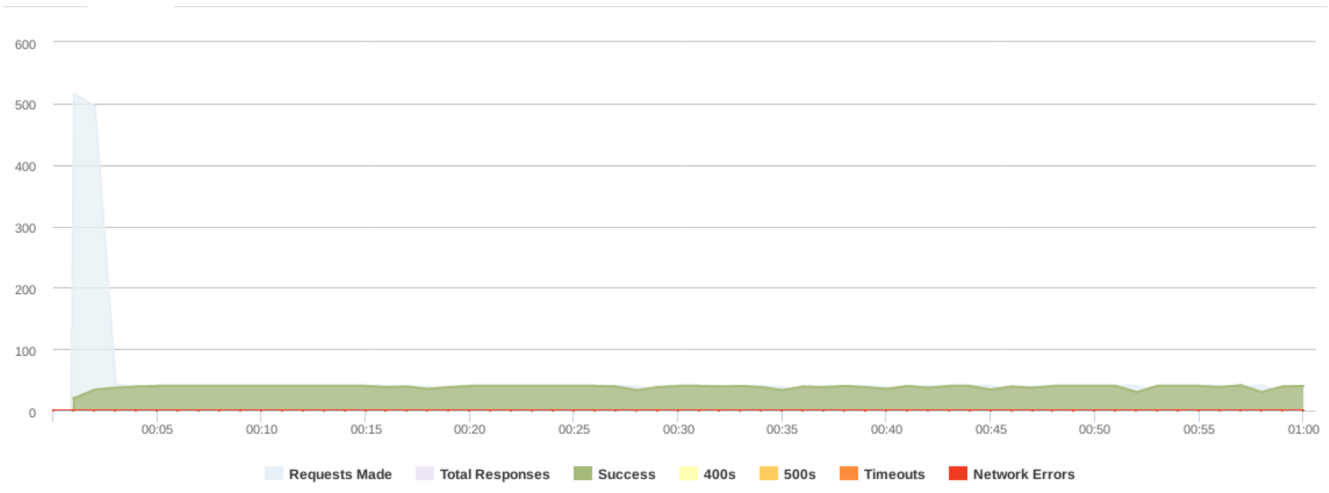

Fig 2. OpenLiteSpeed Stress Test Result, http://bit.ly/365TqPb

In our first experiment, Nginx served 1924 page requests in comparison to OpenLiteSpeed, despite having a huge error rate, the average response time is 4946 milliseconds. On the other hand, OpenLiteSpeed performed slower but was able to serve more page requests.

Test # 2 – Clients per second and cache enabled.

This time following cache plugins, which are best suited for the respected servers, were enabled and the same load test was repeated.

1. WP Fastest Cache for Nginx Server.
2. LiteSpeed Cache for OpenLiteSpeed Server.

TABLE 2

Web Server	No. of clients	Duration (in seconds)	Average Response Time (in milliseconds)	No. of successful requests	Error rate
Nginx	1000	60	3249	7792	41.4%
OpenLiteSpeed	1000	60	90	60000	0%

Result simulation and embedded chart.

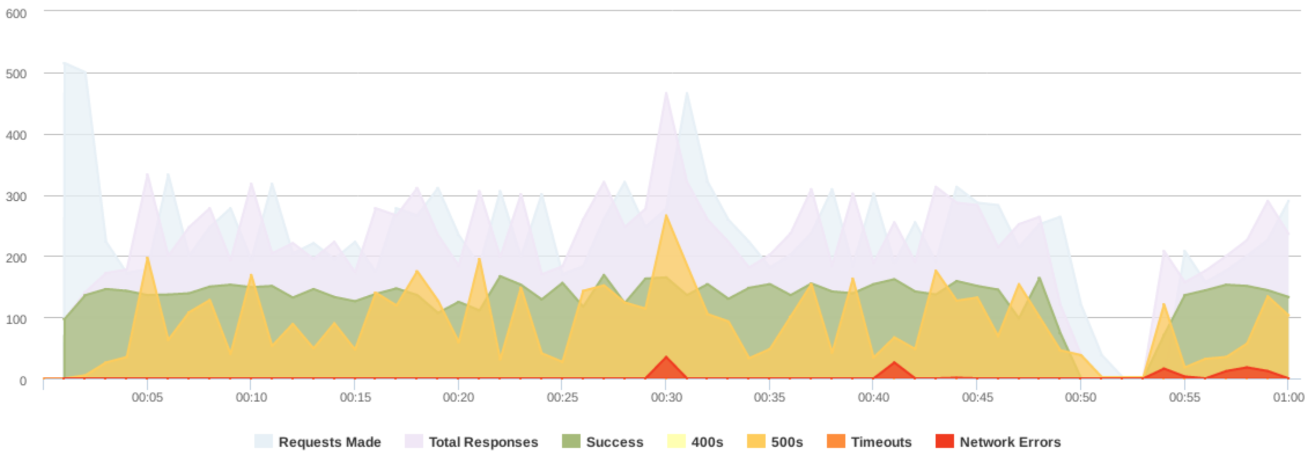

Fig 3. Nginx Stress Test Result, http://bit.ly/2SzkH8V

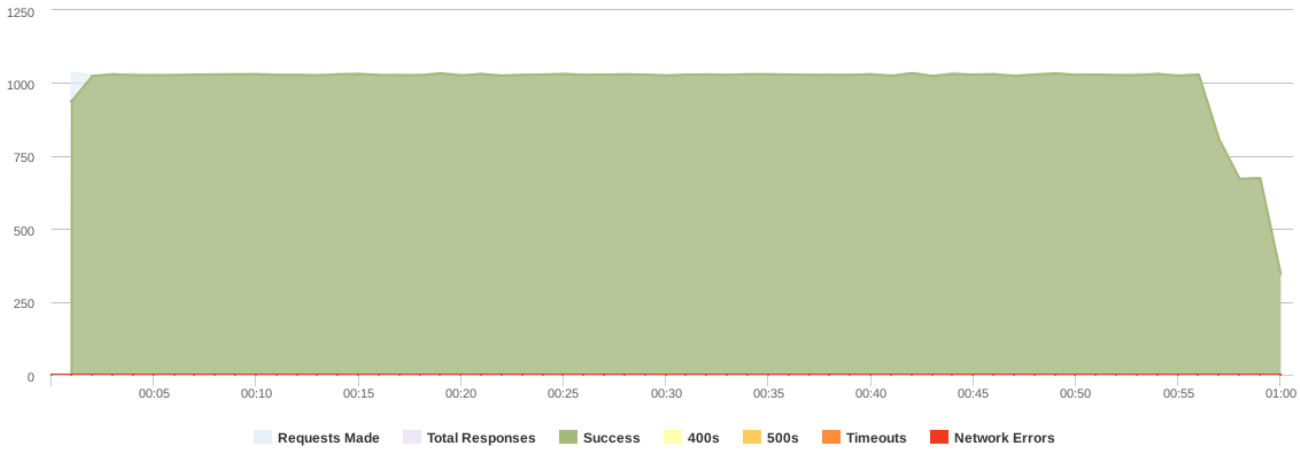

Fig 4. OpenLiteSpeed Stress Test Result, http://bit.ly/2Q3KZOU

Second test with cache storage enabled, presented a big contrast in performance of both the web servers. OpenLiteSpeed was not only able to serve 100% requests but also in an average response time of 90 milliseconds, accounting both the servers with equal amount of resources.

Test # 3 – Nginx server with double the resources and cache enabled.

The following table shows the results after we incremented Nginx server resources 4 times.

1. Nginx server with 4-GB RAM and 2 virtual CPUs
2. OpenLiteSpeed server with 1-GB RAM and 1 virtual CPU

TABLE 3

Web Server	No. of clients	Duration (in seconds)	Average Response Time (in milliseconds)	No. of successful requests	Error rate
Nginx	1000	60	1603	18448	21%
OpenLiteSpeed	1000	60	90	60000	0%

Result simulation and embedded chart.

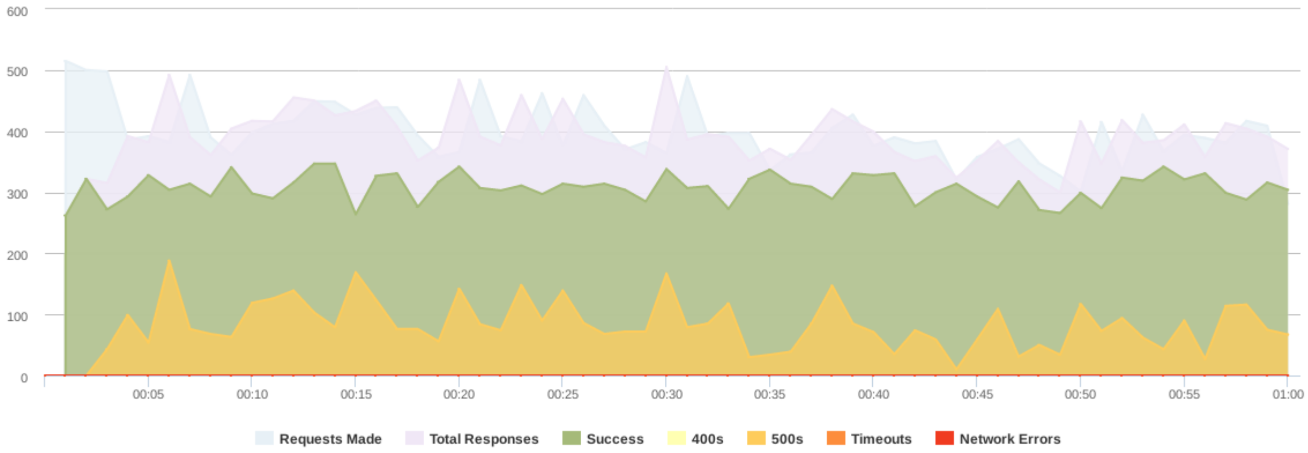

Fig 5. Nginx Stress Test Result, http://bit.ly/2rGwGXm

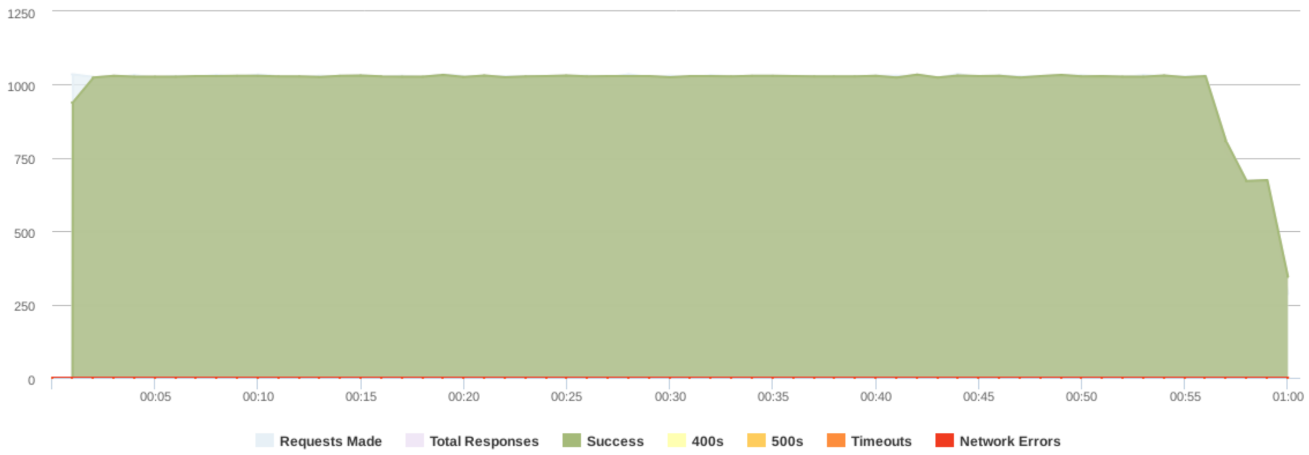

Fig 6. OpenLiteSpeed Stress Test Result, http://bit.ly/2Q3KZOU

We quadrupled Nginx server resources to match the performance of OpenLiteSpeed server but still it was not able to serve half the amount of page requests compared to OpenLiteSpeed server with 4 times less server resources.

3. Conclusion

In this paper we tested two different web servers with resource deviation to analyze their performance under stress. The comparison will help businesses, developers and individuals make critical decision while a web server as per their requirements. It has also been observed that cache memory helps boost performance and serve more traffic with less response time. Furthermore OpenLiteSpeed which is quite new in the web server space performed better than Nginx webserver, a market leader, even with lesser server resources. In future, research to identify the flexibility across multiple programming languages can be considered to measure performance and reliability of both the webservers.

References

Fielding, R.T. and Kaiser, G. (1997). The Apache HTTP Server Project. *IEEE Internet Computing*, 1(4), pp.88–90.

Flexera Blog. (2019). *Cloud Computing Trends: 2019 State of the Cloud Survey | Flexera Blog*. [online] Available at: https://www.flexera.com/blog/cloud/2019/02/cloud-computing-trends-2019-state-of-the-cloud-survey/.

J., C. (2019). *Global digital population 2019 | Statistic*. [online] Statista. Available at: https://www.statista.com/statistics/617136/digital-population-worldwide/ [Accessed 24 Dec. 2019].

Shuler, R. (2019). *How Does the Internet Work?* [online] Stanford.edu. Available at: https://web.stanford.edu/class/msande91si/www-spr04/readings/week1/InternetWhitepaper.htm.

Statista. (2018). *Consumer cloud computing users worldwide 2018 | Statista*. [online] Available at: https://www.statista.com/statistics/321215/global-consumer-cloud-computing-users/ [Accessed 27 Dec. 2019].

Ramakrishnan Ramanathan, B. L., 2014. Resource Optimization Based on Demand in Cloud Computing. Research Journal of Applied Sciences, Engineering and Technology, pp. 1724 - 1731.

Srivastava Priyanshu, K. R., 2018. A Review Paper on Cloud Computing. International Journals of Advanced Research in Computer Science and Software Engineering, 8(6), pp. 17-20.

L, M. and D, A. (2014). Virtualization in Cloud Computing. *Journal of Information Technology & Software Engineering*, 04(02).

Liu, J. (2016). The Improvement and Implementation of the High Concurrency Web Server Based on Nginx. *Computing, Performance and Communication systems*.

OpenLiteSpeed. (2019). *Get OpenLiteSpeed!*. [online] Available at: https://openlitespeed.org/ [Accessed 29 Dec. 2019].

SendGrid, I. (2019). *General FAQs - Loader.io Help Desk*. [online] Support.loader.io. Available at: https://support.loader.io/category/29-frequently-asked-questions [Accessed 29 Dec. 2019].

SendGrid, I. (2019). *Test Results - Loader.io Help Desk*. [online] Support.loader.io. Available at: https://support.loader.io/article/19-test-results [Accessed 29 Dec. 2019].

Vultr. (2019). *SSD VPS Servers, Cloud Servers and Cloud Hosting by Vultr*. [online] Available at: https://www.vultr.com/ [Accessed 29 Dec. 2019].