A Comprehensive Study of Various Techniques of Big Data Analytics used in Healthcare Sector

Abstract

Today we are living in the world of data and data escalated applications are on rising. There are various corporate sectors that generate lot of databases and one of them is healthcare sector. Healthcare sector generates large amount of heterogeneous data sets and we can easily handle these data sets with the help of big data technologies. This paper focuses on big data technologies, applicability of big data technologies in healthcare, their benefits, their challenges and some conclusions.

Keywords: Healthcare, Hadoop, Map reduce, IoT, Cloud Computing.

Introduction

The era of big data has begun in healthcare sector. Healthcare sector has large amounts of heterogeneous data which should be effectively analyzed and managed for selection of appropriate treatment paths and improvement in healthcare system. These large data sets are difficult to process using traditional data processing applications. The need to process, analyze and manage these massive data sets in healthcare can be gratified with the use of big data and big data technologies such as Hadoop, IoT (Internet of Things), Cloud Computing.

Aim of the Study

The healthcare sector generates large amount of data such as physicians notes, admit and discharge summary, insurance details, medical images etc. As the storage, analysis and processing of health related data shifts towards digitization, the aim of this paper is to reveal various factors in successful implementation of big data in healthcare as well as suggests different strategies for successful implementation of big data in healthcare.

Bigdata in Healthcare

Healthcare includes patient's care data such as prescribed machines, lab reports, x-ray reports, diet chart etc. to be managed by the use of big data.

Big data is a collection of large and complex digital data based on 4 V's (Volume, Velocity, Variety, Veracity), followed by three phases – Data Acquisition, Data Organization and Data Processing.

"4V's" of Bigdata in Healthcare

Volume

Means size of data. In healthcare system, volume represents incredible data generated by medical equipments such as radiology images, 3D imaging, biometric sensor readings etc.

The Volume of worldwide health care data estimated to grow 25000 PB (nearly equals to 500 billion 4-drawer file cabinets) by 2020.

Velocity

Means how fast the data is being produced to satisfy the requests and demands for making decision based output.

In healthcare system, with traditional paper files, x-rays etc. velocity of data increases with the data streamed from mounting devices that represents regular monitoring such as blood pressure readings, diabetic glucose measurements etc. And the ability to perform real time analytics against such high volume data in motion revolutionizes health care.

Variety

Means type of data. The data may be structured, unstructured and semi-structured. In healthcare system, structured and semi-structured data includes instrument readings, office medical records etc. And unstructured

Amita Gandhi
Research Scholar,
Deptt. of Computer Science and Applications,
Sunrise University,
Alwar, Rajasthan

Vandana Jindal
Research Scholar,
Deptt. of Computer Science and Applications,
Sunrise University,
Alwar, Rajasthan

data includes handwritten nurse and doctor notes, radiograph films, MRI and other images.

Use of big data in healthcare gives efficient ways to combine and convert variety of data including automating conversion from structured to unstructured data.

Veracity

Veracity or "data assurance" deals with uncertain and imprecise data.

Traditional data warehouses were based on assumptions that the data is certain and accurate. But it is not so in big data analytics.

In healthcare, the data quality issues are extremely important because of life or death decisions based on having accurate information and the quality of healthcare data. The big data analytics and outcomes are error-free and believable.

Volume	Velocity	Variety	Veracity
Data at Rest	Data in motion	Data in many forms	Data in Doubt
TB to XB of data to process	Streaming data	Structured, unstructured	Uncertainty due to data ambiguity

Fig 1: Big Data - 4 "Vs"

Phases of Big Data in Health Care Sector

Big data analysis in healthcare followed by three phases – Data acquisition, Data organization and Data processing, because it necessitates a mechanism on how to collect data, organize these data and process it to derive some meaningful information.

Data Acquisition

Data acquisition is the process of gathering, filtering and cleansing data before the data is placed on some storage location.

Healthcare sector generates large amount of data and Data acquisition in healthcare involves:
1. Collection of data from various sources such as case history, medical images etc. and storing it in HDFS (Hadoop Distributed File System).
2. Removal of junk data, if exists.
3. Filtering of data based on their structure.

Data Organization

To manage large volume of data, healthcare system needs some mechanism to take and process data from original storage location.

Apache Hadoop, a big data technology to process these large volumes of data.

Data Processing

In big data analytics, after the data acquisition and data organization the next step is to query the data.

To respond a query, it necessitates distributed processing of data. Data processing in healthcare system includes generation of reports after processing data. For example, a report on birth rate in a particular location.

Need of Big Data in Healthcare

The big data analytics in healthcare improves the quality of healthcare by taking into account the following:

To Provide Patient Centric Services

By detecting diseases at earlier stages, minimizing drug doses based on clinical data provides faster relief to patients.

To Provide Provider Centric Services

Big data provides real-time access to patient's clinical data and other significant data for improved decision making and facilitate effective, efficient and error free care.

To Predict Spreading Diseases Earlier

By analyzing the patients suffering from a disease in a particular location, earlier prediction of viral diseases before spreading is possible. This helps healthcare professionals to take necessary preventive actions.

To Improve the Treatment Methods

Continuous monitoring of patients vital signs and based on analysis of continuous monitoring, dosages of medications can be changed for faster relief. This helps in providing customized patient treatment.

To Reduce in Healthcare Cost

Use of big data in healthcare reduces the cost of clinical care.

Big Data and Technologies
Hadoop in Healthcare

Hadoop is the most demanding computing technology for big data processing. It fundamentally changed the way of storing and analyzing information. Haddop uses Map reduce as a default algorithm to work with large quantity of data sets. Map reduce divides the work in two phases – mapper phase and reducer phase. Each phase has key or value pairs as an input/output. The key or value pairs are chosen by the user of the program.
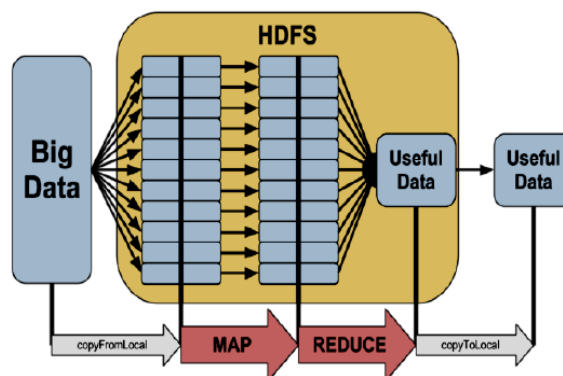
Fig 2: Hadoop's Map Reducing

In healthcare sector, data is growing day by day, which is difficult to manage and process by healthcare organizations. Beyond that 80% of medical data is "unstructured" and Hadoop is the only

91

technology that allows healthcare to store data in its original form.

Hadoop facilitates distributed parallel processing on large volume of data. Hadoop also provides Hadoop Image Processing Interface (HIPI) to analyze the large number of medical data from various sources and extracts the necessary data to give right diagnosis.

Challenges and Suggestion

Even though big data analytics and hadoop for healthcare optimize the cost there are some challenges like

1. Difficult to understand and use by a non technical medical professional.
2. Privacy and security issue in real time capturing of data.
3. Scalability of data, includes capability to handle growing amount of data.
4. Also, Hadoop process use-cases into batch processing and healthcare sector needs real time interactivity with data.

Cloud based technologies and cloud based hadoop can be improves the performance of healthcare system.

Cloud Computing in Healthcare

Cloud computing allows use of latest technologies at low cost or we can say at pay-per-use and with minimum resources for clients.

Cloud computing is a technology that supports ubiquity, it supplies on demand access to network for sharing computing resources (for example - networks, storage, application, services etc.) without service providers interaction.

In healthcare system, cloud computing provides quick access to medical history of each person at any location can speed up diagnosis and treatment quality and saving lives. Also it can help patients to gain access to their medical history from anywhere in the world via Internet.

Challenges and Suggestion

Yet cloud computing improves access to information, lowers the cost, increases interoperability but still privacy and security challenges exists.

The use of Private clouds ensures the security of data and communication between various departments of healthcare sector.

IoT (Internet of Things) in Healthcare

Internet of things or we can say "Things on internet" is one of the notable technology which has the ability to change and affect the way people work, learn and live. IoT connects everyday objects (like smart phones, sensors, TVs etc.), considered them as smart object and allows them to communicate with each other through internet technologies physically or virtually.

The IoT in healthcare systems improves access to care, increases quality of care and reduce cost of care. Internet of things used by clinical care to monitor physiological condition of patients through sensors by collecting and analyzing their information

and then sending analyzed patient's data remotely to processing centers to make suitable actions. For example, monitor an aging family member, continuous heart rate monitoring, medical fridges, patient surveillance etc.

Challenges and Suggestion

Since IoT changes the clinic centric healthcare to patient centric healthcare, both patients and providers stand to benefit from IoT in healthcare but there may be a possibility of overloading of physician with too much data, results in distracting the treatment paths.

The technique of Internet of Things Analytics (IoTA) in which the raw data is converted into a useful and medically relevant data using the techniques like data extraction and data analytics reduces the problem of storing and analyzing large amount of data.

Conclusion

This paper presents usage and inclusion of some technologies based on area of big data used in different health care sectors. Most health systems can do plenty today with big data, including meeting most of their analytics and reporting needs.

References

1. http://www.ibef.org/industry/healthcare-india.aspx. 2014
2. http://www.informationweek.com/big-data/. 2014
3. Yanglin Ren, Monitoring patients via a secure and mobile healthcare system, IEEE Symposium on wireless communication, 2011.
4. http://hadoop.apache.org/. 2013
5. White, T. Hadoop: the Definitive Guide (2nd Edition) [M]. O'Reilly Media, 2010.
6. Weiyi Shang. Zhen Ming Jiang. Hemmati, H. Adams, B. Hassan, A.E. Martin, P. 2013. Assisting developers of Big Data Analytics Applications when deploying on Hadoop clouds, Software Engineering (ICSE), 35th International Conference. hipi.cs.virginia.edu/ 2014/ 2014
7. cs.ucsb.edu/ 2014
8. Manogaran, G., Thota, C., Kumar, M.: MetaCloudDataStorage architecture for Big Data security in cloud computing. Procedia Comput. Sci. 87, 128–133 (2016)
9. Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: state-ofthe-art and research challenges," Journal of Internet Services and Applications, vol. 1, pp. 7-18, May 2010.
10. P. Mell and T. Grance, "The NIST Definition of Cloud Computing," NIST Special Publication 800-145, September 2011
11. Dave Evans. April 2011. The Internet of Things:How the Next Evolution of the Internet Is Changing Everything, Cisco.
12. G. Kortuem, F. Kawsar, D. Fitton, and V. Sundramoorthy, "Smart objects as building blocks for the internet of things,"Internet Computing, IEEE, vol. 14, pp. 44-51, 2010.