

SECURITY WITH BIG DATA IN CLOUD COMPUTING

Ashutosh Kumar,

*Department of Computer Science & Engineering,
Institute of Technology, Gopeshwar*

ABSTRACT

In recent year, big data and cloud computing are the significant issues in an associations. It enables computing assets to be provided as data innovation service with high productivity and viability. The present big data is the one of the serious issue that analysts attempt to comprehend it. The primary spotlight is on security issues in cloud computing that are related with big data. Big data applications are an extraordinary advantage to relatives, industry, organization and various huge scale and slight scale enterprises. The another important issue is the method by which to increase an ideal security for big data in cloud computing. The potential answers for the issues in cloud computing security and Hadoop. Cloud computing assumes a fundamental job in ensuring data, applications and the related framework with the assistance of approaches, advancements, controls, and big data instruments. cloud computing, big data and its applications, points of interest are probably going to speak to the most encouraging new outskirts in science.

Keywords: *Cloud computing, Big data, Hadoop, Map reduce, HDFS, Big data security, CSA.*

1. INTRODUCTION

Big data is the term for data sets, so great and complex that it ends up hard processing utilizing usual data the board instrument or prepare application. The data and to differentiate designs it is essential to safely store, oversee and share a lot of composite data. Cloud accompanies an explicit security challenge, for example the data administrator maybe won't have any control of where the data is put. Hadoop is an open-source usage of Google MapReduce, including a single record framework, gives to the application software engineer. the deliberation of the guide and the decrease. With Hadoop it is simpler for associations to take a few to get back some mixture on the huge volumes of data being produced every day, and yet can also make issues identified with security, data get to, checking, high accessibility and business congruity. Google has presented MapReduce system for prepare a lot of data on item tools. Apache's Hadoop convey document framework (HDFS) is developing as an supreme programming segment for cloud computing connected together corresponding parts, for example, Map Reduce. Cloud Security Alliance (CSA) model blueprints where the data is being handled and put away, and incorporates the big data sources, preparing bunch and endpoint buyers of the data (frameworks, cell phones, and so forth.), alongside the cloud conditions. Security and protection issues are improved by speed, volume, and verity of big data, for example, vast scale cloud frameworks, various varieties of data sources ad groups, spill nature of data obtaining and high volume between cloud movement.

2. BIG DATA

Big Data is the word used to express massive volumes of structured and unstructured data that are so large that it is very difficult to process this data using traditional databases and software technologies. The term "Big Data" is believed to be originated from the Web search companies who had to query loosely structured very large distributed data. The three main conditions that mean Big Data have the following properties:

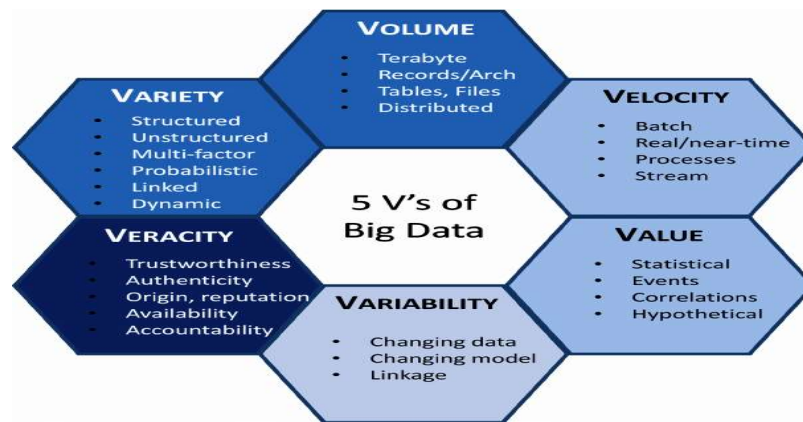


Figure 1 5 Vs of Big data

Volume: Many factor put in towards increasing Volume streaming data and data collected from sensors etc.

Variety: Today data comes in all types of formats emails, video, audio, transactions etc.

Velocity: This way how fast the data is being created and how fast the data needs to be process to meet the demand.

The other three dimensions that need to consider with respect to Big Data are Veracity, Variability and Complexity.

Variability: Along with the Velocity, the d peaks. This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to hold and manage the data effectively.

Veracity: The quality of the data being capture can vary greatly. Accuracy of analysis depends on the reality of the source data.

Complexity: Complexity of the data also needs to be consider when the data is coming from multiple sources. The data must be linked, matched, cleansed and transformed into required formats before actual processing.

When making an attempt to understand the concept of Big Data, the words and “Hadoop” cannot be avoid.

3. HADOOP

Hadoop, which is a free, Java-based programming framework supports the processing of large sets of data in a distributed computing background. It is a part of the Apache project sponsored by the Apache Software Foundation. Hadoop cluster uses a Master/Slave structure. Using Hadoop, large data sets can be processed across a cluster of servers and applications can be run on systems with thousands of nodes involving thousands of terabytes. Distributed file system in Hadoop helps in rapid data transfer rates and allows the system to continue its normal operation even in the case of some node failures. This approach lower the risk of an entire system failure, even in the case of a major number of node failure. Hadoop enables a computing solution that is scalable, cost effective, flexible and fault tolerant. Hadoop Framework is used by popular companies like Google, Yahoo, Amazon and IBM etc., to support their applications involving enormous amounts of data. Hadoop has two main sub projects – Map Reduce and Hadoop Distributed File System (HDFS).

4. CLOUD ISSUES AND CHALLENGES

Cloud computing comes with various security issues because it encompasses many technologies including networks, databases, operating systems, virtualization, resource scheduling, transaction management, load balancing, concurrency control and memory management. Hence, security issues of these systems and technologies are applicable to cloud computing. For example, it is very important for the network which interconnects the systems in a cloud to be secure. Also, virtualization model in cloud computing results in several security concerns. Data security not only involves the encryption of the data, but also ensures that appropriate policies are enforced for data sharing. In addition, resource allocation and memory management algorithms also have to be secure. The big data issues are most extremely felt in certain industries, such as telecoms, web marketing and advertising, retail and

financial services, and certain government activities. The data explosion is going to make life difficult in many industries, and the companies will gain considerable advantage which is capable to adapt well and gain the ability to analyze such data explosions over those other companies. Finally, data mining techniques can be used in the malware detection in clouds.

The challenges of security in cloud computing environments can be categorized into Four issues:

Network level: The challenges that can be categorized under a network level deal with network protocols and network security, such as distributed nodes, distributed data, Inter node communication.

Authentication level: The challenges that can be categorized under user authentication level deals with encryption/decryption techniques, authentication methods such as administrative rights for nodes, authentication of applications and nodes, and logging.

Data level: The challenges that can be categorized under data level deals with data integrity and availability such as data protection and distributed data.

Generic types: The challenges that can be categorized under general level are traditional security tools, and use of different technologies.

5. BIG DATA PRIVACY AND SECURITY

Big Data is one of the most talked about technology. 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 data, often fall short. 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.

Top 10 challenges as shown in figure 2 below.

The Expanded Top 10 Big Data challenges have evolved from the initial list of challenges presented at CSA 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 well-mannered solutions based on the threat model.

Implementation: implanting the solution in existing infrastructures.



Figure 2 Top 10 Challenges of CSA's Big Data Working Group

The information security practitioners at the Cloud Security Alliance know that big data and analytics systems. The group's latest 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 justifying 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.

Real-time system monitoring technique that works well on minor volumes of data but not very large datasets. The increasing number of devices, from smart phones to sensors, producing data for analysis. General misunderstanding "surrounding the diverse legal and policy restrictions that lead to ad hoc approaches for ensuring security and privacy. Given the very large data sets that contribute to a Big Data implementations, there is a virtual certainty that also 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: effortlessly protects Big Data environment at the file system and volume level. This Big Data analytics security solution allows organization to gain the profit 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 safeguard 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 in order needed to identify the latest in advanced persistent threats (APTs) and other security attacks on your data.

Encryption and Key Management: Data break improvement and agreement regime require encryption to protect data. Vormetric provides the strong, centrally managed, encryption and key management that enables compliance and is clear 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 convene firm fulfillment requirements. private users of all types (including system, network and even cloud administrators) can see plaintext information only if specially enabled to do so. System update and administrative processes continue to work to all comers – 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 compromise accounts and malicious insiders as well as finding access pattern by processes and users that may signify and APT attack in process. Use the Vormetric Toolkit to easily deploy, integrate and manage your Vormetric Data Security implementation with the rest of your big data implementation.



Figure 3. Vormetric Concept Automation:

CONCLUSION

in recent times , researchers focusing their efforts in how to manage , handling and also processing the enormous amount of data as known a Big data deals with three concepts volume , Variety and velocity which requires a new mechanism to manage , processing , storing , analyzing and securing the big data. Its managing and processing of big data have many problems and required more efforts to hold these requirements when deal with big data, security is one of the challenges that occur when systems try to hold the concept of big data. More researches required to beat the security of big data instead of current security algorithms and methods. Cloud environment is widely used in engineering and research aspects; therefore security is an important aspect for organization running on these cloud environments. Using proposed approaches, cloud environments can be secured for complex business operations. By lever aging the work of the CSA working group for big data and focusing explicitly on the key controls that should be in place, enterprise consumers can help to properly evaluate the state of big data communications and applications in their service providers' environment.

REFERENCE

- Ren, Yulong, and Wen Tang. "A SERVICE INTEGRITY ASSURANCE FRAMEWORK FOR CLOUD COMPUTING BASED ON MAPREDUCE." Proceedings of IEEE CCIS2012. Hangzhou: 2012, pp 240 – 244, Oct. 30 2012-Nov. 1 2012
- N, Gonzalez, Miers C, Redigolo F, Carvalho T, Simplicio M, de Sousa G.T, and Pourzandi M. "A Quantitative Analysis of Current Security Concerns and Solutions for Cloud Computing.". Athens: 2011., pp 231 – 238, Nov. 29 2011- Dec. 1 2011
- Hao, Chen, and Ying Qiao. "Research of Cloud Computing based on the Hadoop platform.". Chengdu, China: 2011, pp. 181 – 184, 21-23 Oct 2011.
- Y, Amanatullah, Ipung H.P., Juliandri A, and Lim C. "Toward cloud computing reference architecture: Cloud service management perspective.". Jakarta: 2013, pp. 1-4, 13-14 Jun. 2013.
- A, Katal, Wazid M, and Goudar R.H. "Big data: Issues, challenges, tools and Good practices.". Noida: 2013, pp. 404 – 409, 8-10 Aug. 2013.
- Lu, Huang, Ting-tin Hu, and Hai-shan Chen. "Research on Hadoop Cloud Computing Model and its Applications.". Hangzhou, China: 2012, pp. 59 – 63, 21-24 Oct. 2012.
- Wie, Jiang , Ravi V.T, and Agrawal G. "A Map-Reduce System with an Alternate API for Multi-core Environments.". Melbourne, VIC: 2010, pp. 84-93, 17-20 May. 2010.