



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 Issue: IV Month of publication: April 2018

DOI: <http://doi.org/10.22214/ijraset.2018.4668>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Big Data and Its Applications in Healthcare

Sayali Avinash Joshi¹, Prof. Flavia Gonsalves (Lobo)²

^{1, 2}MET's Institute Of Computer Science, Bandra Mumbai University

Abstract: *In the last few decades, the healthcare industry has grown tremendously. With the growth of healthcare sector, a huge amount of data is generated on a daily basis. The data comes from different sources because of the Electronic Health Record (EHR) systems implemented at hospitals. These different sources have caused the conventional relational databases difficult to manage such huge data with huge volume. The rapidly expanding field of big data plays an important role in the evolution of healthcare industry. Big data provides different tools to gather, manage and analyze large volumes of structured, unstructured and diverse data generated by the current healthcare systems. In this paper, we will discuss the impact of big data on the healthcare sector, the opportunities, and challenges in the implementation of big data and the tools to implement big data in the healthcare industry.*

Index Terms: *Big Data, Healthcare, Big data analytics, Electronic Healthcare Records (EHR), Datasets*

I. INTRODUCTION

The healthcare industry has grown extensively over the past few years. The healthcare industry is broadly composed of hospitals, druggist, pharmacist, pathologist, radiologist as well as any another web services based applications which are related to healthcare and management [1, 2]. The healthcare industry in the past has generated a large amount of data. Different types of data sources are involved in the generation of such huge amount of data. Most of the data is stored in a hard copy form. However, in today's age, there is the rapid digitization of such huge amount of data.

"Big Data" as a concept is not new in today's world, however, its definition is constantly changing. Big Data can be defined as any dataset which cannot be represented efficiently and cannot be interpreted using traditional database systems which are currently in use. Dataset coming from banking, healthcare, weather, e-commerce, social media fields can be classified as big data [3]. Three important characteristics of big data are volume, velocity, and variety also known as three V's[3]. In healthcare, big data refers to Electronic Health Records (EHR) which is quite large and complex. This kind of data is difficult to manage with traditional hardware and software. The execution of big data is, however, the most difficult task. Many researchers today suggest installation of big data tools in the standalone systems. Big data generally consists of large voluminous data. The processing and execution of such data can be carried out in distributed nodes. Therefore, to take decisions related to health in a better way, an open source distributed processing platforms for data, such as Hadoop or Map Reduce are required.

Big Data and its practices in health and medical science become even more noteworthy due to the use of social media and networks (for e.g. Facebook and Twitter), sensory or digital technology, and mobile devices with smartphone apps and personal sensor health data along with real-time digital data accumulations [4,5]. Big Data Science is emerging in today's world and is also valuable in areas other than healthcare such as financial, government, educational, etc. So, this paper deals with the application of big data and its analytics in healthcare and also focuses on the challenges faced by big data.

II. BIG DATA AND ITS SOURCES

Big data is a term which is used to describe a huge amount of data which can be structured or unstructured. Big data analytics is time-sensitive and is used to take a faster decision based on a large amount of diversified data. The Big Data Analytics also plays a crucial role for both small and medium-sized organizations so to increase their business.

Industry analyst Doug Laney originally coined the concept of big data while referring to the challenge of data management [6]. According to that, there are three important dimensions of the big data concept illustrated below [7].

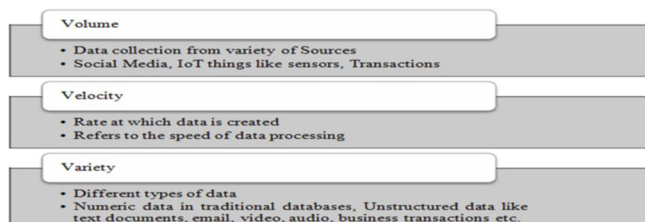


Figure 1 - Three Vs of Big Data

A. Big data sources

Big data can be generated from different social media sources such as Facebook, Instagram, Twitter, etc. These sources generate terabytes of data on a daily basis. Devices such as the laptop, computers also generate a large amount of data. The Internet of Things (IoT) devices such as sensors are also generating a tremendous amount of data. The data coming from many sources can be categorized as follows:

- 1) *Social media data*: Data generated from Facebook, Twitter, blogs, etc. It also includes health plan websites, mobile apps, etc.
- 2) *Device-to-device generated data*: Includes readings from devices such as sensors, etc.
- 3) *Human-generated data*: This can be structured or unstructured data. For e.g. papers, doctor notes, etc.
- 4) *Biometric data*: Data in the form of x-rays, fingerprints, retinal scans, etc.

III. BIG DATA ANALYTICS (BDA) AND HEALTHCARE

Information and Communications Technology (ICT) plays a very important role in the improvement of healthcare for individuals as well as the community. It helps in the improvement of healthcare efficiency and also prevents medical errors. ICT supported health mechanism known as eHealth helps to serve society in a better way.

One of the characteristics of the healthcare system is its data richness. In every country, the healthcare authorities are making it compulsory to digitize the process involved in the healthcare industry [8]. In most of the hospitals, registration of every new patient is recorded in the Electronic Registration System (ERS). The data recorded in ERS then needs to be issued to a secure chip-based data by using which the records can be updated in different departments.

Big data analytics is the process of exploring huge data sets which contain a variety of data types to reveal hidden patterns, unknown correlations, customer preferences, market trends, etc. [9]. Big Data Analytics in healthcare can be defined as a set of methodologies and procedures which convert raw data into meaningful and useful information. This information makes decision making faster and more effective. Data is generated from a variety of sources which needs to be aggregated. The data goes through a variety of processes like extraction, cleaning, transformation, and loading. This leads to the generation of useful data which can be used by users. The data created at a particular location is then collaborated to achieve objectives of Content-Based Retrieval system (CBR) and accurate analytics which provides fast and cost-effective services to patients and the healthcare system [10].

IV. TOOLS OF BIG DATA IN HEALTHCARE

We will discuss the different big data tools which are used in the healthcare system [11]:

- 1) *Hadoop System*: This provides the basic storage for Hadoop cluster (HC), and divides the data into two parts and also performs distribution for different nodes or servers.
- 2) *Map Reduce*: It allows the interface to scatter the subtasks and extract the outputs. After execution of each task, MapReduce performs the tracking of every node and server.
- 3) *Hive*: It is the runtime Hadoop version which combines SQL with Hadoop and also allows SQL programmers to develop (HQL).
- 4) *Zookeeper*: This allows synchronization for server clusters. BDA uses this service to the parallel process of big data clusters.
- 5) *HBase*: This is a column-based database management system, which tops on Hadoop system by using non-SQL approach.

V. ADVANTAGES AND OPPORTUNITIES OF BIG DATA IN HEALTHCARE

This section describes the various advantages and opportunities of big data in healthcare:-

- 1) *Reduction in the cost of healthcare*: Big data helps in reducing the cost of medical treatments. Analysis of the data gives the healthcare providers information about the population at risk for illness. Big data provides accurate information on where education and prevention are needed to people at lower cost. Treatment is more evidence-based using big data technology [12]. Sharing the data between the healthcare providers and the doctors or physicians as they examine the patients can further reduce the need for duplicate tests and also improve patient care.
- 2) *Preventive care made stronger*: As goes the famous proverb, prevention is better than cure. With the introduction of big data, it is easy to capture, analyze and compare patient symptoms earlier so that preventive care can be offered to patients in a better way.
- 3) *Protecting patient's identity*: Privacy is the primary concern. Insurers are using big data analytics to detect medical fraud and identity thefts of patients. Medical records contain various personal information. In this system, data access is limited to the authorized users. Also, data is stored in a highly secured server.

- 4) *Clinical Decision Support (CDS)*: CDS helps to improve the quality of the healthcare services by enhancing the outcomes. The aim of the system is to provide the right information to right people and properly customized healthcare management process. This system allows doctors, patients, chemists to know specific medical related information.
- 5) *Disease management*: This system helps to analyze various diseases. The analytics help to improve the accuracy to find out the information which helps to improve the outcome so that medical solutions can be determined. This requires data support from various organizations and individuals.
- 6) *Patient-centric medical approach*: With the help of big data which supports prescriptive analysis, patient-centric medical approach can be developed. It explores the previous disease-management system, approach to cure the patient, symptoms of the disease, etc. By exploring and analysis of the mentioned things, the result will be clear to treat a patient based on the symptoms instead of a general disease management system.
- 7) *Evidence-based medicines*: Using this approach, doctors can match symptoms to a larger patient database in order to come to an accurate diagnosis faster and more efficiently. It involves combining and analyzing of structured and unstructured data from EMRs, financial and operational data, clinical data, etc to match the treatment with the result, predict patients at risk or disease and provide more efficient care to the patients.

VI. CHALLENGES OF BIG DATA IN HEALTHCARE

Following are some of the challenges which make big data analytics difficult to use in healthcare:

- 1) *Protecting patient's identity*: One of the most difficult challenges in healthcare's big data is the policies that protect the privacy of patient's data. Many laws protect the privacy of the patient's data and do not reveal the patient's identity which makes big data analytics difficult. However, sometimes healthcare providers are unwilling to share their patient's data because of market competition as the doctors or physician may not want their competitors to know which procedures and treatments are being used. Some of the datasets are publicly available but these data sources are typically historical data or limited to government payer [13].
- 2) *Managing data*: Hospitals which have a large potential for a big data evolution have not kept up with other industries. They need to develop the ability to assess, analyze and manage big data, including the flexibility to add, compute power and storage capacity quickly when needed for which the data centers need to be transferred to the cloud.
- 3) *The Requirement for expert knowledge*: Big data systems require data scientists with special experience to support design, implementation and continued use. The McKinsey Global Institute estimates that there will be more than 100,000 person shortage till 2020. It means that 50-60% of data scientist positions may go vacant. Data scientist are required to possess skills such as communication, leadership, creativity, etc.
- 4) *Data Aggregation*: Data coming from different sources is in an unstructured format. These unstructured data may include images, graphs, doctor's notes, etc. The nature of structured data is heterogeneous. This may lead to a huge problem during the data aggregation process.
- 5) *Availability and disaster recovery*: It is important for the healthcare industry to maintain the availability of services and also provide authorized access to the data, even if the data center is destroyed due to some accident. Moving all or important parts of the data center to the cloud, across various locations, can help.

VII. CONCLUSION

Big data analytics has the potential to transform the way healthcare providers use sophisticated technologies to gain insight from their clinical repositories and make better decisions [14]. As we discussed in this paper, there are several opportunities or big data in healthcare. In near future, there will be widespread implementation of big data analytics across the healthcare industry. Doctors and nurses will be able to determine diseases and risks easily as well as provide needed treatment at the right time. This study will push both doctors and patients together to adopt new technique and collaborate together to reach a high level of connection between the medical staffs and patients in order to keep the system up to date and gather high-quality data.

Generally, in healthcare industry data analytics is a very important topic, since all the advantages we mentioned could lead to a better choice of medical practice and prevent illnesses. The big data storage and solutions would provide an efficient solution in contrast to the traditional storage solutions. The big data solution will definitely save millions of lie and also improve patient services.



REFERENCES

- [1] Raghupathi, Wullianallur, and VijuRaghupathi. "Big data analytics in healthcare: promise and potential." *Health Information Science and Systems* 2.1, 2014.
- [2] Srinivasan, U.; Arunasalam, B., "Leveraging Big Data Analytics to Reduce Healthcare Costs," in *IT Professional*, vol.15, no.6, pp.21-28, Nov.-Dec. 2013
- [3] Apache Hadoop. (2014, September 10) [online]. Available: <http://hadoop.apache.org>.
- [4] Bachrach Y, Kosinski M, Graepel T, Kohli P, Stillwell D (2012) Personality and patterns of Facebook usage. *WebSci '12: Proceedings of the 3rd Annual ACM Web Science Conference* pp: 24-32.
- [5] Bollen J, Mao H, Zeng XJ (2010) Twitter mood predicts the stock market. *Journal of Computational*.
- [6] Doug Laney, *Application Delivery Strategies*, retrieved from <http://blogs.gartner.com/DougLaney/files/2012/01/ad949-3D-DataManagementControllingDataVolumeVelocityandVariety.pdf> on December 20, 2015
- [7] Big Data, retrieved from http://www.sas.com/en_us/insights/big-data/what-is-big-data.html on December 20, 2015
- [8] Acampora, G.; Cook, D.J.; Rashidi, P.; Vasilakos, A.V., "A Survey on Ambient Intelligence in Healthcare," in *Proceedings of the IEEE*, vol.101, no.12, pp.2470-2494, Dec. 2013.
- [9] Big data analytics definition retrieved from <http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics>
- [10] Jun-ping Zhao, "Electronic health in China: from digital hospital to regional collaborative healthcare," in *Information Technology and Applications in Biomedicine*, 2008. ITAB 2008. International Conference on, vol., no., pp.26-26, 30-31 May 2008
- [11] Sagiroglu, S.; Sinanc, D., "Big data: A review," *Collaboration Technologies and Systems (CTS)*, 2013 International Conference, vol., no., pp.42-47, 20-24 May 2013
- [12] Big Data Offers Big Opportunities in Healthcare, retrieved from <http://www.villanovau.com/resources/bi/big-data-healthcare-opportunities/#.VnfRArZ95kg>
- [13] White SE, A review of big data in health care: challenges and opportunities, *Open Access Bioinformatics* 2014, 6:13-18, 2014
- [14] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4341817/>



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)