

Data Protection in Health Care using Big data Analytics

Priyanka Sharma
Professor
Lingaya's University
priyankasharmaid2@gmail.com

P.V.S. Avinash
Student
Lingaya's University
princeavinash1997@gmail.com

P.Y.S Chakradhar Reddy
Student
Lingaya's University
yuvachakri@gmail.com

Abstract

In recent times, privacy became more concern in many sectors. Mainly, privacy is more important aspect in healthcare industry. The paper describes the nascent field of big data analytics in healthcare, discuss the benefits, storage of patient records in big data, data privacy in healthcare, data protection techniques, discusses the challenges and offers conclusions. As the privacy is one of the most important things that patients holds on, we presented the current methods used the patient's privacy in healthcare industry. We focused more on the security and privacy aspects of medical data in a Big Data environment.

Keywords: Big data, Merits, Security, Privacy, HIPAA.

1. Introduction:

Aging populations and lifestyle changes pretense increasing pressures on healthcare systems around the world. The data is stored in traditional manner like hard copy form i.e., record keeping, compliance & regulatory requirements, and patient care generated large amount of data. Today's world is focusing towards rapid digitization of large amount of data. In healthcare industry, we use a large amount of person information and medical data that includes patient details and physician's written notes, prescriptions, medical images, laboratory, pharmacy, insurance and other administrative data. These produces a massive quantity of data, it takes long time to analytics this data. Big Data analytics can revolutionize the healthcare industry. It can improve operational efficiencies, help to predict and plan responses to disease epidemics, improve the quality of monitoring of clinical trials. Analysts predict that big data can, and will, be used to reduce the cost of healthcare delivery.

Reports say data from the U.S. healthcare system alone reached, in 2011, 150 Exabytes. At this rate of growth, big data for U.S. healthcare will soon reach the Zettabyte (10^{21} GB) scale and, not long after, the Yottabyte (10^{24} GB).

By definition, Big data analytics has offered a new way to healthcare organizations to develop actionable insights, organize their future visualization, boost up the outcomes and reduce time to value. This approach is also helpful to provide insightful information to the healthcare enterprises regarding their planning and the measurements. The evaluated results can further help enhance the decision making capacity of the top management.

Healthcare data is increasingly being digitized, as electronic health records, medical images, and prescriptions. This data redundancy could be of great benefit to the healthcare sector but the data's ease of use and accessibility threatens highly the patient's privacy. The medical data's privacy is then an important factor that healthcare adopters of big data need to consider seriously. Privacy though is a very important side that people always require in many sectors, and it's not less important in the healthcare industry, where information only gets even more personal and sensitive that any privacy breach becomes an issue of life or death.

2. Big Data Analytics in Healthcare:

It is Not Just About Building Bigger Databases. Big data is not about the technologies to store huge amounts of data. It is about creating a flexible infrastructure with high-performance computing, high performance analytics and governance, in a deployment model that makes sense for the organization.

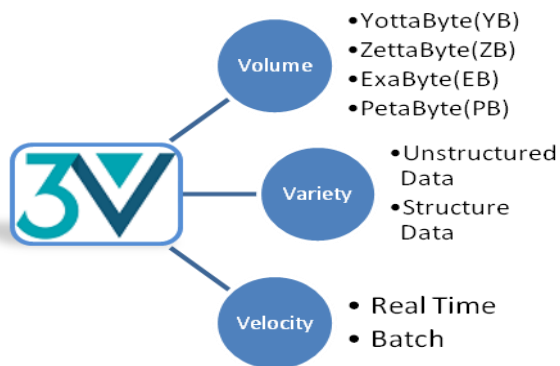


Fig 1: 3V's of big data.

There are three main "dimensions" to Big Data, commonly referred to as the 3V's

1. Volume = quantity, from terabytes to zettabytes
2. Variety =unstructured, semi-structured and structured

3.Velocity = from any-time batch processing to real-time streaming.

Volume

90% of all data ever created, was created in the past 2 years. From now on, the amount of data in the world becomes twice for every two years. By 2020, we will have 50 times the amount of data as that we had in 2011. The sheer volume of the data is enormous and a very large contributor to the ever growing digital world is the Internet of Things with sensors all over the world in all devices creating data every second

Velocity

The Velocity is the speed at which the data or information is created, stored, analyzed and visualized. In the past, we use batch processing, it was normal to receive an update from the database every night or even every week. Computers and servers required significant time to process the data and update the databases. In the big data, data is created in real-time or near real-time. With the availability of Internet connected devices, these devices can pass-on their data the moment it is created.

Variety

In the past, all data that was created was structured data, it neatly built-in columns and rows but those days are over. Currently, 90% of the data that is generated by today's world is unstructured data. Data today comes in many different formats: structured data, semi-structured data, unstructured data and even messy structured data. The wide variety of data requires a different approach as well as different methods to store all raw data.

3. Merits of big data in healthcare

Big Data can help healthcare facilities run smarter; instead, here is our list of ways that Big Data can benefit your healthcare facility.

3.1. Optimizing facility performance: In this, the more you know about your facility, the better it runs. Big data technology allows facility managers to do side-by-side examinations of performance reports in areas ranging from power monitoring to security services. It also goes without saying that as facility performance goes up, so will patient safety, patient fulfillment, and facility ratings.

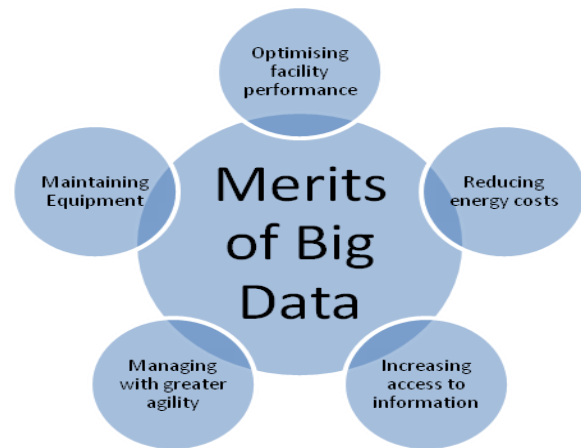


Fig 2: Merits of big data

3.2. Reducing energy costs: This is another case where more information naturally leads to better performance. Increased monitoring capabilities as part of a Big Data solution make it so healthcare facility staff and managers can identify areas with high energy use, and determine where it is really required and where energy is being wasted, and then take steps to fight against those inefficiencies. A decrease in overall energy use comes with the benefit of decreased energy costs. It's also likely to create a more environmentally-conscious facility.

3.3. Increasing access to information: With the cloud storage technologies constantly improving and expanding Big Data can be accessed from just about anywhere. This is a huge benefit to doctors who can access electronic medical records anytime, anywhere to improve patient care. It's also beneficial for facility management, operators, and engineers, who can discover and deal with problems slightly.

Big Data is even changing how patients choose their healthcare providers. The internet has allowed immediate access through sites such as Yelp, Health Grades, Hospital Safety Score, hospital ratings and patient reviews on doctors.

3.4. Managing with greater agility: Access to live updates and alerts allows facility managers and hospital staff to respond quickly and efficiently to the problems and concerns. With digital hospital solutions that send the right information to the right people in real-time, users can make informed decisions and get ahead of the issue.

3.5. Maintaining equipment: Hospitals and healthcare facilities have a significant physical and technological infrastructure, and in order to ensure smooth

and problem-free facility operation, it's very important to keep equipment in good working order. Big Data allows equipment diagnostics and analytics to run in real time. Equipment performance issues can be addressed immediately to ensure best performance, and to prevent larger repairs or system failures.

With impacts like these, the healthcare industry should consider taking advantage of Big Data. Effective use of Big Data will improve efficiency, access to information, facility performance, and assist to bring the healthcare facilities of today into the ranks of those of the future.

4. Big Data Collection methods

The process of dealing with big data is quite different from handling traditional data. Big Data is handled at various stages namely, collection, storing, organizing, analyzing etc. with a motive of deciphering useful insights useful to make business decisions.



Fig 3: Data collection process in big data.

4.1. Collection: This stage includes collection of data from several types of data sources, data marts and data warehouses.

4.2. Storing: This stage includes storing the data into distributed database systems and servers. The data is stored in such a manner that for each and every data stored, the backup is simultaneously created. This stage involves setting up physical infrastructure or cloud for data storage.

4.3. Data Organization: This stage involves categorizing and arranging the data on the basis of structured, unstructured and semi-structured data which is easy to access and analyze using big data technologies such as Hadoop Distributed File System (HDFS).

4.4. Data Analysis: After the data has been organized (stored and arranged properly), this stage includes deriving the data and applying statistical & business analytics concepts to carve out the hidden insights from the data helpful for decision making.

4.5. Data Visualization: Once the hidden insights have been carved out from data, the very next stage is to represent it. The representation is usually done using Data Visualization. Since, not everyone feels easy with numbers, but everyone can be able to understand easily the pattern when represented on graph.

5. Data Privacy in health care

Privacy is also among major concerns in big data with regard to healthcare applications. The health related data are always private in nature. Privacy issues rise from many reasons. It may be personal belief, social and cultural environment and other general public/private reasons. They have emphasized that if the issues associated with privacy are not honestly debated and open ways there is a risk that there will be a public backlash which will lead to mistrust and consequently the technology will not be used for the many valuable applications where it can provide major benefits. Whether the data are obtained with the consent of the person or without it due to the need by the system, misuse or privacy concerns may restrict people from taking advantage of the full benefits.

There are major questions raised by people from time to time. For example, authors in have raised questions about guarding the privacy of an individual such as, where should the data of health be stored, and who can view a patient's medical record. There are also some questions such as to whom should this information be disclosed to, without the patient's permission and who will be responsible for maintaining these data in case any problem arise, who will be held accountable. There are several main issues that should be determined in order to protect privacy as well as to some extent the security of the information.

In normal circumstances, there are only few users of the data like the physicians, nurses and other clinical/technical staffs. This limits the number of users in the system. Well defined rules and firm guidelines regarding use of data for these users can limit the concerns for privacy. But it should also be noted that in some cases such as emergency, disasters or remote patient monitoring may require disclosure of information to other people in order to serve the patients in need. So the system must be flexible enough and users should be made to accept or compromise to some extent. Still measures must be placed to make the users of the private healthcare data responsible for their actions or else these people may not care about the privacy concerns of an individual which can lead to bad implications on the social life of the person concerned. Authors in have argued that without appropriate privacy safeguards the information that may go into the public domain straight away, which is potentially undesirable for number of reasons. People may not want some personal data be available in public domain. It is also essential that these data should not fall into the hands of people with malicious intent and hence person.

Privacy measures

The role of wireless communications in healthcare applications is expected to become more prominent with a

growing mobile society and with the deployment of mobile and wireless networks. Hence it is a better idea to be ready for such situations before the time comes for it. Educating people about the future ahead can make them more relaxed as well.

6. Data protection techniques:

Currently, there are multiple techniques used to ensure the patient's privacy, the main ones are:

6.1 De-identification: is a method that claims to protect the disclosure of confidential information, by removing identifying information from medical records. There are two methods proposed by the HIPAA to de-identify medical information:

Safe-harbor: requires the removal of specific identifiers of the individual or of relatives, employers or household members of the individual from the data: names, addresses, dates, telephone numbers, health plan beneficiary numbers, e-mail addresses, social security numbers, medical record numbers, certificate/license numbers, vehicle identifiers and serial numbers, device identifiers and serial numbers, IP addresses, biometric identifiers, full face photographic images.

Statistical method: a person with appropriate training verifies that enough identifiers have been detached.

6.2 Notice and consent: The patient has the right to know and approve to how his data will be used and by who, by giving him/her notice of an entity's information practices before collecting any personal information.

6.3 Notice and consent successor: The patient can willingly choose a third party and associate themselves with it, in order to inherit a set of privacy preference profiles that the healthcare organization can use every time the patient's data need to be disclosed.

6.4 Deletion and non-retention: when data is no longer in use, it is a good practice to delete it. This technique can be used within healthcare organizations to reduce the privacy breaches, but when big data gets involved, the question remains which data is in use and which data is not, because what seems to be valueless today may bring big value tomorrow, besides, given the distributed and unnecessary nature of data storage, there's no guarantee that no useful data has survived the destruction.

7. Conclusion and Future Work:

In today's world, many organizations such as educational system, banking, and other insurance companies etc. are using big data for analysis purpose. In healthcare industry also, big data plays a prominent role for analysis purpose. Big Data analytics can revolutionize the healthcare industry. It can improve operational efficiencies, help to predict and plan responses to disease epidemics. In spite of its advantages, big data in healthcare has raised new challenges in the area of privacy and security. As privacy is one of the major concerns in the healthcare industry in recent times. The main reason behind the privacy problem is a huge amount of patient's information such as patient records, regulatory requirements etc. is freely available directly or indirectly in the form of digital information. So,

there is the loss of privacy because anyone can access the data and this may cause harm to the patients. In this paper, we presented the current methods used the patient's privacy in the healthcare industry. We focused more on the security and privacy aspects of medical data in a Big Data environment.

In healthcare, big data is not only used in privacy aspect but also it is used for generating and sharing large amount of personal health data with the help of consumer devices such as smart watches and wristbands that regulate the sleeping patterns, exercise, heart rate, calorie consumption an much more.

References:

- [1] Big data analytics in health care: promise and potential <http://www.hissjournal.com/content/2/1/3>
- [2] Big Data in Healthcare Hype and Hope <https://www.ghdonline.org/>
- [3] Data Analytics for Healthcare: Creating understanding from big data
- [4] Privacy Concerns Are Slowing Big Data Adoption Within Health care <http://www.smartdatacollective.com/jonathanbuckley/323971/>
- [5] Big Data Privacy in Healthcare <http://creativecommons.org/licenses/by-nc-nd/4.0/>
- [6] 6 best ways to protect against health data breaches <http://www.healthcareitnews.com/news/6-best-ways-protect-against-health-data-breaches>
- [7] Big data and analytics in healthcare <http://link.springer.com/article/10.1007/s10796-016-9641-2>
- [8] Significant Benefits of Big Data Analytics In Healthcare Industry <http://www.builtinla.com/blog/significant-benefits-big-data-analytics-healthcare-industry>
- [9] Healthcare big data: privacy and security workgroup <http://www.hipaajournal.com/healthcare-big-data-privacy-and-security-workgroup-gives-preliminary-report-8029/>
- [10] AHIMA. The state of HIPAA privacy and security compliance. <http://library.ahima.org/xpedio/>
- [11] Journal of Biomedical Informatics Security and privacy in electronic health records: A systematic literature review <http://www.elsevier.com/locate/yjbin>
- [12] Wullianallur Raghupathi and Viju Raghupathi: "Big data analytics in healthcare: promise and potential."
- [13] Jasmeen Gill , Shaminder Singh, Devdutt Baresary: "Big Data: Big Innovations in Healthcare."
- [14] Priyanka K, Nagarathna Kulennavar: " A Survey On Big Data Analytics In Health Care".
- [15] Prof. A.I.Petrenko, Dr .N.V.Roenko "Personal healthcare platform for chronic diseases with mobile self-management support"

[16] Peter Groves,Basel Kayyali,David Knott,Steve Van Kuiken:"The 'big data' revolution in healthcare: Accelerating value and innovation"

<http://www.mckinsey.com/h>

[17] A look at challenges and opportunities of Big Data analytics in healthcare.

<http://ieeexplore.ieee.org/document/6691753>[19][https://www.ijmijournal.com/article/S1386-5056\(03\)00033-9/](https://www.ijmijournal.com/article/S1386-5056(03)00033-9/)

[18]<https://www.healthcatalyst.com/big-data-in-health-care-madesimple>

IJSER