



BIG DATA ANALYTICS FOR HEALTHCARE INDUSTRY: IMPACT, APPLICATIONS, AND TOOLS

Ms. Chandana Malavade

Assistant Professor, Department of Information Technology Pragati College of Arts & Commerce, Dombivli (E)

Abstract :

In recent years, huge amounts of structured, unstructured, and semi-structured data have been generated by various institutions around the world and, collectively, this heterogeneous data is referred to as big data. The health industry sector has been confronted by the need to manage the big data being produced by various sources, which are well known for producing high volumes of heterogeneous data. Various big-data analytics tools and techniques have been developed for handling these massive amounts of data, in the healthcare sector. In this paper, we discuss the impact of big data in healthcare, and various tools available in the Hadoop ecosystem for handling it. Hadoop is an Apache open source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming models. The Hadoop framework application works in an environment that provides distributed storage and computation across clusters of computers. Hadoop is designed to scale up from single server to thousands of machines, each offering local computation and storage. In this paper we also explore the conceptual architecture of big data analytics for healthcare which involves the data gathering history of different branches, the genome database, electronic health records, text/imagery, and clinical decisions support system.

Key words: *big data; healthcare; Hadoop; Map Reduce.*

Copyright © 2022 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC 4.0) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

Introduction :

Every day, data is generated by a range of different applications, devices, and geographical research activities for the purposes of weather forecasting, weather prediction, disaster evaluation, crime detection, and the health industry, to name a few. In current scenarios, big data is associated with core technologies and various enterprises including Google, Facebook, Instagram and IBM, which extract valuable information from the huge volumes of data collected. An era of open information in healthcare is now under way. Big data is being generated rapidly in every field including healthcare, with respect to patient care, compliance, and various regulatory requirements. Big Data is emerging as an opportunity for organizations. Now, organizations have realized that they are getting lots of benefits by Big Data Analytics. As the global population continues to



increase along with the human lifespan, treatment delivery models are evolving quickly, and some of the decisions underlying these fast changes must be based on data. Healthcare shareholders are promised new knowledge from big data, so called both for its volume as well as its complexity and range. Pharmaceutical-industry experts and shareholders have begun to routinely analyze big data to obtain insight, but these activities are still in the early stages and must be coordinated to address healthcare delivery problems and improve healthcare quality. Early systems for big-data analytics of healthcare informatics have been established across many scenarios, e.g., the investigation of patient characteristics and determination of treatment cost and results to pinpoint the best and most cost-effective treatments. Health informatics is described as the assimilation of healthcare sciences, computing sciences and information sciences in the study of healthcare information. Health informatics involves data acquisition, storage, and retrieval to provide better results by healthcare providers. In the healthcare system, data is characterized by its heterogeneity and variety as a result of the linking of a diverse range of biomedical data sources including, for example, sensor data, imagery, gene arrays, laboratory tests, free text, and demographics. Most data in healthcare system (e.g., doctor's notes, lab test results, and clinical data) is unstructured and is not stored electronically, i.e., it exists only in hard copies and its volume is increasing very rapidly. Currently, there is a major focus on the digitization of these vast stores of hard copy data. The revolutions of data size are actually creating a problem in order to achieve this goal. The various terminologies and models that have been developed to resolve the problems associated with big data focus on solving four issues known as the four Vs, namely: volume, variety, velocity, and veracity. The various classes of data in healthcare applications include Electronic Health Records (EHR), machine generated/sensor data, health information exchanges, patient registries, portals, genetic databases, and public records. Public records are major sources of big-data in the healthcare industry and require efficient data analytics to resolve their associated healthcare problems. According to a survey conducted in 2012, healthcare data totaled nearly 550 peta bytes and will reach nearly 26000 petabytes in 2020. In light of the heterogeneous data formats, huge volume, and related uncertainties in the big-data sources, the task of realizing the transformation of raw data into actionable information is daunting. Being so complex, the identification of health features in medical data and the selection of class attributes for health analytics demands highly sophisticated and architectural specific techniques and tools.

Big Data Analytics in Health Informatics :

The main difference between traditional health analysis and big-data health analytics is the execution of computer programming. In the traditional system, the healthcare industry depended on other industries for big data analysis. Many healthcare shareholders trust information technology because of its meaningful outcomes—their operating systems are functional and they can process the data into standardized forms. Today, the healthcare industry is faced with the challenge of handling rapidly developing big healthcare data. The field of big data analytics is growing and has the potential to provide useful insights for the healthcare system. As noted above, most of the massive amounts



of data generated by this system is saved in hard copies, which must then be digitized. Big Data can improve healthcare delivery and reduce its cost, while supporting advanced patient care, improving patient outcomes, and avoiding unnecessary costs. Big data analytics is currently used to predict the outcomes of decisions made by physicians, the outcome of a heart operation for a condition based on patient's age, current condition, and health status. Essentially, we can say that the role of big data in the health sector is to manage data sets related to healthcare, which are complex and difficult to manage using current hardware, software, and management tools. In addition to the burgeoning volume of healthcare data, reimbursement methods are also changing. Therefore, purposeful use and pay based on performance have emerged as important factors in the healthcare sector. In 2011, organizations working in the field of healthcare had produced more than 150 Hexabytes of data, all of which must be efficiently analyzed to be at all useful to the healthcare system. The storage of healthcare related data in EHRs occurs in a variety of forms. A sudden increase in data related to healthcare informatics has also been observed in the field of bioinformatics, where many terabytes of data are generated by genomic sequencing. There are a variety of analytical techniques available for interpreting medical, which can then be used for patient care. The diverse origins and forms of big data are challenging the healthcare informatics community to develop methods for data processing. There is a big demand for technique that combines dissimilar data sources. A number of conceptual approaches can be employed to recognize irregularities in vast amounts of data from different datasets. The frameworks available for the analysis of healthcare data are as follows:

- **Predictive Analytics in Healthcare :** For the past two years, predictive analysis has been recognized as one of the major business intelligence approaches, but its real world applications extend far beyond the business context. Big data analytics includes various methods, including text analytics and multimedia analytics. However, one of the most crucial categories is predictive analytics which includes statistical methods like data mining and machine learning that examine current and historical facts to predict the future. Predictive methods which are being used today in the hospital context to determine if patient may be at risk for readmission. This data can help doctors to make important patient care decisions. Predictive analysis requires an understanding and use of machine learning, which is widely applied in this approach.
- **Machine Learning in Healthcare :** The concept of machine learning is very similar to that of data mining, both of which scan data to identify patterns. Rather than extracting data based on human understanding, as in data mining applications, machine learning uses that data to improve the program's understanding. Machine learning identifies data patterns and then alters the program function accordingly.
- **Electronic Health Records :** EHR represents the most widespread health application of big data in healthcare. Each patient has his/her own medical records, with details that include their medical history, allergies diagnosis, symptoms, and lab test results. Patient records are shared in both public and private sectors with healthcare providers via a secure information system. These files are modifiable, in that doctors can make changes over time and add new medical test results, without the need for paper work or duplication of data.



Four Vs of Big Data in Healthcare :

Four primary attributes (shown in Fig. 1) that are associated with big data: volume, velocity, variety, and veracity.

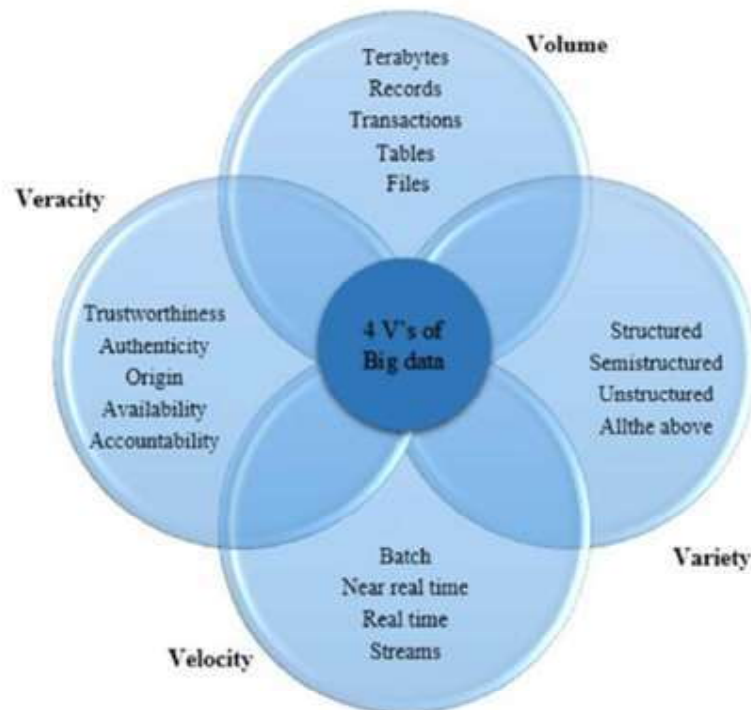


Fig. 1 : Big data's four Vs in healthcare.

- Volume:** Big data is a term to referring to huge volumes of collected data. There is no fixed threshold for the volume of this data. Typically, the term is used with respect to massive-scale data which must be managed, stored, and analyzed using traditional databases and data processing architecture. The volume of data generated by modern IT and the healthcare system has been growing and is driven by the reduced costs of data storage and processing architectures and the need to extract valuable insights from data to improve business processes, efficiencies, and services to consumers.
- Velocity:** Velocity, which represents primary reason for the exponential growth of data, refers to how fast data is collected. Healthcare systems are generating data at increasingly higher speeds. In the volume and variety of the structured or unstructured data collected, the velocity of the generation of this data after processing requires a decision based on its output.
- Variety:** Variety refers to the form of the data, i.e., unstructured or structured, text, medical imagery, audio, video, and sensor data. Structured data information includes clinical data (patient record data), which must simply be collected, stored, and processed by a particular device. Structured data comprises just 5% to 10% of healthcare data. Unstructured or semi-structured data includes e-mails, photos, videos, audios, and other



health related data such as hospital medical reports, physician’s notes, paper prescriptions, and radiograph films.

- **Veracity:** The veracity of data is the degree of assurance that the meaning of data is consistent. Different data sources vary in their levels of data credibility and reliability. The outcomes of big- data analytics must be credible and error-free, but in healthcare, unsupervised machine learning algorithms make decisions that are used by automated machines based on data that may be worthless or misleading^[4]. Healthcare analytics are tasked with extracting useful insights from this data to treat patients and make the best possible decisions.

Impact of Big Data on the Healthcare System :

The potential of big data is that it could revolutionize outcomes regarding the most suitable or accurate patient diagnosis and the accuracy information used in the health informatics system. As such, the investigation of huge amounts of information will have a powerful effect on medicinal services framework in five respects, or “pathways” (shown in Fig. 2). Improving outcomes for patients with respect to these pathways, as described below, will be the focus of the healthcare system and will directly impact the patient.



Fig. 2 : New pathways that have impact on big data.

- **Right Living:** Right living refers to the patient living a better and healthier life. By right living, patients could manage themselves by making the best decisions for themselves, based on the utilization of information mining better choices and enhancing their wellbeing. By choosing the right path for their daily health, regarding their diet, preventive care, exercise, and other activities of daily living, patients can play an active role in realizing a healthy life.



- **Right Care :** This pathway ensures that patients receive the most appropriate treatment available and that all providers obtain the same data and has the same objectives to avoid redundancy of planning and effort. This aspect has become more viable in the era of big data.
- **Right Provider :** Healthcare providers in this pathway can obtain an overall view of their patients by combining data from various sources such as medical equipment, public health statistics, and socioeconomic data. The accessibility of this information enables human service providers to conduct targeted investigations and develop the skills and abilities to identify and provide better treatment options to patients.
- **Right Innovation :** This pathway recognizes that new disease conditions, new treatments, and new medical will continue to evolve. Likewise, advancements in the provision of patient services, for example, upgrading medications and the efficiency of research and development efforts, will enable new ways to promote wellbeing and patient health via national social insurance system. The availability of early trial data is important for stakeholders. This data can be used to explore high-potential targets and identify techniques for improving traditional clinical treatment methods.
- **Right Value :** To improve the quality and value of health-related services, providers must pay careful and ongoing attention to their patients. Patients must obtain the most beneficial results identified by their social insurance system. Measures that could be taken to ensure the intelligent use of data includes, for example, identifying and destroying data misrepresentation, manipulations, and waste, and improving resources.

Conclusion :

In this paper, we have provided an in-depth description and a brief overview of big data in general and in healthcare system, which plays a significant role in healthcare informatics and greatly influences the healthcare system and the big data four Vs in healthcare. We also proposed the use of a conceptual architecture for solving healthcare problems in big data using Hadoop-based terminologies, which involves the utilization of the big data, generated by different levels of medical data and the development of methods for analyzing this data and to obtain answers to medical questions. The combination of big data and healthcare analytics can lead to treatments that are effective for specific patients by providing the ability to prescribe appropriate medications for each individual, rather than those that work for most people. As we know, big data analytics is in the early stage of development and current tools and methods cannot solve the problems associated with big data. Big data may be viewed as big systems, which present huge challenges. Therefore, a great deal of research in this field will be required to solve the issues faced by the healthcare system.

References :

1. A. Gandomi and M. Haider, Beyond the hype: Big data concepts, methods and analytics, *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.
2. A. O’Driscoll, J. Daugelaite, and R. D. Sleator, “Big Data”, Hadoop and cloud computing in genomics, *Journal of Biomedical Informatics*, vol. 46, no. 5, pp. 774–781, 2013.



3. C. L. P. Chen and C. Y. Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on big data, *Information Sciences*, vol. 275, pp. 314–347, 2014.
4. M. Herland, T. M. Khoshgoftaar, and R. Wald, A review of data mining using big data in health informatics, *Journal of Big Data*, vol. 1, no. 1, p. 2, 2014.
5. D. H. Shin and M. J. Choi, Ecological views of big data: Perspective and issues, *Telematics and Informatics*, vol. 32, no. 2, pp. 311–320, 2015.
6. B. Saraladevi, N. Pazhaniraja, P. V. Paul, M. S. Basha, and P. Dhavachelvan, Big data and Hadoop-A study in security perspective, *Procedia Computer Science*, vol. 50, pp. 596– 601, 2015.

Cite This Article:

Ms. Chandana Malavade, (2022). Big Data Analytics for Healthcare Industry: Impact, Applications, and Tools. Aarhat Multidisciplinary International Education Research Journal, XI (II), 297-303