

Improving Healthcare-Big Data Analytics for Electronic Health Records on Cloud

Sreekanth Rallapalli and Augustin Minalkar

Botho University, Gabarone, Botswana

Email: {rallapalli.sreekanth, augustin.minalkar}@bothouniversity.ac.bw

R. R. Gondkar

Brindavan College, Bangalore, India

Email: rrgondkar@gmail.com

Abstract—Increase in storing the Electronic Health Records (EHR) of patients has developed a large scale datasets. Visual analytics plays an important role in decision making process because humans have ability to quickly gain insight through visual analysis. Existing visual analysis tools and techniques do not properly fit the big data. There is a need to design an efficient data visualization tools. These tools should consider the quality of data and presentation to facilitate navigation. Visualization can assist in descriptive, predictive and prescriptive type of analytics. Users like to visualize data processed in clouds have the same experience and feel as though data were processed locally. The batch-job model provided by most cloud environments would now include MapReduce as a backend to features provided in interfaces with which users are familiar. In order to improve the healthcare for patients health care organizations should provide accurate data to the right patient and at the right time. Big data analytics will benefit most of the components of health care system. In this paper we first study the big data impact on healthcare and then analyze how data analytics can improve patient care and reduce costs. Data analytics of Electronic Health Records on cloud can reduce much more costs to the Healthcare organization.

Index Terms—datasets, visual analysis, MapReduce, EHR, cloud, big data

I. INTRODUCTION

With the number of hospitals, clinics, administrators, generating huge amounts of data and these data need analyses to extract useful information. Visualization tools provide to cope up with these increasing amounts of data. Quality of data is to be considered by the visual tools to facilitate the interaction between the data and the system. The difficulty of Visual Analytics with reference to Big Data is extracting useful information and to visualize the data when many items are displayed in the screen.

In order to advance the visual extensibility, dimension reduction, clustering, machine learning and data mining tools can act as an computational methods. These methods are very much useful for the raw data which has

been generating from various sources. These methods do not support the visual analytics for the Electronic Health records as the data is very huge [1]. Health care organizations need to derive more value from these data by using suitable visualization-based data discovery tools. As per David McCandless, by visualizing the information we turn it into a landscape that can be explored with our eyes. It is like an information map. There are many tools available currently which can visualize from real world applications [2]. Healthcare practitioners have to make decisions based on rich set and real time information which can help doctors; physicians determine the most efficient course of action for their patients. The data analytics on the Electronic health records also help to discover new treatment options and potentially save lives of patients by using the latest treatment options which are made available on time.

II. IMPACT ON HEALTHCARE FROM BIG DATA

In order to improve the life expectancy of the people, there is a need to analyze the health data of all the population in a geographical area in an efficient manner. As per the institute for health technology transformation 150-200 Exabyte of healthcare data was recorded in recent years [3]. Efficient analytics on the health care data will yield to generate the new projects which are linked with improving the life expectancy and quality of health care. The possibility of outbreak of particular infections can be known by analyzing the data related to the people who are affected to these infections. Big data analytics on the health care records can alert the doctors, physicians and other healthcare practitioners with an information that a patient with the specific symptoms need an quality medical care.

Healthcare sector need to improve the operational efficiency and increase profitability by improving the quality of healthcare and thereby reducing the costs. By utilizing the Big Data technology healthcare sector can efficiently and effectively report their activities. Government sectors can benefit by knowing the reports of unnecessary medical bills and can reduce the

readmissions for healthcare providers [4]. Big data can optimize costs by reducing the number of preventable readmissions, improving physician and worker performance and drive greater operational efficiencies.

As per the MCKinsey Global Institute estimates that big data in healthcare has potential value of more than \$300 billion USD an year [5].

III. BIG DATA IMPROVES PATIENT CARE

Data source from patient EHR along with other sources of the patient information have to be analyzed so that the patient can have the prescribed treatment. Big data will improve the patient care by analyzing the data source coming from various sources. Healthcare providers can achieve the personalized treatment of the patients by having a better view of the patient data which are arriving from various sources. As now the data will be moved to cloud it is easy for healthcare providers to collect and analyze the patient medical data. Big data also help to track the patients in a better way beyond the hospital premises. As per the Institute of Health Technology and Transformation 80 percent of the Electronic Health Data is Unstructured. Big Data Hadoop along with NoSQL databases can help in storage, processing, analyzing and extraction of huge unstructured clinical data [6].

IV. DATA ANALYTICS WITH BIGDATA FOR EHR ON CLOUD

Big data refers to tools that can manage the large and complex data sets which are difficult to manage by using traditional management tools. Big data is not just about the size of data but it can help to analyze complex, noisy, heterogeneous voluminous data and aims to provide some insights on previously unanswered questions [4]. There are few challenges for analyzing the Healthcare data. This includes making the unstructured clinical notes meaningful in the right context. Efficiently handling large volumes of medical image data and extracting useful information. Patient's data have to be captured through sensors, their changing behaviors data need to be captured. Big data analytics of Electronic Health records on cloud will help the health care organizing reduce their costs and improve outcomes through effective decision making. The main goal of the Big data analytics in healthcare is to provide right information and treatment for the patients at the right time. It can also help to improve the personalized care to patients. It can potentially benefit all the components of the health care system from provider, payer, patient and management.

Big data analytics [7] can help the health care organizations to identify the patients who will be admitted to a hospital within the next year by utilizing the historical claims data. Due to unnecessary hospital admissions billions of money was spent and this can be minimized. The main goal of the analytics is to identify the patients who are at high risk of having medical issues and providing them a proper treatment. Another

challenging goal is to find the efficient algorithms where it can find the number of days that a patient will spend in a hospital during coming years.

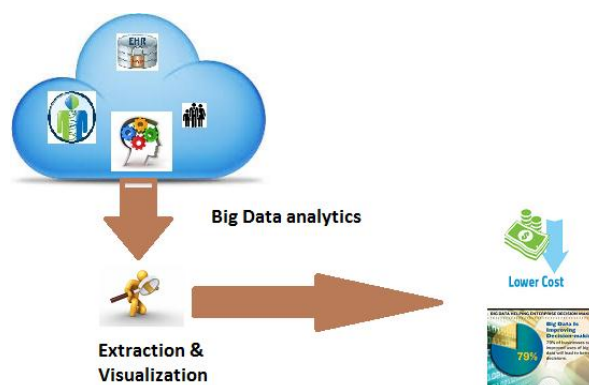


Figure 1. Big data analytics

V. AREAS OF ELECTRONIC HEALTH RECORD DATA MANAGEMENT

Electronic health records have developed large scale data sets over a period of time. These data must be efficiently analyzed to extract knowledge hidden in it. In other words data mining on these records will help the health care organizations to examine large, relational databases from multiple angles. This will help to identify the correlations and various patterns matching a particular patient's medical history.

In order to obtain a better and clear picture of patients medical history descriptive analytics will be useful. Electronic health records can be analyzed effectively to find out the patients with similar medical cases and provide specialized treatments.

In order to analyze the future medical case of the patients current data and historical data will be helpful for predictive analytics. For better decision making for the patients prescriptive analytics will be efficient. An efficient big data platform and powerful applications which support the platform is required to perform all the above said analytics that can help to achieve the goals of the big data analytics in healthcare sector.

VI. HADOOP FRAMEWORK FOR PROCESSING HEALTH CARE DATA ON CLOUD

Healthcare data analytics is rapidly emerging with huge potential for organizations to provide healthcare by reducing the costs and improving healthcare decisions. Analytics help in gaining the information to improve decision making by using advanced data mining tools. A healthcare information and management system uses big data analytics for operational excellence. As Electronic Healthcare records [8] are unstructured in nature, big data adoption is gaining importance in processing and visualizing the data. Big data utilizes hadoop framework to process the large data sets in distributing computing environment.

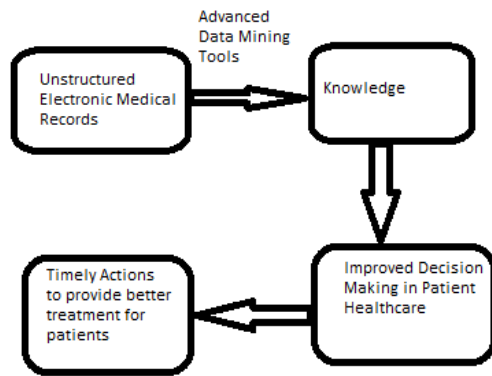


Figure 2. Building blocks for intelligence platform

As healthcare data is collected from different sources, and most of the medical data is unstructured it is suggested that all the data to be brought to centralized location. Healthcare Data can be from Electronic medical records, Health information exchanges, and clinical data from laboratories, patient claims data from claim systems and many other such systems.

Hadoop is an open source system that supports data intensive distributed applications [9]. As said earlier EHR data is unstructured and it is often needed to query such large unstructured data from distributed data sets. MapReduce [10] can be used to program jobs that can execute queries to hadoop based databases such as Hive, Hbase, Cassandra, MongoDB. Fig. 3 shows the framework to process Healthcare data using commodity server cloud on Hadoop environment. Health care data from various sources like drug research, social media, patient records from clinics, hospitals, Laboratory test results, claim data, home monitoring data, data from mobile applications is collected and then all these data are moved to cloud environment. The data is then moved to Hadoop environment for reducing. MapReduce programming can be used to reduce these huge medical data into data chunks for better results. The data is then further put into commodity server clouds. The data in server is then moved to end user Health care systems for better visualization of patient data.

Visual analytics can be performed on the data using various available tools which helps the doctors,

physicians and other people in healthcare for better patient data. This will help the healthcare sector to provide specialized treatment for the patients. Also moving the data to cloud will reduce cost to the organizations.

VII. RESULTS

Hadoop uses commodity hardware and by using it in health care data storage the cost of storage can be reduced. Efficient storage and computing can be performed on a single platform thereby reducing the cost and enabling efficient data processing with minimal downtime. The following table shows the time taken to sort the Electronic health records using Hadoop architecture.

TABLE I. TIME TAKEN TO SORT RECORDS USING HADOOP ARCHITECTURE

No of Electronic Health Records	Size	Time to Sort
14 Billion	1400 GB	60 Sec
15 Billion using MapR	1.5 TB	59 Sec

VIII. CONCLUSION

In this paper we study how big data analytics impact on health care. Analytics on health care data collected from various sources are gathered into cloud platform. EHR data is processed on cloud platform using Hadoop platform which uses commodity cloud servers. These results are summarized and then sent to healthcare information system for effective visualization. Big data analytics will help the health care organizations to reduce costs and improve effective decision making. It also helps health care organizations to Genome Processing and DNA Sequencing, personalized treatment planning, assists diagnostics, fraud detection, monitor patient vital signs. Cloud computing is an platform for storage of Health care data minimizing the costs to the organizations. Hadoop architecture helps to reduce the cost by using the commodity hardware for storage and computing. Health care organizations should effectively use the cloud services and analyze the data using big data analytics.

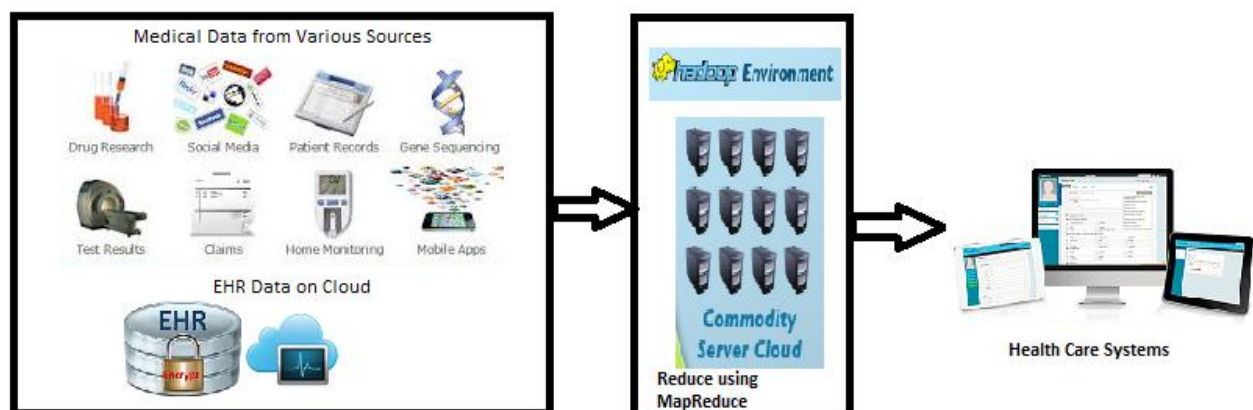


Figure 3. Framework to process healthcare data using commodity server clouds on Hadoop environment

REFERENCES

- [1] D. Sommer, R. L. Sallam, and J. Richardson, "Emerging technology analysis: Visualization-based data discovery tools," June 17, 2011.
- [2] D. Fisher, "Incremental, approximate database queries and uncertainty for exploratory visualization," in *Proc. 1st IEEE Symposium on Large-Scale Data Analysis and Visualization*, 2011.
- [3] IHT² Institute for Health Technology Transformation. [Online]. Available: <http://ihealthtran.com>
- [4] J. M. Sen and C. K. Reddy, *SIAM International Conference on Data Mining*, Austin Tx, 2013.
- [5] [Online]. Available: www.mckinsey.com/.../big_data_the_next_frontier_for_innovation
- [6] "Data analytics can improve patient care and reduce costs," *HP Big Data and Health Care*, 2014.
- [7] S. Connolly and S. Wooldge, "Harnessing the value of big data analytics," *Teradata*, 2013.
- [8] W. Raghupathi and S. Kesh, "Interoperable electronic health records design: towards a service-oriented architecture," *e-Service Journal*, vol. 5, pp. 39-57, 2007.
- [9] HDFS. (2009). Hadoop distributed file system. Architecture. [Online]. Available: <http://hadoop.apache.org/common/docs/current/hdfsdesign.html>
- [10] J. Dean and S. Ghemawat, "Mapreduce: A flexible data processing tool," *Commun. ACM*, vol. 53, no. 1, pp. 72-77, 2010.



Testing.

R. R. Gondkar received his PhD from Kuvempu University, India. He is currently working as Professor & HoD, Brindavan college Bangalore. He has published more than 40 papers in national and international conferences and journals. He is Research guide to various universities in India. He is guiding Phd students in area of Cloud computing, Data mining, Artificial intelligence, Image processing and Software



Augustin Minalkar received his Masters Degree in Computer Applications from Mysore University in the year 2002. He has 11 years of teaching experience in universities in India and Botswana. He is currently HoD of Department of FoC, Botho University. He has attended three international conferences. His area of interest in research is big data, Data warehousing and Data mining, RDBMS and Networking.



Sreekanth Rallapalli received his Masters Degree in computer Applications in 2001, and MPhil (Computer Science) in 2006 from Bharathidasan University, India. He is currently pursuing his PhD from Bharathiyar University, India. He has 15 years of work experience in teaching various universities in India, Zhejiang University, China, AlMusanna college of technology, Sultanate of Oman. Currently he is servicing as Senior

Lecturer, Botho University, Gabarone, Botswana, Africa. He has published more than 20 research papers in National and International Journals and conferences. His area of research include Big Data analytics, Cloud Computing.