Security Challenges in Big Data

Miss. Debalina Nandy[1] Mr. Renish J Padariya[2] Miss Tosal Bhalodia[3]
[1,2,3]Assistant Professor
[1,2,3]Department of Computer Engineering
[1,3]Atmiya Institute of Technology & Science, Rajkot, Gujarat [2]Om Engineering College, Junagadh, Gujarat

Abstract— The biggest challenge for big data from a security point of view is the protection of user's privacy. Big data frequently contains huge amounts of personal identifiable information and therefore privacy of users is a huge concern. Because of the big amount of data stored, breaches affecting big data can have more devastating consequences than the data breaches we normally see in the press. This is because a big data security breach will potentially affect a much larger number of people, with consequences not only from a reputational point of view, but with enormous legal repercussions. While using big data a significant challenge is how to establish ownership of information. If the data is stored in the cloud a trust boundary should be establish between the data owners and the data storage owners.

Key words: Analytics of Big Data, Data Security, Data Source, Big Data Framework

I. INTRODUCTION

Nowadays, organizations are collecting and processing massive amounts of information. The more data is stored, the more vital it is to ensure its security. A lack of data security can lead to great financial losses and reputational damage for a company. As far as Big Data is concerned, losses due to poor IT security can exceed even the worst expectations. As the amount of data being collected continues to grow, more and more companies are building big data repositories to store, aggregate and extract meaning from their data. Big data provides an enormous competitive advantage for corporations, helping businesses tailor their products to consumer needs, identify and minimize corporate inefficiencies, and share data with user groups across the enterprise. Big data presents a tremendous opportunity for enterprises across industries. By tapping into new volumes and varieties of data, scientists, executives, product managers, marketers, and a range of others can start making more informed plans and decisions, discover new opportunities for optimization, and deliver breakthrough innovations [1].

A. The Infrastructure and Technology

Compared with a single high-end database server, distributed environments are more complicated and vulnerable to attack. When big data environments are distributed geographically, physical security controls need to be standardized across all accessible locations. When data scientists across the organization want access to information, perimeter protection becomes important and complicated to ensure access to users while protecting the system from a possible attack. With a large number of servers, there is an increased possibility that the configuration of servers may not be consistent – and that certain systems may remain vulnerable [3].

An additional big data security challenge is that big data programming tools, including Hadoop and NoSQL databases, were not originally designed with security in mind. For example, Hadoop originally didn't authenticate services or users, and didn't encrypt data that's transmitted between nodes in the environment. This creates vulnerabilities for authentication and network security. NoSQL databases lack some of the security features provided by traditional databases, such as role-based access control. The advantage of NoSQL is that it allows for the flexibility to include new data types on the fly, but defining security policies for this new data is not straightforward with these technologies [2].

II. MASSIVE SCOPE IN BIG DATA SECURITY

To establish comprehensive big data security, executives and administrators have to address the following areas:

- Data sources: To most fully exploit the advantages of big data, organizations leverage various forms of data, including both structured data in a range of heterogeneous applications and databases and unstructured data that comes in a number of file types. Organizations may leverage data from enterprise resource planning systems, customer relationship management platforms, video files, spreadsheets, social media feeds, and many other sources. Further, more data sources are added all the time. Today, you don't know where new data sources may come from tomorrow, but you can have some certainty that there will be more to contend with and more diversity to accommodate. These big data sources can include personally identifiable information, payment card data, intellectual property, health records, and much more. Consequently, the data sources being compiled need to be secured in order to address security policies and compliance mandates [2].
- Big data frameworks: Within the big data environment itself—whether it's powered by Hadoop, MongoDB, NoSQL, Teradata, or another system—massive amounts of sensitive data may be managed at any given time. Sensitive assets don't just reside on big data nodes, but they can come in the form of system logs, configuration files, error logs, and more [5]
- Analytics: The ultimate fruit of a big data initiative is the output, the analytics that help the business optimize and innovate. This information can be presented in dashboards and reports, and accessed via on-demand queries. In some businesses, big

data analytics represent the most sensitive asset of all, intelligence that provides a critical competitive differentiator—and a huge competitive exposure if it falls into the wrong hands [5].

III. APPLYING BIG DATA SECURITY ANALYSIS

Big data creates a single view of the enterprise while automating security response tools, which promises to dramatically change the way IT professionals approach security:

- Security management – The convergence of security information and event management (SIEM) with real-time network monitoring provides a unified approach to security management. All the information needed to inform security can now be combined with big data analytics to correlate thousands of events per second without additional hardware [4].
- Identify and access management (IAM) – Big data security analysis allows the enterprise to continuously adapt identity controls, enabling situation-aware IAM tools that can provide secure access on demand [3]
- Fraud detection and prevention – Analyzing massive amounts of behavioural data makes it possible to distinguish between legitimate and fraudulent activity [3].

Governance, risk management, and compliance (GRC) – Big data security analysis also unifies and enables real-time access to all the factors affecting business GRC. Analyzing large volumes of data will promote smarter decision-making that mitigates risk.

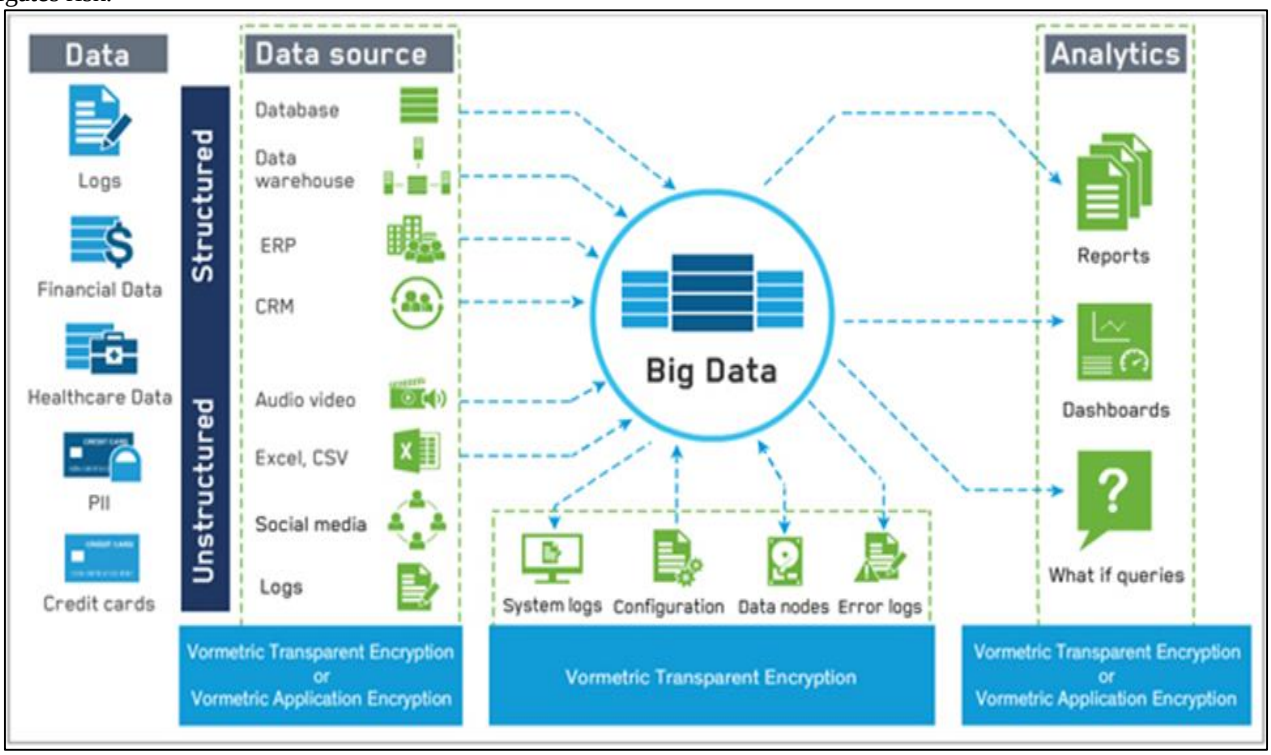

Fig. 1: Data Source Analytics

IV. LIMITATIONS OF TRADITIONAL ENCRYPTION APPROACHES

While some vendors offer big data encryption capabilities, these offerings only secure specific big data nodes, not the original data sources that are fed into the big data environment or the analytics that come out of the environment. Further, these big data encryption offerings don't even secure all the log files and configuration information associated with the big data environment itself. Ultimately, with these disparate approaches to big data security, IT teams have to contend with fragmented key and policy management, which adds administrative effort, while making it difficult to apply standards consistently. Further, these point approaches also tend to introduce a significant performance hit, which can present significant issues in processing-intensive big data environments [4]. Vormetric solutions for big data security enable organizations to maximize the benefits of big data analytics—while maximizing the security of their sensitive data and addressing the requirements of their compliance office. The Vormetric Data Security Platform offers the granular controls, robust encryption, and comprehensive coverage that organizations need to secure sensitive data across their big data environments—including big data sources, big data infrastructure, and big data analytic results. By delivering a single security solution that offers coverage of these areas, Vormetric enables security teams to leverage centralized controls that optimize efficiency and compliance adherence. The Vormetric Data Security Platform offers capabilities for big data encryption, key management, and access control—featuring several product offerings that share a common, extensible infrastructure. Further, the solution generates security intelligence on data access by users, processes [6].

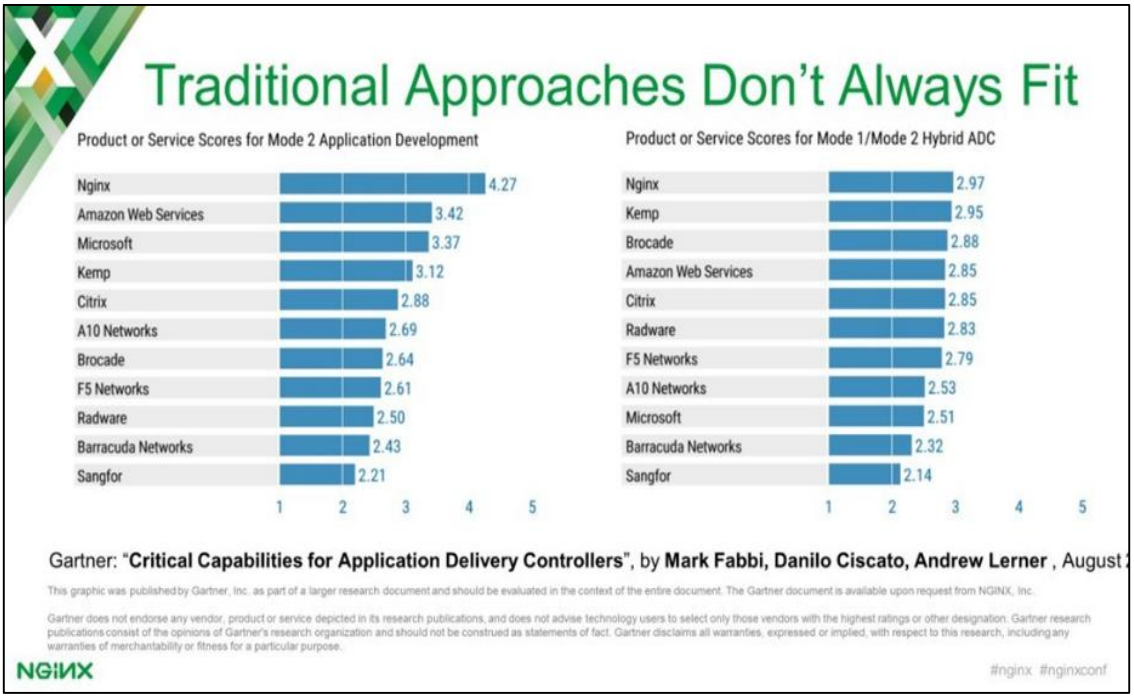

Fig. 2: Traditional Approaches of Data Analytics

To protect your operation, it's crucial that you understand big data security issues before moving forward on a big data implementation. While most professionals have at least a vague notion about the importance of good data security practices, IT and Security professionals are the ones who know the true magnitude of harms done by enterprise security incidents [5].

V. CONCLUSION

More intimidating still, consider the fact that the magnitude of security risks is often proportional to the amount of data that's vulnerable to attack. The advent of big data storage and processing capabilities has introduced implicit risks for big security breaches-Risks that must be secured before a business implements a big data project at scale. In this fourth post of BMC's big data series, you'll get an overview on the main big data security issues you should be aware of. lmost every big data security issue that's common to enterprise-wide implementations can be traced back to design omissions in the original Hadoop distribution. Not that Hadoop's original design was faulty or bad, it just was not designed to be used in an enterprise data environment. Enough time has passed, however, that effective adaptations and solutions have been developed to address these security concerns. The trick is to first understand what the potential weaknesses are, and then verify that you have taken proper precautions to protect those weaknesses. As interest in big data has exploded, organizations have rushed to grab competitive advantage by deploying analytics pipelines that exploit this newly available resource.Many projects have been set up in a "skunkworks" environment, often by data science teams. While this has accelerated the time to market for new features, it has created a potential security nightmare that organizations are gradually waking up to [8].

REFERENCES

[1] G, H., Chen, G., Ooi, B.C., Tan, K.L., Zhang, M.: In-memory big data management and processing: A survey. IEEE Transactions on Knowledge and Data Engineering (2015)
[2] Gandomi A, Haider M. Beyond the hype: big data concepts, methods, and analytics. Int J Inf Manag. 2015; 35(2):137–44.
[3] Cambria E, Rajagopal D, Olsher D, Das D. Big social data analysis. In: Akerkar R, editor. Big Data Computing. Boca Raton, Florida: Chapman and Hall/CRC; 2013. p. 401–14
[4] Ishikawa H. Social big data mining. Boca Raton: Taylor & Francis Group, CRC Press; 2015.
[5] Tole AA, et al. Big data challenges. Database Syst J. 2013; 4(3):31–40.
[6] Manovich L. Trending: the promises and the challenges of big social data. Debates Digit Humanit. 2011; 2:460–75.
[7] Roger Schell Security"A Big Question for Big Data" in 2013 IEEE International Conference on Big Data
[8] Ang Yang,Xianghan Zheng "Type Based Keyword Search For Securing Big Data" in 2013 International Conference On Cloud Computing and Big Data"
[9] Sultana, S., Shehab, M., Bertino, E.: Secure provenance transmission for streaming data. IEEE Trans. Knowl. Data Eng.25 (8), 1890–1903 (2013)
[10] Katina Michael, Keith W. Miller Big Data: New Opportunities and New Challenges Published by the IEEE Computer Society 0018-9162/13/$31.00 © 2013 IEEE
[11] Ueli Huang and Xiaojiang Du"Achieving Big Data Privacy via Hybrid Cloud" in 2014 IEEE INFOCOM Workshops: 2014 IEEE INFOCOM workshop on security and privacy in Big Data

[12] Min Sheng Lin,Chien Yi Chiu,Yuh Jye Lee and Hsing Kuo Pao"Malicious URL Filtering A Big Data Application in 2013 IEEE International Conference on Big Data"
[13] Kogge, P.M., (20-24 May, 2013), big data, deep data, and the effect of system architectures on performance Szczuka, Marcin (24-28 June, 2013)," How deep data becomes big data
[14] Dona Sarkar, Asoke Nath, "Big Data a Pilot Study on Scope and Challenges nternationalJournal of Advance Research in Computer Science and Management Studies (IJARCSMS, ISSN: 2371-7782), Volume 2, Issue 12, Dec-31, Page: 9-19(2014)
[15] Grosso, P; de Laat, C.; Membrey, P, (" Addressing big data issues in Scientific Data Infrastructure")