International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 03 | Mar -2017 www.irjet.net p-ISSN: 2395-0072

PRIVATE HIDDEN DATA FOR HEALTH CARE

Zainulabedin Shaikh, Smita Dhawane, Shadab Khan, Swapnil Kharat

Prof. Pavan Kulkarni, Dept. Of Computer Engineering, Trinity College of Engineering and Research, Pune University, Pune, India

---

ABSTRACT - *This paper provides an understanding of how we can collect healthcare data digitally like patient's records, doctor's details and how we can produce useful data from these records by using analytics techniques and tools which will help patients as well as doctors to save time and money. It generates huge amount of heterogeneous data from diverse sources like patients laboratory test result, medical equipment, health insurance data, drug research, social media, genome study, clinical report, health records, transactions from multiple hospitals. Data analytics is used to manage this data, make it useful for retrieval. Therefore the concept of "big data" can be applied. Big data is characterized as tremendously huge data sets that can be estimated computationally to find trends, patterns, and associations, querying, visualization, predictive analytics and information privacy on large wide spread collection of data.*

Key Words: Big Data, Hadoop, Hive, Data Mining, Clustering, Tomcat, Big Data Analytics.

1. INTRODUCTION

Big data is defined as large data sets or complex data where traditional data processing applications are insufficient to deal with them.

Big data is describes as the huge size of data – both structured and unstructured – that helps a business on a daily basis. But important thing is not the amount of data collected. The important thing is what organization do with data. Big data can be analyzed for insights that lead to better decisions and strategic business moves.

Big data is being generated by the whole thing around us at all times. Every social media and exchange digital process produces it. Systems, mobile devices and sensors transmit it. Big data is incoming from multiple sources at startling variety, velocity and volume. To remove meaningful value from big data, you need analytics capabilities skills and optimal processing power.

2. LITERATURE SURVEY

2.1 BIG DATA IN HEALTHCARE: CHALLENGES AND OPPORTUNITIES

Nowadays huge amounts of healthcare data is generating by use of Mobile phones, patients, sensors, researchers, hospitals, organizations and providers.

To make people's lives healthier and easier healthcare systems find, manage, analyze and collect information, by contributing and understand new therapies and diseases but also to predict outcomes at earlier stages and make real-time decisions. This explain the benefits of big data to healthcare and discover how it empowers patients and improves treatment, researchers and providers.

We define the skill of authenticity mining in collecting large amounts of data to recognize predict outcomes, detect people's habits and illustrate the profits of big data analytics.[2]

2.2 PREDICTIVE BIG DATA ANALYTICS IN HEALTHCARE

Nowadays in world the enormous set of data is produced from different organizations all over the world. This large and heterogeneous data is called Big Data. Big Data Analytics provides tremendous visions to different organizations especially in healthcare.

The traditional database designs cannot face the challenge with huge data, which is generating in organizations today, and it creates a big havoc.

In achieving predictive analysis in the healthcare domain Big Data plays an important role. Big Data can handle huge explosion of data, which is found in medical organizations.

This paper provides an overview of retrieving and storing methods, Big Data techniques and tools used in healthcare clouds, part of Big Data Analytics in healthcare and discusses the benefits, faces challenges and offers solutions.[2]

2.3 SECURITY SOLUTIONS FOR BIG DATA ANALYTICS IN HEALTHCARE

Today, data is a strategically important asset, structure and the goal is to maximize the value of their knowledge. Increasing number of companies are using technology to collect and analyze petabytes of data including click stream data, web logs and social media content to gain better perception about their customers and their business.

Big Data analytics provides large competitive advantage for corporations, supporting the businesses, tailor their products to consumer requirement.in healthcare big data refers to electronic health data sets that are related to patient healthcare and well-being.

In healthcare sector the security and privacy issues of big data are a big concern as data is bound by international regulations like the Health Insurance Movability and Answerability Act (HIPAA), The Health Knowledge Technology for System and Medical institution Health (HITECH), HCSC, FTC (Federal Trade Commission) etc.

In Hospital Information System the applications that are impacted by big data are genetic science, pharmacovigilance, patient care etc. The aim of this paper is to present different viable security solutions to support the potential of big data concern to healthcare in a highly regulated environment.[3]

3. SYSTEM OVERVIEW

3.1 SYSTEM ARCHITECTURE:

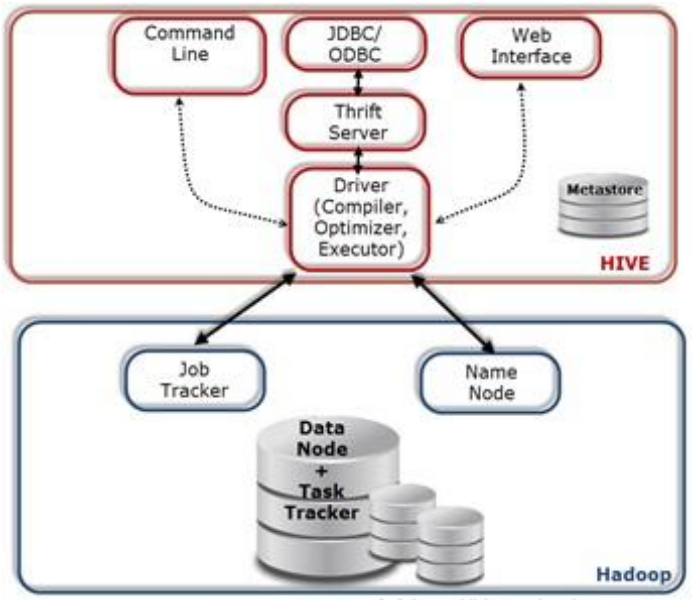

Fig: System architecture

3.2 HIVE

Hive is a data warehouse infrastructure tool to procedure structured data in Hadoop. It exist on top of Hadoop to makes querying, summarize Big Data and analyzing easy.

Primarily Hive was greatly developed by Facebook, later the Apache Software Foundation took it to develop it further as an open source under the name Apache Hive. It is used by different companies. For example, Elastic MapReduce is used by Amazon

Sqoop, Hive, and Pig are tools of Hadoop system that are used to help Hadoop modules. Sqoop: It is used to export and import data to and from between HDFS and RDBMS. Pig: It is (a procedural language platform) used to develop a script for MapReduce operations.

3.3 HADOOP

Hadoop is Java-based programming framework that offer supports to the processing and storage of tremendously large data sets in a distributed computing environment. It is sponsored by the Apache Software Foundation.

Hadoop is used to run applications on systems with thousands of hardware nodes, and to handle thousands of terabytes of data. Distributed file system allows swift data transfer rates among nodes and in case of a node failure it allows the system to continue operating.

This method lowers the risk of disastrous system failure and unpredicted data loss, even if a significant number of nodes become inactive.

3.4 TOMCAT

Tomcat is an application server from the Apache Software Foundation that run Java servlets and renders Web pages that contain Java Server Page coding. Described as a "reference implementation" of the Java Servlet and the Java Server Page contract.

Tomcat is the result of an open practice of developers and is reachable from the Apache Web site in both binary and source versions. Tomcat can be used with other Web servers or as a single product with its own inside Web server, including Apache, Netscape Enterprise Server, Microsoft Internet Information Server (IIS), and Microsoft Personal Web Server. Tomcat requires a Java Runtime Project Situation that descriptor to JRE 1.1 or later.

Apache Tomcat is an open source webserver product of Apache Relation like Apache HTTP server.

It is used to deploying JSP applications and Java Servlet. To spread any application in Tomcat we can simply create a war file and spread them.

Apache Tomcat, frequently mentioned as Tomcat Server, is an open-source Java Servlet Container developed by the Apache Software Foundation (ASF). Tomcat utilize several Java EE specifications including Java Servlet, Java Server Pages (JSP), Java EL, and WebSocket, and provides a "pure Java" HTTP web server environs in which Java code can run.

4. IMPLEMENTATION

Our project will provide healthcare services, it contains modules such as:
1. Hospital login
2. Patient login
3. Doctor login
4. Lab
5. Nearest Hospitals

1. HOSPITAL:-

The hospital manages all the details of patients. The management has adds the patient, doctor and also check or view the details of doctors, checks the reports, checking the details of visiting doctors , checking the details of patient etc.

2. PATIENT:-

The patient can take an appointment of respective doctors. He can check doctors availability.

3. DOCTOR:-

The doctor has checks the patient at the time of checking taking the details of patient and gives the prescription to the patient.

4. LAB:-

Lab checks the prescription according to the doctor has suggest the patient what type of medicine and reports required for patient. Lab also stores the information of the patient.

4.1 MAP REDUCE (HADOOP) TECHNIQUE:

Patients can check their previous prescription of doctor in their login form.

Hospital can generate patient admit bills by using map reduce.

Lab can maintain particular patient diseases related report so patient can fetch by their account, where data can be stored and retrieved by using HIVE database.

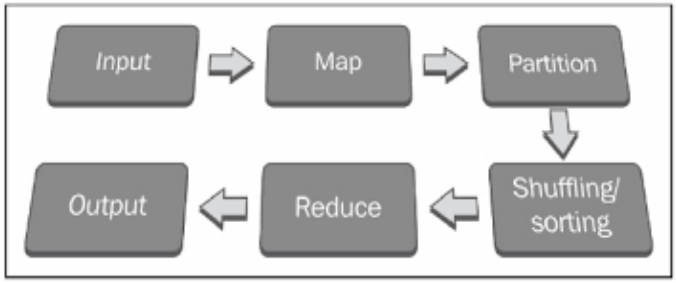

Fig: Map reduce flow.

5 MATHEMATICAL MODEL

In this sytem server will find out the nearest hospital depending upon various variables related to the emergency. It will autimaticallt find the hospitals nearby patient. After successful searching it will suggest the nearest hospital to the client or patients. Our application also provide the service which leads to know the any hospital near by patient.Also tells, the shortest distance between source and destination. Our application provide alarm / notification which will tell user specified destination while travelling in the form of alarm

System S= {EMSApplication}
SystemS = {S1, I, δ,O}
S1= {Search,GoogleMaps}
I = {V,SD}
δ = Function
O = Output
I1 = V -» Variables.
I2 = SD -» SourceAndDestinationArray.
[1] I1 = {Area,EmergencyType}
δ 1 = I1-»O1
O1 = {H1,H2,H3….,Hn}
H -»Hospital in that area which satisfies criteria.
[2] I2 = {Source;Destination}
δ 2 -» Cal
Cal = {Source,Destination}
R = {R1,R2,R3….,Rn}
R1 = {Source,Node1,Node2,Node3….,Noden}
--Source = Distance [Source]+n
Σ DistanceBetween [Source,Nodei]
D = {D1,D2,D3,…….,Dn}
[R = Routes;D = Distance]
MinDistance = Min (D1,D2,D3……..,Dn)
O2 = {MinimumDistance}
[3] I3 = {OH1,OH2,OH3..,OHn}
OH-»OptimalHospital
Distance = Sort (OH1,OH2, OH3……..,OHn)

6. SYSTEM REQUIREMENTS

SOFTWARE:

Operating System: Fedora, Ubuntu, (VM Ware)
Hadoop, Tomcat7, Eclipse, Java.

HARDWARE:

RAM: Min 8 GB
Hard disk: Min 1TB
Processor: core i5, i7 Min 2.50 GHz

7. CONCLUSION

Here we have introduced big data for healthcare system to make life easy for patients and doctors. The increase of diseases and viruses will be handled by this system.

ACKNOWLEDGEMENT

This research has been detailed in the structure of the project "Private hidden data for health care". Expanding our sincere and heartfelt thanks to our guide, Prof. Pavan Kulkarni, for providing us with the right guidance and for showing the right way, precious suggestions and guidance. Lastly we would like to thank our family & friends for the assist and confidence they have given us during the course of our work.

REFERENCES

1) "Big data in healthcare: challenges and opportunities" authors: hiba asri , hajar mousannif, hassan al moatassime, thomas noel.

2) "Predictive big data analytics in healthcare" authors: a. rishika reddy, p. suresh kumar

3) "Security solutions for big data analytics in healthcare" smitha rao, s.n. suma, m. sunitha

4) "Parallel Algorithms for Mining Large-Scale Rich-Media Data" Author :E.Y. Chang, H. Bai, and K. Zhu.

5) "Collective Mining of BayesianNetworks from Distributed Heterogeneous Data" Author :R. Chen, K. Sivakumar, and H. Kargupta.

6) "Disco: Distributed Co-Clustering with Map-Reduce: A Case Study Towards Petabyte-Scale End-toEnd Mining," Author: S. Papadimitriou and J. Sun

7) C. Ranger, R. Raghuraman, A. Penmetsa, G. Bradski, and C. Kozyrakis, "Evaluating MapReduce for Multi-Core and Multiprocessor Systems,"