Big Data: An Ultimate Solution in Health Care

Kiran Baby
Computer Science and Engineering
Mar Baselious College of Engineering and Technology
Trivandrum, India

Aswathy Ravikumar
Computer Science and Engineering
Mar Baselious College of Engineering and Technology
Trivandrum, India

ABSTRACT
Digital healthcare promises for better, efficient and cost effective solutions to existing health problems. The huge amount of data produced per hospital every year is increasing at an alarming rate, and there is a requirement for data to be preserved for decades for future references and study purposes. Big Data is the only solution for secure and efficient storage of data. This paper is a study on existing Big Data technologies and how they can be utilized in health care sector for efficient data storage and, inevitably better utilization of stored data for treatments in future.

General Terms
Big Data, Security, Health Care.

Keywords
Digital healthcare; alarming rate; preserved; Big Data technologies; treatments

1. INTRODUCTION
Protecting, safeguarding and extending the life span of human lives have always been the ultimate aim behind all the discoveries and inventions. In this computerized era, where everything went digital, health industry is not far behind. Digital healthcare promises for better, efficient and cost effective solutions to existing health problems [1]. The amount of data produced in a hospital per patient includes the medication reports, medical diagnosis, lab results, insurance and billing details. It is predicted that the average amount of data per hospital will increase from 167TB to 665TB in 2015, because of the enormous growth of medical images and electronic medical records. This data is increasing in volume each and every second and cannot be deleted or destroyed as soon as the patient is discharged, rather need to be stored for decades so that medical professionals can improve patient care and reduce treatment costs by extracting relevant clinical information from vast amounts of data, and data can be preserved for future reference or study purposes. Big Data becomes the only option in this situation.

Big Data is a term that refers to any collection of data sets, so large and complex, such that it rules out the option of processing by traditional data processing applications. By utilizing Big Data there is a huge availability of data in respective fields, thereby increasing the expectations in making use of these large volumes of data to improve the existing treatments, reduce the cost, better outcomes. The ability to process multiple high-speed data streams from multiple locations in real time can ultimately improve the healthcare efficiency and patient outcomes [2].

This paper is organized as follows. In section 2, the different Big Data technologies are described. A close study on Big Data solutions existing in e-Health sector is discussed in section 3. The final section concludes with a comparison on different existing Big Data technologies for e-health and summarizes the benefits of each.

2. LITERATURE SURVEY

2.1 Storm
Storm is a free and open source distributed real time computation system for processing fast, large streams of data [14].Storm adds reliable real time data processing capabilities. Storm topology takes streams of data and processes these streams in different ways by subdividing them according to the requirement. The source of data streams are termed “Spouts” and logic of processing these streams are termed “Bolts” as in Figure 1.

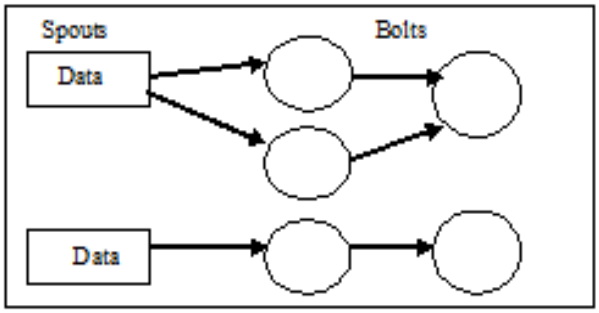

Figure 1.Storm data processing

There are mainly five characteristics that make Storm ideal for real-time data processing workloads. Storm is fast; capable of processing 1 million 100 byte messages per second per node [14]. They are highly scalable by using parallel calculations that can run across enormous number of machines. Storm is fault tolerant; even if a specific node dies, the work will be restarted on another node. They provide reliability by avoiding unnecessary replaying of messages and at the same time guarantee that each unit of data will be processed atleast once. Easiness of operation is another major advantage, once deployed, always easy to use. However, partitioning of data into streams of data is not always feasible as the data types may include scan results, images, which when partitioned will result in loss of data; therefore Storm is of limited application when health services are concerned.

2.2 Hadoop
Hadoop is an open-source software framework for processing large datasets [15]. Hadoop works on the principles of Batch processing. Hadoop framework provides both reliability and data motion to applications [2]. It is widely used due to its high performance, scalability, and low-cost option. Hadoop framework contains different modules as in Figure 2. Hadoop Common; is useful for other Hadoop modules as they contain libraries and utilities. Hadoop Distributed File System (HDFS), it is a machine with high bandwidth across the cluster, used for storing data. Hadoop Yarn is responsible for managing computed resources and proper scheduling for user applications. Hadoop MapReduce, is a large scale data processing programming model.

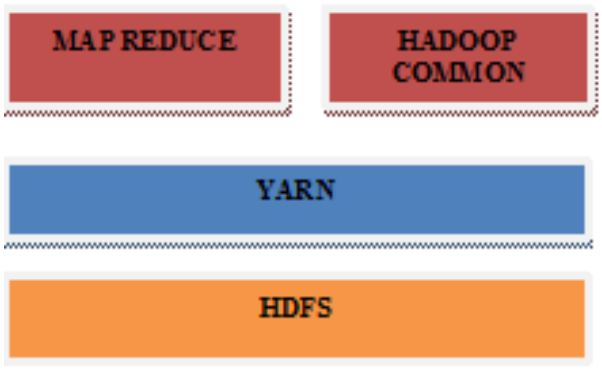

Figure 2.Major Modules of Hadoop

Hadoop is widely useful for generation of reports and future research works, as batch processing feature in Hadoop is very useful for offline processing, however it cannot be used for real time processing of e-health data [2].

2.3 Mumps

Massachusetts General Hospital Utility Multiprogramming System (Mumps) is a Big Data programming tool [12]. Mumps is a multiple user, highly scalable and designed to handle huge databases [2].Mumps are widely used mainly in areas requiring high reliability and availability like banking sector and in hospitals. It provides simple data abstraction, where reduction of data values into simpler forms and all these data can be structured into multiple dimensional arrays [2]. Mumps is a language combined with a database engine [2]. Its application is limited to areas requiring high performance and high demand databases that must support scarce or scattered data.

3. BIG DATA TOOLS IN E-HEALTH

In traditional model, data is loaded into a single warehouse; however this is not practical when data is used up by different enterprises like labs, insurance institutions and regulatory agencies where each one of them requires specific data and not all the stored related data. Big Data will break the traditional model [4]. Data federation is a solution, in which the Big Data architecture is based on collection of nodes within and outside the enterprise where data is accessed through a layer, integrates the data and logic [5].Proper segmentation of data for each requiring firm is a main concern in designing Big Data solutions for e-health. Security and privacy of data is another challenge in storage of high sensitive data.

Since data from different sectors related to health are stored into a single window, extracting and correlating the existing data to form any future conclusions is difficult. These challenges give rise to the thought that there need to be a means to interrelate the data.

In this fast growing computer era, technology is getting updated in a short span of time, so is the increasing number of new types of data, hence the Big Data solution to be used in an e-health sector should be flexible enough to deal with transporting and storing data from additional sources. The importance of any Big Data solution in health industry is the fast and efficient solutions to life threatening situations; hence any Big Data solution must be accurate enough to carefully query the entire relevant data with proper constraints and provide an optimum solution in very less response time.

3.1 BDEHS

Bdehs (Big Data e-Health Service) provides data federation and aggregation [3][5][6].A data flow in this system is mapped into a stream of data with additional processing stages. Data formats are initially converted into a common format provided security policies are signed along with the flows [2]. Data Federation is applied into its e-Health adaptation gateways [3][4], which is responsible for the data to flow into respective processing nodes having the data processing logics like filtering, iterative analysis , aggregation and so on [2]. Data sinking gateways are also present in this system which is used in handling situations like adverse treatment or drug effects, so that they are immediately brought into the notice of the officials. This exit gateways also helps in generation of data reports and can formulate solutions that require immediate decisions like in case of a detecting the spread patterns of an epidemic[2]. Major functional blocks of Bdehs service are shown in Figure 3.

Bdehs ensures Data Consistency, by eliminating any local differences and increasing the quality of data being used up. Bdehs allows health care researchers to study a large population for clinical studies, latest epidemic spread patterns and how to ensure localizing those spread patterns by early detection, by aggregating regional and global data [2]. Another benefit of combining regional and global data is identify the best practices for better and improved results in case of surgeries or any other critical treatments as they contain data from different regions and countries with best success rates.

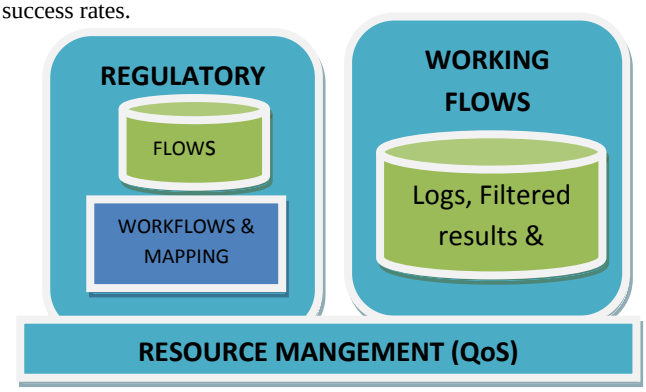

Figure 3.Bdehs Block Diagram

There are several other key capabilities for Bdehs.

3.1.1 Promises Security Monitoring

E-Health data includes patient's identification, medication details, diagnosis reports and so on. Storage and manipulation of this data is violation of one's privacy, hence this system deals with all such privacy concerns. When data is moved into Big Data engines, patient identity information's are removed according to Bdehs security policies. Added advantage of this ID removal is access to de-identified data thereby improving levels of self reporting as well as data sharing [2].The issue of using only relevant data or selecting and dealing with only needed data is now possible . It is possible to assign access priorities to individuals or organizations.

3.1.2 Operational Management

There are QoS Service Managers to control and coordinate the entire Big Data services. They monitor the entire transactions by keeping logs and reports of performance feedbacks [2]. Since there are large varieties of e-Health application flows, the only way to maintain them is by creating Bdehs service profiles [2]. These profiles can then specify the priorities,

usage parameters and all necessary categories, like describing specific interfaces needed by the participants for exchanging health related information [2]. Bdehs is flexible enough to sustain any future evolution in terms of schemas or structures or data types for transporting, manipulating and storing data.

3.2 Artemis

Successful implementation of Big Data in intensive care units has a big scope in health industry, as it leads to early detection and prevention of deadly medical conditions. The major challenges addressed by an ICU (Intensive Care Units) include, enormous amount of data being generated and the limited scope of the memory of the devices mounted on these ICU's which can store data approximately for 72 hrs only. The major devices used include ECG (electrocardiography) which measures the heart activity, SIPs (smart infusion pumps) which stores drug and nutrition infusion data, can produce more than 60 different types of data every 10 seconds [7]. Both paper and electronic charting mechanisms allow storage of only one value per hour of a heart rate [7].Currently the medical staff goes for a qualitative than a quantitative approach to judge the stability or instability of the patients [9]. However second to second activities need to be monitored for early detection and prevention of deadly situations.

Artemis is a clinical decision support mechanism which is an online platform [7]. The block level architecture of the Artemis platform is shown if Figure 4. It supports acquiring data and storage of this data for the purpose of on-line real-time analytics, retrospective analysis and in data mining [8].

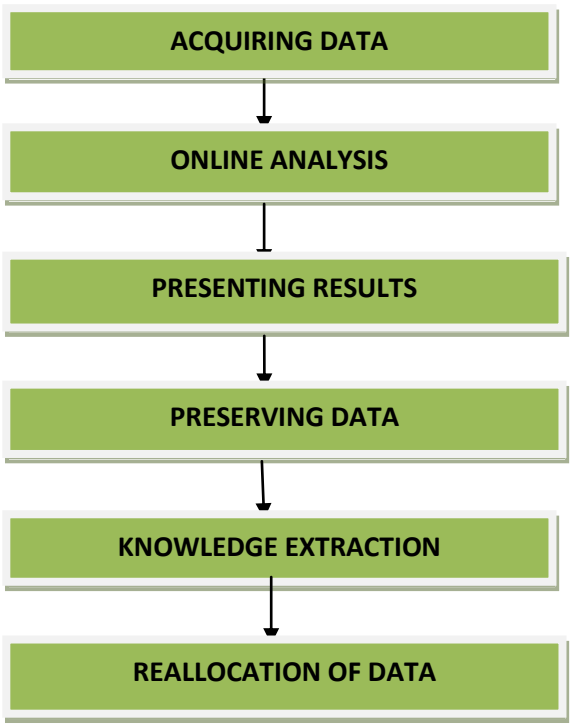

Figure 4.Artemis Block level Architecture

Acquiring data component continuously inputs data streams from medical devices and other patient relevant data like lab results, medication details, diagnosis reports, and insurance details and then forwards the data into an online analysis component, which is not necessarily present in the hospital , rather it can be anywhere , may be in a different country too. Online analytics component is responsible for aggregation, partitioning and storage in respective formats so as it can be accessible by other relevant medical staff or researchers and can be utilized for future reference. Artemis stores the original data together with the newly generated data, this is done in the preserving block [7].The Knowledge extraction component selects only relevant data and preserving only those data and avoids personnel details of the patient thereby preserving privacy. Reallocation or redeployment component provides with new algorithms to the online analysis component [7].

Ability to process different types of data streams from different patients across the globe and synchronize them all by an online analytic component can bring about drastic change in the healthcare and patient outcomes [9]. However preserving the privacy of all this stored data should also be an important factor of the system developers.

4. CONCLUSION

All the Big Data technologies discussed in the previous sections provide means and measures for efficient storage, retrieval and manipulation of enormous amount of data in a cost effective way. The usefulness of these technologies in health industry is however not fully explored. A comparative study on major features needed to use them in health sector is shown in Table1.

Table 1. Comparison of various Big Data Technologies

Big Data Technology	Major Components			
	Speed	***Online Processing***	***Scalability***	***Reliability***
Storm	High	Low	Medium	High
Hadoop	Medium	Low	Low	Medium
MapReduce	Low	Low	Medium	Medium
Bdehs	High	High	Medium	Medium
Artemis	High	High	Medium	High

Storm, Hadoop and MapReduce provides efficient offline processing of data but incapable of real time processing of data hence its usage for online application is limited; Storm has high speed processing because of its parallel computation capability. Hadoop also have reasonable processing speed because of its batch processing feature, while MapReduce has low processing speed as compared to other technologies. Storm, Hadoop and MapReduce ensures reasonable scalability and reliability but Big Data services like Bdehs and Artemis, which were solely developed for e-health proves to serve better in terms of processing, easiness of operation, reliability and scalability as they have facilities to keep logs and filtered results in case of Bdehs and online analysis and data preserving components in case of Artemis. The study on different Big Data innovations in Healthcare ensures for a better, efficient and cost effective healthcare solutions. The discussed Bdehs systems as well as the Artemis online analytical system both have wide areas of scope. Currently Artemis is limited to usage only in NICUs (Neonatal Intensive Care Units), which can in future deploy to other ICUs and clinics. Another promising application in terms of health sector is in the increasing trend of remote monitoring where specialized medical practitioners can provide consultation for surgeries in real-time. Research and work is also needed in developing similar platforms with more features and security and proper training need to be given to the medical practitioners or the care givers for a smooth operation of such systems in hospitals.

5. REFERENCES

[1] W. Liu and E.K. Park, "e-Health AON (Application Oriented Network)", IEEE International Conference on Computer Communication Networks, BMAN Workshop, Nausa, Bahamas, July 2013.

[2] Dr.W.Liu and Dr E.K Park, "Big Data as an e-Health Service", California State University-Chico, International Conference on Computing and Networking ,Published 2014.

[3] W. Liu, E.K. Park and Udo R. Krieger, "e-Health Interconnection Infrastructure Challenges and Solutions Overview", IEEE HealthCom-2012, the 14th IEEE International Conference on e-Health Networking, Applications &Services , Beijing , China, October 2012.

[4] W. Liu and E.K. Park, "Emerging Platform for Healthcare IT Services", IEEE International Conference on Computer Communication Networks 2010, WiMAN Workshop, Zurich,Switzerland, August 2010.

[5] W. Liu and E.K. Park, "e-Healthcare Security Solution Framework", IEEE International Conference on Computer Communication Networks, MobiPST-2012, Munich, Germany, August 2012.

[6] W. Liu and E.K. Park, "e-Health Service Characteristics and QoS Guarantee", IEEE International Conference on Computer Communication Networks, Workshop on Context-aware QoS Provisioning and Management for Emerging Networks, Applications and Services, Maui, HI, August 2011.

[7] Carolyn McGregor, "Big Data in Neonatal Intensive Care", University of Ontario Institute of Technology, Canada, IEEE Eng in Medicine and Biology Magazine, Published in June 2013.

[8] M. Blount et al., "Real-Time Analysis for Intensive Care: Development and Deployment of the Artemis Analytic System," IEEE Eng. in Medicine and Biology Magazine, vol. 29, no. 2, 2010, pp. 110-118

[9] C. McGregor et al., "Next Generation Neonatal Health Informatics with Artemis", User Centered Networked Health Care, A. Moen et al., eds., IOS Press, 2011, pp. 115-119

[10] US Department of HHS, "Health Information Technology: Initial Set of Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology", January 2010.

[11] US Committees on Energy and Commerce, Ways and Means, and Science and Technology, "Title IV - Health Information Technology for Economic and Clinical Health Act", January 16, 2009.

[12] ANSI Standard, "MUMPS Database and Language " implementation information available in http://mumps.sourceforge.net/

[13] P. Jacobs and T.W. Noseworthy, "National Estimates of Intensive Care Utilization and Costs: Canada and the United States," *Critical Care Medicine*, vol. 18, no. 11, 1990, pp. 1282-1286.

[14] http://storm-project.net/.

[15] http://hadoop.apache.org/