International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V69I6P109

Volume 69 Issue 6, 57-60, June 2021

Original Article

A Survey on 7V's of Big Data Security and Privacy

Reshma[1], Sanath Kumar[2], Mithun S[3]

[1]*Programmer, Department of Computer Science, M.G.M College, Udupi, Karnataka, India.*
[2]*Programmer, Department of Computer Science, M.G.M College, Udupi, Karnataka, India.*
[3]*System Administrator, Department of Computer Science, M.G.M College, Udupi, Karnataka, India.*

Received Date: 08 May 2021
Revised Date: 12 June 2021
Accepted Date: 23 June 2021

Abstract *- Big data plays important role in present day every organization, company, business, and even for newly startup companies. Big Data helps to create a new growth opportunities for company, organization and business etc. hence in the past many years data producing large-scale. This Big data is highly scalable, storage and processing of data with unique mathematical calculation. Big data handle the vast amount of data, process and storage of that data is efficient. During the process of data collection, storage, personal information will be leaked hence security and privacy of the data is more important than everything. This paper discuss about the big data, its character, security and privacy of big data Then, we present some possible solution to Big data security and privacy.*

Keywords *- Big data, Database, Data privacy, Scalable, Security.*

I. INTRODUCTION

The term Big Data is a process of collecting data. That data volume is huge, it grows exponentially with time. This data size is large and more complex because of this traditional method cannot store or process data efficiently [1].So that advanced tools are used to store the data and processing the data. In every sector work becomes digitalized. Present Pandemic situation digital work getting more important because of these reason Data creation and collection quickly exceeds the bound with time. The data has been expanding every 2 years since 2011 [12].

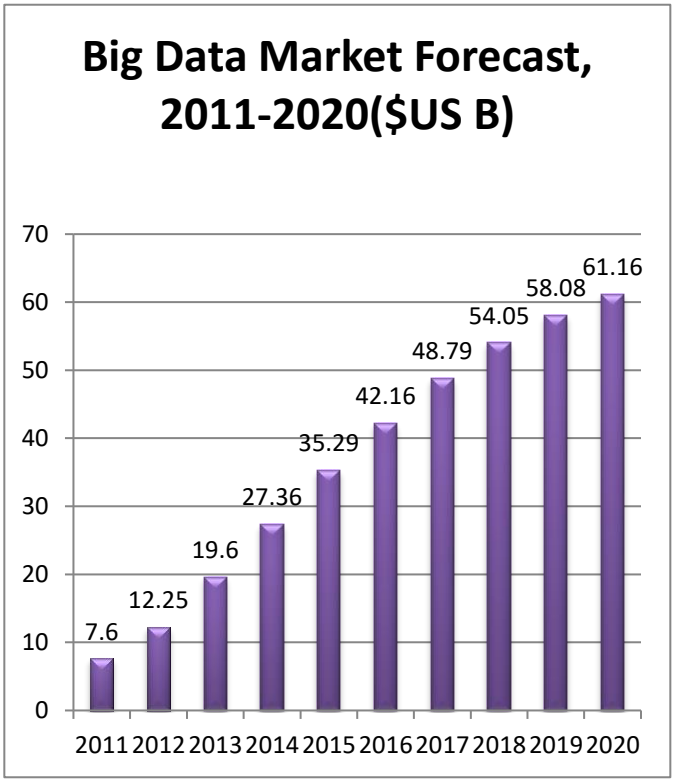

Fig. 1 Worldwide data analysis market review on market revenue segment forecast 2011-2020

This graph show Big Data applications and analytics is estimated to grow from $7.6 in 2011 to $61.16B in 2020, attaining a CAGR of 26 %. Big Data market worldwide includes all the data warehouse, Industries, and Services which is projected to grow from $35.29 in 2015 to $61 in 2020 [5].

Table 1. Information contained in big data market forecast

Information stored and collected by end Users	Provided information by third parties
• Volume of the data. • Storage Capacity of the data warehouse. • Application and Analysis of the data. • Computing the data.	• Raw data providers • Professional Services information taken by the organization, companies or business. • The Strength of Network Security

II. CHALLENGES OF SECURITY AND PRIVACY IN 7Vs

Here we discuss the effect of big data characteristics on security and privacy. According to the definition of big data [1], big data characteristics classified into "7Vs", i.e. Volume, Value, Velocity, Veracity, and Variety, visualization, and Value.

Table 2. Features Of 7V's

Application	Features
SECURITY AND PRIVACY 7 V's	• Volume • Value • Veracity • Variability • Variety • Velocity • Visualization

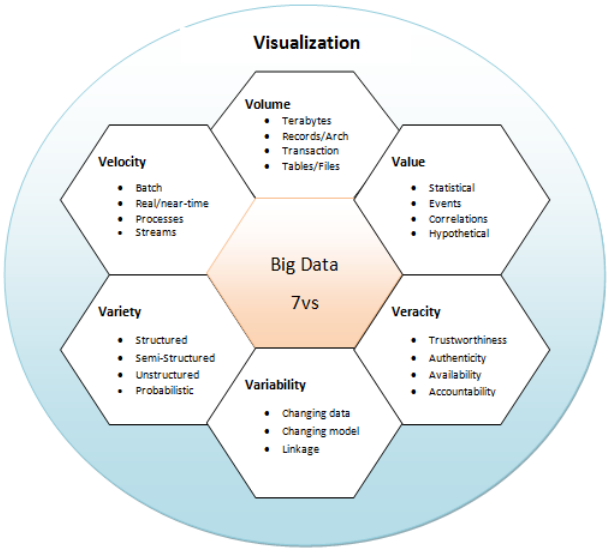

Fig. 2 Big data characteristics

A. Volume

The term "volume" is a size of the data. This size is unimaginable and to calculate this data we need a unfamiliar numerical terms. Doug Laney says In every day, world produces 2.5 quintillion bytes of data that is 2.3 trillion gigabytes. These amounts of huge data generated in every second by different fields [7]. For the data provider it difficult to handle these data. By using these data we can predict the behavior or identification of the person, this prediction help to get individual privacy. As the volume of data increases risk of information leakage also high. Some of method such as monitoring, regular tracking etc are not enough. When the data is vast amount to implement of that data is complicate [12].

B. Velocity

The term Velocity defines how quickly and the speed of the data increase in volume that can be accessed and processed. Considering the example, social media posts, audio files, YouTube videos, images that are uploaded in every milliseconds, so that data should be accessible as early as possible [6]. The data are warehoused before analysis, hence need for real-time processing of these enormous volumes increase, such as the millions e-mail, billions of tweets and number of streaming videos in YouTube platform, that are the video are uploaded every minute in the day. Real-time processing gives less storage requirements it provides more responses, it gives genuine accurate and profitable responses [8], [7]. Fast growing and iterating data requires non-relational databases, thus distributed programming frameworks should have been developing in a manner where security and privacy are kept in mind. Besides, the hacker can launch Advanced Persistent Threats(APT) more easily, the APT are very hard to find and resolve the threats in a traditional protection strategy [12],[11].

C. Variety

Variety means types of data sources and format. Data format may be structured, semi-structured, and unstructured [11]. In the present situation, data generated in big really big quantities. These data unstructured, this kind of unstructured data may be video, audio, images, text files. These data consisting of natural language, hash tags, geographical data, multimedia, sensor events and many more, the fetching of meaning from such diversity difficult[6].

D. Variability

Variability refers that change in data rapidly. Variability mainly focuses about understanding and interpreting the correct meanings of raw data. This property has become challenging because usage of digital media increases, which is the main reason for peak in data loads. The boundless variability of Big Data presents a different way of encrypt challenge if one is to take advantage of the data value fully. [6], [7].

E. Veracity

"Veracity" refers to applicability, trustworthiness abnormality and other quality properties of data, if the data being analyzed are inaccurate or incomplete then it is useless. This kind of situation comes from where data streams originate from different sources presenting a variety of formats with differing multimedia signal-to-noise ratios. Within the period of time data arrives at Data analysis stage, where the data are kept and gather together errors that are difficult to sort out. Veracity of the final analysis is filtered without first cleaning up the data it works with [11], [7].

F. Visualization

In Big Data processing system, a core task is to change the vast scale of it into something easily under stable and actionable. The best and easiest way to understand and interpret is converting it into graphical formats. Graphical formats are easily understandable. However, due to the attributes of velocity and variety Spreadsheets and even three-dimensional visualizations are often not up to the task. There may be a multiple of spatial and temporal parameters and relationships between them to condense into visual forms. Solving these problems is the main force behind AT&T's Nanocubes visual representation package [7].

G. Value

It is final stage of big data. In this stage tsunami of data transform into business. This data easily attracts by hacker. Hacker can get a sensitive information by hack the database ,and cost of the hack is decreased for hacker. There are lots of effort, resources used in above 6V's hence every user sure that organization get a value form data or not [6],[12].

III. FIVE BIG DATA SECURITY CHALLENGES:

Here we discussed five challenges of big data.

A. Framework Distribution

For the faster analysis big data frameworks distribute data processing task in many system, this also helps to get accurate result. Hadoop, is a popular open-source framework. It's used in distributed data processing and storage [2].

B. High Speed of NoSQL Databases

Row and column of tabular schema method used in Traditional relational databases, hence These Traditional relational databases difficult to handle the big data because of its high scalability and variety in structure, more complexity. To overcome of these limitations on-relational databases designed. On-relational databases, also known as NoSQL databases

Tabular schema method not used in Non-relational databases. It computed based on the type of data. Because of these reason Non-relational databases much better than Traditional relational database. Non-relational databases flexible than Traditional relational database so that Organizations use Non-relational databases but they have to establish database with additional security measures and trusted environment [2].

C. Sensitive Data Mining

Data mining is most import in big data. To find the pattern in unstructured data some data mining tool are used. big data contain a personal and financial detail .hence data mining so that, to protect from external and internal threats companies need to add extra security layers [2].

D. Vulnerabilities in Endpoint

At endpoint devices Cybercriminals can manipulate data and there are transmitting that manipulates data to data lakes. Consider example, Cybercrimals can hack manufacturing systems that system detect malfunctions using sensor. Once access done , Cybercrimals shows fake results in sensor. We can solved this problem with fraud detection technologies [2].

E. Struggles of Access Controls

Companies or organization are made restriction to access of sensitive data. Consider a example in medical field that records include personal information. People like medical researchers they need to use this data still do not have access permission. For that in many organizations allow granular access. granular access means that individuals have permission for see that needful details. granular access won't work in Big data technologies. [2].

IV. SOLUTIONS TO MAJOR BIG DATA SECURITY CHALLENGES

In this paper we listed some general security :

A. Check out the Cloud Providers

If the big data stored in cloud you should be aware of the providers .whether they have sufficient authenticity security tools and also check the providers whether they have regular security audits and when adequate security standards are not met by provider on that situation they agree to give penalties [10],[9].

B. Secure your Sensitive Data

You have to protect you primary data and also data that come at end of analytics. For the sensitive data you have to use encryption so that no information leakage problem happen [9].

C. Should Create a Appropriate Access Control Policy

Create a policy such that policy do not allow access to unauthorized users. Also that unauthorized person cannot access the data from internal sources and external sources. only authorized person can access the data[9] ,[10].

D. Communications Protection

Data in transit should be adequately protected .You have to sure about that data is confidential and integrity [9].

E. Monitor your Real-time Safety

Aware of your data .Only a authorized person can monitor the data and access to the data. Threat intelligence should be used to prevent unauthorized access so that unauthorized person not access any data whether it internal or external source [9].

V. BIG DATA SECURITY TOOL

In company, organization or any other field to protect their data they use many security tools. But In this paper we are discuss about the encryption. Encryption is simple tools It is a used to protect sensitive information. It encrypt the data when data encrypted that data is fruitless for cybercriminals. They don't get a key to unlock it. Encryption also is critical for many modern network applications, such as ensuring the security of most WIFI and intranet connections of data technology [3].

VI. CONCLUSION

In the present day technology the big data is a significant issue for company and other field because of technological revolution. In our paper, the discussion is about challenges of security and privacy in 7Vs.Then challenges of big data physically and also data stored in cloud and we listed some of general security solution ..There are many of solution for the technical purpose but in our paper we have listed out the explanation of importance of encryption role in big data.

REFERENCES

[1] https://www.guru99.com/what-is-big-data.html retrieve on 14 April 2021.

[2] https://www.dataversity.net/big-data-security-challenges-and-solutionsretrieve on 21 April 2021.

[3] https://www.sisense.com/glossary/big-data-security/retrieve on 2 may 2021.

[4] https://impact.com/marketing-intelligence/7-vs-big-data/ retrieve on 2 may 2021.

[5] https://www.kdnuggets.com/2015/04/wikibon-big-data-market-forecast-2020.htmlretrieve on 5 may 2021.

[6] https://www.yourtechdiet.com/blogs/7vs-big-data/#:~:text=The%20term%20volume%20here%20defines%20big%20data%20as%20%E2%80%9CBIG%E2%80%9D.&text=Because%20of%20this%2C%20now%20the,being%20generated%20on%20YouTube%20itself.retrieve on 6 may 2021.

[7] https://bigdatapath.wordpress.com/2019/11/13/understanding-the-7-vs-of-big-data/ retrieve on 6 may 2021.

[8] https://dataconomy.com/2014/05/seven-vs-big-data/ retrieve on 14 may 2021.

[9] https://www.newgenapps.com/blog/big-data-security-challenges-solutions-problems-security/retrieve on 17 may 2021.

[10] https://www.f-secure.com/en/consulting/our-thinking/big-data-security-challenges-and-solutions retrieve on 17 may 2021.

[11] Haina Ye, Xinzhou Cheng, Mingqiang Yuan, LexiXu, JieGao, and Chen Cheng, A Survey of Security and Privacy in Big Data, IEEE, November 2016 (978-1-5090-4100-8).

[12] Saba Ahmad, Mohd Awais Azam, J.Angelin Blessy,A Survey on Security And Privacy Of BigData, International Journal of Engineering Research in Computer Science and Engineering (IJERCSE), 5(4)(2018).

[13] Surendiran,R., Secure Software Framework for Process Improvement. SSRG International Journal of Computer Science and Engineering (IJCSE), l3(12) (2016) 19-25. ISSN: 2348 – 8387.