

# Big Data Performance Analysis In Apache And Internet Information Services

Mohammed M. A<sup>1</sup>, Bartholomew E<sup>2</sup>

Department of computer science, Faculty of science, University of port harcourt, nigeria

## ABSTRACT

*The goal of this study is to identify the impacts of big data on server performance, how it is being used and the amount of processing power it consumes from the server. Big data handling system individually on different servers using integrated software tools. In this study, data were captured and analyzed in order to gain actionable insights from the data at a much lower cost. The study uses the metrics (Response Time, Throughput, Computer Resources, Availability, Data Compressibility and Bandwidth) and main techniques which include Analysis modeling, Simulation and Measurement. The results of this analysis offer researchers with tool to predetermine true expectations*

*from the big data. This will allow for more evaluation of big data and may direct future research on its performance on web servers. By identifying big data performance on servers, we will eliminate the assumption of server barriers. The performance analysis covers the volume of data, variety of types of data and the speed at which the data must be must processed. The findings may be useful for business people, consumer companies, massive big data companies, media, private sector, retail banking, government, individuals and researchers. This was further implemented using VB.NET.*

**Keywords:** Big Data, Server, Performance, Analysis, XAMPP, SSMS, VB.NET, APACHE, IIS, NGINX.

## Introduction

Big Data could be defined as the data that surpass the processing ability of schematic database. Such data is excessively large, speedy in nature, and could not fit the organization of database design. To benefit from that type of data, different approaches should be improvised. But the significance is restricted to its accessibility which is linked to processing capabilities and performance of the servers that process it. We would like to ask and answer certain questions to aid our research. What server is the best in aiding fast processing of a medicine data, for instance, to provide us valuable information needed to tackle a serious health issue? Irrespective of the loyalty we have for a server or its platform, once we have such information we needed to migrate. This can also quickly process the information. That is what Big Data is really all about. What is truly transformative about Big Data is the ease with which we can now use data. No longer do we need complex software that takes months or years to set up and use. Nearly all the analytics power needed is available through simple software downloads or in the cloud. It is the combination of these capabilities that is truly changing the game when it comes to making it easy to use and apply data big data. But as researchers we needed to pulse and ask ourselves simple questions; the cloud and the internet are powered by various

servers among which include Apache, NginX and Internet Information Service (IIS). The performance of which affect the Big Data operations hence a detail analysis of the performance will not only point to new directions but offer developers with tool to predetermine their true expectations from the Big data. A server is any combination of hardware or software designed to provide services to clients [1]. The word performance means “How well is the server doing the work it is supposed to do”. Performance in Servers can be viewed as comparing relative to other servers or the same server before/after changes. It is characterized by the amount of useful work accomplished by the software compared to the time and resources used. A good server performance can be evaluated in measurable, technical terms, using one or more of the metrics [2] listed below.

- i) Short **response time** for a given piece of data.
- ii) High **throughput** (rate of processing data)
- iii) Low utilization of **computer resource(s)**.
- iv) **High availability** of the server system or application.
- v) Fast (or highly compact) **data compression** and decompression.
- vi) High **bandwidth** /short **data transmission** time.

So why has Big Data become the focus of server performance analysis? Big Data has broken into the mainstream due to three trends coming together. First, multiple high-profile consumer companies have ramped up their use of Big Data. For instance Facebook uses Big Data to track user behavior across its network. The company makes new friend recommendations by figuring out whom else the user know. The more friends you have, the more likely you are to stay engaged on Facebook. Business networking site LinkedIn uses Big Data to connect job seekers with job opportunities. With LinkedIn, headhunters no longer need to cold call potential employees. They can find and contact them via a simple search.

Similarly, job seekers can get a warm introduction to a potential hiring manager by connecting to others on the site. Facebook on NASDAQ, LinkedIn on NYSE. These companies and Google are consumer companies on the surface, they are really massive Big Data companies at the core. Third, business people, who are active users of Amazon, Facebook, LinkedIn, and other consumer products with data at their core, started expecting the same kind of fast and easy access to Big Data at work that they were getting at home. If Internet retailer Amazon could use Big Data to recommend books to read, movies to watch, and products to purchase, business users felt their own companies should be able to leverage Big Data too [3].

Servers are at the core of this modern revolution. The Google search engine itself is designed to work seamlessly across hundreds of thousands of servers. If more processing or storage is required or if a server goes down, Google's engineers simply add more servers. Some estimates put Google's total number of servers at greater than a million [4]. But improved server performance is likely to save organizations billions in terms of server additions for a specific Big Data job handling. In 2012, the Obama administration announced the Big Data Research and Development

Initiative, to explore how big data could be used to address important problems faced by the government. The initiative is composed of 84 different big data programs spread across six departments. Big data analysis played a large role in Barack Obama's successful 2012 re-election campaign.

### Materials and Methods

This research work focused on the impacts of big data performance on the web servers. We studied various servers' functionalities, techniques, technologies, to enhance the literature and our scope, focusing on the big data performance analysis on Apache and IIS web servers. This was done by using certain metrics as performance indicators. Due to the structure of the big data and the chosen servers, the Evaluation model (as shown in Figure 3.1) was chosen as a process model for this project. Here, performance of each web server was evaluated against big data as it was sent to the server. Consequently, reports were generated based on the tests conducted on the servers. Data were captured and analyzed, comparison were made. We implemented our designed approach using VB.NET as Programming Language, XAMPP and Sql Server Management Studio (SSMS) respectively as development tools which serves as platforms for Apache and IIS.

Big data is an evolving term that describes any voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for information [5]; Gartner opined big data as high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization [6]. Big data performance analysis as presented in this research work when compared with the existing data processing methodology enabled researchers to determine the right server in handling big data.

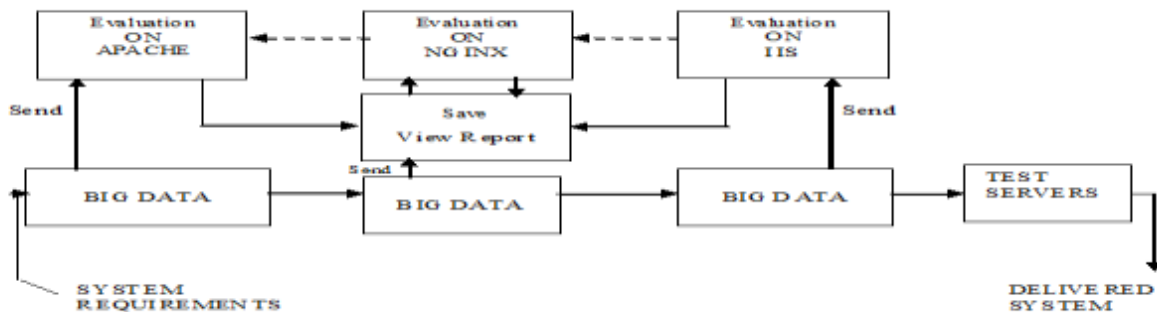


Figure 3.1: An illustration of the Evaluation model

### Results

The result obtained after selection of the program module of the Interface which test servers, paved way for comparison (as shown in Figure 3.2). At this point, data were captured and analyzed which provides actionable insight from the data. We were able to

evaluate and compare the performance of existing competitive servers, which are assumed to be available for performance measurements. In the area of server management, the server whose result yields higher metrics in terms of response time, memory usage and job accomplished enabled us determine the impacts of big data on the performance of servers.

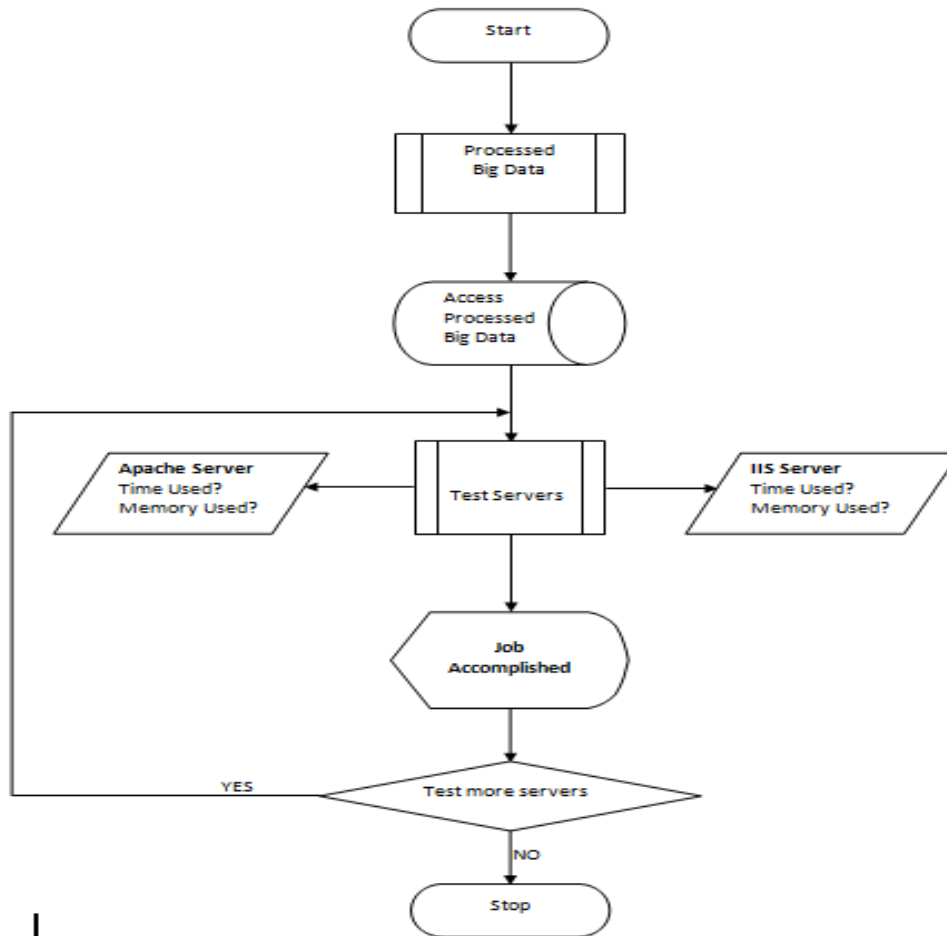


Figure 3.2: Showing how big data performance analysis was conducted

### Discussion

With our proposed system, big data performance on servers could be improved using performance indicators, by taken adequate measures in the event that outputs generated from the implementation contrary to the objectives. Consequently, consumers companies, business people, private sectors and end users would not only save cost but also optimize times and resources when using big data on servers.

### Conclusion

In conclusion, the new approach to big data performance analysis on web servers was developed using open-source software (XAMPP and SSMS) and implemented with VB.NET being a simple, multi-paradigm, structured, object-oriented, modern and event-driven high level programming language. This enabled us as researchers to determine the right server in handling big data. Accuracy in big data may lead to more confident decision making. And better decisions can result in greater operational efficiency, cost and risk reduction.

### **References**

- [1] Hilbert, A. B., & Martin (2015). "Big Data for Development: A Review of Promises and Challenges. Development Policy Review". Retrieved 2015-10-07 from martinhilbert.net.
- [2] Arnold, O. A. (1994). Computer Performance Analysis with Mathematica, Academic Press, USA.
- [3] David F. (2015) Big Data BootCamp, What Managers need to know to benefit from Big Data Revolution, Apress USA.
- [4] David F. (2015) Big Data BootCamp, What Managers need to know to benefit from Big Data Revolution, Apress USA.
- [5] Margaret Rouse (2014). Definition of Big Data. Retrieved October 2014 from <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data>
- [6] Laney and Douglas (2012) "The Importance of 'Big Data': A Definition". Gartner.