

# Security and Privacy Challenges in Big Data Environment

Dr. Alok Singh Chauhan

**Abstract** — Big Data and its applications, advantages are likely to represent the most promising new frontiers in science. Big Data applications are an incredible advantage to associations, business, organizations and numerous huge scale and little scale ventures. Big Data stays a standout amongst the most discussed innovation inclines in 2013. In any case, lost among all the energy about the capability of Big Data are the genuine security and protection challenges that take steps to moderate this force. Security and privacy issues are magnified by the three V's of big data: Velocity, Volume, and Variety. These components incorporate factors, for example, enormous scale cloud frameworks, assorted variety of information sources and configurations, gushing nature of information securing and the inexorably high volume of between cloud movements. In this paper, an attempt to study and reviewed the big data security & privacy challenges and show the effect of these challenges. This paper also describes some solutions and practices to overcome these challenges.

**Index Terms** — Big Data, 3V, Security & Privacy Challenges

## I. INTRODUCTION

Big Data is a gathering of informational indexes so huge and complex that it winds up hard to process utilizing close by database the executives apparatuses or conventional information handling applications. Big Data is information that surpasses the handling limit of regular database frameworks. The information is too huge, moves excessively quick, or does not fit the structures of your database designs. To pick up an incentive from this information, you should pick an elective method to process it. [14]



Figure 1: An Artificial view of Big Data

Big Data implies extremely a Big Data; it is an accumulation of huge datasets that can't be prepared utilizing customary registering strategies. Huge information isn't simply information; rather it has turned into a total subject, which includes different apparatuses, methods and structures. Thus, Big Data includes huge volume, high velocity and extensible variety of data. The data in it will be of three types. [16, 17]

1. Structured Data: Relational data.
2. Semi Structured Data: XML data.
3. Unstructured Data: Word, PDF, Text, Media Logs.

There are some common characteristics of big data [20], such as:

1. Big data integrates both structured and unstructured data.
2. Addresses speed and scalability, mobility and security, flexibility and stability.
3. In Big Data the affirmation time to information is essential to expel a motivation from various data sources, including mobile phones, radio repeat unmistakable confirmation, the web and a creating once-over of motorized substantial

advances.

## II. BIG DATA BACKGROUND

Big Data is the word used to describe massive volumes of structured and unstructured data that are so large that it is very difficult to process this data using traditional databases and software technologies. The term "Big Data" is companies who had to query loosely structured very large distributed data. The three main terms that signify Big Data have the following properties:

- 1. Volume:** Many factors contribute towards increasing Volume streaming data and data collected from sensors etc.
- 2. Variety:** Today data comes in all types of formats emails, video, audio, transactions etc.
- 3. Velocity:** This means how fast the data is being produced and how fast the data needs to be processed to meet the demand.



Figure 2: IBM 3 V's of Big Data

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

**4. Variability:** Along with the Velocity, the data flow can be highly inconsistency with periodic peaks.

**5. Complexity:** Multifaceted nature of the information additionally should be viewed as when the information is originating from different sources. The information must be connected, coordinated, scrubbed and changed into required configurations before genuine preparing.

## III. WHAT COMES UNDER BIG DATA?

Big Data includes the information delivered by various gadgets and applications. Given underneath are a

portion of the fields that go under the umbrella of Big Data. [21]

**1. Black Box Data:** It is a component of helicopter, airplanes, and jets, etc. It catches voices of the flight team, accounts of mouthpieces and headphones, and the presentation data of the air ship.

**2. Social Media Data:** Web based life, for example, Facebook and Twitter hold data and the perspectives posted by a large number of individuals over the globe.

**3. Power Grid Data:** The power grid information holds data devoured by a specific hub concerning a base station.

**4. Transport Data:** Transport information incorporates model, limit, separation and accessibility of a vehicle.

**5. Stock Exchange Data:** The stock trade information holds data about the 'purchase' and 'sell' choices made on an offer of various organizations made by the clients.

**6. Search Engine Data:** Search engines recover bunches of information from various databases.

#### IV. BIG DATA SECURITY & PRIVACY CHALLENGES

These days, associations are gathering and handling gigantic measures of data. The more information is put away, the more imperative it is to guarantee its security. An absence of information security can prompt incredible monetary misfortunes and reputational harm for an organization. To the extent Big Data is concerned, misfortunes because of poor IT security can surpass even the most noticeably terrible desires.

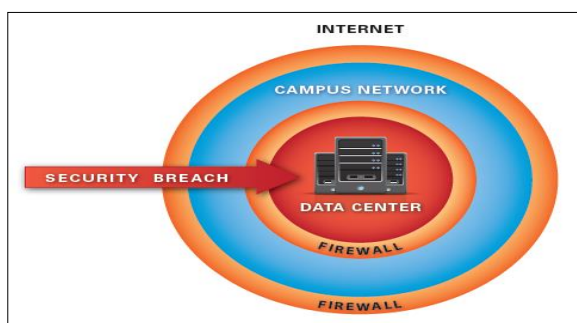


Figure 3: Security Breach in Big Data or Data Centre

Some main challenges are as follows:

1. The greatest test for Big Data from a security perspective is the insurance of client's protection. Big Data oftentimes contains gigantic measures of individual recognizable data and in this manner security of clients is a tremendous concern.
2. When delivering data for enormous information, associations need to guarantee that they have the correct harmony between utility of the information and protection.
3. While utilizing Big Data a critical test is the way to set up responsibility for. On the off chance that the information is put away in the cloud a trust limit ought to be build up between the information proprietors and the information stockpiling proprietors.

4. Most circulated frameworks' calculations have just a solitary dimension of assurance, which isn't suggested.
5. Unethical IT experts rehearsing data mining can assemble individual information without approaching clients for consent or advising them.
6. Access control encryption and associations security can wind up dated and out of reach to the IT pros who depend on it.
7. Recommended nitty gritty reviews are not routinely performed on Big Data because of the tremendous measure of data included.
8. Due to the measure of Big Data, its causes are not reliably observed and followed.
9. When a framework gets a lot of data, it ought to be approved to stay dependable and exact; this training doesn't generally happen, be that as it may.
10. Automated information exchange requires extra safety efforts, which are regularly not accessible.

#### V. SOLUTION TO ENSURE THE BIG DATA SECURITY AND PRIVACY

1. The most significant answer for guaranteeing that information stays secured is the satisfactory utilization of encryption. For instance, Attribute-Based Encryption can help in giving fine-grained get to control of scrambled information.
2. Anonymizing the information is likewise imperative to guarantee that protection concerns are tended to. It ought to be guaranteed that all touchy data is expelled from the arrangement of records gathered.
3. Real-time security checking is additionally a key security part for major information and its task. It is significant that associations screen access to guarantee that there is no unapproved get to.
4. It is additionally significant that risk insight is set up to guarantee that increasingly advanced assaults are distinguished and that the associations can respond to dangers as needs be.

#### VI. BEST PRACTICES FOR BIG DATA SECURITY

Big Data is a sensibly new idea and along these lines there isn't a rundown of best practices yet that are extensively perceived by the security network. Anyway there are various security suggestions that can be connected to huge information:

1. **Vet your cloud suppliers:** If you are putting away your enormous information in the cloud, you should guarantee that your supplier has sufficient assurance components set up. Ensure that the supplier completes occasional security reviews and concur punishments on the off chance that those satisfactory security norms are not met.
2. **Create an adequate access control policy:** Create arrangements that enable access to approved clients as it were.

3. **Protect the data:** Both the crude information and the result from examination ought to be enough ensured. Encryption ought to be utilized as needs be to guarantee no delicate information is spilled.
4. **Protect communications:** Data in travel ought to be enough secured to guarantee its secrecy and trustworthiness.
5. **Use real-time security monitoring:** Access to the information ought to be observed. Risk insight ought to be utilized to counteract unapproved access to the information.

#### VII. BEST PRACTICES FOR BIG DATA PRIVACY

The prescribed procedures for working with huge information are as yet developing; however there are beforehand exercises that can help move this guarantee of advancement forward without relinquishing the protection of individual information.

1. The initial phase in powerful endeavor of huge information is to turn out to be exceptionally skilled in obtaining and overseeing cloud administrations, which are viewed as an essential for enormous information to be financially savvy. There must likewise be continuous observing and reviews of cloud benefits alongside any important measurements that show dimensions of information respectability, secrecy and accessibility.
2. The next strategy to empower better utilization of huge information is to actualize combined capacity. Combined capacity is increasingly effective and will decrease the probability of mistakes that impact information quality or exactness.
3. Another best practice is to appropriately disinfect information, as it keeps away from some of the previously mentioned protection issues.

#### CONCLUSION

We have entered an era of Big Data. Big Data depicts an all encompassing data the executive's procedure that incorporates and coordinates numerous new kinds of information and information the executives close by conventional information. Huge information can be investigated for bits of knowledge that lead to better choices. Through better examination of the huge volumes of information that are getting to be accessible, there is the potential for making quicker advances in numerous logical teaches and improving the benefit and achievement of numerous endeavors. However, security and privacy challenges described in this paper must be addressed before this potential can be realized fully and to make Big Data processing and computing infrastructure more secure. Our hope is that this paper will spur action in the research and development community to focus on the barriers to greater security and privacy in Big Data platforms.

#### REFERENCES

- [1] Gartner, <http://www.gartner.com/it-glossary/bigdata>.
- [2] Big Data and Current Cloud Computing Issues and Challenges, Volume 4, Issue 6, June 2014 ISSN: 2277 128X, [www.ijarcsse.com](http://www.ijarcsse.com)
- [3] A Review on the Role of Big Data in Business, IJCSMC, Vol. 3, Issue. 4, April 2014, pg.446 – 453 ISSN 2320-088X
- [4] BIG DATA: A Review, IJCSMC, Vol. 3, Issue. 7, July 2014, pg.106 – 110, ISSN 2320-088Xs
- [5] Russom, "Big Data Analytics", TDWI Research, 2011.
- [6] Sawant, Nitin, and Himanshu Shah. "Big Data Application Architecture." Big Data Application Architecture Q & A. Apress, 2013. 9-28.
- [7] Chen, H., Chaing, R.H.L. and Storey, V.C. (2012) Business Intelligence and Analytics: From Big Data to Big Impact, MIS Quarterly, 36, 4, pp. 1165-1188
- [8] V. Gopalkrishnan, D. Steier, H. Lewis, and J. Guszczka. Big data, big business: bridging the gap. In Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, Big-Mine '12, pages 7{11, New York, NY, USA, 2012. ACM.
- [9] C. Parker. Unexpected challenges in large scale machine learning. In Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, BigMine '12, pages 1-6, New York, NY, USA, 2012. ACM.
- [10] J. Gama. Knowledge Discovery from Data Streams. Chapman & Hall/Crc Data Mining and Knowledge Discovery. Taylor & Francis Group, 2010.
- [11] J. Gantz and D. Reinsel. IDC: The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. December 2012.
- [12] <https://cloudsecurityalliance.org/media/news/csa-releases-the-expanded-top-ten-big-data-security-privacy-challenges/>
- [13] <http://www.datacenterknowledge.com/archives/2016/01/19/nine-main-challenges-big-data-security/>
- [14] <https://www.bartleby.com/essay/Big-Data-Processing-Capacity-Of-Conventional-Database-PKHRHLN9MWW>
- [15] <http://datumresource.com>
- [16] <https://matrumandirit.blogspot.com/2017/01/big-data-introduction-part-2.html>
- [17] [https://www.tutorialspoint.com/hadoop/hadoop\\_big\\_data\\_overview.htm](https://www.tutorialspoint.com/hadoop/hadoop_big_data_overview.htm)
- [18] <https://blog.cloudsecurityalliance.org/big-data/>
- [19] <https://www.scribd.com/document/317109557/laetsd-Mapreduce-Streaming-Over-Cassandra-Datasets>
- [20] [https://www.researchgate.net/publication/276199542\\_Security\\_Issues\\_Associated\\_with\\_Big\\_Data\\_in\\_Cloud\\_Computing](https://www.researchgate.net/publication/276199542_Security_Issues_Associated_with_Big_Data_in_Cloud_Computing)
- [21] <https://www.quora.com/What-is-the-difference-between-Hadoop-and-big-data>
- [22] <https://blog.cloudsecurityalliance.org/big-data/>
- [23] <https://cloudsecurityalliance.org/articles/csa-releases-the-expanded-top-ten-big-data-security-privacy-challenges/>
- [24] <http://javatechnocrat.com/education.php>
- [25] <https://www.mwrinfosecurity.com/our-thinking/big-data-security-challenges-and-solutions/>
- [26] <https://www.datacenterknowledge.com/archives/2016/01/19/nine-main-challenges-big-data-security>
- [27] <http://community.mis.temple.edu/mis520817/2017/03/29/nine-main-challenges-in-big-data-security/>
- [28] [https://www.tutorialspoint.com/hadoop/pdf/hadoop\\_big\\_data\\_overview.pdf](https://www.tutorialspoint.com/hadoop/pdf/hadoop_big_data_overview.pdf)
- [29] <http://www.ijcsit.com/docs/aceit-conference-2016/aceit201611.pdf>
- [30] <http://iosrjournals.org/iosr-jce/papers/Vol18-issue6/Version-4/C1806041318.pdf>
- [31] [https://www.researchgate.net/publication/276199542\\_Security\\_Issues\\_Associated\\_with\\_Big\\_Data\\_in\\_Cloud\\_Computing](https://www.researchgate.net/publication/276199542_Security_Issues_Associated_with_Big_Data_in_Cloud_Computing)

AUTHORS' DETAIL

---

**Dr. Alok Singh Chauhan**

Assistant Professor,  
IT Department, IMS, Ghaziabad (University Courses  
Campus), India  
Email: alok.chauhan@imsuc.ac.in

CITE THIS ARTICLE AS :

Alok Singh Chauhan, " Security and Privacy  
Challenges in Big Data Environment,"  
International Journal of Technology and Science,  
vol. 6, Issue. 1, pp. 13-16, 2019