INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

REVIEW ON BIG DATA SECURITY IN CLOUD COMPUTING

MISS. ANKITA S. AMBADKAR[1], PROF. H. R. DESHAMUKH[2]

1. Department of Computer Science and Engineering, IBSS College of Engineering, Amravati, Maharashtra, India.
2. Professor, Department of Computer Science and Engineering, IBSS College of Engineering, Amravati, Maharashtra, India.

Abstract: Big data is the term for a collection of data sets which are large and complex, it contain structured and unstructured both type of data. Data comes from everywhere, sensors used to gather climate information, posts to social media sites, digital pictures and videos etc. This data is known as big data. In this paper, we discuss security issues for cloud computing, Big data, Map Reduce and Hadoop environment. The main focus is on security issues in cloud computing that are associated with big data. We also discuss various possible solutions for the issues in cloud computing security. Cloud computing security is developing at 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.

Keywords: Cloud Computing, Big Data, Hadoop, Map Reduce, HDFS (Hadoop Distributed File System)

Corresponding Author: MISS. ANKITA S. AMBADKAR

Access Online On:

www.ijpret.com

How to Cite This Article:

PAPER-QR CODE

INTRODUCTION

The term 'Big Data' appeared for first time in 1998 in a Silicon Graphics (SGI) slide deck by John Mashey with the title of "Big Data and the Next Wave of Infra Stress". The term big data refers massive amount of information companies and government collect about us and our surrounding. Every day we create 2.5 quintillon bytes of data-so much that 90% of the data in the world today has been created in the last two year alone. The data produced nowadays is estimated in the order of zetta bytes, and it is growing around 40% every year. A new large source of data is going to be generated from mobile devices and big companies as Google, Apple, Facebook, Yahoo are starting to look carefully to this data to find useful patterns to improve user experience. "Big data" is pervasive, and yet still the notion engenders confusion. The term Big Data refers to large-scale information management and analysis technologies that exceed the capability of traditional data processing technologies.[1] Big Data is differentiated from traditional technologies in three ways: the amount of data (volume), the rate of data generation and transmission (velocity), and the types of structured and unstructured data (variety). Big Data is a heterogeneous mix of data both structured (traditional datasets –in rows and columns like DBMS tables, CSV's and XLS's) and unstructured data like e-mail attachments, manuals, images, PDF, medical records such as x-rays, ECG and MRI images, forms, rich media like graphics, video and audio, contacts, forms and documents. In order to analyze complex data and to identify patterns it is very important to securely store, manage and share large amounts of complex data. Cloud comes with an explicit security challenge, i.e. the data owner might not have any control of where the data is placed. Hence it is required to protect the data in the midst of untrustworthy processes. Since cloud involves extensive complexity, we believe that rather than providing a holistic solution to securing the cloud, it would be ideal to make noteworthy enhancements in securing the cloud that will ultimately provide us with a secure cloud.

II. Cloud Computing

Cloud Computing is a technology which depends on sharing of computing resources than having local servers or personal devices to handle the applications. In Cloud Computing, the word "Cloud" means "The Internet", so Cloud Computing means a type of computing in which services are delivered through the Internet.

Following fig.1 show the cloud computing

The goal of Cloud Computing is to make use of increasing computing power to execute millions of instructions per second. Cloud Computing consists of a front end and back end. The front end includes the user's computer and software required to access the cloud network. Back end consists of various computers, servers and database systems that create the cloud. The user can access applications in the cloud network from anywhere by connecting to the cloud using the Internet. 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, this technology requires to install a single software in each computer that allows users to log into a Web-based service and which also hosts all the programs required by the user. There's a significant workload shift, in a cloud computing system.

III. BIG DATA

Big Data starts with large-volume, heterogeneous, autonomous sources with distributed and decentralized control, and seeks to explore complex and evolving relationships among data. These characteristics make it an extreme challenge for discovering useful knowledge from the Big Data. The term "Big Data [5]" is companies who had to query loosely structured very large distributed data.

1. V's in Big Data

Doug Laney was the first one talking about 3V"s in Big Data Management

Volume: The amount of data. Perhaps the characteristic most associated with big data, volume refers to the mass quantities of data that organizations are trying to harness to improve decision-making across the enterprise. Data volumes continue to increase at an unprecedented rate.

Variety: Different types of data and data sources. Variety is about managing the complexity of multiple data types, including structured, semi-structured and unstructured data. With the explosion of sensors, smart devices and social collaboration technologies, data is being generated in countless forms, including: text, web data, tweets, audio, video, log files and more.

Velocity: Data in motion. The speed at which data is created, processed and analyzed continues to accelerate.

The other two dimensions that need to consider with respect to Big Data are Variability and Complexity [4].

 Variability: Along with the Velocity, the Periodic peaks.

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

2. Need of security in big data

By using online big data application, a lot of companies can greatly reduce their IT cost. However, security and privacy affect the entire big data storage and processing, since there is a massive use of third-party services and infrastructures that are used to host important data or to perform critical operations. Cloud computing introduces risks to any sensitive data it touches. These risks largely arise from the need to entrust data protection to a third party cloud provider. Different nodes in the environment may be controlled or administered by different untrusted parties, and could be vulnerable to attacks from other cloud tenants, malicious insiders or external adversaries. When data owners release control of their data to a cloud environment, they require guarantees that their data remains appropriately protected. Today, these guarantees are typically legal promises that the cloud provider makes to the owner as outlined in a service level agreement (SLA). Cryptography allows data owners to protect their data proactively instead of relying solely on legal agreements that are difficult to monitor or enforce. The scale of data and applications grow exponentially, and bring huge challenges of dynamic data monitoring and security protection. Unlike traditional security method, security in big data is mainly in the form of how to process data mining without exposing sensitive information of users. Besides, current technologies of privacy protection are mainly based on static data set, while data is always dynamically changed, including data pattern, variation of attribute and addition of new data. Thus, it is a challenge to implement effective privacy

protection in this complex circumstance. In addition, legal and regulatory issues also need attention. The challenge of detecting and preventing advanced threats and malicious intruders, must be solved using big data style analysis. These techniques help in detecting the threats in the early stages using more sophisticated pattern analysis and analyzing multiple data sources. The use of big data by government agencies will not change this; rather it may add an additional layer of complexity in terms of managing information security risks. Big data sources, the transport and delivery systems within and across agencies, and the end points for this data will all become targets of interest for hackers, both local and international and will need to be protected.

3. Hadoop

Hadoop, which is a free, Java-based programming frame work, supports the processing of large sets of data in a distributed computing environment. Hadoop cluster uses a Master/Slave structure [9]. 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 tera bytes. 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. Hadoop Framework is used by popular companies like Google, Yahoo, Amazon and IBM etc., to support their applications involving huge amounts of data. Hadoop has two main sub projects – Map Reduce and Hadoop Distributed File System (HDFS).

4. Big data advantages

In Big data, the software packages provide a rich set of tools and options where an individual could map the entire data landscape across the company, thus allowing the individual to analyze the threats he/she faces internally. This is considered as one of the main advantages as big data keeps the data safe. With this an individual can be able to detect the potentially sensitive information that is not protected in an appropriate manner and makes sure it is stored according to the regulatory requirements.

IV. ISSUES AND CHALLENGES

The challenges of security in cloud computing environments can be categorized into network level, user authentication level, data level, and generic issues [4]. 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 acutely felt in certain industries, such as telecoms, web

marketing and advertising, retail and financial services, and certain government activities. Cloud computing comes with numerous 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.

V. CONCLUSION

Cloud environment is widely used in industry and research aspects; therefore security is an important aspect for organizations running on these cloud environments. Using propose approaches, cloud environments can be secured for complex business operations. Big Data is not a new concept but very challenging. It calls for scalable storage index and a distributed approach to retrieve required results near real-time. It is a fundamental fact that data is too big to process conventionally. It is an urgent need that computer scholars and social sciences scholars make close cooperation, in order to guarantee the long-term success of cloud computing. This approach may finally overcome the lack of efficiency of all current protocols and lead to practical solutions for data security in the cloud. Although Big data and Cloud computing is a new phenomenon which is set to revolutionize caution must be exercised in the way we use the Internet. The use of big data could provide sufficient benefit to a small to medium sized company to the extent that the business would be willing to commit resources to implement big data technology in-house, Cloud computing implementation of big data has the potential to become a frontrunner in promoting a secure, virtual and economically viable IT solution in the future.

VI. REFERENCES

1. Ren, Yulong, and Wen Tang. "A service integrity assurance framework for cloud computing base on the map reduce". Proceedings of IEEE CCIS2012. Hangzhou: 2012, pp 240 –244, Oct. 30 2012-Nov. 1 2012

2. Hao, Chen, and Ying Qiao. "Research of Cloud Computing based on the Hadoop platform.". Chengdu, China: 2011, pp. 181 – 184, 21-23 Oct 2011.

3. 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.

4. 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.

5. P. Mell and T. Grace, "The NIST definition of cloud computing," NIST Special Publication 800-145, 2011

6. P. R , Anisha, Kishor Kumar Reddy C, Srinivasulu Reddy K, and Surender Reddy S. "Third Party Data Protection Applied To Cloud and Xacml Implementation in the Hadoop Environment With Sparql."2012. 39-46, Jul – Aug. 2012.

7. "Securing Big Data: Security Recommendations for Hadoop and NoSQL Environments. "Securosis blog, version 1.0 (2012)

8. 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.

9. K, Chitharanjan, and Kala Karun A. "A review on hadoop — HDFS infrastructure extensions." JeJu Island: 2013, pp. 132-137, 11-12 Apr. 2013.

Available Online at www.ijpret.com