

Big Data in Healthcare: Challenges and Security



Engineering

KEYWORDS : Big data; problems; solutions; security;tools;

Arpitha R Information technology, RVCE, Bengaluru, India

Mandara K R Information technology, RVCE, Bengaluru, India

Smitha G R Information technology, RVCE, Bengaluru, India

ABSTRACT

Amount of data we generate every day and the data collected in everything we do is pretty big. This is what we call "Big Data" and it is helping businesses and every industry to become more efficient and productive. Data in healthcare is growing day by day. Healthcare consists of large amount of unstructured data- like insurance details, electronic health record, sensor data etc.

The main challenge in healthcare systems is how to find, collect, analyse and manage information to make people's lives healthier. The Surgeons would not like to stay in village to fulfil their livelihood requirements so this big data can help the doctors in the village area to refer into multiple records and to give the best treatment to patient. Big data repositories could be an attractive target to malicious insiders and outsiders, so major issue of big data is privacy and security. This paper discuss about the problems of big data and analyzing the solutions to use it efficiently. It also discuss about various security threats and provide solutions to enhance the security.

Introduction

World's population is increasing and there is change in the life style of human being, due to this the modes of treatment is rapidly changing in the health care sector. The main key resource behind this improved medical service is Big Data.

Big Data is a trending technology and is characterised by different kind of data source. It enables the hospitals to gather, store, manage and analyse large amount of data and also effectively deals with complex, heterogeneous data. Big data enables hospitals to analyse a mix of structured, semi-structured and unstructured data which ranges to terabytes, petabytes so managing this large amount of data is very challenging.

The data in health care comes from various sources which includes MRI Scan reports, Blood test reports, CT scan reports, X-Rays images, ECO, blood pressure readings. Big data provides tools like HIVE, PIG to handle structured, unstructured and semi-structure data.

During recent days, the main key behind the success of medical field is to understand as much about the patient more possible so that the signs of future diseases or any other serious illness can be detected in early stage providing better health care. This is basically Prevention is better than cure policy.

USES OF BIG DATA IN HEALTH SECTORS :

The main application area of big data includes the health care sector so let's have a brief look at how big data is helping the health care industry in their progress. Big data in healthcare sector has following uses:

Electronic data:

In today's world all the health care data is made electronic so the data transfer for exchange of knowledge will be easy and there is no chances of misplacing the patient's record.

History Of records:

Maintaining the history of patient's record gives better insight about the patient's diseases so the doctor can have accurate idea about the patient's condition and can provide appropriate care.

Advance Prediction:

By matching the patient's report with the already stored data, any signs of future disease of the patients can be detected which increases the quality of medication.

Benefit to Researchers:

Having such a large amount of patient's data, it provides a sophisticated way for the researchers to use those data for their research.

Innovation:

Using the patient's data that is stored across various location of the world, it helps in new innovations to prevent various diseases and to provide better global health.

CHALLENGES OF BIG DATA IN HEALTH CARE ORGANIZATIONS

As data is growing tremendously, the analysis of big data and making a decision out of it takes large amount of time. Mining a data from millions of rows of data is a challenging task. Following are the main challenges of using Big Data in healthcare industries: [1]

Integration of heterogeneous data:

In healthcare, data is collected from different sources which includes data from medical records, patient scans and images. This collected data is of different formats so aggregating and storing all these data is a complex issue.

Need for Speed:

Today's world needs valid data for analysis but at the same time, getting the relevant data as quickly as possible is also important. In Healthcare industry delay in making decisions leads to less efficient patient care.

Skills:

As big data repository contains data from different sources. Understanding the data and to know how to shape it is a requirement. To handle this huge amount of data, we require certain skills. It is hard to find the data analysts and data scientists and other expertise of particular field.

Large Investment:

A big investment is required for companies to acquire staff (data scientists), resources and also to buy data analytics tools and technologies.

Compatibility:

Addressing the growing healthcare data is a major concern. The medical industries already have Legacy data systems and integrating the data from new technologies is a big issue.

Retrieving meaningful Results:

When dealing with extremely large amounts of information or a variety of categories of information, retrieving the exact result is complex issue but as it is concerned with the life of patient, there should be some means to extract the best result when queried.

Data Sharing:

To share the data throughout the clinical community, how to transfer huge amount of data in a cost effective way is a serious question.

HANDLING BIG DATA

Big data technology provides a process for managing and getting value out of huge amount of data that the organizations have to deal with. In particular, they aim to process the unstructured data which is collected from different sources. There are several tools to handle structured, unstructured and semi-structured data.[2][3]

Hadoop:

Hadoop framework allows user to store and process huge amount of data (big data) in a distributed environment across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Hadoop is an open-source framework.

HDFS:

HDFS (Hadoop distributed file system) is a java based file system that will provide reliable and scalable data storage; it is mainly designed to handle huge clusters of commodity servers. Hadoop distributed file system is fault-tolerant, data is stored in minimum of three datanodes, if any one datanode fails, and data can be retained by other datanodes.

MapReduce:

MapReduce is a framework provided by Google, which mainly consist of two tasks, map task and reduce task. A MapReduce splits the input data-set into different separate sets of data,

Job of *map task* is to process *data set(task)* and all task executed in parallel. The framework sorts all the outputs of the maps, which are provided as input for *reduce tasks*. Both the input and the output of the job are stored on disk. This framework provides faster results.

D. HBase

HBase Provides random access to large amount of structured data. It also provides real-time read/write access to data in the Hadoop File System. We can store the data in HDFS either directly or using HBase. Customer can read/ accesses the data in HDFS using HBase.

E.Hive:

Hive is a data warehouse infrastructure that is built on top of Hadoop framework for providing efficient data summarization, query, and analysis. Hive deals with structured data and it is mainly used data analyst.

F.Pig:

Pig was developed by yahoo. It is platform analyzing large

data sets. Pig is mainly used by Researchers and Programmers and it deals with semi-structured data.

G.Schema-less databases, or NoSQL databases:

Several NoSQL databases includes key-value stores and document stores, which focus on the storage of structured and unstructured data in a model other than tabular format. They achieve performance gains by propagating the changes to all nodes and by providing updated results for the query.

V. BIG DATA SECURITY CHALLENGES:

Since data is growing day by day, number of people accessing this data is also increasing, so we need to protect data sensitive medical data from misuse. Following are the big data security challenges

A. Ownership of data

Almost all the companies are using big data technology to handle and store huge amount of data. Data classification and data ownership is a problem in most of the companies, so it is necessary to identify owners of the output.

B. CIA property:

Information security architecture must include properties like, data availability, data confidentiality and data integrity. Currently many NoSQL databases do not provide encryption, which is one of security problem.

C. Access provision:

As the medical data is sensitive and the data is about the life of a patient, data privacy and providing access to this data to only authorized person is serious security issue.

D. Maintaining real Time Security:

Predicting the threat sources, detecting malware and attacks on medical data is a serious challenge. As the attacks are carried out with a sophisticated mechanism, detecting them and preventing them is of high complexity.

E. Distributed Computing Environment:

Distributed data processing in big data architecture is prone to data leakage and requires high level data security. In distributed system, all the participant systems must have proper data protection.

VI. BIG DATA SECURITY SOLUTIONS:

Managing large amounts of data increases the magnitude and risk of data potential. So, providing a solution to secure sensitive data and enabling the analysis of data for meaningful output, is necessary for any Big data initiative. Some of the primary requirements for the solution are:

A .Limit access:

Controls the access to electronic health records to only authorized persons and provide the access only when it is needed. Protect the access by means of unique identity and strong password. Limit the access to facilities of the system that houses the patient's sensitive data.

B. Training to staff:

Train the staffs about the security rules and procedures. Knowledge the staff about the consequences of data leakage and ask them to follow the security rules without fail. Do the background checks well before hiring the staff.

C. Up to date systems:

Update the firewalls, Anti-virus software and Encrypt all the computers and routers.

D. Security Policies:

Build strong security policies and procedures and also make the violation of those rules as offence.

C. Log Maintenance:

Maintain a log of every transactions and administrative activity to validate the usage and to provide forensic system log.

VII. CONCLUSION

Big data is a boon to health care industry with all patients' data stored at different locations and enabling the doctors to examine the stored data and to provide better healthcare to patients. In this paper, we review some of the storage challenges and security issues in healthcare industry and we also shed light of solutions for handling those challenges to meet the growing demand of securing healthcare data. Overcoming all these challenges makes big data driven health care a successful venture.

References

1. Eiko Kai,Ashir Ahmed," Technical Challenges in Providing Remote Health Consultancy Services for the Unreached Community" 27th International Conference on Advanced Information Networking and Applications Workshops, 2013
2. Prabha Susy Mathew, Dr. Anitha S. Pillai, "Big Data Solutions in Healthcare: Problems and Perspectives" , IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems ICIECS'15
3. Jigna Ashish Patel, Priyanka Sharma, " Big data for Better Health Planning" IEEE International Conference on Advances in Engineering & Technology Research (ICAETR - 2014)"
4. Hiba Asri,Hazar Mousannif,Hassan Al Moatassime, Thomas Noel," Big Data in healthcare: Challenges and Opportunities"2015
5. Smitha Rao,Suma S N,Sunitha M," Security Solutions for Big Data Analytics in Healthcare", Second International Conference on Advances in Computing and Communication Engineering 2015