A Review of Big Data Analytics in Healthcare

M.Roopa[1] Dr.S.Manju Priya[2]
[1]Research Scholar [2]Assistant Professor
[1,2]Department of Computer Science
[1,2]Karpagam Academy of Higher Education Coimbatore, India

Abstract— The Big Data is changing the world. Big data is irregular and consumes more real-time analysis. Big data analytics has emerged from two distinct concepts - Big data and Analytics .Big data analytics plays a vital role in health care industries because health care data set contains lots of information about patient, it also contains family history, personal information, genomics, clinical reports, public health reports, hospital procedures such as bills, other data and these data is for only person but a hospital may have more patients data which differ from each other so all these data are enormously large and cannot be handled by regular software or hardware. Understanding the unstructured clinical notes in the right index is the challenge. Some of the applications are very useful in predicting the stage of diseases and certain tools which are created for providing service in health care. This paper states big data analytics and its features, advantages and challenges.

Key words: Big Data, Analytics, Health Care, Hadoop, Jaql

I. INTRODUCTION

Big Data in the beginning aimed the dimensions of data that could not be processed efficiently by traditional database methods and tools. The Health care has numerous of data. Nowadays it is rising to the peak and it is vast and fast. The data in the health care is to improve the quality of the care given to the patients, and also at hand so countless conflicts intensifying when we deal with numerous of data, especially in analyzing those data. Health care information is a collection of biological science and computer science. There are numerous current areas of research within the field of Health care. The purpose is gathering of immense amounts of data and provides right intervention to the right patient at the right time [1]. In health care industries the big data indicates the data of patient which itself is a large in volume and has many types of different information's. Exploring huge data sets hold a mixture of data called big data this data which is to find out concealed patterns, unfamiliar correlations, market trends, preferences of customer and other useful business information is called big data analytics. The healthcare industry traditionally has generated huge amounts of data, obsessed by record keeping, fulfillment & narrow requirements, and patient care [2].This paper deals with the role of big data in healthcare.

II. BIG DATA

Big data bring up to data volumes in the range of hexa bytes and afar. [19].such volumes beat the competence of current on-line storage systems and processing systems and it is described with three characteristics: (as shown in fig 1)

- Volume – bulky amounts of data originated and engaged;
- Velocity – occurrence and speed, of which data are obtained, created and distributed;
- Variety – assortment of data types and formats from a variety of sources.

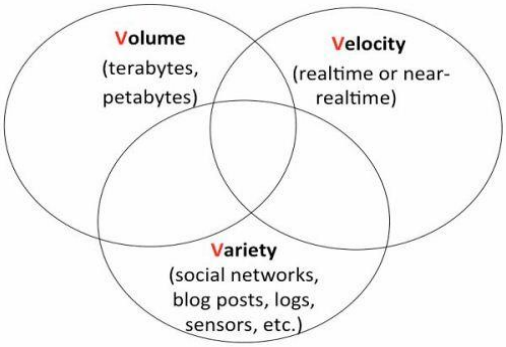

Fig. 1: Three Vs of Big Data [24]

The size and involvedness of big data makes it complex to use habitual database management and data processing tools. Data is being created in much shorter, from hours to milliseconds. Volume is that there is lot of data-zeta byte in future it will be yotta byte (1024 gigabytes). The main challenge is being capable to categorize, situate, analyze and combine different data in large, it will be a structured data set. The variety will be many types of data, both structured and unstructured data. The variety is based on the database which will be of often planned, daily, weekly, or monthly updates of data. Even the five minutes old data is not used to cross the busy street. The data should be updated and should be trusted one.

The big data has many implementations for health care, research, cloud, payers and other concerns. Now the model of health care is processed on reciprocal quantity due to its enlarging. The data should be of accurate, privacy, consistency, facility, security [8].

III. TYPES OF DATA FOR HEALTH CARE

The types of data projected to be used in big data analytics for health care include: (as shown in figure-2)

- Clinical data – among all data 80 per cent of health data is clinical data and which is unstructured as documents, images, clinical or prescribed notes; [6]
- Publications – clinical investigate documents and medical reference material;
- Clinical references – text-based carry out guidelines and health invention (e.g., drug information) data;
- Pharmaceutical data – discovering new drug, samples, doses and Medicine prescriptions.
- Genomic data – it is major amounts of new gene sequencing data;
- Streamed data – secure home systems, tele-health, handheld and sensor-based wireless or smart devices are new data sources and types;
- Web and social networking data – purchaser utilization of Internet, data from search engines and social networking sites;

- Business, organizational and external data – administrative data such as billing and scheduling and other non-health data [11].
- Life style management data- Based on legacy gene prediction of some disease is possible so in such cases a patient can be eliminated from getting a disease so in order to do this storage of family data s is essential and for a group this will be a huge data

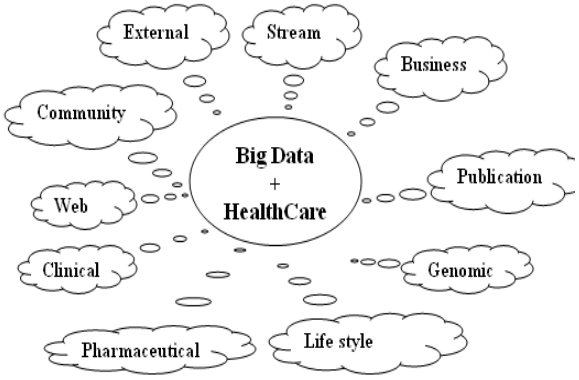

Fig. 2: Types of data in Healthcare indusry

- Community data- larger data and this is used for future reference on patient care, evaluating risk of certain disease and used as a case studies.

IV. Opportunities for BDA (Big Data Analytics) in Health Care

Big data analytics delivers a novel come up to analytics in Health Care.

A. Clinical Decision Support

BDA technologies that shift throughout bulky capacity of data, distinguish, sort out and expand knowledge from it, and then envisage outcomes or urge alternative treatments to clinicians and patients at the point of care.

B. Pharmaceutical-Industry

Big data Analytics has more influences on pharmaceutical industries, day by day a new drug is discovered and the chain of compounds used for the drug preparations are stored in large size which results in big data set. Medicine prescription for a particular disease varies from person to person for example dose is prescribed according to the weight of a person so all people may not have the same weight. Association Rule Mining is applied and also predictive algorithms provide ways to manage data in Pharmaceutical-industry.

C. Personalized Care

Predictive data mining or analytic solutions with the intention of custom-made care (e.g., genomic DNA (Deoxyribo Nucleic Acid) series for cancer care) in authentic instance to emphasize best practice medications to patients. These methods may tender premature finding and outcome before a patient develops disease symptoms. [10].

D. Public and Population Health

BDA (Big Data Analytics) solutions that can mine web-based and social media data to predict flu outbreaks based on region, social content and query activity. BDA (Big Data Analytics) can also sustain clinicians and epidemiologists performing a survey across patient populations and care venues to facilitate spot disease trends. [10].

E. Clinical Operations

BDA (Big Data Analytics) can support initiatives such as wait-time management, where it can mine huge amounts of chronological and unstructured data look for patterns and model various scenarios to predict events that may affect wait times before they actually happen [10].

F. Policy, Financial and Administrative

BDA (Big Data Analytics) can bear decision makers by integrating and analyzing data linked to key presentation indicators.

G. Internet Transactions

By 2015, more than three billion people will be online. Billions of online purchases, stock trades, social networking exchanges, Internet searches and extra transactions ensue every day, collectively with incalculable mechanized transactions. Each creates a number of data points collected by retailers, banks, credit card issuers, credit agencies, social networking and search engine service providers and others [22].

H. Mobile Devices

The use of mobile phones are increasing rapidly every year messages calls and other data are generating data for every single minute. Transmitting location, GPS (Gobal Position System) data are also rapidly increasing it may also contain lots of information's. Several Mobile Applications are also being developed for health care management and for life style management. (as shown in figure-3)

I. Social Networking and Media

Several blogs and pages are increasing data by data. Sharing of pages, messages often result in huge data storage in social network sites handling such data are more tedious than normal data. Health related talks are frequent in TV shows storing episodes also result in big data.

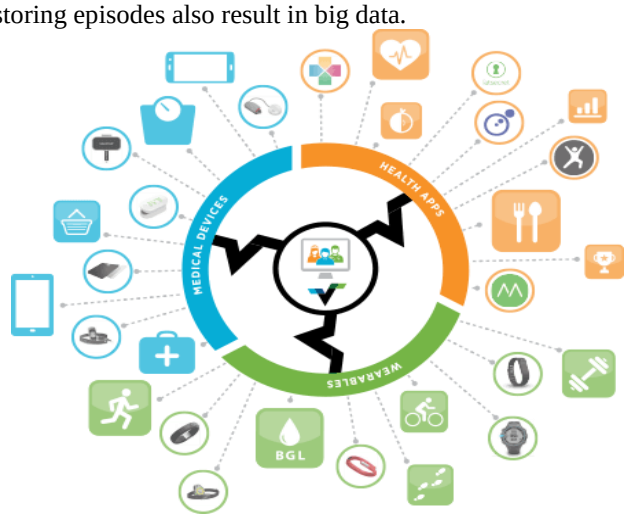

Fig. 3: Apps and medical devices in Healthcare [25]

J. Networked Devices and Sensors

Electronic strategy of all sorts of servers and other IT hardware, smart energy meters and temperature sensors, patient monitors and aides all create semi-structured log data that record every action.

V. PLATFORMS AND TOOLS FOR BIG DATA ANALYTICS IN HEALTH CARE

Many of the tools that are used to manage big data are Open source and the enterprises are facing a high demand in tools and platforms.

Some of the commonly used tools and platforms in Health Care are

A. *Hadoop*

This is the one of the main tool for Big data analytics. It helps in managing records, claim process and also it can be used under The Apache distributed data, several Data Mining Algorithms [17].

B. *Hadoop Distributed File System*

Data in a Hadoop cluster is compressed into negligible pieces also called blocks and stretch all throughout the cluster. Thus, the map and reduce functions can be executed on minor subsets of your larger data sets, and this produces the scalability that is needed for big data processing. (As shown in figure-4) [20].

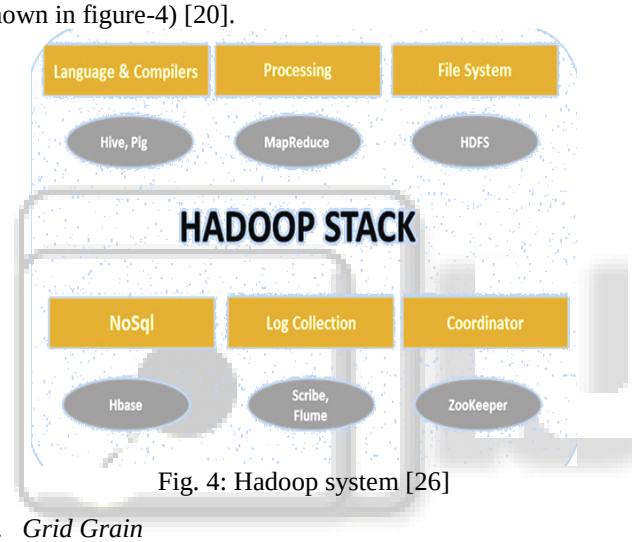

Fig. 4: Hadoop system [26]

C. *Grid Grain*

It is applied with Hadoop Distributed File system and it is an alternative for Map Reduce which provides faster realistic data in memory processing [11].

D. *Map Reduce*

Map Reduce is heart of Hadoop and developed by Google. Map Reduce program is composed of a Map () procedure that performs filtering and sorting. [11].It essentially refers to two separate and dissimilar tasks that Hadoop programs carry out.

E. *PIG and PIG Latin*

Pig is made up of two components. Where the first component is the language itself, which is called Pig Latin and the second, is a runtime environment where Pig Latin programs are executed.

F. *Hive*

Hive allows SQL developers to write Hive Query Language (HQL) statements that are comparable to standard SQL statements. HQL statements are wrecked behind by the Hive service into Map Reduce jobs and executed diagonally a Hadoop cluster (as shown in figure-4) [20].

G. *MongoDB*

MongoDB (DataBase) is a NoSQL database which provides replication, full index support and document-oriented storage for health care records.

H. *Jaql*

Jaql facilitate you to progress both structured and nontraditional data and was contributed by IBM to the open source community [11]. Jaql is a efficient, declarative query language that is intended to practice huge data sets. For parallelism, Jaql rewrites high-level queries, when suitable, into "low-level" queries consisting of Map Reduce jobs. [21]

I. *HBase*

HBase is a column-oriented database management system that runs on peak of HDFS. (as shown in figure-4). They are Linear and modular scalability [11]. It has Automatic and configurable sharing of tables and habitual fail above sustain linking Region Servers. It provides convenient base classes for backing Hadoop MapReduce jobs with Apache HBase tables and comprehensible Java API for client access [20].

VI. CONCLUSION

Today it's the world of digitalization mostly all processes are converted to digital data but in the healthcare when compared to other sectors the process is still carried out. Big data analytics has the prospective to convert the way in healthcare. Health care is a data affluent province. The data with extra intricate keep sprouting in healthcare in leading to more opportunity in big data analytics. The amount of data being digitally collected and stored is vast and intensifying hastily. The foremost eventual ambition is safeguarding protection, eminence, confidentiality, establishing standards, and that is improved in tools. The purpose of big data analytics is at a budding arena. The hasty enlargement of tools and platforms results in Data management.

REFERENCES

[1] Jimeng Sun, Chandan K. Reddy: Big Data Analytics for Healthcare,IBM white paper, SIAM International Conference on Data Mining, Austin, TX, 2013.pg:4-9
[2] Raghupathi W: Data Mining in Health Care. In Healthcare Informatics: Improving Efficiency and Productivity. Edited by Kudyba S. Taylor & Francis; 2010. pg: 211–223.
[3] Priyanka Ketal, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014., pg:5865-5868.
[4] Murdoch, Travis B., and Allan S. Detsky. "The inevitable application of big data to health care." JAMA 309.13 (2013): pg: 1351-1352.
[5] Groves, Peter, "The 'big `data'revolution in healthcare." McKinsey Quarterly (2013).pg:1-13.
[6] Peter Grooves,Kayyali Basel, David Knott, and Steve Van Kuiken. "The big-data revolution in US health care: Accelerating value and innovation." Mc Kinsey & Company (2013).pg:3-8.
[7] Bollier, David, and Charles M. Firestone. The promise and peril of big data. Washington, DC, USA: Aspen Institute, Communications and Society Program, 2010.pg:25-31.

[8] Transforming Health Care through Big Data Strategies for leveraging big data in the health care industry. 2013.pg:7-16.
[9] Dembosky A: "Data Prescription for Better Healthcare." FinancialTimes, December 12, 2012.pg:65-69.
[10] Rajendra Akerkar. Data Intensive in Healthcare, TMRF white paper, TMRF report 03-2012 .pg:5-11.
[11] Big Data Platforms, Tools, and Research at IBM Ed Pednault CTO, Scalable Analytics Business Analytics and Mathematical Sciences, IBM Research. IBM Corporation. 2011.pg:3-14.
[12] Anand Loganathan, Ankur Sinha, Muthuramakrishnan V., and Srikanth Natarajan. A Systematic Approach to Big Data Exploration Hadoop Framework , The International Journal of Information & Computation Technology. Volume 4, Number 9 (2014), pg:869-878.
[13] Kuo Lane Chen, Huei Lee, The Impact of Big Data on the Healthcare Information Systems, Transactions of the International Conference on Health Information Technology Advancement, Center for Health Information Technology Advancement, 2013.pg:43-45.
[14] Wullianallur Raghupathi and Viju Raghupathi, Big data analytics in healthcare: promise and potential, Health Information Science and Systems, 2014, PMCID: PMC4341817.
[15] Rebecca Hermon, Patricia A H Williams, Big data in healthcare: What is it used for?, Australian eHealth Informatics and Security Conference,2014. pg: 40-45.
[16] Intel: Big Data Analytics. 2012. Website: http://www.intel.com/content/dam/www/public/us/en/documents/reports/data-insights-peer-research-report.pdf
[17] Zikopoulos PC, Eaton C, DeRoos D, Deutsch T, Lapis G: Understanding Big Data – Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill: Aspen Institute; 2012
[18] Martin M: Big data/social media combo poised to advance healthcare.HPC Source 2013, pg: 33-35.Website: www.scientificcomputing.co m/digital-editions/2013/04/hpc-source-big-data-beyond.
[19] Stephen Kaisler, Frank Armour, J. Alberto Espinosa, William Money, Big Data: Issues and Challenges Moving Forward, 46th Hawaii International Conference on System Sciences, 2013.pg:187-193.
[20] Apache organisation for tools: Website: http://hadoop.apache.org/Version2.0/Published:2015-07-23.
[21] Kevin Beyer, Vuk Ercegovac, Rainer Gemulla Andrey Balmin, Mohamed Eltabakh Fatma Ozcan Eugene Shekita, : Jaql: A Scripting Language for Large Scale Semi-Structured Data Analysis, IBM Almaden Research Center, September 3, 2011.pg:1272-1290.
[22] Abdul Raheem, Kumar gillela , Dr.C.Venugopal,"The future revolution on big data", International Journal of Advanced Research in Computer and communication Engineering,vol:2,issue 6,june 2013.PG:2446-2451.
[23] P. Zikopoulos, C. Eaton, D. deRoos, T. Deutsch,and G. Lapis. IBM Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data.McGraw-Hill Companies, Incorporated, 2011.pg:51-81.
[24] http://www.infoq.com/articles/stream-processing-hadoop
[25] http://nuviun.com/content/healthcares-big-data
[26] https://technocents.wordpress.com/2014/03/24/hadoop-ecosystem

