

Big Data Analytics for the Optimal Delivery of Primary Health Care in China

Muni Kumar N
Department of CSE
SKIT, Srikalahasti
A.P, India

Zongyu Li
Dalian, Liaoning
Province,
China

Jingwen Du
Dalian, Liaoning
Province,
China

Yinghui Zhang
Dalian, Liaoning
Province,
China

Abstract: Health care is one of the greatest concerns for any country with no exception to China. The country has a large healthcare demand gap due to ageing population, growing urbanization, proliferating lifestyle diseases, rapidly increasing consumer wealth, and the advancement of universal health care insurance. The inefficiencies and inequalities in the public health care access have pushed forward the need for creative thinking and innovative solutions to strengthen the same. Implementation of Electronic Medical Record (EMR), which is the digitized version of medical record which contains notes and documents about patient's identity, examination, therapy, treatments and other services that have been given to the patient. The exponential rise in data in health care has brought lot of challenges in terms of data transfer, storage, computation and analysis. The increase in the number and size of medical records provides new opportunities for identifying patterns and behaviors of the patients and helps the practitioners to make better use of the existing data. This paper discusses the use of Big Data Analytics and Hadoop to reveal the impact of the analytics to render better healthcare services to every citizen of china with optimal cost.

Keywords: Health care in china, Primary Health Care, Data Analytics, Big Data, Analytics in Health Care, Hadoop.

1. INTRODUCTION

China has an exponential increase in the Health care spending. Healthcare Spending are primarily driven by consumer's rapidly increasing incomes and the government's public health care reforms. It is estimated that, the China's health care spending are 5.4 percent of gross domestic product (GDP) in 2013, which is still much lower as compared to Organization for Economic Co-operation and Development (OECD) countries. The country also has a large health care demand gap due to an ageing population, growing urbanization, proliferating lifestyle diseases, rapidly increasing consumer wealth, and the advancement of universal health care insurance. Though, all these elements are driving overall health care market expansion, continued development in health care and better delivery cannot proceed without heavy investment by and strong support from the Chinese government. Public hospitals are the most common and dominant care providers in China. Currently, they rely heavily on medicine charges for funding. In long run, This automatically weakens the public social security system which leads to medical resources allocation inefficiency and prompts patients to switch to private hospitals.

Health strategies costs[1] in China will almost triple to more than \$1 trillion annually by 2020, driven by an aging population and government efforts to broaden insurance coverage making it the second-largest healthcare economy after the United States. More than 95% of China's 1.4 billion citizens have access to health insurance. However, benefits are rather limited and there are wide gaps in access to

healthcare between urban centers, with its far-superior hospitals, and rural centers, where living standards are considerably lower. To address these disparities, the Chinese government has been embarking on a variety of health reform initiatives like Regional medical and telemedicine services. Community medical and health service systems will still remain a priority investment area. In addition, closer connections will be established between medical services and medical insurance, management of population, and pension. While health reform is having a positive impact on China's citizens, these initiatives are creating a more difficult environment for multinational companies that want to expand into the China market. More now than ever, IT suppliers will need local partners to compete effectively. With the implementation of Telemedicine services in china, the huge gap in health care provision and quality can be minimized in the rural parts of the country and better health care delivery can be guaranteed with the combination of telemedicine and big data analytics.

To address China's huge irregularities in health care provision and quality, the government is anticipated to increase spending in rural areas for the establishment of clinic with sophisticated equipments, supply of sufficient drugs and for the patient insurance. Research studies reveal that, Ninety-six (96%) percent of China's population is covered by some form of medical insurance either through national health insurance or a rural co-operative medical insurance system. However, the coverage is very less with low premium contributions and patients continuing to pay large outpatient fees.

Private and foreign hospitals in china can bring leading medical technologies and advanced management, clinical practices, and service models to China. Also, Chinese officials have encouraged the development of private health insurance as a supplement to the public scheme. In fact, some local governments have formed partnerships with private companies to manage their public insurance plans. The expansion of private hospitals and private insurance should generate growth opportunities to related industries along the health care value chain, such as senior care, medical tourism, health management, medical devices, mobile health (mHealth) and other health information technology (HIT). The downside is that most of these new health services are not currently within the scope of social insurance coverage; the advancement of commercial insurance in China could help to mitigate these difficulties.

This paper is organized as follows: section II discusses about the Literature Review with focus on the Health care delivery in china, section III discusses about big data and its characteristics, section IV focuses on various big data technologies that can be applied to health care, section V deals with the application of big data analytics to health care and section VI concludes the work.

2. LITERATURE REVIEW

This literature aims to highlight the range of medical facilities provided by the health care in China. These literatures have emphasized upon the fact that Chinese hospitals and medical centers are categorized into three major types such as primary, Secondary or Tertiary[2]. According to Hou, these facilities are involved in the provision of preventive care, nominal health care and rehabilitation examination of different types of patients [3]. In the process of discussing the primary care quality among health care structures in Tibet, China, Wenhua Wang, Shi and Yin have highlighted that still differences in township and country hospitals is visible for better primary care[4]. The author reported that primary care in China was aimed to provide first contact and inclusive health services by lowering cost to different geographic areas but currently due to structural problems, the primary care system is facing with imbalanced situations. In contrary, Rohde, Cousens and Chopra have identified that current compressive primary care system in China based on dependable referral system is better than the countries with selective primary care [5].

On the other hand, “China allocates 5.6% to its total budget to health care sector. Further, China spends 35% of its allotted budget on the primary health and 65% of its budget on the secondary and tertiary health care”. The study gave credit to the new china health reforms in bringing a remarkable transformation in the primary health and balanced approach. Likewise, researches have further pointed out that secondary and tertiary cares are provided in public hospitals, institutions, and specialist outpatient clinics (SOPC). These care facilities are involved in surgery and emergency services.

In order to investigate further that how health care is currently being provided to people in China, past authors have also covered information about the types of insurance available for patients in China. According to The State Council Information Office of the People's Republic of China, Chinese health care system is based on multi-layered insurance system. The system offers range of supplementary medical insurance and commercial health insurance to the urban and rural populations for providing security and opportunities to avail different types of cares. In a research conducted by Hu and Ljungwall [6], authors have mentioned different names of the medical insurance systems introduced in China for helping their population. These include urban employee basic medical insurance (UEBMI), urban resident basic medical insurance (URBMI), rural cooperative medical system (RCMS). The report further highlighted that UEBMI is funded through the mix of the social pooling and individual medical savings while funds for URBMI is gathered from the households and only up to certain extent, government provides subsidies for it. Likewise, RCMS is also the outcome of combined individual payments, financial assistance from communal businesses and state-sponsored subsidies. The point directs that Chinese government does not bear the entire burden of health care services at present.

Additionally, it can be examined from the overview of past literature that at present preventive care is the vital component of Chinese health system but current reforms do not consider it effectively. The government has introduced anti-smoking initiatives for dealing with the diseases caused through drugs intake. Likewise, health-ageing programs, established of community support centre for cancer patients and immunization are also visible in the past studies. Xin has further added in the discussion by stating that reimbursements are also vital source of the preventive care in the China to increase affordability of patients[6]. Similarly, another example of preventive care in China can be examined from the HIV/STD prevention programs initiated by the health care system for increasing communication and awareness about the sex education among the population [7].

Thus, based on the findings gathered from different studies in the literature review, it can be examined that current provisions and health care facilities in China are multi-dimensional, which has taken account of different types of patients (primary, secondary, tertiary, preventive and catastrophic care patients. These services are to reformed also cover both financial and non-financial aspects.

3. BIG DATA AND ITS CHARACTERISTICS

Big data is the amount of data created in the world up to and including 2005 is now created every 2-days. Big data is a platform for importing, storing and analyzing data to uncover information not previously known. This explosion of the data, changing the way people think about everything. From the cutting edge scientific research to the monetization of social

media and exchanging the way people think about healthcare analytics too. However, the health care has not kept pace with Big Data

Historically, data was being generated and accumulated by workers. In other words, employees of companies were entering data into a computer system. But, then the things are evolved with the Internet and now users could generate data on their own. For eg. In the websites like Face book, Google, Amazon etc. all the users are signing up and entering the data themselves, which is scalable and it is much larger than the first in magnitude. Regarding scalability, it is scaled up from just employees entering the data to users entering their own data, so all of a sudden the amount of data being generated and accumulated was very higher than it was historically. Now, the third level is, the machines are accumulating data. The buildings all over the cities are full of monitors that are monitoring humidity and temperature, there are satellites around the earth that are monitoring the earth 24-hours a day taking pictures accumulating data. Once machines started accumulating data, the data has become much voluminous. Back in the good old days, people used to use relational databases to process the data and bring it to the processor that the CPU processes the data. But, now there is so much data that overwhelms the power of CPU and the CPU cant process as the data is very huge. So, now what the people are doing is, they are bringing multiple processors to the data and do parallel processing. So, now the data is being processed in a whole bunch of different places at the same time. Before, the data is being brought to the processor, but now the processor is being brought to the data to processing. In the first case, data is brought to one CPU, But now the we could bring in infinite number of CPU's to infinite number of individual servers and do parallel processing. So, now the data has grown to larger orders of magnitude to higher and now we have a way to process higher magnitude data as well, that the technological shift. Some of the technologies that are allowing this to happen are Hadoop, Map Reduce, NOSQL.

3.1 Definition

Big Data is a collection of large and complex data sets which are difficult to process using common database management tools or traditional data processing applications. According to zdnet.com, "Big data refers to the tools, processes and procedures allowing an organization to create, manipulate and manage very large data sets and storage facilities".

Big data is being generated by everything around us at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Big data is arriving from multiple sources at an alarming velocity, volume and variety. To extract meaningful value from big data, you need optimal processing power, analytics capabilities and skills. Big data is changing the way people within organizations work together. It is creating a culture in which business and IT leaders must join forces to realize value from all data. Insights from big data can enable all employees to make better decisions—deepening customer engagement, optimizing operations, preventing threats and

fraud, and capitalizing on new sources of revenue. But escalating demand for insights requires a fundamentally new approach to architecture, tools and practices.

The big data are generated from online transactions, emails, videos, audios, images, click streams, logs, posts, search queries, health records, social networking interactions, science data, sensors, mobile phones and their applications[8].

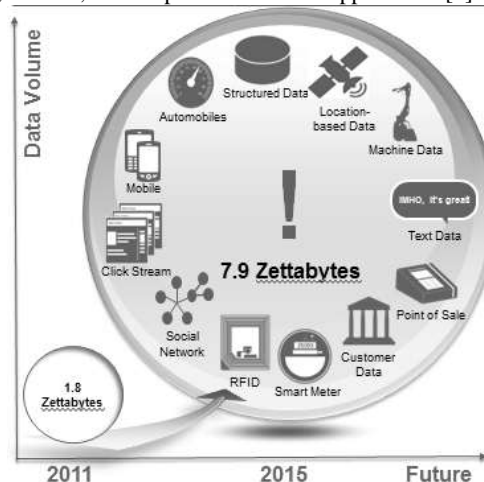


Fig.1 Voluminous Data generated through different sources

Big data is a platform for importing, storing and analyzing data to uncover information not previously known. This explosion of the data, changing the way people think about everything. From the cutting edge scientific research to the monetization of social media and exchanging the way people think about healthcare analytics too. However, the health care has not kept pace with big data.

The large Chinese health care system needs to harness healthcare's "big data" and analyze a complex set of data, including electronic medical records and sensor data. This enables clinicians to access and analyze healthcare big data to ascertain quality, determine best practice, assess treatment strategies and identify patients at risk. The promises and potential of big data in transforming digital government services, governments, and the interaction between governments, citizens, and the business sector, are substantial. From "smart" government to transformational government, Big Data can foster collaboration; create real-time solutions to challenges in agriculture, health, transportation, and more; and usher in a new era of policy- and decision-making. Big Data raise a large number of information management issues, primarily in the areas of privacy, security, accuracy, and archiving, spanning major issues such as personally identifiable information, security of government data and information, and the accuracy of publicly available data. By fostering collaborations and economic development through private-public partnerships, government agencies appear to be tacitly endorsing the privacy, security, and other policies employed by those private sector entities.

3.2 Characteristics of Big Data

Big data generally refers to the social network data from the micro-blogging sites like Twitter, LinkedIn and social media platforms like Face book, Traditional enterprise including transactional data, web store transactions etc. and machine generated / sensor data like call data records, smart meters, manufacturing sensors, trading systems, traffic data, air data etc. which keeps on increasing without the human intervention. Big data is not only driven by the exponential growth of data but also by changing user behavior and globalization. Globalization provides competition among the participants in the market. As a result, organizations are constantly looking for opportunities to increase their competitive advantage by using better analytical models.

The typical characteristics of the Big data are:

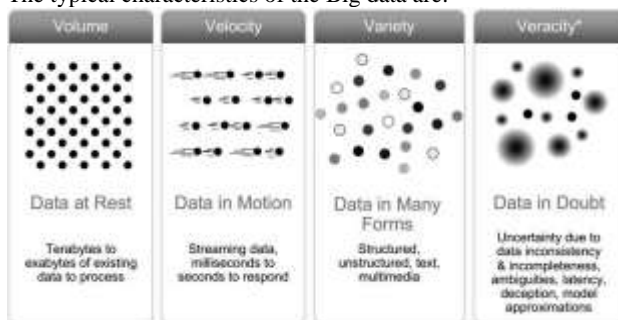


Fig 2. Characteristics of Big data

3.2.1 Volume: The amount of data being generated by various health care monitoring systems connected to patient over a period of time will be very huge in size. This voluminous data will become difficult for the conventional systems to handle and manage.

3.2.2 Velocity: Velocity is the speed at which the data is being generated like streamed data from various smart devices into social media and also camera streamed data which stores the data in motion from huge number of closed circuit cameras. In health care systems, the data will be generated by the sensors connected to the patients and will be generated with a high velocity.

3.2.3 Variety: The Variety refers to the various formats of data like structured, semi-structured or un-structured. The data that will be generated at the health care organization will be in different forms like text, documents, emails, audio, video, text messages, graphs, diagnostic reports etc.

3.2.4 Veracity: Veracity refers to the biases, noise and abnormality in data. Veracity in data analysis is the biggest challenge when compared to other characteristics like volume and velocity.

3.2.5 Value: Value refers to what type of value does the big data generates. This characteristic deals with the various ways of presenting results values like statistical, correlation, hypothetical etc.

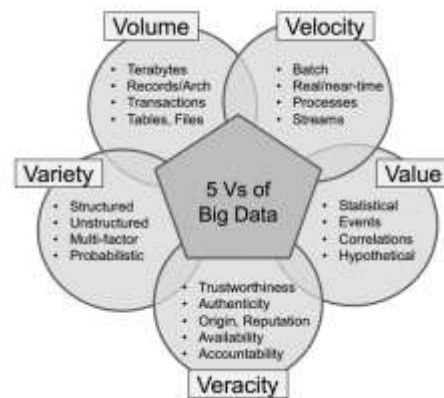


Fig.3: Characteristic of Big Data. (5 Vs of Big Data)

4. BIG DATA TECHNOLOGIES

4.1 Hadoop: Apache Hadoop [9] is an open source software project that enables the distributed processing of large data sets across clusters of commodity servers. It is designed to scale up from a single server to thousands of machines, with a very high degree of fault tolerance. Rather than relying on high-end hardware, the resiliency of these clusters comes from the software’s ability to detect and handle failures at the application layer.

Hadoop is an open source platform that is developed under General Public License (GPL) and it is developed by the developers across the world and it is free to use. It organizes the parallel processing and this is the software that allows parallel processing to happen and the second is called MapReduce. The Hadoop and Map Reduce are working together for the processing and analytics of Big Data. The companies like Google, Amazon and many others are using these processes together to discover unbelievable insights that is hidden inside all of that Big data.

Components of HADOOP

- **YARN - Yet Another Resource Negotiator (YARN)** assigns CPU, memory, and storage to applications running on a Hadoop cluster. The first generation of Hadoop could only run MapReduce applications. YARN enables other application frameworks (like Spark) to run on Hadoop as well, which opens up a wealth of possibilities.
- **HDFS - Hadoop Distributed File System (HDFS)** is a file system that spans all the nodes in a Hadoop cluster for data storage. It links together the file systems on many local nodes to make them into one big file system.

Features of Hadoop

- **Scalable–** New nodes can be added as needed, and added without needing to change data formats, how data is loaded, how jobs are written, or the applications on top.
- **Cost effective–** Hadoop brings massively parallel computing to commodity servers. The result is a

sizeable decrease in the cost per terabyte of storage, which in turn makes it affordable to model all your data.

- Flexible–Hadoop is schema-less, and can absorb any type of data, structured or not, from any number of sources. Data from multiple sources can be joined and aggregated in arbitrary ways enabling deeper analyses than any one system can provide.
- Fault tolerant– When you lose a node, the system redirects work to another location of the data and continues processing without missing a flight beat.

4.2 Map Reduce

Map Reduce [10] is a software framework for easily writing applications which process vast amounts of data (multi-terabyte data-sets) in-parallel on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner. A Map Reduce job usually splits the input data-set into independent chunks which are processed by the map tasks in a completely parallel manner. The framework sorts the outputs of the maps, which are then input to the reduce tasks. Typically both the input and the output of the job are stored in a file-system. The framework takes care of scheduling tasks, monitoring them and re-executes the failed tasks.

Map Reduce is a programming model and an associated implementation for processing and generating large data sets with a parallel, distributed algorithm on a cluster. Typically the compute nodes and the storage nodes are the same, that is, the Map Reduce framework and the Hadoop Distributed File System are running on the same set of nodes. This configuration allows the framework to effectively schedule tasks on the nodes where data is already present, resulting in very high aggregate bandwidth across the cluster.

The Map Reduce framework consists of a single master Job Tracker and one slave Task Tracker per cluster-node. The master is responsible for scheduling the jobs' component tasks on the slaves, monitoring them and re-executing the failed tasks. The slaves execute the tasks as directed by the master.

4.3 Hadoop Distributed File System (HDFS)

The Hadoop Distributed File System (HDFS) is a distributed file system designed to run on commodity hardware. It has many similarities with existing distributed file systems. However, the differences from other distributed file systems are significant. HDFS is highly fault-tolerant and is designed to be deployed on low-cost hardware. HDFS provides high throughput access to application data and is suitable for applications that have large data sets.

Name nodes and Data nodes: HDFS has a master/slave architecture. An HDFS cluster consists of a single Name node, a master server that manages the file system namespace and regulates access to files by clients. In addition, there are a number of Data nodes, usually one per node in the cluster, which manage storage attached to the nodes that they run on.

HDFS exposes a file system namespace and allows user data to be stored in files. Internally, a file is split into one or more blocks and these blocks are stored in a set of Data nodes. The Name node executes file system namespace operations like opening, closing, and renaming files and directories. It also determines the mapping of blocks to Data nodes. The Data nodes are responsible for serving read and write requests from the file system's clients. The Data nodes also perform block creation, deletion, and replication upon instruction from the Name node.

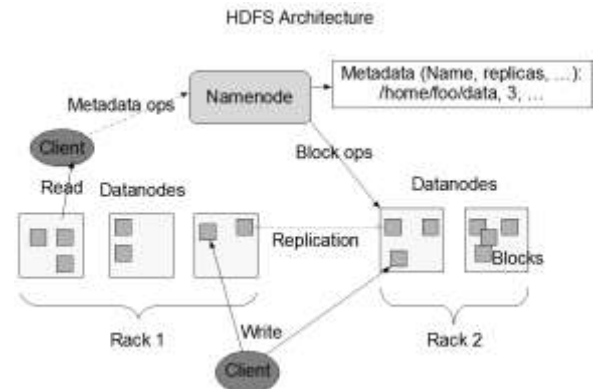


Fig 4: HDFS Architecture

The NameNode and DataNode are the pieces of software designed to run on commodity machines. These machines typically run a GNU/Linux operating system (OS). HDFS is built using the Java language; any machine that supports Java can run the Name node or the Data node software. Usage of the highly portable Java language means that HDFS can be deployed on a wide range of machines. A typical deployment has a dedicated machine that runs only the Name node software. Each of the other machines in the cluster runs one instance of the Data node software. The existence of a single Name node in a cluster greatly simplifies the architecture of the system. The Name node is the arbitrator and repository for all HDFS metadata. The system is designed in such a way that user data never flows through the Name node.

5. CRITICAL ISSUES IN HEALTH CARE

The following are the critical issues to be considered to deliver optimal health care in rural parts of the china and bridge the gap between quality and affordability in government hospitals. These ideas will enable us to access the services on par with the private super specialty hospitals. Further, the implementation of these issues will provide cheaper, better and easier health care facilities to the citizens of India.

1. **e-Health File:** The creation of a e-Health care file for each patient, where all health care providers and patients themselves were able to submit information (with the consent of the patient). Both subjective data, symptom diaries, lab data, image diagnostics, pathology reports etc., could be filed. To overcome the information overload from the massive amounts of data, Big Data Analytics could be employed for the processing of the

data and obtain the desired results with great accuracy in reasonable time.

2. **Creating awareness with chronic diseases:** The system must identify and create awareness among the people with the common chronic diseases at particular areas, through which we can prevent diseases. These chronic diseases are responsible for the 75% of health care spending due to lack of awareness and prior care.
3. **e-Prescribe:** Paper based prescriptions are archaic and lead to several miseries each year due to errors in prescription. But if every doctor is provided with an electronic prescription system, it would improve safety by making prescriptions easier to read and providing instant checks on drug interactions, dosages, and a patient's medication history.
4. **Electronic Medical Records:** Medical Experts agree that electronic medical records (EMRs) are a must for the better health care in India. But, at present only few hospital are maintaining EMR's, mainly because of cost, privacy issues, and the lack of one compatible, easy-to-use infrastructure.
5. **Stop Unnecessary Treatments:** Doctors should avoid trial and error type of medication. The problem must be examined thoroughly by performing the required diagnostic tests during the preliminary days of disease. The right treatment should be suggested at the first visit only which avoids the disease to become more critical. Most of the issues are arising with the misdiagnosis and wrong treatment during the early stages.
6. **Reduce Infant Mortality:** The Infant Mortality rate in our country is substantially large compared to other countries. Though, the government has several schemes for the pregnant women, those were not yielding better results due to the lack of proper medical care. If the proper care is taken towards the pregnant women, definitely the new-born baby will be healthy in all aspects to make India healthy.
7. **Tele Medicine:** Doctors can often diagnose or prescribe without seeing the patient. The patient has to physically appear before the nearby health center, where the nurses or health workers will diagnose at first level, note the symptoms and informs the high level specialist doctors about the case. After examining the reports, the specialist doctor suggests the treatment through health worker which reduces costs and creates satisfaction by virtual communication of patients and doctors to discuss medication changes and test results through an online system.

5.1 Tele-Medicine & EMR

Telemedicine can be defined as the use of electronic communication technology to exchange patient information and provision of health care services at remote locations. Global Telemedicine has gone far beyond providing health care services alone. It is now being extensively used also for education, research and management of data. It is, however, paradoxical that despite

India's strength in information technology, the use of telemedicine is still at a fairly nascent stage especially in the public health sector.

Use of Information and Communication Technology (ICT) in health can be broadly in four areas viz. Education, Research, Referral, and Management of Data.

- **Health and Education :** When ICT is applied to medical education, it is possible to make high quality education available pan India seamlessly.
- **Hospital Management System :** It is common experience that work places in the health systems are not adequately governed especially in remote areas. The management of medical colleges and district hospitals can be strengthened significantly with the application of Hospital Management System.
- **Health Research :** ICT can potentially transform the medical scene in India, by bringing about a sea-change in medical research. From traditional clinical research to the modern synthetic biology-based research, the opportunity is immense. Work on problems such as Cancer prevention, screening, diagnosis, and therapy can benefit from inter disciplinary cooperation. Medical fraternity has availed such benefits when MRI and Nuclear Imaging was integrated into medicine a few decades ago.
- **Managing Health Data :** Electronic Medical Records (EMR), is a fundamental pre-requisite in using ICT seamlessly in healthcare. While EMR is available in several forms, size, shape, and format, the medical community with a specific mandate should standardize EMR, create and establish ICT platforms for using EMR based systems, for universal benefits.

6. ROLE OF BIG DATA IN HEALTH CARE

We live in the age of big data. The amount of data created in the world up to and including 2005 is now created every two days. Big data is a platform for importing, storing and analyzing data to uncover information not previously known. This explosion of the data changing the way people think about everything. From the cutting edge scientific research to the monetization of social media and exchanging the way people think about healthcare analytics too. However, the health care has not kept pace with big data. Big Data Healthcare is the drive to capitalize on growing patient and health system data availability to generate healthcare innovation. By making smart use of the ever-increasing amount of data available, we find new insights by re-examining the data or combining it with other information. In healthcare this means not just mining patient records, medical images, diagnostic reports etc., for insights, diagnoses and decision support device, but also continuous analysis of the data streams produced for and by every patient in a hospital, at home and even while on the move via mobile devices [11]. Even today the majority of health care analytics is performed by doing monthly data refreshes in relational databases that produce pre-processed reports. A fair gap is often missing lab

test is often 45 days old, as the data flow move from batched data fields to real time fields from transactional systems and streaming data from analytical modeling devices. This old model of analytics will fail. Analysis will need to be done on that spot moment not in the pre-processed form. Data refreshes need to be done in real-time not once in a month. The data analysis tools of today are likely yellow pages phone book in the era of Internet Search Engine. They are becoming more obsolete with each passing day. The traditional health care analytic tools are built on tools developed by IBM in 1970, more than 40 years ago.

If all the three parties (payer, provider, pharmaceutical company) [12] work collaboratively and share data/insight, disease management programs will become cost-effective and deliver improved patient outcomes at a scale that will further optimize overall health-care cost structure.

The term “e-health” [13] defined by WHO: “a new term used to describe the combined use of electronic communication and information technology in the health sector”. e-health is the main driver for three significant changes within the health care environment[13]:

1. Patients to become better informed
2. Patients to become more active and empowered in their health care
3. Healthcare to become more efficient.

It is referred in the Cognizant 20-20 insights by Cognizant, [14] Big data solutions attempt to cost-effectively solve the challenges of large and fast-growing data volumes realize its potential analytical value. For instance, trend analytics allow you to figure out what happened, while root cause and predictive analytics enable understanding of why it happened and what it is likely to happen in future. All healthcare constituents – patients, payers, providers, groups, researchers, governments etc. – will be impacted by big data, which can predict how these players are likely to behave, encourage desirable behavior. These applications of big data can be tested, refined and optimized quickly and inexpensively and will radically change healthcare delivery and research. The healthcare domain has been an easy target for people who seek easy money by using fraud methods [15]. Healthcare fraud is expected to continue to rise as people live longer. The white paper by trend analytics [15] reveals that healthcare fraud prevention has resulted savings of nearly \$4.1 billion in 2011. A big data platform has ability to sift through a huge amount of historical data in relatively shorter amount of time, so that the business transactions can use fraud detection on real time. Though, the big data analytics in healthcare plays a crucial role to provide better health care services, provide analysis on the historical data to uncover hidden information, the big data analytics has the challenges like Heterogeneity and Incompleteness of data, scale, timeliness, privacy and Human Collaboration. The future research is all about to overcome the challenges and use big data analytics in healthcare to uncover the knowledge from the raw unstructured data.

Opportunities for big data in health care

Big Data has a large potential to contribute in many areas of the health care Industry. At the moment, there are some good initiatives, but this is not enough to keep up with the demand of health care service and the rising costs.

The opportunities for Big Data in health care are:

1. Electronic Health records (EMR/EHR) which serve the customer/patient
2. Structuring data and information for service optimization
3. Accurate information about patients can reduce mistakes
4. Cost optimization through efficiency of new e-health services
5. Increased customer satisfaction (better e-health services)
6. Analysis of big datasets for R&D purposes

7. CONCLUSION

Big Data Analytics in Hadoop's implementation provides systematic way for achieving better outcomes like availability and affordability in healthcare service to all population. In this paper we discussed about the implementation of Electronic Medical Records (EMRs) in China. Further, this paper discusses about various ways of transforming health records (EHRs) to useful and analyzed result, this analysis will make the patient to understand the complications to occur and thereby the patient will be cautious towards his health status or health behavior. The goal of this research deals with the study and application of big data analytics to health care industry.

8. FUTURE WORK

In the future, the present research work will be extended with the implementation of Hadoop, HDFS and MapReduce to handle huge amount of semi-structured/ un-structured health care data and also provide the real analytics through which the patients and doctors will enjoy the real benefit of Hadoop and Big Data Analytics.

9. ACKNOWLEDGMENTS

We would like to thank Almighty, our professors, friends and our family members who helped and supported us in getting this research work done with in time.

10. REFERENCES

- [1] Shi, Leiyu and Douglas A. Singh. *The Nation's Health*. London: Jones & Bartlett Learning, 2011.
- [2] Hou, Thomas. *The Chinese Primary Care System: Its Evolution, Challenges and Legal Aspects of Reform*. Research Paper. Pennsylvania: University of Pennsylvania, 2009.
- [3] Wenhua Wang, et al. "Primary Care Quality among Different Health Care Structures in Tibet, China." *BioMed Research International* (2015): 1-8.
- [4] Rohde, J, et al. "30 years after Alma-Ata: has primary health care worked in countries?" *The Lancet*, 372(9642) (2008): 950-961.

- [5] Hu, Shanlian and Christer Ljungwall. *China's Healthcare System - - Overview and Quality Improvements*. Ostersund: Swedish Agency for Growth Policy Analysis, 2013.
- [6] Xin, Li. "The Effect of China's New Cooperative Medical Scheme on Rural Utilization of Preventive Medical Care and Rural Households' Health Status." *Master of Pacific International Affairs* (2011): 1-20.
- [7] Lieber, Eli, et al. *Sociocultural Contexts and Communication about Sex in China: Informing HIV/STD Prevention Programs*. Los Angeles: UCLA/NPI , 2014.
- [8] Sagiroglu, S.; Sinanc, D., "Big data: A review," *Collaboration Technologies and Systems (CTS)*, 2013 International Conference on , vol., no., pp.42,47, 20-24 May 2013
- [9] <http://www-01.ibm.com/software/data/infosphere/hadoop/> accessed on 30-08-15
- [10] http://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html accessed on 30-08-15
- [11] Edwin Morley-Fletcher, "Big Data Healthcare, An overview of the challenges in data intensive health care", Discussion Paper.
- [12] SoumendraMohanty, MadhuJagadeesh, and HarshaSrivatsa,"Big Data Imperatives", Apress publications. 2013
- [13] Manuel Zwicker, Juergen Seitz and NilminiWickramasinghe, "Identifying Critical Issues for Developing Successful E-Health Solutions", *Proceedings of the Pacific Asia Conference on Information Systems (PACIS 2012)*. Paper 33. <http://aisel.aisnet.org/pacis2012/33>
- [14] Bill Hamilton , "Big Data is the Future of Healthcare", *Cognizant 20-20 Insights*, September 2012.
- [15] VenkatarreddyKonasani, MukulBiswas and Praveen Krishnan Kolet, " Healthcare Fraud Management using Big Data Analytics", A Whitepaper by Trendwise Analytics.
- [16] *Fraud Detection Using Data Analytics in the Healthcare Industry - Discussion white paper*