

Big Data and Cloud Computing Issues

Awodele .O
Department of
Computer Science
and Information
Technology, School
of Computing and
Engineering
Sciences, Babcock
University, Ilishan-
Remo, Ogun State

Izang A.A.
Department of
Computer Science
and Information
Technology, School
of Computing and
Engineering
Sciences, Babcock
University, Ilishan-
Remo, Ogun State

Kuyoro S.O
Department of
Computer Science
and Information
Technology, School
of Computing and
Engineering
Sciences, Babcock
University, Ilishan-
Remo, Ogun State

Osisanwo F.Y
Department of
Computer Science,
School of Computing
and Engineering
Sciences, Babcock
University, Ilishan-
Remo, Ogun State

ABSTRACT

Big data is a key concept that cannot be overlooked in the IT world considering the prominent increase in data, and data related services, it is important to explore this field and look at ways to improve data service delivery especially in the cloud. Cloud computing on the other hand helps in tackling the issue of storage and data service.

This research focus on the two key concept big data and cloud computing and some of the issues and challenges that are inherent with the deployment of cloud services and big data. The shows study that security challenges is among the most prominent issue in cloud and big data services. The plumbing issue and some other issues such as the issue of the cost to run cloud services in handling big data were observed. Also the issues of service level agreement which gives an organization the assurance of enjoying all services rendered by the organization running the cloud services.

After considering some of the issues associated with big data and cloud computing, some solution were suggested towards improving the two key concepts which will go a long way in increasing the adoption rate of cloud computing by organizations. It is important for organizations to consider the nature of how their data will grow in the future before deploying any cloud service in their business.

Keywords

Big data, cloud computing, virtualization, and Data Services.

1. INTRODUCTION

In today's business environment, the importance of data cannot be overemphasized as economic and social activities have long relied on data to survive. But today the increased volume, velocity, variety, social and economic value of data gestures a paradigm shift towards a data-driven socioeconomic world. In parallel with the continuous and significant growth of data; better data access, availability of powerful ICT systems, and ubiquitous connectivity of both systems and people has come. This has led to intensified activities around Big Data and Big Data Value. Powerful tools have been developed to collect, store, analyze, process, and visualize huge amounts of data. Open data initiatives have been launched to provide broad access to data from the public sector, business and science [1].

In the Journal of Science 2008, "Big Data" is defined as "the representation of the progress of human cognitive processes, which usually includes data sets with sizes beyond the ability

of current technology, method and theory to capture, manage, and process the data within a tolerable elapsed time". So also ,

the definition of big data as given by the Gartner defined it 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" [2].

Big Data can be described as a massive volume of structured and unstructured data which are so large and very difficult to process this data using traditional methods and recent software technologies. [3].

Furthermore, Big data is the elusive, all-encompassing name given to enormous datasets stored on enterprise servers for example, data stored in Google (which organizes 100 trillion Web pages), Facebook (with 1 million gigabytes of disk storage, its data keep increasing on a daily basis), and YouTube (which contains 20 petabytes of new video content per year).

Big data is also used in science, for scientific applications such as weather forecasting, earthquake prediction, seismic processing, molecular modeling, and genetic sequencing. Many of these applications require servers to run with tens of petabytes of storage, such as the Sequoia (Lawrence Livermore) and Blue Waters (NCSA) supercomputers [4].

The three main terms that generally signify Big Data are:

- i. Volume: This has to do with the amount of data generated on a daily basis which is so large and keeps increasing with time.
- ii. Variety: Today data is created in different type, form and formats such as emails, video, audio, transactions etc.
- iii. Velocity: This has to do with the speed it takes to produce data and how fast this data produced needs to be processed on time to meet individual demand.

The other two properties that need to be critically consider when talking about Big Data are Variability and Complexity as depicted by [3].

- i. Variability: this goes along with velocity, and it has to do with how inconsistent the flow of data can be with respect to time and far it can go.
- ii. Complexity: the complexity of the data must be considered especially when we have multiple source

of data. The data must be rearranged in such a format that will be suitable for processing.

Recent trends in Technologies today has not only supported the collection of large amounts of data but is has also utilizes

the management of such data effectively. And such technologies also supports the daily transactions made all over the world ranging from Bank transactions, Walmart customer transactions, and social media transactions such as the once generated from Facebook, twitter, Instagram, YouTube etc.



Figure 1: Big Data [5]

Some common characteristics of big data as highlighted by [5] includes the following:

- a) Big data incorporates both structured and unstructured data.
- b) Big data discourses the following speed and scalability, mobility and security, flexibility and stability.
- c) In big data the time it takes to retrieve information is critical to be able to emphasize how important various data sources are which may include mobile devices, radio frequency identification, tablets, the web and a growing list of automated sensory technologies etc.

There are various definitions used to define cloud computing however, most researchers agree with [6] definition who defines Cloud Computing as “a model for enabling ubiquitous, 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. This definition gives Cloud Computing, a more general description.

However, Cloud Computing is seen as a technology that depends on sharing of computing resources than having local servers or personal devices to handle the applications of users. In Cloud Computing, the word “Cloud” can be used interchangeably with “Internet”, so therefore Cloud Computing means a type of computing in which services are delivered through the Internet or network of servers. The main aim of Cloud Computing is to make use of computing power to execute a large pool of instruction per second. Cloud Computing uses networks of a large group of servers with specialized connections to distribute data processing among the servers. Instead of installing a software suite for each computer, with the use of cloud computing technologies, it is possible to install a single software in a computer which happens to be the host computer and users will log into a web-based service that host all the programs to be accessed by the user. There is a significant workload shift, in a cloud computing system which helps local computers to lessen the burden of having to host a lot of programs and applications.

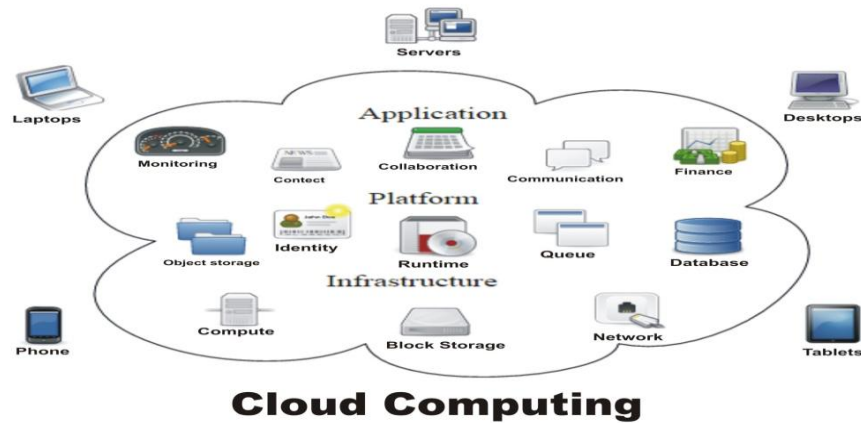


Figure 2: Cloud computing scenario [5]; [6]

Cloud computing simply performs a desired computation (mostly on big data) on a remote server that a subscriber has configured and controls, rather than on the subscriber's local desktop PC or tablet. The leading commercial cloud computing provider's includes: Amazon EC2, Microsoft Azure, and Google Compute Engine (still in beta). The service charge, charged by Cloud computing providers for using their platforms go as little as \$0.10 per CPU-hour for renting MIPS, memory, and disk space and other services [7].

Cloud servers can house up to a few hundred thousand processor cores, plus many terabytes of disk storage hence it has a high computational power. Cloud computing also offers virtualization technology that gives users the ability to select any of the following: operating systems, applications, and network interconnects additional software flexibility for their modest rental fee. Therefore the solution to big data lies in cloud computing technologies.

2. RELATED WORKS

So many works have been done as regards big data and cloud computing issues and challenges to mention a few. [8] conducted a study on the security issues and challenges in the cloud and they looked at the advantages of using cloud services which include scalability, resilience, flexibility, efficiency and outsourcing non-core activities. Furthermore, they highlighted that Cloud computing offers an innovative business model for organizations to adopt IT services without upfront investment. Despite the potential gains of using cloud services, organizations are slow in accepting it due to security issues and challenges associated with it. Security is one of the major issues which hamper the growth of cloud and the main issue is lack of trust from both party because the idea of handing over important data to another company is worrisome to the subscribers of cloud services; such that the consumers need to be vigilant in understanding the risks of data breaches in this new environment. [9] Looked at the security issues for cloud computing, big data, Map Reduce and Hadoop environment. The main focus is on security issues in cloud computing looking at how important data is for effective running of any business, therefore cloud computing security is developing at a rapid pace which includes computer security, network security, information security, and data privacy. Cloud computing plays a very vital role in protecting data, applications and the related infrastructure with the help of policies, technologies, controls, and big data tools. Moreover, cloud computing, big data and its applications, advantages are likely to represent the most promising new frontiers in

science; it is therefore necessary to protect such important infrastructure.

[10] carried out a study on the awareness and adoption of cloud computing in Nigeria by small and medium scale enterprises in Lagos State and the found out that of all the challenges facing cloud computing adoption are security, privacy, lack of liability of providers in case of security incidents, and difficulty of migrating to the cloud (legacy software) that ranked high.

[11] Carried out a study on big data and current cloud computing issues and challenges by looking at a detailed analysis of between big data and cloud computing security issues and challenges focusing on the cloud computing types and the service delivery types. However, the researcher highlighted that big data entails a huge commitment of hardware and processing resources, thereby making adoption costs of big data technology unaffordable to small and medium sized businesses. Cloud computing on the other hand offers a lot of advantages to small and medium business especially but yet the rate of its adoption is still low due to security threat associated with cloud technology.

From all the related works reviewed, one paramount problem is the issue of security associated with big data technology and cloud computing which cannot be overemphasized. Hence this paper will critically examine this issue with other related issues associated with big data and cloud computing and proffer some solutions to the aforementioned problems.

3. ISSUES AND CHALLENGES

Some of the most common issues associated with big data and cloud computing includes but not limited to:

The Plumbing Problem: this problem arises due to the rate at which data is being created and stored every day. The digital universe will approximately double every two years, or 41% per year, [12] and it is rising significantly faster than the bandwidth of network connections. In 2012, there was just 11% growth in wired speeds, compared to an average connection rate of 2.8 Mbits/s [13] the growth of connections is not keeping up with the growth in data. In 2020, according to IDC, the digital universe will comprise 40,000 Exabyte's, and 68% of that will either be created or consumed by end users (versus businesses). Netflix and similar video-on-demand services occupied 30% of Internet bandwidth in December 2012. Similarly, YouTube received 72 hours of new video every minute, which required 17 petabytes of new storage in 2012 [14]

Mobile devices will both consume and generate much of this data. By the end of 2012, mobile devices generated 25% of Internet traffic. According to Cisco, video will account for 86% of all wireless traffic by 2016. Mobile devices also generate lots of sensor data, such as GPS location data and patient monitoring. Thus, they are the primary source of the machine-to-machine (M2M) traffic that comprises the Internet of Things. The IDC report forecasts that machine-generated data will represent 42% of all data by 2020, from the 11% in 2005 [12]

Security: Another important issue is security of data stored in the cloud as evident in [8], [10] and so many other articles. These Cloud computing technology comes with numerous security issues and this could be due to the fact that it encompasses many technologies which may include networks, databases, operating systems, virtualization, resource allocation, containerization, resource scheduling, transaction management, load balancing, concurrency control, managing contents distribution in a content delivery network (CDN) and memory management. Hence, security issues of these systems and technologies exist in cloud computing. For example, the security of the network that interconnects the systems in the cloud must be much secured. Also, containerization and virtualization paradigm in cloud computing bring about several security concerns. For example, the mapping of containers and virtual machines to the physical machines has to be done in a secured way [5].

The security issues associated with cloud computing devices and environments can be categorized into the following: network level, user authentication level, data level, and generic issues as depicted by [5], [9].

- Network level: The challenges associated with network level will include issues with network protocols and network security, such as distributed nodes, distributed data, Internode communication.
- User Authentication level: The issues and challenges associated with user authentication level includes encryption/decryption techniques, authentication

methods which may include issues with administrative rights for nodes, authentication of applications and nodes, logging etc.

- Data level: The issues and challenges associated with data level will include data integrity and availability issues such as data protection and the distribution of data.
- Generic types: The issues and challenges associated with general level security issues includes issues with traditional security tools, and use of different technologies

On the other hand, for big data security challenges, they are overblown by the three key characteristics of big data which are volume, variety, and velocity. Some of the unique treats that causes security vulnerabilities in big data are: Large-scale cloud infrastructures, diversity of data sources and formats, as well as the streaming nature of data acquisition and high volume inter-cloud migration.

Cloud Security Alliance in 2012 identified 10 big data security issues which are:

- Secure computation in distributed programming frameworks
- Security best practices for non-relational data bases
- Secure data storage and transactions logs
- End-point input validation/filtering
- Real-time security monitoring
- Scalable and composable privacy-preserving data mining and analytics
- Cryptographically enforced data centric security
- Granular access control
- Granular audits
- Data Provenance

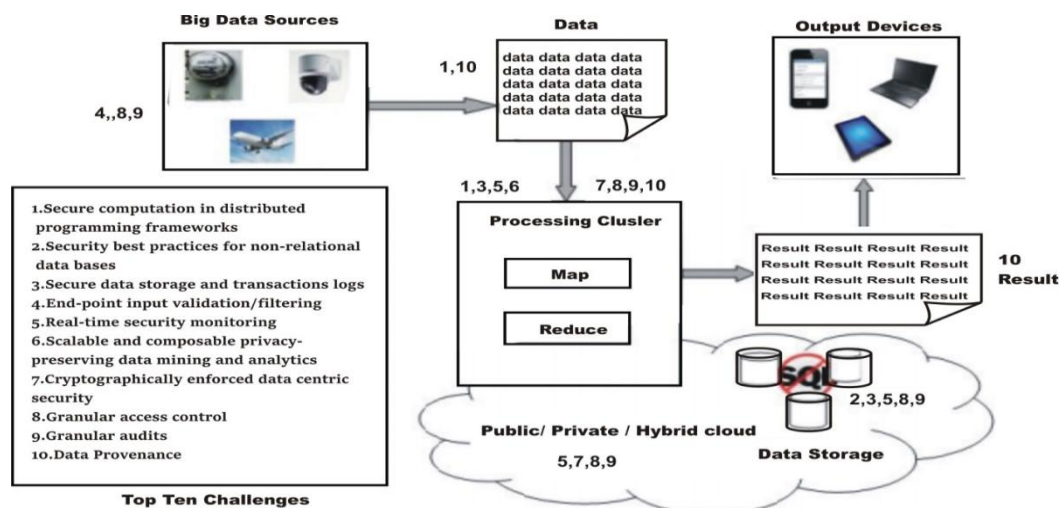


Figure 3: Top ten security and Privacy Challenges in Big data Ecosystem [15]

And in 2013 the Cloud Security Alliance (CSA) categorized this 10 security issues into four main categories which are:

Infrastructure Security, Data Privacy, Data Management and Integrity as depicted in figure 4.

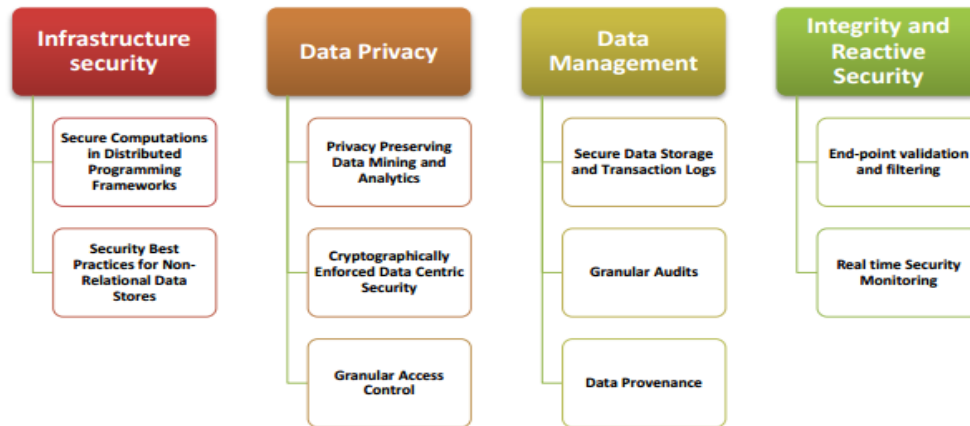


Figure 4: Classification of the top ten challenges [16]

Some other cloud computing issues in big data as depicted by [11] include

- i. **Costing Model:** - this involves the cost of transferring an organization's data to and from the different types of cloud they operate such as public and community Cloud and the cost per unit of computing resource used by such organizations is likely to be higher. This problem becomes very obvious if the consumer uses the hybrid cloud deployment model where the organization's data is distributed amongst a number of public/private (in-house IT infrastructure) /community clouds.
- ii. **Charging Model:** - This covers the charges charged by cloud providers on the pool of resources been used by subscribers and such elastic resource pool has made the cost analysis a lot more complicated than regular data centers, which often calculates their cost based on consumptions of static computing. Moreover, an instantiated virtual machine has become the unit of cost analysis rather than the underlying physical server. For Software-as-a-Service cloud providers, the cost of developing multitenancy can be very expensive. These may include: re-design and redevelopment of the software that was originally used for single-tenancy., cost of providing new features that allow for intensive customization, performance and security enhancement for concurrent user access, and dealing with complexities induced by the above changes.
- iii. **Service Level Agreement (SLA):**- Since cloud consumers do not have control over the resource pool of information been managed by the cloud providers, they have to ensure that the quality, availability, reliability, and performance of these resources when consumers have migrated their core business functions onto their entrusted cloud is in place. It is therefore vital for consumers to obtain guarantees from providers on service delivery. This is done typically through what is called service level agreement. (SLA) which is an agreement negotiated between the providers and consumers.

4. SOME SUGGESTED SOLUTIONS

Below are some suggested solutions to some of the highlighted problems:

- **Shipping disk drives to cloud computing:** The solution to the plumbing problem may be shipping disk drives to

cloud computing which will help to a great extent. For example, Amazon's AWS Import/Export service will receive shipped disk drive and transfer data to a local AWS server [7].

- **Use of Data mining techniques:** Also for the problem of malware diffusion, data mining techniques can be used for **malware** detection in clouds to tackle that problem perpetrated by cybercriminal which happens to be a security and data privacy threat.
- **Use of Access control techniques:** in order to secure the infrastructure of Big Data systems, the distributed computations and data stores must be secured. [16] highlighted that to secure the data itself, information dissemination must be privacy-preserving, and sensitive data must be protected through the use of cryptography and granular access control techniques. Managing the enormous volume of data necessitates scalable and distributed solutions for both securing data stores and enabling efficient audits and data provenance.

Finally, solving security and privacy challenges associated with big data and cloud computing technologies can require addressing this three issues as listed below as highlighted by [15]:

- i. **Modeling:** here it requires validating a threat model that will cover most of the cyber-attack or data-leakage scenarios by the cybercriminals
- ii. **Analysis:** finding tractable solutions based on the threat model formalized
- iii. **Implementation:** implementing the solution in existing infrastructures and technologies then performing a comparison of it with the threat models.

5. CONCLUSION

This work looked at big data; cloud computing and the major challenges and issues inherent in this two concepts which include but not limited to: the plumbing issues, security challenges at different level, the issue with cost, and the problem of service agreement. It was observed from the study that data will keep on increasing as the year run by so it is very important to make adequate arrangement on how to secure such vital information. Cloud environment is widely used in industry and research aspects; therefore security is an

important aspect for organizations running on these cloud environments considering the fact that the best place to keep such big data is in the cloud, it is important to make sure that the service, resources and devices in the cloud are readily available to handle the crave demand for big data. Some suggested solutions to the identified problems inherent with big data and cloud computing that will go a long way to improve its adoption and usage were provided in this work.

Furthermore for the future trend of the ever increasing data which is expected to be doubling on a yearly basis, research should continue in this two areas to see how the two key concepts can be improved and how the issues and challenges can be subdued to the barest minimum. Encouraging progresses have been made in the area of big data and cloud computing, but much work still needs to be done. Therefore it is important to continually improve the security strategies for these two important concepts in other to have adequate service delivery.

6. REFERENCES

- [1] Al Wegener (2013) Big Data Plumbing Problems Hinder Cloud Computing Mar 8, 2013 retrieved from <http://www.electronicdesign.com/communications.com> accessed Sep, 10 2015.
- [2] Douglas and Laney (2008), "The importance of 'big data': A definition," 2008. "Big data: science in the petabyte era," *Nature* 455 (7209): 1, 2008.
- [3] Katal, A., Wazid M, and Goudar R.H. (2013) "Big data: Issues, challenges, tools and Good practices." Noida: 2013, pp. 404 – 409, 8-10 Aug. 2013.
- [4] Ji, Changqing., Li, Yu., Qiu, Wenming., Awada, Uchechukwu., Li, Keqiu (2012) Big Data Processing in Cloud Computing Environments, 2012 International Symposium on Pervasive Systems, Algorithms and Networks, 1087-4089/12 \$26.00 © 2012 IEEE DOI 10.1109/I-SPAN.2012.9 pg 17-23
- [5] Venkata N. I., Sailaja A., and Srinivasa R. R. (2014) Security Issues Associated With Big Data in Cloud Computing, *International Journal of Network Security & Its Applications (IJNSA)*, Vol.6, No.3, May 2014 DOI: 10.5121/ijnsa.2014.6304 45
- [6] Mell P., Grance T., 2011 NIST Special Publication 800-145: The NIST Definition of Cloud Computing. Available at: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800145.pdf>
- [7] [AWS (2012) "AWS Import/Export," Amazon Web Services, retrieved from <http://aws.amazon.com/importexport/> accessed Sep 12 2015
- [8] Kuyoro S. O., Ibikunle F. & Awodele O. (2011), Cloud Computing Security Issues and Challenges, *International Journal of Computer Networks (IJCN)*, Volume (3) : Issue (5) : 2011 247
- [9] Saranya A, MuthuKumar, V.P. (2015) Security issues associated with big data in cloud computing *International Journal of Multidisciplinary Research and Development* Volume: 2, Issue: 4, 580-585 April 2015 www.allsubjectjournal.com e-ISSN: 2349-4182 p-ISSN: 2349-5979
- [10] Omotunde A. A.1, Izang, A. A., Awoniyi O. C., Omotunde, B. K., & Mensah, Yaw. A. (2015) Cloud Computing Awareness and Adoption among Small and Medium Scale Businesses (SMB) in Nigeria, *International Journal Of Multidisciplinary Sciences and Engineering*, VOL. 6, NO. 6, JUNE 2015 ISSN: 2045-7057 www.ijmse.org 32
- [11] Iyanda, O. A., (2014) Big Data and Current Cloud Computing Issues and Challenges, *International Journal of Advanced Research in Computer Science and Software Engineering* Volume 4, Issue 6, June 2014 ISSN: 2277 128X Research Paper Available online at: www.ijarcsse.com
- [12] IDC (2012) "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East," IDC, www.emc.com/leadership/digitaluniverse/view/index.htm
- [13] Akamai (2012) The State of the Internet," Akamai retrieved from <http://www.akamai.com/stateoftheinternet> accessed September 12 2015
- [14] G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall, and W. Vogels, "Dynamo: amazon's highly available keyvalue store," in ACM SIGOPS Operating Systems Review, vol. 41, no. 6. ACM, 2007, pp. 205–220
- [15] Cloud Security Alliance CSA (2012) Top Ten Bid Data Security and Privacy Challenges. Retrieved from <http://webseices.thesba.com> accessed Sep 14, 2015
- [16] Cloud Security Alliance CSA (2013) Expanded Top Ten Bid Data Security and Privacy Challenges. Retrieved from <http://webseices.thesba.com> accessed Sep 14, 2015