A Survey on Security And Privacy Of Big Data

[1] Saba Ahmad,[2]Mohd Awais Azam,[3]J.Angelin Blessy
[1][2][3] MTech Computer Science and Engineering Galgotias University

Abstract: In present scenario data is most important part for every organization, company , business, and even for newly startup. Data helps in many ways such as taking important decisions related to their growth. When the data becomes too large, increases beyond the limit and become more complex to handle then we called it BIG DATA. Big data deals with scalability, unique computations, storage and processing challenges. Big data handle the wide amount of data, store it and further process that data in an efficient manner. So that to make important decisions information can be retrieved from the wide amount of data. Even then many companies and organizations are avoiding the use of big data because of the security and some privacy issues that need to be addresses.

Index Terms-Big Data, scalability, security, privacy.

I. INTRODUCTION

The term "Big Data" is used to define enormous volume data, both structured and unstructured in nature. The vast amount of data is impossible to process using traditional software technologies. It requires-wearily parallel software running on tens, hundreds, or even thousands of servers. Data creation and collection quickly exceeds the bound in the digital world of today. The data has been flexure every 2 years since 2011 . It is foreseen that the data will grow 300 times, from 130exabytes in 2005 to 40,000exabytes in 2020.
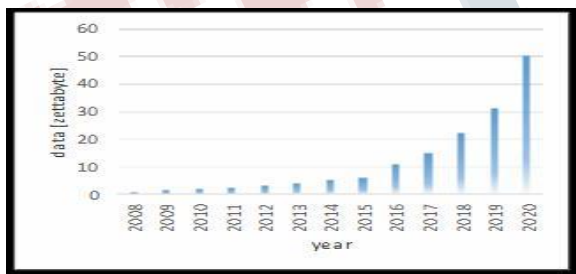

Figure 1. Data generation in various years

As a result of this technological revolution, the big data is becoming progressively a significant issue in the sciences, enterprises and governments.
Big data is a data set, which is tricky to store, share, capture, visualize and analyze on it with current technologies. Even though such difficulties, if you can make out with big data, it furnishes you with generating revenue, better services, executive efficiency, defining needs, strategic decisions, identifying new trends, and developing new products all of which is included in the Data Science. In addition, data science studies parallel and distributed processing, graph analysis, search ranking,

stream processing, clustering, association analysis, machine learning algorithms and dimensionality reduction.

SECURITY AND PRIVACY CHALLENGES PROVOKE BY 5Vs:

In this section, the discussion starts with understanding the impact of big data characteristics on privacy and security.
According to the principle and definition of big data [6], the characteristics of big data are classified into "5Vs", i.e Volume, Velocity, Value, Veracity, and Variety.
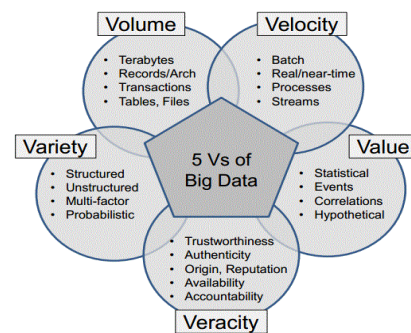

Figure 2. 5vs of Big Data

"Volume" refers to the size of data. There is enormous amount of data generated by sensors, organizations and individuals every second in every field. It is closely impossible for data providers to control supervise all the data they "actively" or "passively" provide to others. By using these data any one's behavior or identification can be predicted, which may further derive the individual privacy.

Accordingly, the vast volume of data increases the risk of information leakage. Besides, existing infrastructure strategies, such as monitoring, regular tracking, security scanning technology or auditing are not sufficient any more. Because it is costly and complicated to implement those methods on the large-scale big data scenario [7].

2) "Velocity" shows the high frequencies and continuousness of data. Velocity is a combined data management process and data infrastructure that addresses different affair that are visible after the addition and creation of big data objects. Velocity is directly related to the integrated data infrastructure and architecture in delivering and managing data to recipients as quickly as possible. Fast generating and iterating data needs non relational databases, thus distributed programming frameworks should have been developing with privacy and security in mind [8]. Besides, the hacker can easily launch advanced persistent threats (APT), while it is hard to be detected by the traditional protection strategy.

3) "Value" refers to the outputs that achieved from vast amount of data sets. The highly intensely integrated data and potential value attracts hacker [9]. Hackers can get more sensitive information and large number of data sets by successfully attacking the database, and thus the cost of the attacked is decreased.

4) "Veracity" refers to the applicability, bias, trustworthiness, noise, abnormality and other quality properties of data [10].
In addition to the rapidly growth in varieties and velocities of data, flow of data can be highly inconsistent with periodic peaks. Daily, seasonal and event-triggered peak data loads can be challenging to handle. Even more so with the unstructured data involved. [11]

5) "Variety" refers to the diversity of data formats and sources. The data formats consist of structured, unstructured and semi-structured ones, while the file type includes figures, texts and videos. Large-scaled cloud infrastructure, as an alternate method to handle and store data, makes the traditional storage and management measures invalid [12]. Consequently, not only the infrastructure security facing massive data up to PB (Petabyte) level, but also data management methods addressing data provenance, needs to be considered.

APPLICATIONS OF BIG DATA:

1. IMPROVING SECURITY AND LAW ENFORCEMENT:
Big data helps in improving the various security features. As we all know that our Indian Government uses the data to track criminal record of & mitigate crime [13], for such policies they start making Adhar card. By that Adhar card the government can easily get the information like fingerprint and other personal details of each citizen, so they can easily control the crime & make our country safer.

2. Understanding And Targeting Customer in e-commerce.
E-commerce is a very huge and main area of Big data use today. Big Data helps in understanding the customer needs and requirements by e-commerce. It also help for address the customer for which organizations are going to make the products.

3. IMPROVING SCIENCE AND RESEARCH:
Big Data bring the numerous new possibilities by which science and research is presently transforming which is very helpful to invent the new things. Before that it was very difficult to process and analysis the vast amount of data of different formats but now with the help of Big Data it is easy to work on this [14].
For example-CERN the nuclear physics lab with its vast Hadron collider, the world's largest and most powerful particle accelerator. Experiments to unlock the secrets of our world, how it started and works-generate vast amount of data. The CERN data centers has 65,000 processors to analyze its 30 petabytes of data [15].

4. UNDERSTANDING AND OPTIMIZING BUSINESS;
It is very important to understand the requirement of the business and then optimize the process so that the production and profit both can be increased by using technologies. The process of production how much they require to manufacture and how much they require to cut the production. All these decisions are made basis on the previous data. Also analyze the supply chain & use the sensor track goods & delivery vehicles and optimizing the route by using live traffic data [16].

REASONS TO USE BIG DATA:

It manage a petabyte of data or more.
It has distributed redundant data storage.
Big Data has extremely fast data insertion.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

Big Data also has orchestration and central management.
Is hardware agnostic.
Can leverage parallel task processing.
Is extensible where its basic capabilities can be altered and augmented.

BIG DATA SECURITY CHALLENGES:
Non-relational databases (No SQL) are severely emerging, making it hard for safety answers to preserve up with name for.
When a machine receives a huge quantity of facts, it has to be validated to stay accurate and honest; this workout doesn't continuously get up, however.
Most distributed systems computations have mild an unmarried level of protection, which isn't always endorsed.
Automated data switch needs more protection capabilities, which might be frequently not to be had.
Due to the size of the Big Data, it's origin aren't continuously tracked and monitored [17].
Suggested positive audits aren't robotically done on Big Data because of the enormous quantity of statistics concerned.

HOW CAN BIG DATA SECURITY BE IMPROVED
Cloud Computing knowledge experts are given as genuine with that most affordable way to brighten the security of Big Data is through the chronic increase of the antivirus agency. A multimedia of antivirus organizations, imparting a spread of solutions, offers a better protection opposition to Big Data protection threats. Refreshingly, the antivirus corporation is continuously touted for its openness. Big Data protection threats, and agency leaders continuously paintings collectively to handle with new malicious software application attacks, providing more gains in Big Data protection.
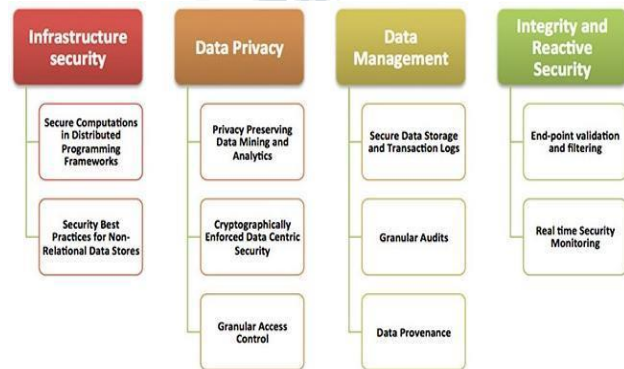

Fig 1.3 Big data security challenges.

These are the some extra pointers to reinforce-Big Data security;
Should take care of software protection, in location of tool safety.
Isolate gadgets and servers containing important facts.
Introduce real-time security data and event control.
Provide proactive and reactive protection [19].

COMPARISON OF SURVEYS ON SECURITY ISSUES IN BIG DATA

Research	Topics		
	Infrastructure Security	Data Privacy	Data Management
[10], [13]		✓	✓
[5],[14]	✓	✓	✓
[22]	✓	✓	
[23]		✓	✓
[24] ,[25]		✓	
[26],[27]		✓	
[28]		✓	

EXISTING PRIVACY PRESERVING TECHNIQUES

Research	Big Data Process Layer/Interface			
	Data Collecting	Data Storing	Data Mining	Data Publishing
[10], [13]	✓	✓	✓	✓
[5],[14]	✓	✓	✓	✓
[22]	✓	✓		
[23]	✓	✓		
[24] ,[25]			✓	
[26],[27]				✓

Privacy preserving aggregation
Operations over encrypted data
De-identification
Privacy-preserving big data analytics is still challenging due to either the issues of flexibility and efficiency or re

identification Risks. However, compared with privacy-preserving aggregation and operations over encrypted data, de-identification is more feasible for privacy-preserving big data analytics if we can develop efficient and privacy-preserving algorithms to help mitigate the risk of re-identification. With these two points in mind, future research work on big data privacy should be directed toward efficient and privacy preserving computing algorithms in the big data, and these algorithms should be efficiently and output correct results while hiding raw individual data. In such a way, they can reduce the re-identification risk in big data analytics and mining.

CONCLUSION:

In this paper, discussion initiated from the basics, security and privacy challenges provoke by 5Vs of big data. All the Application areas, where the big data is really helping in this digital world is discussed. In this paper how can we improved security for big data is discussed.

REFERENCES

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.

T. Vijey, A. Aiiad, "Big Data Security Issues Based on Quantum Cryptography and Privacy with Authentication for Mobile Data Center", Procedia Computer Science, vol. 50, pp. 149–156, 2015.

B. Matturdi, X. Zhou, S. Li, F. Lin, "Big Data security and privacy: A review", Big Data, Cloud & Mobile Computing, China Communications vol.11, issue: 14, pp. 135 – 145, 2014.

C.L.P. Chen, C.Y. Zhang, "Data Intensive applications, challenges, techniques and technologies: A survey on Big Data", Information Sciences, vol. 275, pp . 314-347, 2014.

D. S. Terzi, R. Terzi, et al., A Survey on Security and Privacy Issues in Big Data[C], in Proc. 2015 IEEE International Conference on Internet Technology and Secured Transactions(ICITST' 2015), 2015. N. Miloslavskaya, M. Senatorov, A. Tolstoy, S. Zapechnikov, "Information Security Maintenance Issues for Big Security-Related Data", Future Internet of Things and Cloud (FiCloud), pp. 361 – 366, Barcelona, 2014.

B. Saraladevi, N. Pazhaniraja, P. Victer Paul, M.S. Saleem Basha, P. Dhavachelvan, "Big Data and Hadoop-A Study in Security Perspective", Procedia Computer Science, vol. 50, pp. 596 – 601, 2015.

B. Matturdi, X. Zhou, et al., Big Data Security and Privacy: A Review [J], China Communications Magazine, 2014.

NIST, NIST Big Data Interoperability Framework: Volume 4, Security and Privacy[R], National Institute for Standards and Technology, 2015, http://dx.doi.org/10.6028/NIST.SP.1500-4. B. Matturdi, X. Zhou, et al., Big Data Security and Privacy: A Review[J],China Communications Magazine, 2014. M.Yang, X.Zhou, et al., Challenges and Solutions of Information Security Issues in the Age of Big Data[J]. China Communications Magazine, Mar. 2016. ISO/IEC JTC 1, Big Data [R], Preliminary Report, 2014. K. Hu, D. Liu, et al., Research on Security Connotation and Response Strategies for Big Data[J], Telecommunications Science, vol.2, pp.112-117, Feb. 2014.

M.Yang, X.Zhou, et al., Challenges and Solutions of Information Security Issues in the Age of Big Data[J]. China Communications Magazine,Mar. 2016.

Raghav Toshniwal* Kanishka Ghosh Dastidar Asoke Nath Big Data Security Issues and Challenges International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Issue 2, Volume 2 (February 2015).

D. Mittal, D. Kaur, et al., Secure Data Mining in Cloud Using Homomorphic Encryption[C], in Proc. 2014 IEEE International Conference on Cloud Computing in Emerging Markets(CCEM' 2014), Oct. 2014.

C. Wang N. Cao J. Li K. Ren W. Lou "Secure ranked keyword search over encrypted cloud data" Proc. Of ICDCS' 102010.

L. Ballard S. Kamara F. Monrose "Achieving efficient conjunctive keyword searches over encrypted data" Proc. of ICICS2005.

D. Boneh B. Waters "Conjunctive subset and range queries on encrypted data" Proc. of TCC pp. 535-554 2007. A. Lewko T. Okamoto A. Sahai K. Takashima B. Waters "Fully secure functional encryption: Attribute-based encryption and (hierarchical) inner product encryption" Proc. Of EUROCRYPT 2010. Data Center Knowledge. 2017. Nine Main Challenges in Big Data Security | Data Center Knowledge. [ONLINE] Available at:

http://www.datacenterknowledge.com/archives/2016/01/1 9/nine-main-challenges-big-data-security/. [Accessed 21 February 2017].

E. Sahafizadeh, and M.A.Nematbakhsh, A Survey on Security Issues in Big Data and NoSQL[J]. Advances in Computer Science: an International Journal(ACSIJ), vol.4, no.16, pp.68-72, Jul. 2015. E. Bertino, Big Data-

Security and Privacy[C]. in Proc. 2015 IEEE International Congress on Big Data, Jun. 2015.

R. Lu, H. Zhu, et al., Toward Efficient and Privacy-Preserving Computing in Big Data Era[J], IEEE Network, Aug. 2014.

K. Saranya, K. Premalatha, et al., A Survey on Privacy Preserving Data Mining[C], in Proc. 2015 IEEE Sponsored 2nd International Conference on Electronics and Communication System(ICECS'15), 2015.

B. C. Fung, K. Wang, et al, Privacy-Preserving Data Publishing: A Survey of Recent developments[J], ACM Comput.Surv., vol. 42, no.4, Jun. 2010.

J. Abawajy, M. I. Ninggal, et al., Privacy Preserving Social Network Data Publication[J], IEEE Comm. Surveys Tutorials, vol. pp, no.x, Mar. 2016.

A. H. Rashid, and N. Yasin Privacy Preserving Data Publishing: Review[J], International Journal of Physical Sciences, vol.10 pp. 239-247, Apr. 2015.