International Journal of **Advances in Computer Networks and Its Security– *IJCNS***
Volume 4 : Issue 3 [ISSN 2250 – 3757]
Publication Date : 30 September, 2014

Big Data Analytics on Cloud-Technical, Non-Technical and Security Issues

[Ali Garba Jakwa, Alhaji Idi Babate and Wada Abdullahi]

***Abstract*- Technology is increasingly being used by organizations as a tool for resource management in our complex social system. This paper presents an investigation on the technical issues and practical applications associated with Big Data analytics on cloud as a new and evolving computing technology. Big data analytics on clouds is a strategy empowered by later developments in technologies that drives almost every part of our present day social order, including mobile services, retail, manufacturing, financial services, life sciences, and physical sciences. The argument put forward was how Big Data becomes more adopted in the mainstream, the technical and non-technical security issues associated with it might increase and might hinder the growth of Big Data analytics on the Cloud computing environment. Therefore, we address and highlight the main security issues in Big Data analytics on cloud with a special focus on how and what to do to secure an efficient protection on the loss of confidentiality, integrity, and availability of data.**

***Keywords*- cloud computing, big data analytics, online, transaction**

I. Introduction

Technology is increasingly being used by organizations to mediate social/business relationships and social business transactions. We are awash in a flood of data today, in most expensive range of application areas of continuous data collection, decision that were previously beforehand were based on guess work, or on carefully developed model of reality can now be made dependent upon the data itself. Big Data analysis now drives almost every part of our present day social order, including mobile services, retail, manufacturing, financial services, life sciences, and physical sciences (Techterget, 2013).Thus, it's supports the creation of new or improved services, with significant advantages for customers. Big Data analytics on cloud is a data analysis strategy empowered by later developments in technologies and building design (Sanjay et al, 2013). The data driven design had also recognized the promise of Big Data as real. However, there are numerous technical and non-technical, and security issues that come with Big Data analytics on cloud. As Big Data becomes more adopted in the mainstream, these issues could increase and potentially hinder the growth of Big Data analysis on the Cloud Computing environment.

Advances in digital sensors, communications, computation, and storage have created huge collections of data, capturing information of value to business, science, government, and society (Bryant et al, 2008). For example, search engine companies such as Google, Yahoo!, and Microsoft have created an entirely new business by capturing the information freely available on the World Wide Web and providing it to people in useful ways. In short, the term Big Data refers to information that can't be processed or analyzed using traditional processes or tools (Eaton et al, 2010). Increasingly organizations today are facing much more big data challenges. According to IBM the amount of unstructured and multi-structured data within an average organization is about 80% (Savvas, 2011). Taking account the average data growth annually by 59% (Pettey&Goasduff, 2011), this percentage will likely be much higher in a few years. Not only the volume of data is becoming a problem, also the variety and velocity are issues we need to look at (Russom, 2011). This phenomenon clearly described "big data" and is identified as one of the biggest IT trends for 2012 (Pettey, 2012). The societal benefits of big data analytics on cloud services are immeasurable, having transformed how people find and make use of information on a daily basis. The rise of the cloud plays a significant role in big data analytics and likely this role will increase as the cloud is being adopted by growing number of organizations (Armbrust et al., 2010). Cloud computing is an extremely successful paradigm of service oriented computing and provides services at different levels of IT, for example, Infrastructure as a Service (IaaS), Platform as a service (PaaS) and Software as a Service (SaaS) (Agrawal, &Abbadi, 2010). Data platform can deliver ground-breaking capability when it comes to technical, non-technical and security issues and risk modeling with expanded models that are built on more and more identified casual attributes.

This paper investigates what Big Data Analytics on Cloud is, the technical issues associated with this new and evolving computing technology and also the practical applications of it. It also addresses and highlights the main security issues in Big Data analytics on cloud with a special focus on how and what to do to ensure an efficient protection on the loss of confidentiality, integrity, and availability.

II. Research Methodology

Data for this paper were derived from secondary sources: previous researches and analyses of scholars; government documents; newspaper/magazines as well as journal articles that are related to the subject as the study involved an extensive literature review which critically analysed the present status, technical, non-technical and security issues of big data analytics on cloud. And the critical steps to acquiring the know-how to have a successful transition of migrating big data to the cloud, the risk as well as the confidentiality of the application processing tools; and the content analysis approach was utilized for analysis.

III. Big data analytics

According to the Oxford dictionary, analytics is defined as "the systematic computational analysis of data or statistics" or, a better definition, also according to the Oxford dictionary: "information resulting from the systematic analysis of data or statistics". Analytics is very important in many different fields, both scientific and organizational especially for decision making (Golfarelli, et al, 2004). For example, without analytics, the procurement department of a supermarket chain would have a hard time deciding what to buy and in which numbers. Many forms of traditional data analytics exist. Probably the best known is business intelligence conceptualization by Golfarelli and his Colleagues defined big data analytics as:

"The process of turning data into information and then into knowledge. Business intelligence was born within the industrial world in the early 90's to satisfy the managers' request for efficiently analyzing the enterprise data in order to better understand the situation of their business and improving the decision process" (Golfarelli et al., 2004)

Many other forms of data analytics exist: from web traffic analysis to customer satisfaction analysis. The way it creates value differs for each form of data analytics but in the end, almost always try to achieve the same goal: increase revenues, reduce costs or both. For example, web traffic analysis can lead to insights which can be used to improve the usability of a website leading to more satisfied customers resulting in an increase of sales. Continuing with big data analytics, you might ask what is different compared with traditional data analytics. First of all, big data analytics is not the successor of traditional data analytics. It can however support traditional data analytics, also mentioned by Adrian & Chamberlin (2012):

"Big data can enable richer reporting and analysis, complementing, rather than replacing, traditional data warehouses" (Adrian & Chamberlin, 2012)

Many definitions of big data can be found when reviewing the relatively scarce literature. This widespread of definitions in fact tells something about the current state of the global understanding of this theme. It appears that big data is an unclear term, probably because big data is not tangible nor a fixed property. Big data itself is not a disruptive innovation but a sustaining innovation making it harder to describe, define and distinguish big data (analytics) from traditional data (analytics). Sustaining innovations"are innovations that make a product or service perform better in ways that customers in the mainstream market already value" while disruptive innovations "create entirely new markets through the introduction of a new kind of product or service […]" (Christensen &Overdorf, 2000). What is now called "big data" is not the result of a specific innovation of a product or service but rather a combination of revolutionary and evolutionary changes and innovations in different fields (both technical and non-technical fields) (Gantz, 2011, p. 6). These researchers conceptualization of a big data analytics, clearly expose the technical, non-technical issues of big data analysis, hence the research will address this technical issue of the big data analytics as well as the security issues.

A. *The rise of the cloud*

The rise of the cloud plays a significant role in big data analytics and likely this role willincrease as the cloud is being adopted by a growing number of organizations (Armbrustet al., 2010). Cloud computing is an extremely successful paradigm of service orientedcomputing and provides services at different levels of IT, for example, Infrastructure as aService (IaaS), Platform as a service (PaaS) and Software as a Service (SaaS) (D. Agrawal,Das, &Abbadi, 2010). Some advantages of cloud computing, compared to in-house computing, are:

• Infinite computing resources are available on demand;
• No up-front commitment by cloud users; users can start small but think big;
• Pay for use of resources on a short-term basis (e.g. more resources on peak hours);

This proves the fact that organizations tend to "start small but think big". These advantages are useful for big data analytics in several ways. In order to analyse data, there must be data available and as described earlier, data is being created in a much faster way than ever before. Therefore, a lot of storage space is necessary (especially with the "store and analyse" approach). A significant proportion of data organizations own is created by end-users (such as visitors of the organization's website) and hence, cannot be controlled by the organization itself. Vollaard, innovation manager at the Rabobank, mentioned that big data analytics is being rolled out in different phases trying to find out how big data analytics affects the organization and its Processes (Vollaard, 2012). This indicates the need to easily demand more resources from the cloud provider when required. This principle can be found in many organizations and it shows the dynamic advantage of moving big data to the cloud. the system required no expensive data centers to be purchased, installed and maintained leading to huge investment costs.

Rather necessary computing resources can be rent with the possibility to easily scale when needed. This is extremely beneficial for especially small organizations since big data analytics is now also possible for SMEs as Agrawal (2012) also points out:

"A move to cloud with big data analytics as a service is not far-fetched as SMEs will look to eliminate costs from building in-house infrastructure to support big data analysis." (Agrawal T., 2012).

Gantz&Reinsel (2011, p. 2) also supports this. Another great advantage of cloud computing is the way how computing resources can be rented on very short-term basis especially for analysis-related tasks. These tasks are often being done at specific moments. Amazon's evolution of their database system will be used as an example to illustrate the possible benefits of big data analytics in the cloud and the need for similar systems. This database system became a well-known cloud service publicly available. "Amazon runs a world-wide e-commerce platform that serves tens of millions customers at peak times using tens of thousands of servers located in many data centres around the world" (DeCandia et al., 2007). The slightest outage has significant financial technical consequences and with so many servers world-wide hardware fails continuously, highly decentralized storage technology was needed in order to avoid security bridges.

B. *General Picture of Big Data processing in the Cloud Environments*

Organizations of all sizes and budgets now have access to infrastructure via the cloud that enables big data opportunities. Big data and cloud technology go hand-in-hand. Big data needs clusters of servers for processing, which clouds can readily provide (Techtarget, 2013). But what does that look like in reality? To breakdown the landscape highlighting the various items of data from the real world that have come to be accumulated in cloud data centres. For example, information from position sensors of the Global positioning System (GPS) mounted on mobile phone handsets or automobiles and transaction records from store cash registers are stored along with the location and time of their generation, and transferred via networks to data centres, where they are accumulated (Tsuchiya et al, 2012; Farber et al, 2011). These data can be analysed in terms of time series and associated with factors such as purchase behaviour of individuals so as to estimate what action such individuals are likely to take. In this way, it is beginning to become possible to derive valuable information such as estimates of the purchase behaviour of individuals from data, which have so far been no more than records. These massive amounts of data are also called big data. According to a Trial calculation, the amount of event data generated in the U.S. is estimated to be 7 million pieces per second, which adds up to a few tens to hundreds of PBs per month if accumulated as they are without compression (Tsuchiya et al, 2012; Techtarget, 2013). The structure of a system that realizes big data processing on clouds should combine subsystems for judgment; accumulation and analysis with a large number of servers in order to handle massive event streams see Fig. 1 Subjects in the real world include mobile phone handsets and automobiles, and their numbers may be enormous possibly at 100 million units for mobile phones and tens of millions of units for automobiles (Tsuchiya et al, 2012; Farber et al, 2011). One of the characteristics of event generation is that such events may occur unexpectedly, and the amount of events may significantly vary depending on the season or time of day.

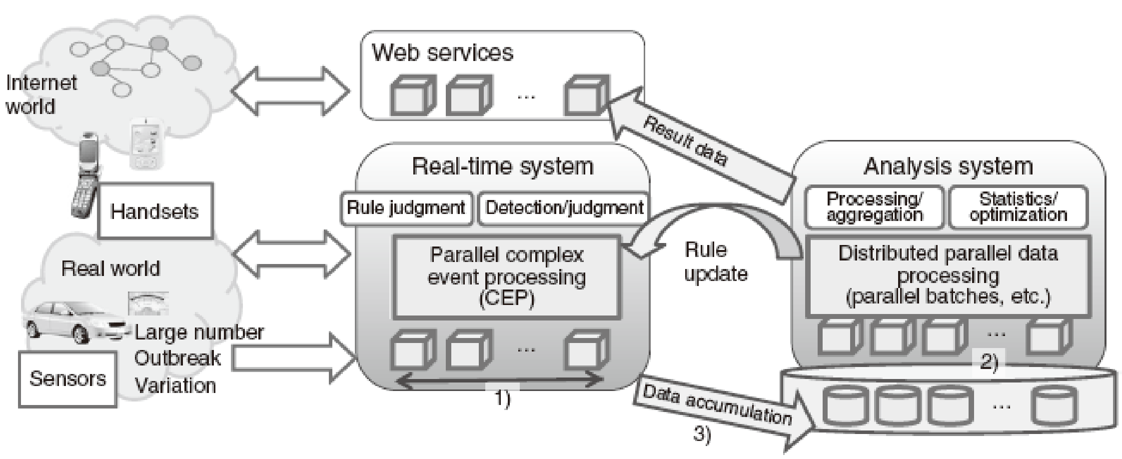

Figure 1: General Picture of Massive Data processing in the cloud

(Program Source: Tsuchiya et al, 2012)

Tsuchiya and his colleagues (2012), further describes, Judgment processing must be capable of handling up to millions of events per second, which is a social system level, and making an immediate judgment without failing to respond to unexpected generation or rapid variation in the number of events. Besides, we need to have a number of servers running in parallel so as to flexibly change system configurations according to the demand without stopping services (Tsuchiya et al, 2012) as shown in figure 1 above.

(big data) in an effort to uncover hidden patterns or unknown correlations. Big data Analytics Applications (BDA Apps) are a new type of software applications, which analyse big data using massive parallel processing frameworks for example Hadoop (Shangy et al (2013).Hadoop was originally developed to support distribution for the Nutch search engine project by Micheal J. Cafarella and Doug Cutting.

Cloud computing has evolved through a series of phases – there was the initial grid (or utility) computing phase, then there was the "application service provision" which was then followed what is now known as SaaS (Mohammed, 2009). According to Mohammed, the most recent evolution of cloud computing is its development with Web 2.0. This was made possible as bandwidth increased in the late nineties. In 1999, salesforce.com *pioneered the concept of delivering enterprise applications via a simple website*. As a result, companies, both mainstream and specialist, started the delivery of Internet-based services and applications (Mohammed, 2009). Following on from that, Amazon Web Services was developed in 2002 – which allowed for many cloud-based services such as computing and storage to be delivered online. In 2006, Elastic Compute Cloud (EC2) was launched by Amazon. EC2 was a commercial web service which enables individuals and small companies to rent computers online to run their own applications (Mohammed, 2009).

Since 2007, cloud computing has become a "hot topic" due to its flexibility to offer dynamic IT infrastructure and configurable software services over the Internet (Wang *et al.*, 2008). The emergence of cloud computing coincides with the development of Virtualisation technologies. Since 2007, Virtualisation technologies have increased and as a result, cloud computing has been observed to have out-paced that of grid computing (Wang *et al.*, 2008). This trend still continues to grow as companies and research community propose and develop this computing paradigm – cloud computing.

C. *Software Evolution: Big data analytics on Cloud*

The process of software evolution can be seen as a never-ending process. Once software is developed, it is maintained, and then repeatedly updated with respect to changes in requirements, processes and methodologies. It is known that 90% of companies' software budget is spent on maintenance and adapting existing software tools than developing new ones from scratch (Brooks, 1995). Big data analytics is the process of examining large amounts of data

D. *Technical Issues of Big data analytics on clouds*

The idea behind big data analytics on cloud computing is that software and hardware services are stored in "clouds", web servers rather than a connection of standalone computers over the Internet. Here, organizations can access the right services and data they require (Aymerich*et al.*, 2008). Another benefit of cloud computing is that of "moving" data to the cloud to allow for access to a user's data anywhere (Aymerich*et al.*, 2008). An important feature that comes with cloud storage of big data is essentially the automation of different management tasks.

It can be noted that a fusion or combination of technologies such as grid computing, autonomic computing (AC) and utility computing has contributed to the evolution of cloud computing. AC is built on the following concepts; self-protection, self-management, healing and configuration. It uses a closed control loop system which allows it monitor and control itself with external input. As the current situation and needs of a system changes, an AC system adapts itself to those dynamical changes – making it self-adaptive as well. This combined with grid computing which was known to be "heterogeneous and geographically detached" (Aymerich *et al.*, 2008), has produced a new computer architecture for cloud computing.

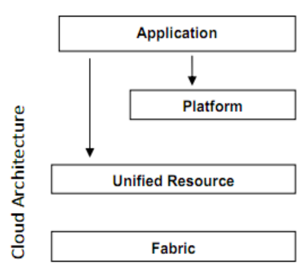

Figure 2: The Cloud Computing Architecture

Foster *et al.* (2008) define a four-layer architect of cloud computing (Figure 2). These layers are Application, unified resource, platform and application. The physical hardware resources (such as storage resources, computing resources and network resources are contained in the *Fabric layer*. Abstracted/encapsulated resources (usually as a result of virtualisation) are contained in the *Unified Resource layer*. These abstracted resources are exposed to the upper layer and also by the end user as integrated resources such as a database system, logical file system, a virtual cluster/computer (Foster *et al.*, 2008). Specialised tools and technologies such as middleware and other services are provided by the *Platform layer* in addition to the resources already contained in the unified resource layer to provide a platform for the development and deployment of applications. Lastly, applications that will run in the cloud are contained in the *Application layer.*

As mentioned earlier there are three different levels of services that are provided by cloud computing (IaaS, PaaS and SaaS)(Agrawal, &Abbadi, 2010). The type of service provision depends on the layer which the service provider wants to make available. However, it is also possible for a service provider to expose services on than one layer.

IaaS allows for the provision of software, hardware and equipment usually at the unified resource layer. PaaS provides users with a *high-level integrated environment to build, test, and deploy* their own built applications. Here, developers are faced with certain technical limitations and restrictions on the type of software tool or application that they can develop in *exchange for built-in application scalability* (Foster *et al.*, 2008).

There are numerous security concerns when it comes to software development over cloud computing. This makes it difficult for certain computing techniques to be incorporated in software development. Additionally, some techniques make programs or software vulnerable in distributed systems and in this scenario transparency is the biggest challenge in moving big data to the cloud, not security (Egan G., 2013).

1) Cross-cutting Concern

Cross-cutting concerns in software development relates to aspects of a program that affect or crosscut other modules or concerns (Abdullin, 2010). Usually, these concerns arise due to difficulty in decomposing them from a program in the developmental stage, which includes the design, coding and implementation phases, as a result can result in the duplication of code (known as scattering) or tangling (these come about when systems have significant dependence on each other) or both . Some examples of cross-cutting concerns include:

- Exception handling
- Validation
- Logging
- Authentication and authorisation

2) Program or Application Clustering

Clustering, in computing, relates to a group of computers or servers dedicated to performing a single task. Software systems are used to configure servers to cluster in application clustering. Servers are connected together by a software program which enables the servers to perform individual tasks like failure detection and load balancing (Bliss, 2010). Here, applications are installed individually on the servers and are pooled in together to handle various tasks when required. It becomes important for the cluster to effectively handle routing of data to and from the cluster (Bliss, 2010). In cloud computing, program clustering helps achieve scalability – the ability of the cloud to appropriate resources to specific tasks i.e. when a task needs more computing resources, it has the ability to recruit more servers or computing power to perform that specific task. The benefit of cloud computing is that it contains hundreds of thousands of connecting computers which makes it easy to distribute work load. There are symmetric clusters where workload is distributed evenly amongst the clustering servers and asymmetric clusters have the ability to reserve particular servers only for use when the main servers fail.

E. *Non-Technical Issues: Cloud Computing*

Big data analytics on cloud computing comes with other non-technical issues or concerns which if not tackled could restrict the growth and evolution of big data analytics on cloud clustering concern.

1) Inadequate Security

Most cloud vendors support what is known as multi-tenancy compute environment by design. What is most important is that, vendors must decide on the right balance between providing essential infrastructure and internal security and the quest for improved cloud computing services. According to Zhang *et al.* (2009), trustworthiness is important when it comes to SaaS services. With SaaS, data privacy and security are the most important factors for end users (also known as tenants).

2) Lack of computability with existing applications

Another major issue currently facing big data and cloud computing clustering is the lack of inherent computability with existing applications. There are, however, efforts to change this. What is observed in order to improve scalability and improve the level of services provided to users, vendors are now providing snippets of existing codes in the case of PaaS. What this means is that new applications are becoming cloud-specific.

3) Lack of Interoperability between Clouds

The lack of a standardisation across platform increases cost of switching clouds and also increases the complexity of code in the event of program migration. Since cloud vendors have different application models, there are vertical integration problems which make it virtually impossible at time to move from one cloud to another. As this is major issue, an organization has to be careful when choosing the right vendor to obtain services from.

4) Other Issues

There is also the issue of service legal arrangement which prohibits organizations from moving from one cloud to another unless certain conditions are met. This increases switching costs for the end user and subsequently, gives more power to the cloud vendor.

a) *Legal Issues*

According to (Longbottom, 2008), the biggest issue concerning big data and cloud computing clustering comes from governments. This is as a result of the borderless global network operations of cloud computing. Unlike grid computing, cloud computing is not geographic-specific. Having no borders makes it difficult for governments to protect or control how data of its people is stored or used elsewhere and also how to tax companies operating services over a cloud. Under taxation, if a company is taxed based on geographical location of its computing operation, it can simply move this to a virtual office in a country with a lower tax rate (Longbottom, 2008).
There are measures being taken to tackle the issue of taxation under big data and cloud computing on a global approach in order to stop companies from exploiting tax advantages. Additionally, there is a recognised need for *harmonised laws* in the global front to police how data is stored and used over the cloud.

b) *Security Issues*

According to Hocenski and Kresimir (2010), one of the main security concerns of cloud computing is that of its immaturity of the technology. The cloud provider and client have both security responsibilities depending on the type of service. In the case of an IaaS service model, the virtualization software security, environment security and physical security rest with the cloud provider. The client or user is responsible for operating system, data and applications. However, in a SaaS model, software services, physical and environment security are the responsibility of the cloud provider.

The main security concern with cloud computing is that of data security. Confidential documents stored on the cloud can become vulnerable to attacks or unauthorised access or modification. There is also the issue of where the data is physically stored i.e. where the data stores are located. Some companies prohibit the storage of their data in certain jurisdictions or countries (King and Raja, 2012).

F. *Applications of Cloud Computing*

One of the reasons for the upward trend of resources committed to cloud computing is that, cloud computing has many real benefits and applications to companies, individuals, research bodies and even government. As the size of data increases, the need for computing power capable of analysing these data increases relatively.
One application of cloud computing is that clients can access their data and applications from anywhere at any particular time. Additionally, clients only need a computing device and an Internet connection to access their data and applications. For example, Dropbox (Strickland, 2011) allows users to store their data on an online cloud and access it using any computing device. Users can also share folders with other users in the same manner. Another example is Google Docs (Strickland, 2011) which allows users to edit, modify documents online without have to move the document around. All modifications are saved on the master document on the cloud. Cloud computing has the possibility of reducing hardware costs for many companies and individuals. Clients can gain access to faster computing power and bigger storage without paying for the physical hardware. Cloud computing gives companies the ability to gain company-wide access to its host of software or applications. Here, a company does not have to buy a licence for every employee to use particular software; rather it can pay a cloud computing company on a usage fee basis (utility computing model). With the introduction of Web 2.0, access to cloud computing has become less OS-dependent. (Strickland, 2011).

IV. Summary and Conclusion

It's clear that marketing, advertising department and educational institutions are among the organizations taking the lead in implementing analytics technology, particularly, big data analytics tools. But why are organizations so interested in big data analytics on cloud. These are some of few technical, non-technical issues this article addresses, the challenges facing the two emerging technologies as well as the compounding intelligent security issues affecting the integrity, confidentiality and availability of the application tools; As Russom of TDWI wrote "one third of organizations (34 percent) do big data analytics today, although it's new". We are awash in a flood of data today by the emerging trends of big data technology.

In this paper, we have given a brief overview of the emerging computing infrastructure for big data processing, and a focus on the three main categories of service provision in cloud computing such as Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS). In SaaS cloud model,

the hardware infrastructure and the provision of the software product is supplied by the service provider.
The challenges include not just the obvious issues of complexity of processing clusters and endpoint consumers of the data (systems, mobile devices, etc.) along with the cloud environments, but also identified the security risks of the computational elements as lack of structure, error-handling and privacy were discussed.

In addition to technical issues, there are real security issues with regards to big data analytics on cloud computing. The most concerning security issue is that of data privacy and integrity in the cloud.

References

[1] Abdullin, R. (2010). "Cross-cutting concern". [Online] Available at: http://abdullin.com/wiki/cross-cutting-concern.html [Accessed 11th Feb 2013].

[2] Adrian, M., & Chamberlin, D. (2012).Orbitz Worldwide Uses Hadoop to Unlock the Business Value of "Big Data."Gartner. Retrieved from https://www.zotero.org/groups/leapforward/items/itemKey/BSHHXBHN

[3] Agrawal, D., Das, S., &Abbadi, El, A. (2010). Big data and cloud computing: new wine orjust new bottles? Proceedings of the VLDB Endowment, 3(1-2), 1647–1648.

[4] Agrawal, T. (2012). 1/3 of US Mid-Market Businesses Using Business Intelligence are Interested in Big Data Analytics; Lack of Expertise is Main Barrier. Techaisle.

[5] Aymerich, F., Fenu, G. and Surcis, S. (2008). "An Approach to a Cloud Computing Network".First International Conference on the Applications of Digital Information and Web Technologies, ICADIWT.

[6] Bliss, H. (2010). "What is Application Clustering?" [Online] Available http://www.wisegeek.com/what-is-application-clustering.htm [Accessed 14th Feb 2013].

[7] Bryant RE., Katz R., and Lazowska. E., (2008) Big-data computing: Creating revolutionary breakthrough in commerce, science, and society Version 8 December, 2008. Online available at: http://www.cra.org/ccc/initiatives(Accessed:09/12/2013)

[8] Brooks, F. (1995). "The Mythical Man-Month". Addison-Wesley, 1975 & 1995.

[9] Curry. S., Kirda E., Schwartz E., and Stewart W. &Yoran A., (2013) Big Data Fuels Intelligence-Driven Security Rapid growth in security Information creates new capabilities to defend against the unknown. Security management and Copliance business unit; RSA, the Security Division of EMC January, 2013

[10] Christensen, C. M., &Overdorf, M. (2000).Meeting the Challenge of Disruptive Change. Harvard Business Review, 78, 66–76.

[11] DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A.,Sivasubramanian, S., et al. (2007). Dynamo: amazon's highly available key-valuestore (Vol. 41, pp. 205–220). ACM.

[12] Dimitirios, Z. and Dimitrios, L. (2012) Addressing cloud computing security issues. *Elsevier*, 28 (3), p.583–592. Available at: http://www.sciencedirect.com.ergo.glam.ac.uk/science/article/pii/S0167739X10002554 [Accessed: 10thDec, 2013].

[13] Eaton. C., Deutcsch T., Deroos. D., Lapis. G., and Zikopoulos. P., (2012) Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. Copy right 2012 by McGraw-Hill Companies online: http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype (Accessed: 10/12/2013)

[14] Farber. M., Ellis. C., and Sullivan J., (2011) Massive Data Analytics and the Cloud a Revolution in Intelligence Analysis Booz Allen Hamilton Strategy and technology consultant online available at: http://www.boozallen.com/media/file/MassiveData.pdf (Accessed: 08/12/2013)

[15] Foster, I., Zhao, Y., Raicu, I. and Lu, S. (2008) "Cloud Computing and Grid Computing 360-Degree Compared" [Online Article] Available http://arxiv.org/ftp/arxiv/papers/0901/0901.0131.pdf [Accessed 14 Feb 2013].

[16] Gantz, J. (2011). Extracting value from chaos.IDC research report IDC research report.

[17] Gantz, J. F., Mcarthur, J., & Minton, S. (2007). The Expanding Digital Universe. Director, 285(6). doi:10.1002/humu.21252

[18] Golfarelli, M., Rizzi, S., &Cella, I. (2004). Beyond data warehousing: what's next in business intelligence? (pp. 1–6). ACM.

[19] Harman, M. and Hierons, R.M.(2006). "An Overview of Program Slicing" [Online] Available at: http://www.cs.ucl.ac.uk/staff/mharman/sf.html [Accessed 10th Feb. 2013].

[20] Hocenski, Ž. and Kresimir, P. (2010) "Cloud computing security issues and challenges ", paper presented at *MIPRO, 2010 Proceedings of the 33rd International Convention*, 24-28 May. IEEE Conference Publications, p.344 - 349 [Online] Available at:http://ieeexplore.ieee.org.ergo.glam.ac.uk/stamp/stamp.jsp?tp=&arnumber=5533317[Accessed 15thNov. 2013].

[21] IBM, 2013 what is hadoop?http://www-1.ibm.com/software/data/infosphere/hadoop/

[22] Karadesh, L. (2012) Applying Security Policies and service level Agreement to Iaas service Model to Enhance Security and Transition. *Elsevier*, 31 (3), p.315–326. Available at: http://www.sciencedirect.com.ergo.glam.ac.uk/science/article/pii/S0167404812000077 [Accessed: 12th Feb, 2013].

[23] Kumar, A. et al. (2012) "Secure Storage and Access of Data in Cloud Computing", paper presented at *ICT Convergence (ICTC), 2012 International Conference on*, 15-17th Oct.. IEEE Conference Publications, p.336 - 339.[Online] Available at: http://ieeexplore.ieee.org.ergo.glam.ac.uk/stamp/stam

p.jsp?tp=&arnumber=6386854 [Accessed: 10thDec, 2013]

[24] Lonbottom, C. (2008) "Obstacles to Cloud Computing". [Online] Available at: http://www.information-management.com/news/10002177-1.html?pg=1 [Accessed 11th Fed 2013].

[25] Pettey, C. (2012, October 18). Gartner Identifies the Top 10 Strategic Technologies for 2012. Gartner.

[26] Pettey, C., &Goasduff, L. (2011, June 27). Gartner Says Solving "Big Data" Challenge Involves More Than Just Managing Volumes of Data. Stamford: Gartner. Retrieved fromhttp://www.gartner.com/it/page.jsp?id=1731916

[27] Rajan. Sr., Ginkel. W., and Sundaresan N., (2013) Cloud Security Alliance Expanded Top Ten Big Data Security and Privacy Challenges April 2013 Cloud Security Alliance – All Rights Reserved Online available at: https://cloudsecurityalliance.org/research/big-data/ (Accessed: 28/11/2013

[28] Russom, P. (2011). Big Data Analytics.TDWI Research.

[29] Savvas, A. (2011, October 25). IBM: Businesses unable to analyse 90 percent of their data. ComputerworldUK. Retrieved August 9, 2012, from http://www.computerworlduk.com/news/it-business/3313304/ibm-businesses-unableto-analyse-90-percent-of-their-data/

[30] Sanjay P. & Moore B., (2013) State of Big Data Analysis in the Cloud Journal of Network and Communication Technologies; Vol. 2, No. 1; 2013ISSN 1927-064X E-ISSN 1927-0658Published by Canadian Center of Science and Education

[31] Shangy.W., Jiangy. Z., Hemmatiy H., Adamsz B., and Hassany A., &Martinx P., (2013) Assisting Developers of Big Data Analytics Applications when Deploying on Hadoop Clouds *Journals of IEEE, ICSE 2013, San Francisco, CA, USA 978-1-4673-3076-3/13 Pp402 - 411*

[32] Shaikh, F. and Haider, S. (2011) "Security threats in cloud computing ", paper presented at *6th International Conference On Internet Technology And Secured Transactions*, Abu Dhabi, 11-14th Dec. IEEE Conference Publications.

[33] TechTarget (2013)."Cloud Computing" [Online] Available at: http://searchcloudcomputing.techtarget.com/definition/cloud-computing [Accessed on 10thDec 2013].

[34] Tsuchiya.S., Sakamoto.Y., Tsuchimoto. Y., and Lee V., (2012) big data Processing in Cloud Environments FUJITSU Sci. Tech. J., Vol. 48, No. 2 April 2012

[35] Vollaard, H. (2012, August 12). Interview Rabobank #1. (N. Mouthaan, Ed.). Utrecht.

[36] Wikipedia, 2013, Apache Hadoop, <http://en.wikipedia.org/wiki/Apache-Hadoop#History

About the Authors:

Ali Garba Jakwa
Federal College of Education (T), PMB 1013 Potiskum Yobe State
Nigeria

Alhaji Idi Babate
Federal College of Education (T), PMB 1013 Potiskum Yobe State
Nigeria

Wada Abdullahi
Federal College of Education (T), PMB 1013 Potiskum Yobe State
Nigeria