International Journal of Advance Research in Science and Engineering
Vol. No.5, Special Issue No. 01, March 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

NEED OF ERA THE GROWING AVALANCHE AND BOON TO HEALTHCARE: BIG DATA

Shivani Kale[1] , Dr. Preeti Patil[2]

[1]PhD Student VTU Belgaum(India)

[2]HOD.CSE KIT's College of Engineering Kolhapur(India)

ABSTRACT

Earlier as the studies shows the death rate of patients were high due to lack of proper diagnosis , due to lack of resources and medicine on time. It was really impossible for doctors to get the information of patient within time using traditional means and analyzing that structured and unstructured data in time to rescue the patient. So the big Data in healthcare is being used to cure disease, predict epidemics, improve quality of life. So the paper focuses on how big data analytics will be helpful to cut down the death rate at certain extent in healthcare.

Keywords: Big Data, security, privacy, sharing, value, infrastructure.

I. INTRODUCTION

Abdunce of health care big data is easy to access because of accessibly-written framework by big data and health analytics. Analysts and decision makers particularly those in hospitals and health systems is recommended for reading the data and health Analytics documents specially those in public health, pharmacy and biotechnology firms, clinical practices, health, accountable care organizations, medical homes, Medicare, health information technology companies, and medical device and diagnostics firms.[1]

II. BIG DATA AND HEALTHCARE

The Big data with hadoop technologies are used to serve the public health with the following objectives:

Public health

1) Disease patterns can be analyzed to track disease outbreaks and can be transmitted to improve condition of public health.

2) Targeted vaccines can be accurately developed with high speed, e.g., choosing the annual influenza strains.

3)The Population can be benefited by identification of needs ,predicting and preventing crises and providing services by turning huge amounts of data into intelligent information.

Big data analytics also contribute to evidence–based medicine

1) Genomic analytics: Genomic analysis can be used as part of the regular medical care decision process by executing gene sequencing more cost effectively and efficiently

2) Pre-adjudication fraud analysis: The fraud, waste and abuse can be reduced by rapidly analyzing large numbers of claim requests.

3) Device/remote monitoring: By capturing and analyzing real-time huge volumes of fast-moving data from in-home devices and in-hospital can be used for safety monitoring and adverse event prediction.

4) Patient profile analytics: By Applying advanced analytics to patient to identify individuals those would benefited from proactive care or lifestyle changes, for example, by taking Preventive care patients at risk of developing a specific disease (e.g., diabetes) can be benefited.[2]

III. ARCHITECTURAL FRAMEWORK OF BIG DATA

The traditional health informatics or analytics project is similar to conceptual framework for a big data analytics project in healthcare. They differed only in execution of their process. A business intelligence tool installed on a single system, such as a laptop or desktop is used for analysis in regular health analytics project. As big data is large by definition, so while processing it is broken down and executed across multiple nodes. The distributed processing concept has existed for decades. So its use in analyzing very big data sets as healthcare providers can be used to gain insight for making efficient and correct health-related decisions. The application of big data analytics in healthcare is encouraged by using open source platforms such as Hadoop, Map Reduce.

The user interfaces of traditional analytics tools and those used for big data are entirely different while the algorithms and models are similar; the health analytics tools of traditional process have become very user friendly and transparent. On the other hand big data analytics tools are more complex, programming intensive, and needs the application of lots of skills. The user-friendliness that vendor-driven proprietary tools possess is lost in the big data analytics tools. [3]

The data itself is complicated. Big data in healthcare can come from internal sources (e.g., clinical decision support systems ,electronic health records, CPOE etc.) and external sources (insurance companies, government sources, laboratories and pharmacies, etc.), often in multiple file extensions or formats (flat files,csv, relational tables, ASCII/text, etc.) and stored at multiple locations (geographic as well as in different healthcare providers' sites) in numerous legacy and other applications (transaction processing applications, databases, etc.).

Data types and sources include

1. Social media data and web: interaction data and click stream from Twitter, Face book, blogs, LinkedIn, and the like. Smartphone apps and health plan websites can also be included, etc.

2. M2M data: Data from meters, remote sensors, and other devices.

3. Big transaction data: The billing records and health care claims are increasingly available in unstructured and semi-structured formats.

4. Data from Biometric: retinal scans, finger prints, genetics, handwriting-ray and other medical images, pulse and pulse-oximetry readings ,blood pressure, and other similar types of data.

5. Human-generated data: The semi-structured and unstructured data such as physicians' notes, email, EMRs, and hardcopy documents. This data has to be pooled for the purpose of big data analytics. In the second component the raw data needs to be properly processed or transformed, for this purpose several options are available. We can use one way of service oriented architectural approach combined with web services. Services are used to call, retrieve and process the raw data. Second way is warehousing of data where the data from different sources is combined and aggregated and made ready for processing, even if the data is not available in real-time.

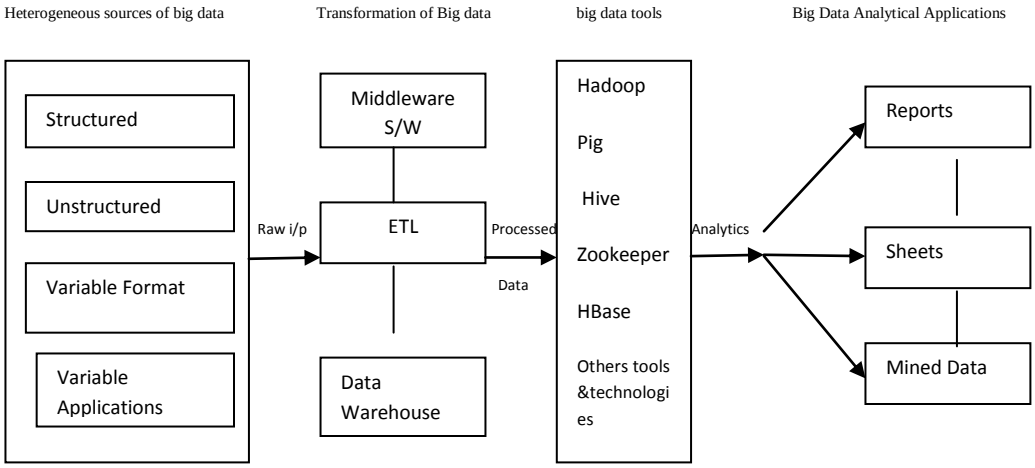


Figure 2: Architectural Framework for Big data

Courtesy: Yuri Demchenko, Canh Ngo, Peter Membrey (System and Network Engineering Group, UvA SNE)

Col: Used By Row: Reqs This	Data Models	Data Mngnt& Lifecycle	BD Infra & Operat	BD Analytics	Big Data Security
Data Models		+	++	+	++
Data Mngnt& Lifecycle	++		++	++	++
BD Infrastr & Operation	+++	+++		++	+++
BD Analytics	++	+	++		++
Big Data Security	+++	+++	+++	+	

Table 1-Interrelation between Big Data Architectural Framework components

IV. BIG DATA IN HEALTHCARE INDUSTRY

Huge electronic health datasets gives birth to the big data technology– With the conventional hardware and software it is difficult to manage these data sets. It impossible to usefully influence all this data by making use of heritage data management methods and tools. Because of the volume of big data as well as due to the different data types and the velocity at which healthcare data need to manage in healthcare is uncontrollable concept. The "Big Data"

problem in the healthcare industry needs to address both the sum total of data related to the patient and their well-being. The contribution of Healthcare informatics is to the development of Big Data analytic technology by posing innovative challenges in terms of clinical decision support ,data knowledge representation, database design and data querying . The modern trend is touching towards rapid digitization of data regardless of the actuality that, most of the records in the health care sector is stored in hardcoded form. Healthcare industry in Big Data promises to support a varied range of healthcare big data management functions like, disease surveillance, and clinical decision support and population health management. As Medical research, Clinical operations, and Treatment courses needs the on an average received 79% of the healthcare unstructured data which is expected to persuade effectively creats a big challenge for the healthcare industry .The volume of big data in healthcare is projected to grow over the coming years and the healthcare industry is expected to grow with changing healthcare refund models thus posing vital challenges to the healthcare Industry. [4]

V. HEALTHCARE DATA SOLUTIONS USING HADOOP

1. Cancer Treatments and Genomics using Hadoop technology.

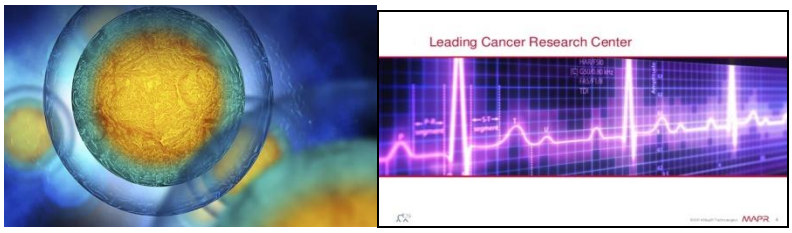


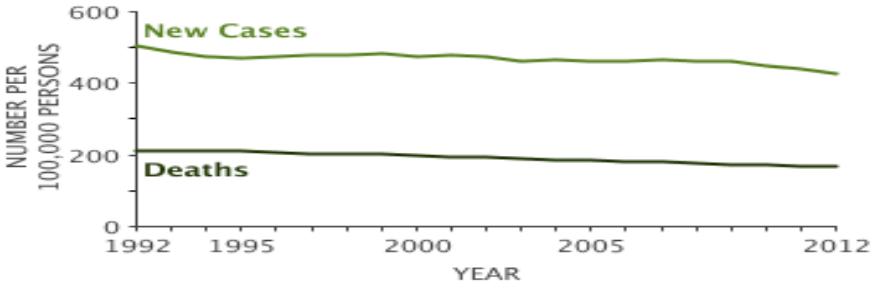
Figure 3: Cancer information (Courtesy CRAY.COM and MAPR leading cancer research center)

Business reports specify that, the human DNA is constituted using about 3 billion base pairs. As we can see in the graph the death rate according to SEER which really makes the health industry to properly analyze the hefty amounts of unstructured data to be structured in a valuable manner if cancer is need to be eradicated.

Table 2: Number of New Cases and Deaths per 100,000 People (All Races, Males and Females), Age Adjusted.

Year	New Cases - SEER 9	New Cases - SEER 13	Deaths -
1975	400.4	-	199.1
1976	407.4	-	202.3
1977	407.8	-	203.0
1978	407.3	-	204.4
1979	412.5	-	204.5
1980	418.1	-	207.0
1981	425.3	-	206.4
1982	424.5	-	208.3
1983	431.1	-	209.2
1984	440.0	-	210.9
1985	448.9	-	211.3
1986	451.4	-	211.8
1987	468.0	-	211.9
1988	463.9	-	212.6
1989	467.8	-	214.3
1990	482.0	-	214.9
1991	503.1	-	215.1
1992	510.7	503.3	213.5
1993	493.4	486.9	213.4
1994	483.6	475.3	211.7
1995	477.1	471.3	209.9
1996	479.3	473.0	207.0
1997	486.6	478.6	203.6
1998	488.5	480.2	200.8
1999	490.7	481.3	200.7
2000	486.5	475.0	198.8
2001	490.2	478.6	196.3
2002	488.0	474.2	194.4
2003	475.7	463.2	190.9

Courtesy: Seer.cancer.gov



Graph 2: Courtesy: Seer.cancer.gov

It is actuality that cancer mutates in unusual patterns and reacts in diverse ways depending on the genetic structure of an individual and that is the principal reason that cancer has not been cured yet. Hence, in sort to cure cancer dieses oncology doctors and researchers invented a solution that, Depending on individual patient's genetics make up patients will need to be given personalized treatment on the type of cancer. So the 3 billion DNA base pairs can be divided and analyzed using Hadoop tools for parallelization and mapping using Map Reduce programs.

The personalized medication for cancer patients is main objective of big data in healthcare to develop.[5]

2. Monitoring Patient Vitals using hadoop technology

There are several hospitals across the world that uses Hadoop to help the hospital staff work efficiently with Big Data. Without Hadoop, most patient care systems could not even imagine working with unstructured data for analysis.

Figure 4 -Image Credit: slideshare.net

In Atlanta children's Healthcare treatment uses a sensor beside the bed which helps them constantly to track patient signs such as respiratory rate, blood pressure and heartbeat. The data produced by these sensors is huge chunks of information that cannot be stored by traditional systems for more than 3-4 days for processing and analysis. So it is not useful to store up and analyze the vital signs of the patients. An alert is to be generated to a team of doctors and assistants in the hospital if there is any change in data pattern received continually. Using Hadoop ecosystem components - Flume, Hive, Sqoop and Spark, etc, it was effectively achieved.[22]

3 Hadoop technologies in the Hospital Network

Use of Big Data in healthcare is to provide to reduce the cost of care measurement, the most excellent clinical support and manage the population of susceptible patients. The company Explorys has seemingly created the prime database in the healthcare industry with over a billion data points using hadoop. The medical experts analyze data in real time from assorted sources such as payroll data, financial data and electronic health records. These insights help the health care provider's and medical practitioners find out the top treatment plans for a group of patient or for an individual patient.[25]

4. Hadoop technology in Fraud Prevention and Detection

In any case minimum of 15% of the Healthcare assurance payments are endorsed to fraudulent claims. So this is really a global multibillion dollar problem. Fake claims are not a new problem but the convolution of the insurance frauds seems to be escalating exponentially resulting in great difficulty for the healthcare insurance companies to handle them. HealthCare insurance companies can be benefited by Big Data Analytics technology to recognize and avoid fraud at an early stage. By making use of real-time and historical data of medical claims with Hadoop technology, insurance companies can get success in developing prognostic models to identify weather data, fraudsters, voice recordings, wages, demographics, and cost of attorneys and call center notes. In the field of Fraud Detection hadoop's potential to store big unstructured data sets in NoSQL databases and using Map Reduce to analyze this data helps in the analysis and discovery of patterns.[23]

VI.CONCLUSION

The paper focuses mainly on the big data analytics which is an optimistic way in its early years for the healthcare area. As in healthcare more and more data is being unruffled, so big data analytics is in demand. Hence it is proved that the percentage of the death rate by cancer can be reduced to approximately 40% by using the big data analytics technique. Resourcefully utilizing the massive healthcare data repositories can raise some urgent returns in terms of patient diagnosis and making cost reasonable. Records with new complexities keep budding in healthcare thus giving to more opportunities for big data analytics

REFERENCES

[1] Q. Zou, X. , Li,, W. Jiang,, Z. Lin, G. Li, and K. Chen,. "Survey of mapReduce frame operation in bioinformatics," Briefings in Bioinformatics, pp. 1-12, 2013.

[2] H. Li and C.Liu, "Biomarker identification using text mining," Computational and Mathematical Methods in Medicine, pp. 1-4, 2012..

[3] W.Fleuren, et al., "Identification of new biomarker candidates for glucocorticoid induced insulin resistance using literature minin," BioDataMining, Vol. 6, 2, pp.1-15, 2013.

[4] T. Golub, et al., "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science," Vol. 286, issue 5439, pp.531-537, 1999.

[5] C. Sotiriou, and L. Pusztai , "Gene-expression Signatures in breast cancer," The New England Journal of Medicine, 360, 8, pp. 790-800, 2009.

[6] V. Tusher, , R. Tibshirani, and G. Chu, G., "Significance analysis of microarrays applied to the ionizing radiation response," Proc. Natl Acad Sci USA, Vol. 98, 9, pp.5116-5121, 2001.

[7] T. Bo,and I. Jonassen, "New feature subset selection procedures for classification of expression profiles," Genome Biology, Vol. 3, issue 4, pp. 1-11, 2002.

[8] A. Subramanian, et al., "Gene set enrichment analysis: a Knowledge-based Approach for interpreting genome-wide expression profiles," Proceedings of National Academy Science, 102, 43, pp.15545-15550, 2005.

[9] R., Curtis, M. Oresic, and A.,Vidal-Puig, "Pathways to the analysis of microarray data," Trends Biotechnology, 23, 8, pp. 429-435, 2005.

[10] H. Chuang, , E. Lee, Y. Liu, , D. Lee, and T. Ideker. "T: network-based classification of breast cancer metastasis," Molecular Systems Biology, 3, 140, 2007.

[11] C. Li and H.Li, "Network-constrained regularization and variable selection for analysis of genomic data," Bioinformatics, 24, 9, pp. 1175-1182, , 2008.

[12] M. Jahid and J. Ruan, "A steiner tree-based method for biomarker discovery and classification in breast cancer metastasis," BMC Genomics, 13 (Suppl 6):S8, pp. 1-9, 2012. .

[13] Y. Zhu, X. Shen, and W. Pan, "Network-based support vector machine for classification of microarray samples," BMC Bioinformatics, 10 (Suppl 1):S21, 2009.

[14] Z. Wei and H. Li, "A markov random field model for network-based analysis of genomic data," Bioinformatics, 23, 12, 1537-1544, 2007.

[15] L.Chen, , J. Xuan, R. Riggins, R. Clarke, and Y. Wang, "Identifying cancer biomarkers by network-constrained support vector machines," BMC Systems Biology, 5, 16, 2011.

[16] Hwang et al., "Robust and efficient identification of biomarkers by classifying features on graphs," Bioinformatics, Vol. 24, 18, pp.2023–2029, 2008.

[17] J. Xia, M. Benner, and R. Hancock, "NetworkAnalyst - integrative approaches for protein–protein interaction network analysis and visual exploration," Nucleic Acids Research. 42, 167-174, 2014.

[18] R. Taylor, "An overview of the hadoop/mapReduce/HBase framework and its current applications in bioinformatics," BMC Bioinformatics, 11(Suppl 12):S, 1-6, 2010.

[19] C. Sansom, "Up in a cloud?. Nature Biotechnology, 28, 1, pp.13-15, 2010.

[20] L. Stein, "The case for cloud computing in genome informatics," Genome Biology, 11:207, 2011.

[21] The 'big data' revolution in healthcare by MCkinsey and company

[22] "Hadoop is transforming healthcare" By HORTONWORKS.

[23] IBM website5 Healthcare applications of Hadoop and Big data.

[24] Nrusimham Ammu, 2 Mohd Irfanuddin "Big Data Challenges"International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 613 - 615 (2013)Special Issue of ICACSE 2013

[25] www.forbes.com/sites/.../2015/04/.../how-big-data-is-changing-healthcare

[26] Big Data Analytics for Healthcare by IBM