International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Big Data Ethics - In Terms of Transparency and Security

Jitender Sharma[1], Nitin Pandey[2]

[1, 2]Amity Institute of Information Technology, Amity University, Sector-125, Noida-201303, India

Abstract: ***In the recent times, Big Data is the most concerned term in the trending technology. Big data is a set of large data which is complex to manage, to capture, to analyze, to store, etc. it is managed by many industries as well as various academic groups. Nowadays, we can store Terabyte or Zeta byte of data but the challenge is how to store the data in our own hand. By giving all the access or maintenance to someone else is it really worth able or not? If the data goes to wrong hand, we don't know what do to, how it will be done because day-by-day big data is becoming bigger. So, there is a need to adopt more transparency & security measures. Most challenging phase of big data is security and transparency. Security and transparency can be measured in 3 parameters those are 3V's of big data. [3] are Velocity, Volume and variety. These three factors include a lot of variables such as Cloud computing infrastructure, Format of data, Streaming of data, Security mechanism, The main objective of writing this paper is to summarize the measurements taken for the security and transparency of big data and examine the big data in term of ethics.***

Keywords: Big data, transparency, security, intrusion deception, Data locker, security breach

1. Introduction

"Big data" appears to be more concerning field day by day.[1] The term describes innovative techniques to distribute, capture, manage, store and analyze petabyte or zeta byte datasets with high-velocity and diverse structures that conventional data management methods are incapable of handling. Big data has demonstrated the capacity to improve predictions, save money, boost efficiency and enhance decision-making in fields as disparate as traffic control, finance, fraud control, weather Forecasting, national security, disaster prevention, business transaction, education, and health care.

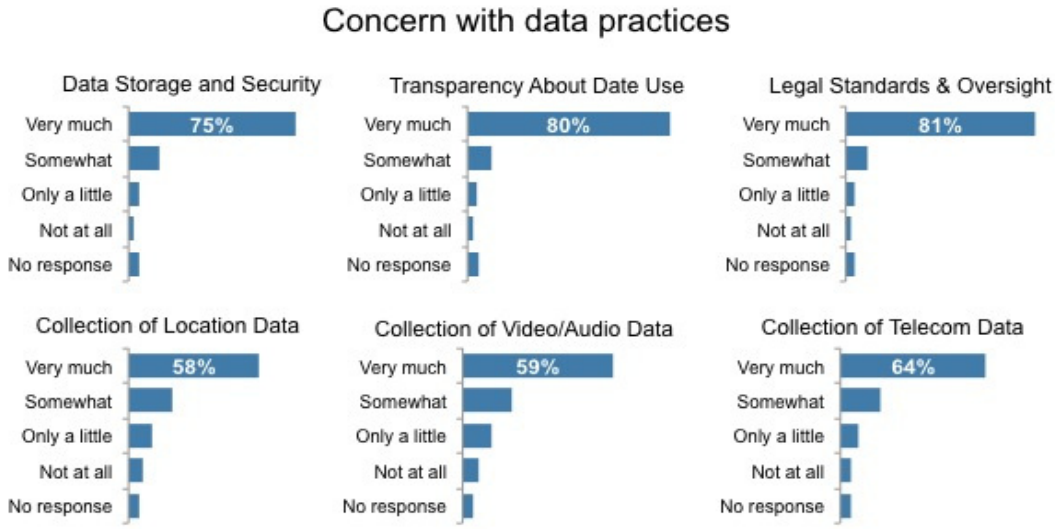


Figure 1: concerning issues in with big data ethics in terms of transparency and security.

As a rising discipline, big data's researchers, developers and users are experiencing many unknowns' facts during the journey of discovery and application. From this survey, there might be most concerning issues in with big data ethics in terms of transparency and security.

2. Ethics of Big data

Big data ethics can be defined in term of transparency and security:-[8]

Transparency here means using the data in a fair manner there should be such specific policies so that no hidden view or condition is charged it should mentioned how can be data is used and where it is sold but the user don't know about it.

Security means who is actually accessing our data & for what purpose? [9] It's mean not only to hide data but there should be rule and certain policies that enforces uses of data. We should know big data used by various online sites & companies has not been superfluous. Every swipe of debit card, credit card or login to social networking sites, every routing on your mobile GPS or anywhere your data or the information about you is collected & used. However, there are many parameters to judge the industry using your data but the means & the reason for their access is unknown to uses. For all that there is need of analytical model. Who can

predict it before you can tell anyone, for example, CRT monitor to LED, android mobiles, small data to big data on cloud computing. On the basis of various data points the result is integrated & then analyzed. But, these data are utilized for only marketing purpose. There is no harm & foul but what to do if you want to escape from all this. We believe that one of the defining fights of our time will be the fight over the control of personal information, the fight will be over whether big data will become a force for freedom or a force which will hidden manipulate us. If the users don't want me to use their data, they must have the right to do that. I want users to be informed & consenting users of the tools we develop.

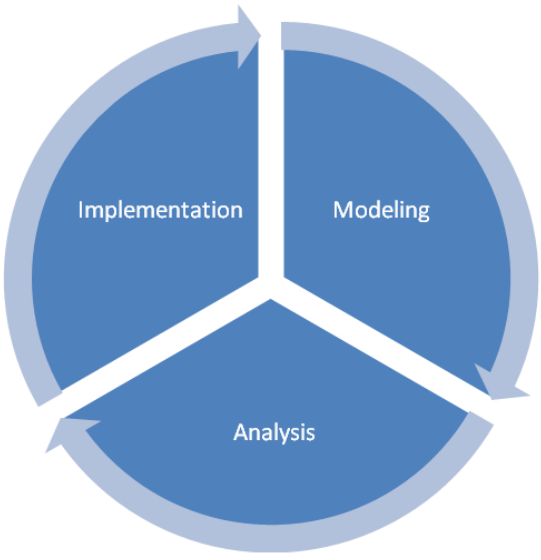
3. How it to Should Be Implemented

There is a need of building trust and should be a transparency mechanisms to our personal data, which can guarantee the people's personal data access. With the growth of big data, its governance should be reliable, scalable, and efficient. We need a modular way or process where we know whatever are getting ,Can control it ,and people can audit it or can know that our data is safe or not and it is being used by right hands it can be implemented in the following way:

- Modeling: create a model to covers all type of security breach or attacks and data-leakage that is called security breach model
- Analysis: finding an analytical solutions based on the security breach model
- Implementation: implementing the analyzed solutions in current infrastructure

In short it might be 3 step processes to reach ethics of big data in term of security and transparency.

- Firstly collect the all security and transparency problem on basis of priority
- Study the published solutions or propose the new solutions
- Mark the problem as challenge if there is no acceptable solution is found



There is need more reliable government policies, more than 45% companies still don't have any concerned data governance standards [4] .There should be some rule and regulations those might be followed by all organization.

- Your policies should be variable because improving in policies time to time will develop more value for your data
- It should be scalable , reliable and efficient [7]
- Each company should discuss their policy
- There should be alignment with data ethics what can do or not

As Dyson says, "Ethics don't change. Circumstances change, but the same Standards apply."[2] That means in case of big data ethics if someone varies only circumstances will be changes but value of ethics might be always static.

4. Related Work

However there are many existing survey and research have been done on big data ethics some of them are following:

- Data locker
- Object as a service
- Cryptographic watermark
- intrusion deception

Data Locker is means to visualize your own data on dashboard where you can text them and share them or can define access who can share it. [10]

Objects as a Service, which encapsulate data into form of digital enclosure it wraps up personal data into digital data [5].User can define permission like data locker but we can limit the user. Cryptographic watermark marks data in form of verifying the origin of data, time and integrity so that anyone access the data time can be checked with real time. Intrusion deception to pretend with fake data to attackers also known as a honey-pot[6]. All these survey or research talk about big data in context of security but still there is no complete solution till yet is found. [11] My paper shows relative study from all these survey and tells what to be done .Big data is two edge processes on one side it's an analytical technique used for massive data, various organization use it, Second availability of real data makes risk of security. [12]

5. Summary

Big data is vast field and using too much so that security is really big challenge every new development or new technique will bring advantage but before it brings disadvantage also. Security breach is major problem with big data there should be a framework model for accessing and authorization of data so that it can be reliable and efficient. Personal data with big data ethics is still a major problem with users however there is no relative solution is so far.

References

[1] https://www.umbel.com/blog/umbel/big-data-ted-talks/
[2] http://www.cmo.com/articles/2014/4/3/big_data_ethics_tran.html

[3] http://www.whitehouse.gov/issues/technology/big-data-review
[4] From "2013 Data Governance Survey" conducted by Rand Secure
[5] researchers at Barcelona UPC University are exploring with Telefonica researchers
[6] Guardtime's KSI
[7] http://www.pcworld.com/article/2048505/nsas-big-data-efforts-need-transparency-privacy-advocates-say.html
[8] http://www.cmo.com/articles/2014/4/3/big_data_ethics_tran.html
[9] https://cloudsecurityalliance.org/media/news/csa-releases-the-expanded-top-ten-big-data-security-privacy-challenges/
[10] http://www.whitehouse.gov/issues/technology/big-data-review
[11] http://blog.digital.telefonica.com/2014/03/20/big-data-transparency-economy/
[12] http://www.scientificcomputing.com/blogs/2014/03/big-workflow-future-big-data-computing

For literature review
[13] Thusoo, A.; Sarma, J. S.; Jain, N.; Shao, Z.; Chakka, P.;Zhang, N.; Antony, S.; Liu, H.; and Murthy, R. 2010. Hive -a petabyte scale data warehouse using hadoop. In Proceedings of the 26th International Conference on Data Engineering ICDE), 996–1005.
[14] Tuchinda, R.; Knoblock, C. A.; and Szekely, P. 2011. Building mashups by demonstration. ACM Transactions on the Web (TWEB) 5(3).
[15] Wu, B.; Szekely, P.; and Knoblock, C. A. 2012. Learning data transformation rules through examples: Preliminary results. In Ninth International Workshop on Information Integration on the Web (IIWeb 2012).

Author Profile

Jitender Sharma is pursuing the MCA from Amity institute of information technology, Amity University, Sector – 125, Noida-201303(UP), India