iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 6 **Issue:** III **Month of publication:** March 2018

**DOI:** http://doi.org/10.22214/ijraset.2018.3744

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887*
*Volume 6 Issue III, March 2018- Available at www.ijraset.com*

# Review on Big Data Security Management

Mr. A. Antony Prakash[1], Mr. S. Prithiviraj [2], Ms. Reeta Mary[3]
[1] *Asst. Professor in Information Tech, St Joseph's College - Tiruchirappalli- 2*
[2, 3] *Student in Information Tech, St Joseph's College - Tiruchirappalli – 2*

***Abstract:*** ***Big data phenomenon arises from the increasing number of data collected from various sources, including the internet. Big data is not only about the size or volume. Big data possess specific characteristics (volume, variety, velocity, and value - 4V) that make it difficult to manage from security point of view. The evolution of data to become big data rises another important issue about data security and its management. NIST defines guide for conducting risk assessments on data, including risk management process and risk assessment. This paper looks at NIST risk management guidance and determines whether the approach of this standard is applicable to big data by generally define the threat source, threat events, vulnerabilities, likelihood of occurrence and impact. The result of this study will be a general framework defining security management on Big Data***

***Keywords: Big data, Security, models, VPN, firewall***

## I. INTRODUCTION

DATA analytics is the key to understand information of knowledge about some certain activities. The process of data analysis includes checking, cleaning, modelling, and transformation of the data. The information gathered from this process is then used for suggestion, summary and support for decision making.

User generated content is responsible for generating a huge volume of data to be analyzed for many purposes, from business to security. Machine to machine communication (M2M) and the Internet of Things also produce a vast amount of data.

Big data has certain characteristics that are called 4V; volume, variety, velocity, and value. Currently, 2.5 quintillion bytes of data are produced daily. The format (and content) of data varies and unstructured. The speed of data creation is faster than the speed of analysis. The use and value of the data also varies. This creates a problem in the analysis and safe guarding of the big data.

One of the biggest problems in big data is security. Some big data initiatives failed due to the unclear security controls. Thus, security is important in big data implementation.

Security can be seen from three aspects; confidentiality, integrity, and availability. Confidentiality is responsible for securing access of big data. Unfortunately, the massive size of data sources and mixing of the sources make it difficult to decide who is granted to access and analyze the data.

Big Data can be seen in the finance and business where enormous amount of stock exchange, banking, online and onsite purchasing data flows through computerized

systems every day and are then captured and stored for inventory monitoring, customer behavior and market behavior. It can also be seen in the life sciences where big sets of data such as genome sequencing, clinical data and patient data are analyzed and used to advance breakthroughs in science in research.

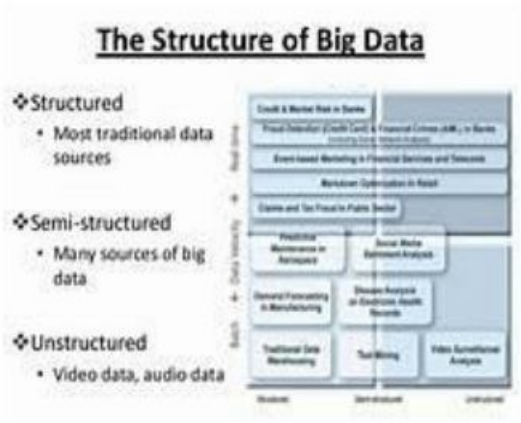
According to, an article about Social Genome: Putting Big Data to work for Population Informatics, security issues such as privacy, confidentiality, access and data integration progressing slowly.

## II. BIG DATA

Big Data is a phrase used to mean a massive volume of both structured and unstructured data that is so large it is difficult to process using traditional database and software technique. In most enterprise scenarios the volume of data is too big or it moves too fast or it exceeds current processing capacity.

Big data refers to a process that is used when traditional data mining and handling techniques cannot uncover the insights and meaning of the underlying data. Data that is unstructured or time sensitive or simply very large cannot be processed by relational database engines. This type of data requires a different processing approach called big data, which uses massive parallelism on readily-available hardware.

©IJRASET (UGC Approved Journal): All Rights are Reserved




Big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization." Cited from Wikipedia.

"Big data is the term increasingly used to describe the process of applying serious computing power – the latest in machine learning and artificial intelligence – to seriously massive and often highly complex sets of information." Cited from 4/2013 the Microsoft Enterprise Insight Blog.

"An easily scalable system of unstructured data with accompanying tools that can efficiently pull structured datasets." Cited from a 4/2013 post on the FCW Blog.

"Big data refers to a process that is used when traditional data mining and handling techniques cannot uncover the insights and meaning of the underlying data. Data that is unstructured or time sensitive or simply very large cannot be processed by relational database engines. This type of data requires a different processing approach called big data, which uses massive parallelism on readily-available hardware." Cited from a Cory Janssen post on Techopedia.com

## III. SECURITY AND BIG DATA

The advent of Big Data has presented nee challenges in terms of Data Security. There is an increasing need of research in technologies that can handle the vast volume of Data and make it secure efficiently. Current Technologies for securing data are slow when applied to huge amounts of data.

Big data has complexities that mostly people and companies are unprepared to deal with. These complexities include security and governance of data in general. Information governance is the capability to create information resource that can be trusted by employees, partners, and customers, as well as government organizations.

Big data comes from many data sources that might have different security and governance policies. A well-defined security strategy has to be applied on whatever information management. Combination of security and governance strategy need collaboration and coordination to share responsibilities across organizations/ parties involved to make sure the accountability is enforced to the data being used.

### *A. Security solution architecture*

The era of big data is producing unprecedented amounts of data points giving us greater insights that drive exciting research, better business decisions, and in many ways, greater value for customers. To achieve these outcomes, organizations need to be able to handle it efficiently, quickly, and because often this data will include sensitive information – securely, all at scale.

©IJRASET (UGC Approved Journal): All Rights are Reserved


*B. Security incident management*

Security incident management is the process of identifying, managing, recording and analyzing security threats or incidents in real-time. It seeks to give a robust and comprehensive view of any security issues within an IT infrastructure. A security incident can be anything from an active threat to an attempted intrusion to a successful compromise or data breach. Policy violations and unauthorized access to data such as health, financial, social security numbers, and personally identifiable records are all examples of security incidents.

*C. Web application security*

Web application security, is a branch of Information Security that deals specifically with security of websites, web applications and web services. At a high level, Web application security draws on the principles of application security but applies them specifically to Internet and Web systems

*D. Content filtering*

Content filtering, in the most general sense, involves using a program to prevent access to certain items, which may be harmful if opened or accessed. The most common items to filter are executables, emails or websites. Content filters can be implemented either as software or via a hardware- based solution.

*E.* Load balancing

The Big Data load balancing technology involves the establishment of cross the cluster network channel, cluster deployment, task scheduling optimization. These should be assessed in a longitudinal study based on network architecture. This paper is to summarize the current situation of load balancing research about Big Data, and provide a basis for management and research of Big Data in the future.

*F.* Traffic optimization

The objective function used in this work is a weighted sum of the delays caused by the signalized intersections. In this paper, we apply generalized 'surrogate problem' methodology that is based on an on-line control scheme which transforms the problem into a 'surrogate' continuous optimization problem and proceeds to solve the latter using standard gradient-based approaches while simultaneously updating both actual and surrogate system states.

*G. Intrusion prevention:*

©IJRASET (UGC Approved Journal): All Rights are Reserved

Intrusion prevention is a preemptive approach to network security used to identify potential threats and respond to them swiftly. Like an intrusion detection system (IDS), an intrusion prevention system (IPS) monitors network traffic. However, because an exploit may be carried out very quickly after the attacker gains access, intrusion prevention systems also have the ability to take immediate action, based on a set of rules established by the network administrator.

*H. Firewall & VPN*

A firewall employs packet filtering to allow or disallow the flow of very specific types of network traffic. IP packet filtering provides a way for you to define precisely what IP traffic is allowed to cross the firewall. IP packet filtering is important when you connect private intranets to public networks like the Internet.

VPN Server and Firewall Configurations

There are two approaches to using a firewall with a VPN server:

1) The VPN server is attached to the Internet and the firewall is between the VPN server and the intranet.
2) The firewall is attached to the Internet and the VPN server is between the firewall and the intranet.

*I. Wireless security*

Wireless security is the prevention of unauthorized access or damage to computers using wireless networks. The most common types of wireless security are Wired Equivalent Privacy (WEP) and Wi-Fi Protected Access (WPA). WEP is a notoriously weak security standard. The password it uses can often be cracked in a few minutes with a basic laptop computer and widely available software tools.

## IV. SECURITY MODEL

Computer security can be a slippery term because it means different things to different people. There are many aspects of a system that can be secured, and security can happen at various levels and to varying degrees. We have stated in previous chapters that information security is made up of the following main attribute

*1) Availability* Prevention of loss of access to resources and data
*2) Integrity* Prevention of unauthorized modification of data
*3) Confidentiality* Prevention of unauthorized disclosure of data

## V. CONCLUSION AND FUTURE SCOPE

Big Data is changing the way we perceive our world. The impact big data has created and will continue to create can ripple through all facets of our life. Global Data is on the rise, by 2020, we would have quadrupled the data we generate every day. This data would be generated through a wide array of sensors we are continuously incorporating in our lives. Data collection would be aided by what is today dubbed as the “Internet of Things”. Through the use of smart bulbs to smart cars, everyday devices are generating more data than ever before. These smart devices are incorporated not only with sensors to collect data all around them but they are also connected to the grid which contains other devices. A Smart Home today consists of an all encompassing architecture of devices that can interact with each other via the vast internet network. Bulbs that dim automatically aided by ambient light sensors and cars that can glide through heavy traffic using proximity sensors are examples of sensor technology advancements that we have seen over the years. Big Data is also changing things in the business world. Companies are using big data analysis to target marketing at very specific demographics. Focus Groups are becoming increasingly redundant as analytics firms such as McKinsey are using analysis on very large sample bases that have today been made possible due to advancements in Big Data. The potential value of global personal location data is estimated to be $700 billion to end users, and it can result in an up to 50% decrease in product development and assembly costs, according to a recent McKinsey report. Big Data does not arise out of a vacuum: it is recorded from some data generating source.

## REFERENCES

[1] Jing, H.: Research on some key technologies of BIGDATA-AS-A-Service. Beijing University of Posts and Telecommunications (2013)Google Scholar

[2] E. Acar, D. M. Dunlavy, T. G. Kolda, and M.Mørup, “Scalable tensor factorizations for incomplete data,” Chemomet. Intell. Lab. Syst., vol. 106, no. 1, pp. 41–56, 2011.

[3] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD:An algorithm for designing overcomplete dictionaries for sparse representation,” IEEE Trans. Signal Processing, vol.54, no. 11, pp. 4311–4322, Nov. 2006.

©IJRASET (UGC Approved Journal): All Rights are Reserved

[4] Michael J. Vandeman, "Is Traffic Signal Synchronization Justifiable?" April 15, 1994.

[5] S. Sagiroglu and D. Sinanc. Big data: A review. In International Conference on Collaboration Technology and

[6] System. International Conference on Collaboration Technology and System, 2013

[7] Chen He Ying Lu David Swanson,"Matchmaking: A New MapReduce Scheduling Technique", Third IEEE International Conference on Coud Computing Technology and Science,2011.

[8] D. Zage, K. Glass, and R. Colbaugh. Improving supply chain security using big data. In International Conference on Intelligence and Security Informatics. IEEE,2013.

[9] M. Zaharia, D. Borthakur, J. S. Sarma, K. Elmeleegy, S. Shenker, and I. Stoica, "Job scheduling for multi-user mapreduce clusters," EECS Department, University of California, Berkeley, Tech. Rep.,Apr 2009.

[10] Vidyullatha Pellakuri, Dr.D. Rajeswara Rao," Hadoop Mapreduce Framework in Big Data Analytics", (IJCTT) – volume 8 number 3– Feb 2014

[11] HDFS Users Guide - Rack Awareness", Hadoop.apache.org. Retrieved 2013-10-17.

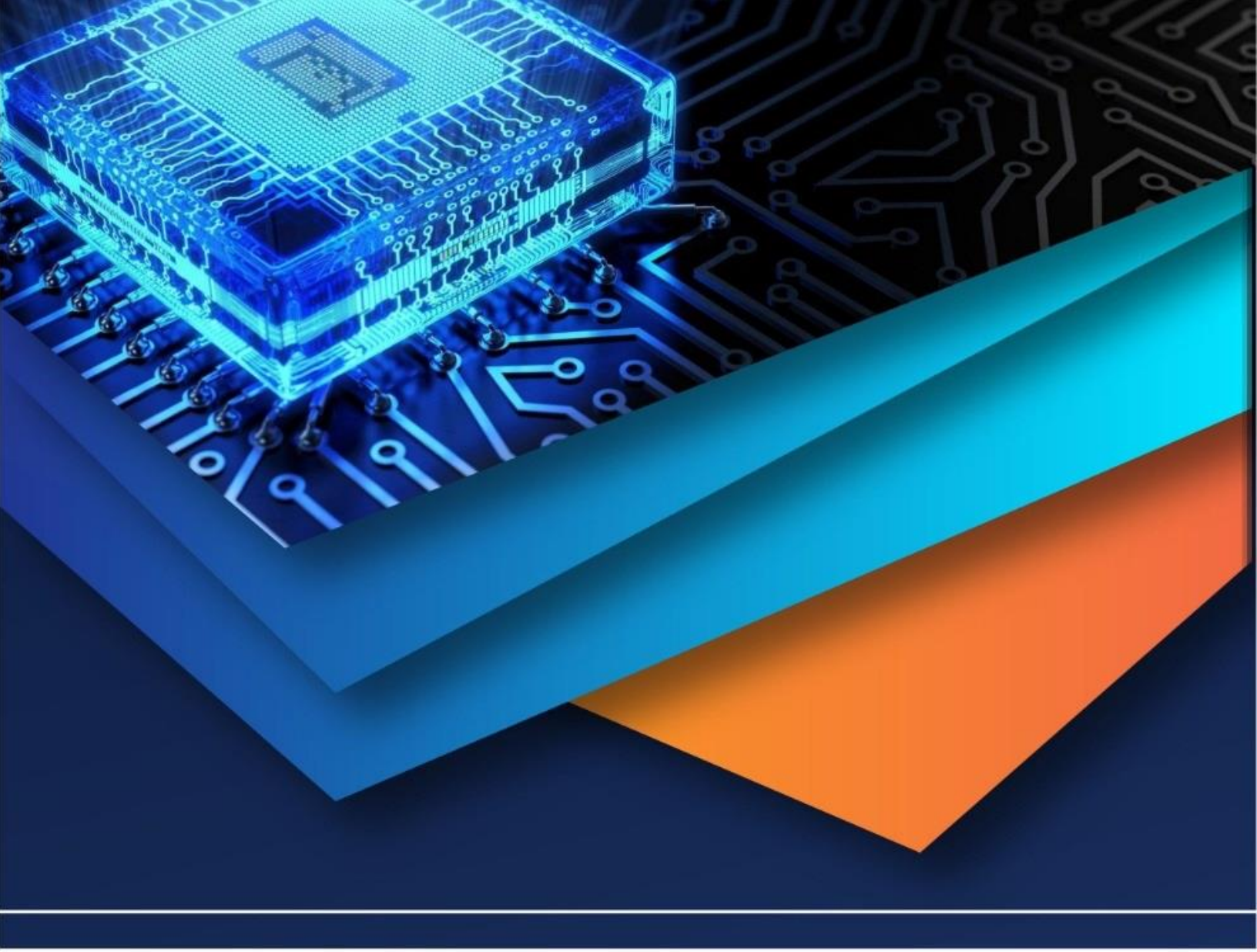
[12] D. Wegener, M. Mock, D. Adranale, and S. Wrobel, "Toolkit-Based High-Performance Data Mining of Large Data on MapReduce Clusters," Proc. Int'l Conf. Data Mining Workshops (ICDMW '09), pp. 296-301, 2009

[13] "Zettaset Launches Version 4 of Big Data Management Solution, Delivering New Stability for Hadoop Systems and Productivity Boosting Features",Zettaset.com 2011-12-06. Retrieved 2012-05-23.

[14] Harshawardhan S. Bhosle, Devendra P. Gadekar : "Big Data Processing Using Hadoop:Survey on Scheduling",IJRS, Volume 3 Issue 10, 2014.

[15] DeWitt & Stonebraker, "MapReduce: A major step backwards", 2008.

©IJRASET (UGC Approved Journal): All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)