

# Big Data Analytics in Healthcare

Ismail Brahim Abbo  
MS (IT) Final Year Student,  
Jain University,  
Bangalore - 69, India.

Dr. Suchithra R ,  
HOD, MS(IT) Dept,  
Jain University,  
Bangalore - 69, India.

**Abstract:** -This paper provides a brief idea about additional value from health information that is used in the healthcare centers via a new information management approach called as big data analytics. Including big data analytics in health sector provides stakeholders. Big data analytics is a growth area with the potential to provide useful insight in healthcare. Even as many scope of big data still present issues in its use and acceptance, such as managing the volume, velocity, variety, veracity, and value, the accuracy, integrity, and semantic interpretation are the greater apprehension in the clinical request. However, such challenges have not deterred the use and the evidence source of big data examination in healthcare. This drives wants to investigate the healthcare information to control and reduce the burgeoning cost of healthcare, as well as to inquire about evidence to improve patient outcomes. This paper describes big data analytics and its advantages and challenges in health care.

**Keywords:-** Big data, Analytics, Healthcare, Hadoop,.

## 1. INTRODUCTION

Previously the industry of healthcare has been generated a large amounts of data, which has been driven by record care, observance & regulatory requirements, and patient care [1]. While the largest part of the data is stored in hard copy form, the current trend is in the direction of rapid digitization of these huge amounts of data. the potential to improve the quality of healthcare delivery temporarily reducing the costs, these massive quantities of data hold the secure of sustaining a wide range of medical and healthcare functions, including with others clinical decision support, sickness examination, and population health management.

Big data is known as a multidisciplinary information processing system. The places where use big data in business, media, government, and in particular healthcare, are increasingly to incorporate with big data into information processing systems [2]. The expensive way to use of the potential of big data in healthcare is an understanding of what the 2.5 quintillion bytes of data, where it can reside, which processed or derived artifacts, and what the description between public and private access is required.

Analytics it helps to optimize the key processes, roles and the functions [6]. It can be leveraged to join both internal and external data. It enables the organizations to meet stakeholder reporting demands, volumes of enormous data managed, the advantages to create market, manage risk, to improve controls and, ultimately, improve the organizational performance by turning information into intelligence.

## 2. THE ISSUES FACING BIG DATA ANALYTICS

The most important review attempts to give foundation for understanding the use of big data in healthcare, to explore the analysis of how big data can be applied to particular areas to achieve the maximum benefit for the targeted study. Big data analytics in the healthcare is the choice of solving the complexity of information system in the healthcare [2]. And here are the four dimensions of big data analytics are categorized separately:

**Volume:** It is the amount of data that produced by the organizations or individuals. All organizations are seeking for the ways to handle the ever growing data volume that's being created every day.

**Velocity:** Is the rapidity and frequency at which the data is produced, captured and shared. Consumers as well as businesses nowadays generate lots of data [5].

**Variety:** It is the quality of new data types that include those from social, mobile and machine resources. Such as Geo-spatial, hardware data points, machine data, radio frequency identification (RFID), search, and web. It also includes unstructured data [5].

**Veracity:** It is defined as the accuracy of data. Incorrect data can cause a lot of problems for organizations. Therefore, the organization required to make certain data is correct.

These four scopes are playing an important role in the issues facing of big data analytics. Big data analytics involves the analysis of high quantity of data that stems from a diversity of sources, including structured and unstructured data [2]. It also refers to the amounts of the data sets contained by the data analysis and the rate in which they are analyzed.

The concept of big data has been introduced to the system of healthcare as a solution to a variety of healthcare allied to information system troubles as health systems grown progressively more complex and expensive [3]. The Estimation that's been suggested in 2012, the healthcare data reached about 500 petabytes. Whereas in the future estimates suggested that by 2020 the healthcare data will be equivalent to 25,000 petabytes. The successful combination of such data using data mining and medical informatics possibly will result in lower costs and improved patient care through well informed decision making [3].

Big data is the brainpower for Electronic Health Records (EHRs), it has the capability to be associated with financial, equipped, and clinical analytic systems, and that may support the evidence-based in healthcare [2]. Evidence based of healthcare involves the systematic reviewing of

earlier clinical data in order to provide the decision makers of information. The evidence suggested that the big data it can be used to detect illness and in exacting support in the clinical genomic analysis of HIV patients (Feldman, 2012). However, big data requires a careful data management in organize to fulfill the goals of big data analytics.

### 3. BIG DATA IN HEALTHCARE

Big data in healthcare can come from internal such as electronic health records, clinical decision support systems, etc. and external sources as government sources, pharmacies, insurance companies etc. often in various formats such as (flat files, relational tables, etc.) and residing at several locations like geographic areas as well as in different providers of healthcare sites, in several legacy and added applications (business deal processing.) [5].

Resources and data types include:

1. *Web and social media data:* Click flow and dealings with data from social media such as Facebook, Twitter, and blogs [5]. It can also include remedial health plan websites, Smartphone applications, etc.
2. *Machine-to-machine data:* Readings from meters, sensors, and other devices [5].
3. *Big transaction data:* The Healthcare claims and other billing of records progressively more presented in semi-structured and unstructured formats [5].
4. *Biometric data:* Fingerprints, genetics, retinal scans, and comparable to this types of data. This also includes X-rays and other medical images [5].
5. *Human-generated data:* The semi-structured and unstructured data such as electronic medical records (EMRs), physician’s notes, email, and paper documents. In recent years, big data analytics has become increasingly evident that multiple streams of data like these can be leveraged with dominant latest collection, aggregation, and technologies of analytics and techniques to improve the transport of healthcare at the stage of individual patients as well as at level of illness and condition-specific populations.

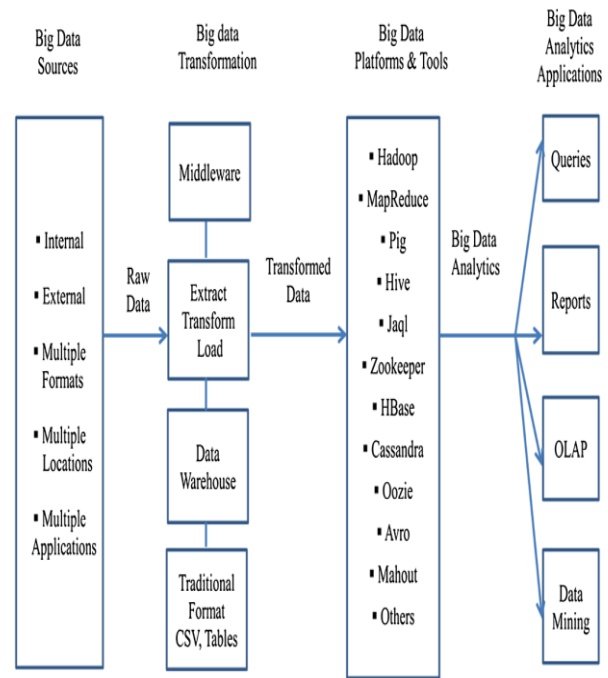


Figure1: Big data analytic’s conceptual architecture [1].

### 4. PLATFORMS / TOOLS FOR BIG DATA ANALYTICS IN HEALTHCARE

Tools	Description
Hadoop Distributed File System (HDFS)	HDFS is a sub-project of the Apache Hadoop project. This Apache Software base project is designed to provide a fault-tolerant file system designed to run on service hardware.
MapReduce	MapReduce is an associated implementation and programming model for processing and generate large data sets through a parallel, distributed algorithm on a cluster.
Hive	Hive it permits SQL programmers to develop Hive Query Language (HQL) statement similar to typical SQL statement [1].
Zookeeper	ZooKeeper is an open source Apache project that is provides a centralized infrastructure and services that allow synchronization across a cluster.
HBase	HBase is an oriented database management system that works on top of HDFS. It is well matched for sparse data sets, which are common in many of big data use cases. It uses a non-SQL approach.
Cassandra	Cassandra is a distributed database system. And it is chosen as a top-level project modeled to handle big data shared across many utility servers.
Mahout	Mahout is yet another Apache project whose objective is to create free applications to distributed and scalable mechanism knowledge of algorithms that support by big data analytics in the Hadoop platform.

## 5. ADVANTAGES OF BIG DATA IN HEALTHCARE

## 7. CONCLUSION:

*Clinical operations:* relative efficacy research to determine more clinically pertinent and cost-effective ways to diagnose and care for patients.

*Research & development:*

1) Analytical modeling to slow destruction and produce a leaner, faster, more targeted R & D pipeline in drugs and devices.

2) the tools of statistical and to improve the algorithms of clinical experiment design and patient prescription to better match treatments to individual patients, accordingly reducing trial failures and speeding latest treatments to market.

3) Analyzing the clinical trials and patient records to identify the result indications and discover poor effects before the products reaches the marketplace.

*Public health:*

1) Analyzing illness pattern and tracking the disease outbreak and spread to improve public health observation and speed response;

2) Faster development of more correctly targeted vaccines, example of choosing the annual influenza strains;

3) Turning the data's large amounts into actionable information that it can be used to identify the needs, and provide services, and predicts and prevent crisis, mainly for the benefit of populations.

## 6. BIG DATA CHALLENGES IN HEALTHCARE.

- Influencing the patient or data correlations in longitudinal records.
- Understanding the formless medical notes in the right content.
- Efficiently managing large volumes of clinical imaging data and extract the potential useful information and biomarkers [4].
- Analyzing genomic data is a computationally determined task and combining with standard medical data and adds extra layers of density.
- Collecting the patient's behavioral data with the help of some sensors; their various common interactions.

Big data analytics in healthcare will become a hopeful field for providing insight from very big data sets and improving the outcomes of reducing the costs. Its potential is so vast; though there have also remained challenges to overcome. Big data analytics has the prospective to change the way of healthcare providers use the technologies to increase insight from their clinical and other repositories of data and make knowledgeable decisions. In the future there will be the quick, extensive implementation and use of big data analytics through the healthcare society. To that end, the several challenges must be represented. As big data analytics becomes more important, the issues such as providing security and privacy, establishing standards and governance, and continuous improvement of the tools and technologies will gain attention. Big data analytics and applications in healthcare are emerging stage of development, but future advances in platforms and tools can be their growing process.

## 8. REFERENCES

- [1] Raghupathi W: Data Mining in Health Care In Healthcare Informatics: Improving Efficiency and Productivity. Edited by Kudyba S. Taylor& Francis; 2010:211-223.
- [2] Commonwealth of Australia. (2013). The Australian public service big data strategy: Improved understanding through enhanced data analytics capability. Canberra: Commonwealth of Australia.
- [3] Originally published in the Proceedings of the 3rd Australian eHealth Informatics and Security Conference. Held on the 1-3 December, 2014 at Edith Cowan University, Joondalup Campus, Perth, Western Australia. This Conference Proceeding is posted At Research Online. <http://ro.ecu.edu.au/aeis/22>.
- [4] Jimeng Sun Healthcare Analytics Department IBM TJ Watson Research Center & handan K. Reddy Department of Computer Science Wayne State University *siam International conference on data mining, austin, tx, 2013.* <http://dmkd.cs.wayne.edu/TUTORIAL/Healthcare>.
- [5] Priyanka Ketal, /IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014, 5865-5868 A Survey On Big Data Analytics In HealthCare. <http://218.248.31.202/journal/article/BIGDATA.pdf>
- [6] Insights on governance, risk and compliance April 2014 Big data. Textbook Changing the way businesses compete and operate