IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Special Issue-8, Apr 2019 E-ISSN: 2347-2693

Big Data Analytics in Health Care System

M. Vijayalakshmi[1]*, S. Kanagasankari[2]

[1]Department of Computer Science, Sri Sarada College for Women, Tirunelveli-11, India
[2]Department of Information Technology, Sri Sarada College for Women, Tirunelveli-11, India

Corresponding Author: vijimarthandam@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7si8.151153 | Available online at: www.ijcseonline.org

Abstract— In Health Sector real time decision making can save a person's life. To treat the patient who has disease the doctor needs to know the history [both family and medical] of the patient. Timely decision making is very much important to increase the quality of health care.The patient details are scattered among various stake holders [doctor, scan center, Insurance, medical store] of health industry.

Availability of patient data at one place can help the doctors to make diagnosis of disease easily. At the time of emergency doctors need not repeat lab test. Accumulation of patient details at one place which is collected from various resources must be stored in the format for analysis and the data extractedfor effective decision making. The accumulated data can be dealt with big data tools and techniques.The application of Big Data Analytics in Health Care helps to lower health service costs while improving patient care process. This paper deals with the dimensions of Big Data, Tools and Techniques of big data and a Model for Health Care using Big Data.

Keywords— Big Data, Health Care, Big Data Analytical Tools.

I. INTRODUCTION

The term "big data" refers to data sets, whose size is beyond the ability of typical database software tools to Capture, Store, Manage and Analyse. Big Data will range from a few dozen terabytes to multiple petabytes [6]. It aims to answer questions that were previously unanswered.

Big Data Analytics can be applied in Healthcare under medical decision support structures, precise analytics applied for patient profile, revised and reformed medicine, performance based pricing for personnel, analysis of disease forms, improve public health, health care forums, health insurance issues, and Research issues to decrease the avoidable deaths[3][4].

Furthermore, healthcare digitization with integrated analytics is one of the next big waves in healthcare Information Technology (IT) with Electronic Health Records (EHRs) being a crucial building block for this vision. With the introduction of EHR incentive programs, healthcare organizations recognized EHR's value proposition to facilitate better access to complete, accurate and sharable healthcare data, that eventually leads to improved patient care.

The explosion of the Internet of Things (IoT) and its ability to provide real-time monitoring and expedited access to care is one of the driving factors for its adoption in healthcare. For instance, harnessing the power of big data analysis and genomic research with real-time access to patient records could allow doctors to make informed decisions on treatments [10] [11].

In every field big data is associated with 6 'Vs' volume, velocity, variety, variability, veracity and value [13]. Healthcare is the prime example of 6 'Vs' of data.

Volume (Data in Rest): It refers to the vast amount of data generation these days from various sources [7].Now a days Health care data are stored in Yotta Bytes which is 1024 Zetta Bytes. It refers to the vast amount of data generation from various sources of Health Sector like Payers(insurance Company), Providers(Doctors, Nurses, Pharmacists, other Care Providers etc.) and Patients(Real time Monitoring).

Velocity (Data in Motion): The Speed of data generation and the Speed at which data moves [14][8]. Healthcare data generated with a high pace from various medical sources such as clinical data from Clinical Support Systems, data from physician's Prescription, laboratory reports, Medical images, Pharmacy information, Administrative data, Insurance data, and Patients data from Electronic Patients Records (EPRs) [12].

Variety (Data in Many Forms): Big data is a combination of different kind of data that includes text and images from MRI, X-ray, Radiology, Sensor data, Videos, Voice recordings, etc.

Variability (Data in Highlight): Inconsistency of the data set can hamper processes to handle and manage it. In terms of health care Variability can occur in all aspects of care delivery including Patient Home Care and Clinical Care. Two main areas in which variability occurs in Health Care are Operational Variability and Knowledge Variability. Operational variability refers to the care delivered by Providers. When Variability occurs Patients can be harmed. For example when the patient is injected with alternate identical medication when two different medications are stored close to each other and in similar packaging. Knowledge variability refers to the death of patient due to overdose of medications. This occurs the medical professional can no longer rely on their training and experiences.

Veracity (Data in Doubt): Since big data is a collection of the vast amount of data, it is very tough to extract out the genuineness of the data or obtain any important information out of it.

Value (Data in Decision making): Value is the important V out of these 6 'Vs' of big data. As the rate at which big data is generated is very high, it doesn't make any sense if we can't extract out any valuable or essential information out of it.

II.TOOLS AND TECHNOLOGY

Health care industry produces huge volumes of structured data, semi-structured data and unstructured data. The aim of Big Data Analytics in Health sector is to reduce the cost with no compromise in the quality of treatment [5][9]. There are various tools which can be used for big data management from data acquisition, data storage to data visualization. Few of the tools which are used for different purposes are described below:

Hadoop: Hadoop developed by apache for ascendable, reliable distributed processing [1][15]. Hadoop mainly consists of HDFS and Map Reduce. The Hadoop distributed file system used to store and process the large amount of data in distributed manner

Map Reduce: This is the programming environment that permits larger jobs implementation scalability against group of server. Map Reduce is used for parallel processing of applications.

Pig: This is another analytical tool which permits the query execution over data stored on a Hadoop instead of a SQL.
Hive: Hive is like SQL, works on the datasets stored in Hadoop.

Spark: Spark supports open source environment which increase the computing power which leads to superiority then Hadoop. Spark is designed for explicit applications like machine learning algorithms and natural language processing. Hadoop provides cluster storage approach, whereas Spark provides scalable data analytics platform with in-memory computing.

Storm: Storm is the implementation of Map-Reduce concept of Hadoop. Storm is used for real-time data analysis. Storm uses no storage concept, which simply tells all about semi-structured, un-structured and structured data together.

HPCC: High Performance Computing Cluster also uses the Map Reduce framework for the analysis of large amount of data. It uses HPCC distributed file system. Major two advantages provided by HPCC over Hadoop is scalability and speed. In support of decision making HPCC is used by Elsevier to boost its logical and critical skills for SciVal.

HBase: A non-relational (NoSQL) databases that runs on top of HDFS. Apache HBase is an open source NoSQL database environment that provides real time read and write access to those large databases.

SAP-HANA: In-memory computing for big data SAP devised a new tool HANA, which processes on block of the data by using advanced parallel architecture and algorithms for faster speed.

Sky Tree: It is a high-performance data analytics and machine learning platform which focuses specifically on big data analytics and handling.

R Tool: R is a well-known programming language and software tool for graphic and statistical computing based data visualization.

Tableau: Tableau is the tool is used to visualize the result in the form of charts, maps, graphs and many other graphics form. A desktop application is available for visual analytics.
Ember Charts: Ember Charts features scatter charts, bar, pie, time series. It's very well-designed and easy touse tool.

Tangle: It is a data visualization tool beyond the visualization, allows the designers and program developers to generate reactive programs that gives a better understanding of data relationships.

III.BIG DATA MODEL FOR HEALTH CARE SYSTEM

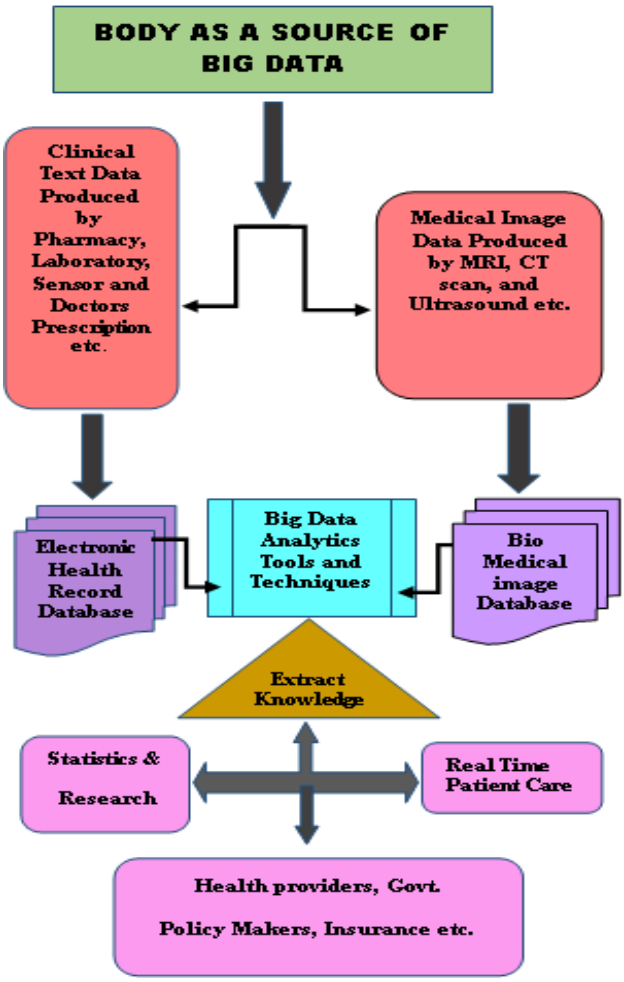

Fig.1. Health Care System

Our model receives (Fig1) patient data from various resources and analysed the information using big data technologies. The information stored in Electronic Health Record can be accessible any time anywhere [2].The outcome of this model can be used by all the stakeholders of the health care industry. There by reducing the clinical cost and helps to build efficient health care system.

IV.CONCLUSION

The big data model for health care system, support doctors to give real time patient care. The insight gained by analysing massive amount of health care data can be used to improve the quality and efficiency of health care system. The major advantage of utilizing Big Data Analytics for health care support system is, integrated health support system with analytics. Thus this system helps in decision making will improve Quality of Life.

REFERENCES

[1]http://hadoop.apache.org/last access 17.01.2014
[2]"EHR incentive programs," 2014. Available: https://www.cms.gov/Regulation-and-Guidance/ Legislation/HER Incentive Programs/index.html.
[3]"Public Law 111-148-Patient Protection and Affordable Care Act," U.S. Government Printing Office (GPO), 2013.
[4] W.Raghupathi and V.Raghupathi,"Big Data Analytics in Healthcare:promise and potential," Health Information Science and Systems,Vol.2,pp.1-10,2014.
[5]H.Chen,R.H.L.Chisng and V.C. Story," Business intelligence and analytics: from big data to big impact," MIS Quarterly,Vol.36,pp.1165-1188,2012.
[6]A.Belle,R.Thiagarajan,S.M.R.Soroushmehr,F.Navadi, D.A.Beard and K.Najarian,"Big Data Analytics in Healthcare,"BioMed Research International,vol.2015, pp.1-16,2015.
[7] Transforming health care through Big Data:"strategies for leveraging big data in health care industry," Institute for health technology information,2013.
[8] J.Bain,U.Topalogun and F.YU,"Towards Large-Scale Twitter Mining for Drug related Adverse Events,"SHB,vol.12,pp.25-32,2012.
[9] W.Raghupathi and V.Raghupathi,"An Overview of Health Analytics,"J.Health Med,Informat.,vol.4, pp.1-11,2013.
[10]Data Driven Healthcare Organizations Use Big Data Analytics for Big Gains, IBM Software,pp.1-8,(2013). http://www03.ibm.com/industries/ca/en/healthcare/documents/data_driven_healthcare_organizations_use_big_data_analytics_for_big_gains.pdf
[11]Health IT Outcomes,"The Power of Real Time Analytics at the point of Care,"December 2015.
[12]Andreu-Perez,J., Poon, C.C., Merrifield, R.D., Wong S.T. and Yang,G.Z.,2015.Big data for health. IEEE journal of biomedical and health informatics, 19(4), pp.1193-1208.
[13]Gantz J.Reinsel D,Extracting value from Chaos IDC iview,2011,pp.1-12.
[14]Mayer –Schonberger V, and Cukier K, Big data: a revolution that will transform how we live, work and think. Boston: Houghtan,Miffin Harcourt;2013.
[15]A.Katal, M.Wazid, and R.Goudar,"Big data issues,challenges,tools and good practices," in the Sixth international conference on contemporary computing,Aug.2013.pp.404-409.

AUTHORS PROFILE

I am Smt.M.Vijayalakshmi,M.SC.,M.Phil.,SET., at present I am working as an Assistant Professor in Department of Computer Science of Sri Sarada College for Women, Tirunelveli-11,Tamilnadu. I have 12 Years experience in the Discipline of Computer Science. My research Broadarea is Big Data Analytics. I am interested in developing Software Models for Health Care Systems.

I am Smt.S.KanagaSankari,M.C.A.,M.Phil., at present I am working as an Assistant Professor in Department of Information Technology of Sri Sarada College for Women, Tirunelveli-11, Tamilnadu. I have 12 Years experience in the Discipline of Computer Science. My research Broad area is Big Data Analytics. I am interested in developing Software Models for Health Care Systems.

