ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Big Data Cloud Computing Safety Measures: A Case Study

By Dr. Ifath Nazia Ghori,
Lecturer, Jazan University – Kingdom of Saudi Arabia.
Email: ghori.ing@gmail.com

ABSTRACT:

Big Data and cloud computing are two important issues in the last two decades; it enables computing resources to be provided as information technology services with high effectiveness and success. Big data is a broad term for data sets so large or complex that traditional data applications are inadequate. Challenges include analysis, capture, search, sharing, storage, transfer, visualization, and information privacy. The term often refers simply to the use of predictive analytics or other certain advanced methods to extract value from data, and seldom to a particular size of data set. Accuracy in big data may lead to more confident decision making. And better decisions can mean greater operational efficiency, cost reduction and reduced risk. Now a day's big data is one of the most problems that researchers try to solve it and focusing their researches over it to get ride the problem of how big data could be managing in the recent systems and managed with the cloud of computing, and the one of the most important issue is how to gain a ideal security for big data in cloud computing, this paper reviews a survey of big-data with cloud computing security and the mechanisms that are used to protect and secure data and also have a privacy for big-data with an available clouds.

Keywords: ***BigData, Large Data, Cloud Providers, Cloud Computing, NAS, Security, data privacy. Data protection***

1. INTRODUCTION

Big data is known as a datasets with size beyond the ability of the software tools that used today to manage and process the data within a dedicated time. With Variety, Volume, Velocity Big Data such military data or other unauthorized data need to be protected in a scalable and efficient way. Information privacy and security is one of most concerned issues for Cloud Computing due to its open environment with very limited user side control. It is also an important challenge for Big Data. After few years later more data globally would be touched with Cloud Computing which provides strong storage, computation and distributed capability in support of Big Data processing. Other considerations are that information privacy and security challenges in both Cloud Computing and Big Data must be investigated. the privacy and security providing such forum for researchers, and developers to exchange the latest experience, research ideas and development on fundamental issues and applications about security and privacy issues in cloud and big data environments.

The cloud helps organizations and enables rapid on demand provisioning of server resources such as CPUs, manage, storage, bandwidth, and share and analyze their Big Data in a reasonable and simple to use. The cloud infrastructure as a service platform, supported by on demand analytics solution seller that makes the large size of data analytics very affordable.
As location independent cloud computing Involving shared services providing resources , software and data to systems and The hardware on demand, actually the storage networking in cloud is a very strong because use driver for high performance.

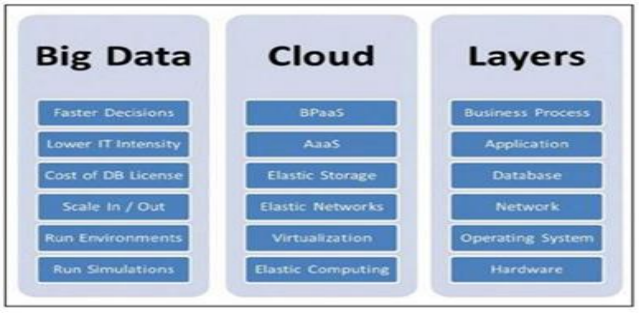
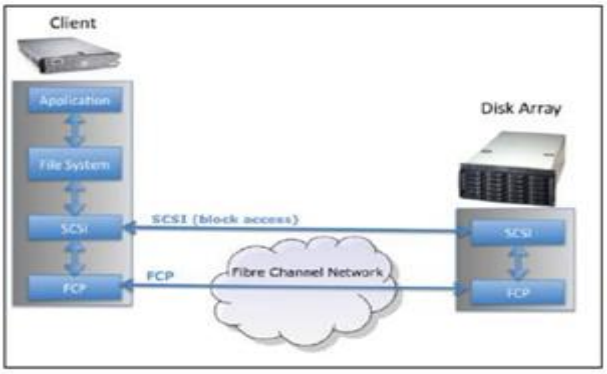


FIG 1. BIG DATA AND CLOUDS

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

For example Arista provides Networks with Specifications and product line of switching solutions as shown in figures 2 and 3 below. However, the requirements of cloud storage needs hypothesized to a group of sub nodes operations performed with some of the units and CPUs advanced.



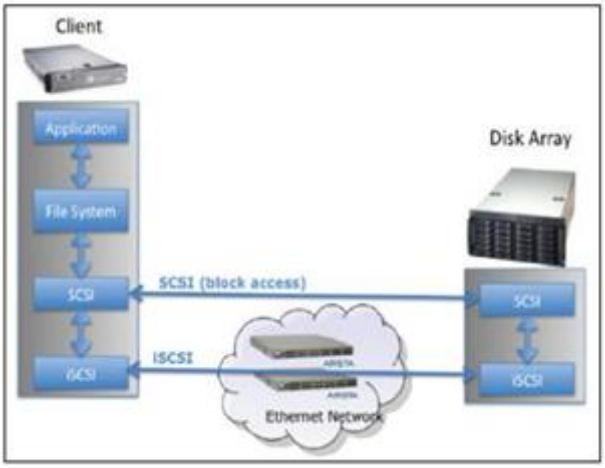


FIG 2. FIBER CHANNEL STORAGE AREA NETWORK FIG 3. ISCSI STORAGE AREA NETWORK

In these above figures Storage Area Networks, the access storage based on block. In network attached storage as in figure 4 below the user access data remotely based on system via network. 10 Gigabit Ethernet prevails as the mainstream technology for Cloud Storage with ISCSI based block storage and Network Attached Storage (NAS). With non-blocking throughput, record density, low latency, and leading total cost of ownership, Arista Networks switches are ideal for cloud storage applications [3].

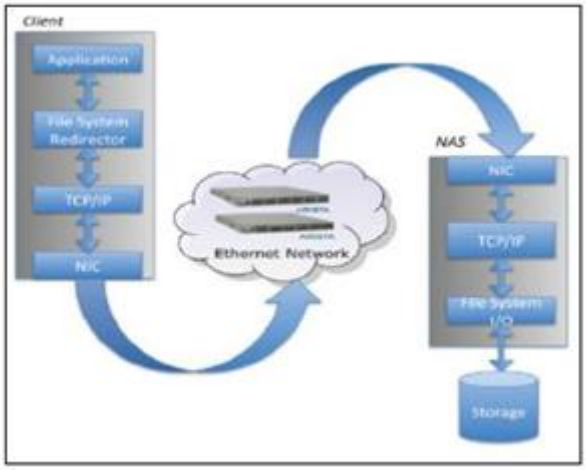


FIG. 4. NETWORK ATTACHED STORAGE

2. BIG DATA : PRESENT AND FUTURE

These days the data are produced from many sources such as social networks, website and sensor network. Also the total of data volume is expanding constantly. However; big data refers essentially to the following data types; Conventional enterprise data such as Customer information in Data Base, the transactions websites companies. Machine generated and Sensors data such as smart meter, manufacturing sensors etc. and Social data such as social network and application platforms like Facebook, LinkedIn, whatsapp, Twitter and YouTube. According to a recent report the most of data unstructured or semi structured and the size of data exists now is doubling in every two years. So between 2013 and 2020 it will go to 44 trillion GBs from 4.4 trillion GBs. Moreover the huge amount of data recorded mostly in nonstandard forms which cannot be analyzed using traditional data models and methods. Big Data today have a wide range of challenges but the opportunities are also exists the right decision making, marketing strategies and improved customer relations, better public services and so on.

By 2015 According to a Gartner report 4.4 million new big data related jobs will be created globally and only one third of these will be filled. So, employment opportunities are enormous in the big data job markets but there are very few training and education offerings

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

focusing on this market. Big data has opened the growing interest to new tools production Beginning with the introduction of Apache Hadoop and Map Reduce and also many open source have been implemented and developed by companies Terradata, Oracle, Cloudera, SAP, IBM, SAS Amazon and many others. Most big data products are mainly based on open-source technologies. Therefore, standards are especially important and needed for interoperability of the hardware and software components of commercial solutions. Lack of official standards also aggravates privacy and security problems.

3. CLOUD COMPUTING IN BIG DATA

The rise of cloud computing and cloud data stores has been a precursor and facilitator to the emergence of big data. Cloud computing is the co modification of computing time and data storage by means of standardized technologies. It has significant advantages over traditional physical deployments. However, cloud platforms come in several forms and sometimes have to be incorporated with traditional architectures. This leads to confuse for decision makers in charge of big data projects, leads to a question of how and which cloud computing is the optimal choice for their computing needs, especially if it is a big data project. These projects regularly exhibit changeable, bursting, or immense computing power and storage needs.

At the same time business stakeholders expect swift, inexpensive, and dependable products and project outcomes. This article introduces cloud computing and cloud storage, the core cloud architectures, and discusses what to look for and how to get started with cloud computing.

4. BIG DATA CLOUD PROVIDERS

Cloud providers come in all shapes and sizes and offer many different products for big data. Some are household names while others are recently emerging. Some of the cloud providers that offer IaaS services that can be used for big data include Amazon.com, AT&T, Rackspace, IBM, and Verizon/Terremark. Currently, one of the most high-profile IaaS service providers is Amazon web Services with its Elastic Compute Cloud (Amazon EC2). Amazon didn't start out with a vision to build a big infrastructure services business.

The cloud computing space has been dominated by Amazon Web Services until recently. Increasingly serious alternatives are emerging like Google Cloud Platform and other clouds that mentioned above. Amazon Web Services compatible solutions, i.e. Amazon's own offering or companies with application programming interface compatible offerings, and Open Stack, an open source project with a wide industry backing. Consequently, the choice of a cloud platform standard has implications on which tools are available and which alternative providers with the same technology are available.

5. BIG DATA CLOUD STORAGE

The cloud storage challenges in big data analytics fall into two categories: capacity and performance. Scaling capacity, from a platform perspective, is something all cloud providers need to watch closely. Data retention continues to double and triple year-over-year because customers are keeping more of it. Certainly, that impacts us because we need to provide capacity.

In Professional, cloud storage needs to be highly available, highly durable, and has to scale from a few bytes to terrabytes. Amazon's S3 cloud storage is the most prominent solution in the space. S3 promises a 99.9% monthly availability and 99.999999999% durability per year. This is less than an hour outage per month. The durability can be illustrated with an example. If a customer stores 10,000 objects he can expect to lose one object every 10,000,000 years on average. S3 achieves this by storing data in multiple facilities with error checking and self-healing processes to detect and repair errors and device failures. This is completely transparent to the user and requires no actions or knowledge. A company could build and achieve a similarly reliable storage solution but it would require fabulous capital expenditures and

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

operational challenges. Global data centered companies like Google or Facebook have the expertise and scale to do this efficiently. Big data projects and start-ups, however, benefit from using a cloud storage service. They can trade capital expenditure for an operational one, which is excellent since it requires no capital outlay or risk. It provides from the first byte reliable and scalable storage solutions of a quality otherwise unachievable. This enables new products and projects with a viable option to start on a small scale with low costs. When a product proves successful these storage solutions scale virtually indefinitely. Cloud storage is effectively a boundless data sink. prominently for computing performances is that many solutions also scale horizontally, i.e. when data is copied in parallel by cluster or parallel computing processes the throughput scales linear with the number of nodes reading or writing.

6. PRIVATE CLOUD

As big data volume, variety and velocity grow exponentially, the enterprise infrastructure must adapt in order to utilize it, becoming increasingly scalable, agile, and efficient if organizations are to remain competitive. To meet these requirements, Synnex Corporation, a leading distributor of IT products and solutions, announced it is the first distributor to offer a fully integrated, big data turnkey private cloud appliance powered by Nebula to the channel. The rack level appliance is a fully integrated private cloud system engineered to deliver workloads for both Apache Cassandra and MongoDB, enabling an open and elastic infrastructure to store and manage big data. The new rack system includes industry-standard servers, the Nebula One Cloud Controller, Apache Cassandra, MongoDB, backed by SYNNEX' powerful distribution model. This big data appliance enables an open and elastic infrastructure to store and manage big data. As an open source NoSQL database technology, Apache Cassandra provides multi-site distributed computing of big data across multiple data centers, while MongoDB is a cross-platform document-oriented open source database system.

Private clouds are dedicated to one organization and do not share physical resources. The resource can be provided in-house or externally. A typical underlying requirement of private cloud deployments are security requirements and regulations that need a strict separation of an organization's data storage and processing from accidental or malicious access through shared resources. Private cloud setups are challenging since the economic advantages of scale are usually not achievable within most projects and organizations despite the utilization of industry standards. The return of investment compared to public cloud offerings is rarely obtained and the operational overhead and risk of failure is significant. Additionally, cloud providers have captured the trend for increased security and provide special environments, i.e. dedicated hardware to rent and encrypt virtual private networks as well as encrypted storage to address most security concerns. Cloud providers may also offer data storage, transfer, and processing restricted to specific geographic regions to ensure compliance with local privacy laws and regulations. Another reason for private cloud deployments are legacy systems with special hardware needs or exceptional resource demand, e.g. extreme memory or computing instances which are not available in public clouds. These are valid concerns however if these demands are extraordinary the question if a cloud architecture is the correct solution has to be raised. One reason can be to establish a private cloud for a period to run legacy and demanding systems in parallel while their services are ported to a cloud environment culminating in a switch to a cheaper public or hybrid cloud.

7. PUBLIC CLOUD

Public clouds share physical resources for data transfers, storage, and processing. However, customers have private visualized computing environments and isolated storage. Security concerns, which entice a few to adopt private clouds or custom deployments, are for the vast majority of customers and projects unrelated. Visualization makes access to other customers' data extremely difficult. Real world problems around public

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

cloud computing are more mundane like data lock-in and fluctuating performance of individual instances. The data lock-in is a soft measure and works by making data inflow to the cloud provider free or very cheap. The copying of data out to local systems or other providers is often more expensive. This is not an impossible problem and in practice encourages utilizing more services from a cloud provider instead of moving data in and out for different services or processes. Usually this is not sensible anyway due to network speed and complexities around dealing with numerous platforms.

8. BIG DATA PRIVACY AND SECURITY

Big Data remains one of the most talked about technology trends in 2013. But lost among all the excitement about the potential of Big Data are the very real security and privacy challenges that threaten to slow this momentum. Security and privacy issues are magnified by the three V's of big data: Velocity, Volume, and Variety. These factors include variables such as large-scale cloud infrastructures, diversity of data sources and formats, streaming nature of data acquisition and the increasingly high volume of inter-cloud migrations. Consequently, traditional security mechanisms, which are tailored to securing small-scale static (as opposed to streaming) data, often fall short [17]. The CSA's Big Data Working Group followed a three step process to arrive at top security and privacy challenges presented by Big Data; Interviewed CSA members and surveyed security practitioner oriented trade journals to draft an initial list of high priority security and privacy problems studied published solutions. Characterized a problem as a challenge if the proposed solution does not cover the problem scenarios. Following this exercise, the Working Group researchers compiled their list of the Top 10 challenges as shown in figure 6 below. The Expanded Top 10 Big Data challenges have evolved from the initial list of challenges presented at CSA Congress to an expanded version that addresses three new distinct issues.

Modeling:

Formalizing a threat model that covers most of the cyber-attack or data-leakage scenarios.

Analysis:

Finding tractable solutions based on the threat model.

Implementation:

Implanting the solution in existing infrastructures.

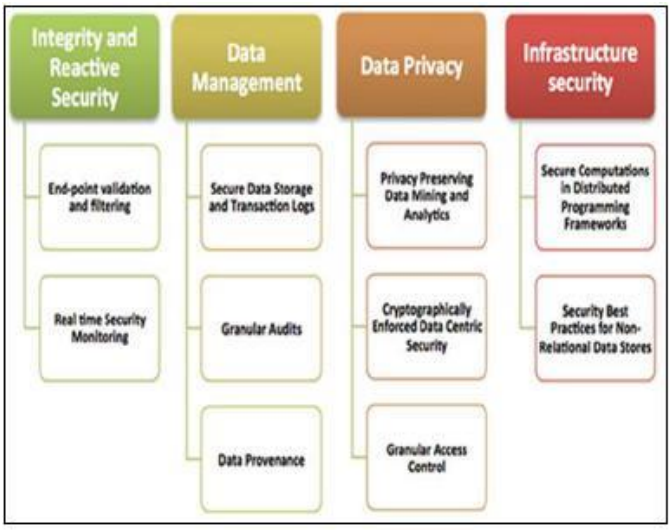


FIG. 5 CHALLENGES OF CSA's BIG DATA WORKING GROUP.

The information security practitioners at the Cloud Security Alliance know that big data and analytics systems are here to stay. They also agree on the big questions that come next: How can we make the systems that store and compute the data secure? And, how can we ensure private data stays private as it moves through different stages of analysis, input and output? The answers to those questions that prompted the group's latest 39-page report detailing 10 major security and privacy challenges facing infrastructure providers and customers. By outlining the issues involved, along with analysis of internal and external threats and summaries of current approaches to mitigating those risks, the alliance's members hope to prod technology vendors, academic researchers and practitioners to collaborate on computing techniques and business practices that reduce the risks associated with analyzing massive datasets using innovative data analytics.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Given the very large data sets that contribute to a Big Data implementations, there is a virtual certainty that either protected information or critical Intellectual Property (IP) will be present. This information is distributed throughout the Big Data implementation as needed with the result that the entire data storage layer needs security protection. There are many types of protection and security used such as Vormetric Encryption, Data Security Platform, Encryption and Key Management, Security Intelligence etc.

Vormetric Encryption: seamlessly protects Big Data environments at the file system and volume level. This Big Data analytics security solution allows organizations to gain the benefits of the intelligence gleaned from Big Data analytics while maintaining the security of their data – with no changes to operation of the application or to system operation or administration.

Data Security Platform: The Vormetric Data Security Platform secures critical data – placing the safeguards and access controls for your data with your data. The data security platform includes strong encryption, key management, fine-grained access controls and the security intelligence information needed to identify the latest in advanced persistent threats (APTs) and other security attacks on your data.

Encryption and Key Management: Data breach mitigation and compliance regimes require encryption to safeguard data. Vormetric provides the strong, centrally managed, encryption and key management that enables compliance and is transparent to processes, applications and users. Fine-grained Access Controls: Vormetric provides the fine-grained, policy based access controls that restrict access to data that has been encrypted allowing only approved access to data by processes and users as required to meet strict compliance requirements. Privileged users of all types (including system, network and even cloud administrators) can see plaintext information only if specifically enabled to do so. System update and administrative processes continue to work freely – but see only encrypted data, not the plaintext source.

Security Intelligence: Vormetric logs capture all access attempts to protected data providing high value, security intelligence information that can be used with a Security

Information and Event Management solution to identify compromised accounts and malicious insiders as well as finding access patterns by processes and users that may represent and APT attack in process.

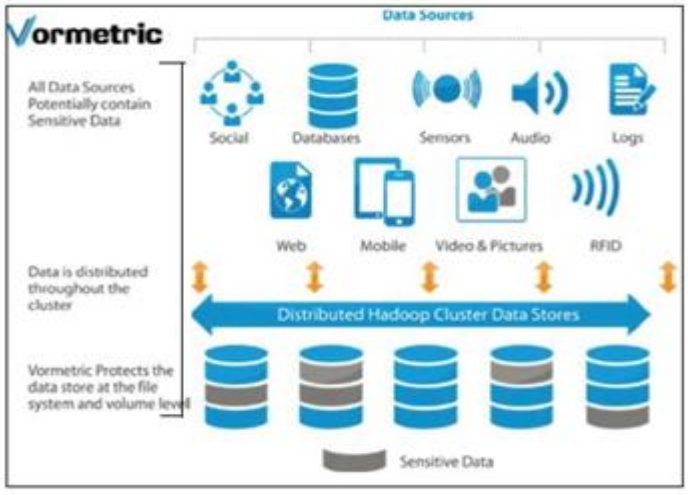


FIG. 6 VORMETRIC PERCEPTION

8. CONCLUSION

Recently, researchers focusing their efforts in how to manage , handling and also processing the huge amount of data as known a Big data deals with three concepts volume , Variety and velocity which requires a new mechanisms to manage , handing out, storing, analyzing and securing the big data . As managing and processing of big data have many problems and required more efforts to handle these requirements when deal with big data, security is one of the challenges that arise when systems try to handle the concept of big data. More researches required to overcome the security of big data instead of current safekeeping algorithms and methods.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

9. REFERENCES

1. Vahid Ashktorab1 , Seyed Reza Taghizadeh2 and Dr. Kamran Zamanifar3 ," A Survey on Cloud Computing and Current Solution Providers", International Journal of Application or Innovation in Engineering & Management (IJAIEM), October 2012 .
2. Addison Snell, Solving Big Data Problems with Private Cloud Storage, Intersect360 Research, Cloud Computing in HPC: Usage and Types,‖ June 2011
3. http://talkincloud.com/cloud-computing/041814/gartner-hybrid-cloud-big-data-analytics-top-smart-government-trends
4. http://www.dummies.com/how-to/content/big-data-cloud-providers.html
5. http://fcw.com/Articles/2012/04/30/FEAT-BizTech
6. Addison Snell, Solving Big Data Problems with Private Cloud Storage, Intersect360 Research, Cloud Computing in HPC: Usage and Types,‖ June 2011
7. http://www.businessweek.com/articles/2013-08-07/the-future-of-big-data-apps-and-corporate-knowledge
8. http://www.techrepublic.com/blog/the-enterprise-cloud/cloud-computing-and-the-rise-of-big-data/
9. http://www.networkworld.com/article/2176086/cloud-computing/big-data-drives-47--growth-for-top-50-public-cloud-companies.html
10. https://en.wikipedia.org/wiki/Big_data
11. https://en.wikipedia.org/wiki/Cloud_computing
12. https://en.wikipedia.org/wiki/Vormetric