Effective Implementation of Data Segregation & Extraction Using Big Data in E-Health Insurance as a Service

[1]A.Arun, [2]G.Kumaresan

[1] PG Scholar, [2]Assistant Professor
Department of Computer Science and Engineering,
Valliammai Engineering College, Chennai, India.

Abstract: **The rise of big data applications where data collection has grown tremendously and is beyond the ability of commonly used software tools to capture, manage, and process within a "tolerable elapsed time". A now day's E-hospital record becomes most important in many countries. There are much standardization putting effort on data exchange and interoperability. E-hospital document provides different applications consider medical research, accounting, billing and commerce intellectual property. E-Health Record decrease human workload, hospital cost and modify individual health management. Due to E-hospital record user can access and store health record like emergency information like blood group, medication history and electronic prescription. In cloud E-hospital document store and process very sensitive patient data and should have a proper privacy framework and security mechanism since the reveal of health record may have social result consequence especially for patients. A prototype of a POC system aims at assisting medical doctors with the interpretation of drug concentration values and conducting clinical studies. It will make a censorious step towards personal medicine, and demo the practicability of performing Therapeutic Drug Monitoring (TDM) using portable devices while decreasing costs and improving treatment efficiency. In the modification, we are analyzing more number of Factors like Disease Types with its Corresponding Reasons, Insurance policy Details with Sanctioned Amount, Family Grade wise Segregation.**

Keywords**: E-hospital, Therapeutic Drug Monitoring, POC, big data.**

I.INTRODUCTION

Big data

Big Data analysis is best left to software programs. When data analysts go straight to the complex calculations, before they perform a simple estimation, they will find themselves accepting wildly measly calculations. Big Data is a relatively new term that came from the need of big companies like Google, Yahoo, and Facebook to analyse big amounts as well in the development research. Big data takes too much time and cost to load into traditional relational database for analysis, New approaches to storing and analysing data have emerged that rely less on data schema and data quality. Big data is opening up new chance for enterprises to extract insight from large volumes of data in real time and across multiple relational and non-relational data types. It is becoming one of the most important technology directions that have the potential for dramatically changing the way organizations use information to enhance the customer experience and transform their business models Instead, raw data with extended metadata is aggregated in a data lake and machine learning and artificial intelligence (AI) programs use complex algorithms to look for quotable percept .Bag data anatomy is often associated with cloud computing because the analysis of bulky data sets in real-time requires a platform like Hadoop to store large data sets across a scattered cluster and Map Reduce to organize, feature and process data from multiple sources.

BIG DATA PLATFORMS AND TOOLS

Hadoop

The Apache scattered data processing software is so permeate that often the terms "Hadoop" and "big data" are used synonymously. The Apache Foundation also sponsors a number of related projects that increase the strength of Hadoop, and many of them are proceeding below. In addition, numerous vendors offer supported versions of Hadoop and related technologies. Operating System: Windows, Linux, OS X.

Map reduce

Originally developed by Google, the MapReduce website describe it as "a programming model and software framework for writing applications that rapidly process vast amounts of data in parallel on large clusters of compute nodes.It's used by Hadoop, as well as many other data processing applications. Operating System: OS Independent.

Grid grain

Grid Grain offers an alternative to Hadoop's MapReduce that is compatible with the Hadoop Distributed File System. It offers in-memory processing for fast analysis of real-time data. Operating System: Windows, Linux, OS X.

Hpcc
Developed by LexisNexis Risk Solutions, HPCC is short for "high performance computing cluster." It claims to offer superior performance to Hadoop. Both free community versions and paid enterprise versions are available. Operating System: Linux.

Storm
Now owned by Twitter, Storm offers distributed real-time computation capabilities and is often described as the "Hadoop of real time." It's highly scalable, robust, and fault-tolerant and works with nearly all programming languages. Operating System: Linux.

II.RELATED WORK

The world is gearing up towards implementation of digital healthcare networking solutions; the solution roadmap for a national e-Health interconnection network infrastructure is in great demand. This paper first identifies the major benefits in interconnection services that are critical to universal deployment. Then it provides an overview of our e-Health AON (Application Oriented Network) approach that offers direct support to the aspects of interconnection services, operational management services, and security control services. [1]This paper focuses on the e-Healthcare application cloud-enabling characteristics and we found the close proximity of the new e-Healthcare architecture and the cloud environment. However there are still challenges in adaptation of a pure cloud solution for digital e-Healthcare. This paper documents those similar characteristics and the new challenges. It also introduces our adaptation architecture design for e-Healthcare Cloud Computing and Networking Solutions. [2]The research efforts for a national eHealth interconnection infrastructure and design guidelines are in great demand. This paper identifies the major challenges in eHealth interconnection network services that are critical to universal deployment. An overview of our solutions framework is summarized with the aspects of interconnection services, operational management services, and security control services.[3]this paper, we propose a model called Temporal & Co reference Topic Modelling (TCTM) to do automatic annotation with respect to the Time Event Ontology (TEO) for the big-size Electronic Health Record (EHR). TCTM, based on Latent Dirichlet Allocations (LDA) and integrated into MapReduce framework, inherently addresses the twin problem of data sparseness and high dimensionality. As a non-parametric Bayesian model, it can flexibly add new attributes or features. Side information associated with corpora, such as section header, timestamp, sentence distance, event distance or disease category in clinical notes makes latent topics more interpretable and more biased toward co referring events. Furthermore, TCTM integrates Hidden Markov Model LDA (HMM-LDA) to obtain the power of both sequential modelling and exchangeability [4] both permanent data info and real-time data flows should pass through the middleware during communication of data sources and application. And by implementing a corresponding data source service wrapper, it's very convenient to add a new data source. At the same time, the middleware uses XML to accomplish data mapping and transmission, so as to solve the incompatibility of data sources schema and ensure platform independence. We deployed the middleware to our E-Health project. Our experience has proved the flexibility and extensibility of the middleware. [5] The No Relational Data Bases are one of the key factors (among the architecture & infrastructure that supports them) of the complete develop of the IoT, and how they must lead the way to the Web of Things. [6] How the No Relational Data Bases are one of the key factors (among the architecture & infrastructure that supports them) of the complete develop of the IoT, and how they must lead the way to the Web of Things. [7]

III.SYSTEM DESIGN

Insurance Agencies are actively participating for the Analysis of Patient's Data and used to Extraction some Useful Information. Analysis of Discharge Summary, Drug & Pharma, Diagnostics Details, Doctors Report, Medical History, Allergies & Insurance policies are made and Useful Data is extracted. In the alteration, we are analyzing more number of causes like Struma Types with its related Reasons, Insurance policy Details with Sanctioned Amount, Family Grade wise Segregation.

Advantage of proposed system

1. More accurate data
2. Improved business decisions
3. Improved marketing strategy and targeting
4. Increased revenue due to increased customer and base and decreased costs.

ARCHITECTURE OF N-TIER

Fig: 1.1 N-Tier Architecture

The discuss in figure1.1 N-Tier Architecture of My SQL Cluster consists of a set of computers, known as hosts, each running one or more processes. These processes, known as nodes, may include My SQL servers (for access to NDB data), data nodes (for storage of the data), one or more management servers, and possibly other specialized data access programs.

SYSTEM ANALYSIS

1. Patient Data Gathering
2. Insurance server
3. Multi Access Control
4. Disease based Data Grouping
5. Discharge Summary & Doctors Report
6. Blood Report with Pharma Analysis
7. Patient Previous History & Personal Data Analysis
8. Big Data Extraction of Useful Information

1. *Patient Data Gathering*

In this module user has to register and these data are collected during the first visit at the outpatient clinics and includes information such as birth date, date of last negative and first positive tests, route of infection, and alcohol and drug usage. Data, including clinical data, are collected on a continuous basis every time the patient is seen by the treating physician. This includes such information as the treatment, symptoms of disease, and laboratory results.

2. *Insurance Server*

In this module we create server in which all the patient information from the hospital to the insurance server. Insurance servers are constructed with the files and the medical claim information is maintained in the insurance server. The data are added in each insurance server, and network construction is made with the entire medical data present in each insurance server. Query is given to the insurance server, so that the insurance server will verify the medical information present in it. And main thing this is allow to access by the big data analyst because he only go do the research and provide the useful information

3. *Multi Access Control*

In this module we design to implement the three different type of record because we designed retrieve three different type information from the user side and they are in the to a lower place point. Once the User creates an account, they are to login into their account and request the Job from the Service Provider. Based on the User's request, the Service Provider will process the User requested Job and respond to them.

➢ ***Insurance Account***

In this account we will add the information about the patient that we retrieve from the hospital i.e. we get the information for the patient applied for the insurance and we will also check their details about the claim bill.

➢ ***Patient Account***

In this patient account we designed this module we make use of the patient because the patient doesn't known about the direction documentation that is whether prescribed detailed medicine is correct or not. So by using this account they can search.

➢ ***Research Account***

In this research module we designed to deploy the big data analysis. Big data is nothing but vast amount of data with unstructured format in this unstructured format data we can valid database so we used this idea of get vast of database from the insurance domain to get useful information in the domain of health care.

4. ***Disease based Data Halogen***

In this spacecraft the big data assayer is going to collect the database about the disease form the insurance server. This can be done once the insurance company accept to allow their customer database to the big data assayer so big data analyst will separate the data based on the disease. By doing this we can easily the get the disease information and list customer applied for the disease which can be indentified and they can be analyzed for the future use.

5. ***Discharge Summary & Doctors Report***

In this module we developed a design to insert the discharge summary and doctors report that was given to the patient. In this module we have the information like date of admission of the patient ,date of discharge of the patient ,discharge diagnoses, reason the patient admitted ,patient problem in briefed, consultant name, doctor prescriptions about medicine these are information must be loaded into the insurance server

6. ***Blood Report with Pharma Analysis***

WBC 8.6, 29haemoglobin 13.4, hematocrit 39.8, platelets 207,000, MCV 91.6, neutrophil percentage of 72.6%. Sodium 133, potassium 4.7, chloride 104. Blood urea, nitrogen of 18 and creatin of 1.1. PT 17.4, INR 1.6, PTT 33.The patient had a chest x-ray, which showed cardiomegaly with atherosclerotic heart disease, pleural thickening and small pleural effusion, a left costophrenic angle which has not altered when canvas to prior examination, COPD pattern. The patient also had a head CT, which showed atrophy with old ischemic changes. No acute intracranial findings.

7. ***Patient Previous History & Personal Data Analysis***

The system is fashioned to equal data that accurately interpret the state of the patient at all times. It allows for an entire patient history to be viewed without the need to track down the patient's former medical record volume and succour in guarantee data is accurate, appropriate and legible. It reduces the chances of data replication as there is only one modifiable file, which means the file is invariably up to date when viewed at a later date and obliterate the issue of lost forms or paperwork. Due to all the information being in a single file, it makes it much more useful when demodulate medical data for the examination of assert able tendency and long term changes in the patient.

8. ***Big Data Extraction of Useful Information***

In this module we utilize big data, in this big data we will have lot or vast amount of data that may wanted or unwanted information in simple the information in the big data are unstructured. So in this compartment the insurance server is going allow permission to access the server by the big data analyst. The big data analyst get the all the information which mention above and demodulate the database by the technique of map reaction arrangement to get useful information which is useful for both insurance and patient.

IV.RESULT

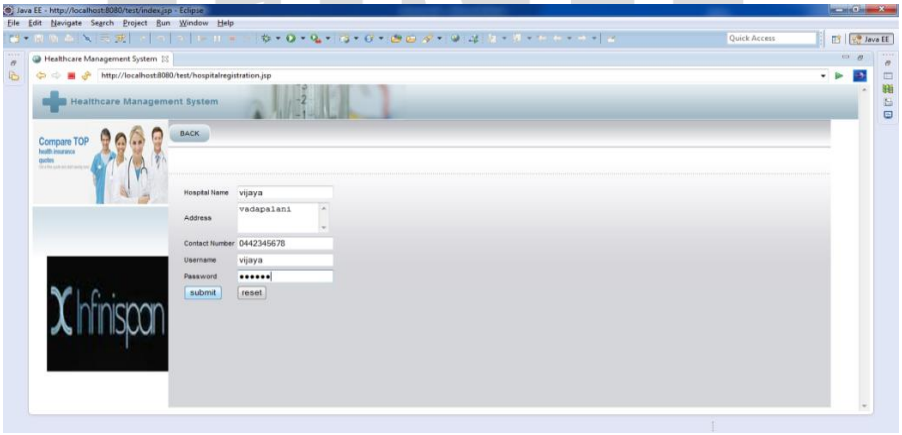

Fig1.2 Registration for Hospital Registration, Insurance Registration, Doctor Registration

The discus about figure: 1.2 hospital, Insurance, Doctor Registration process. And enter the own details and store in Database

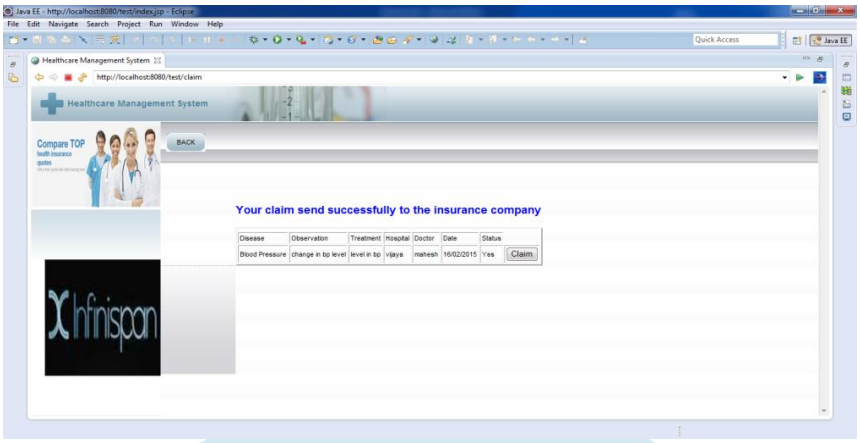

Fig1.3 Claim as Send to the Insurance Company

The discus about figure: 1.3 claims as send to the Insurance Company. That company claim the Insurance money through hospital.

V. CONCLUSION

We have presented a new BDeHS (Big Data for Health Service) approach to bring healthcare flows and data message management brilliance into the scheme for better branch processing, operational management and regulatory compliance. By asserting security control into the service layer with native e-Health features our solutions ensure regulatory compliance. When number of such servers is being used for processing huge amount of data, the processing work is divided into multiple tasks and assigned dynamically to every server embroiled. This alleviates the system to work faster as the accessible nodes share the job of processing such voluminous data in a short span of time. This is possible due to parallel processing of data.
To enhance the performance further these companies have developed customized frameworks that take care of parallel data processing in an efficient fashion. There are many such frameworks existing in the real world.

REFERENCES

[1] W. Liu and E.K. Park, "e-Health AON (Application Oriented Network)", IEEE International Conference on Computer Communication Networks, BMAN Workshop, Nausa, Bahamas, July 2013.

[2] E.K. Park and W. Liu, "e-Healthcare Cloud Computing Application Solutions", IEEE ICNC2013, International Conference on Computing, Networking and Communications, San Diego, CA, January 2013.

[3] W. Liu, E.K. Park and Udo R. Krieger, "e-Health Interconnection Infrastructure Challenges and Solutions Overview", IEEE HealthCom-2012, the 14th IEEE International Conference on e-9877

Health Networking, Application & Services, Beijing, China, October 2012.

[4] C.G. Chute, "Obstacles and options for big-data applications in biomedicine: The role of standards and normalizations", 2012 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).

[5] D. Li, C. Tao, H. Liu and C. Chute, "Ontology-Based Temporal Relation Modelling with Map Reduce Latent Dirichlet Allocations for Big EHR Data", Second International Conference on Cloud and Green Computing (CGC) , 2012.

[6] X. Lu, H. Tang, W. Cheng and T. Zhang , "Heterogeneous Data Source Middleware for Android E-Health Application", Eighth International Conference on Mobile Ad-hoc and Sensor Networks (MSN), 2012.

[7] M. Diaz, G. Juan, O. Lucas and A. Ryuga, "Big Data on the Internet of Things: An Example for the E-health, "Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012.

[8] W. Liu and E.K. Park, "e-Healthcare Security Solution Framework", IEEE International Conference on Computer Communication Networks, MobiPST-2012, Munich, Germany, August 2012.

[9] W. Liu and E.K. Park, "e-Health Service Characteristics and QoS Guarantee", IEEE International Conference on Computer Communication Networks, Workshop on Context-aware QoS Provisioning and Management for Emerging Networks, Applications and Services, Maui, HI, August 2011.

[10] J. Yang, D. Tang and X. Zheng, "Research on the distributed electronic medical records storage model", International Symposium on IT in Medicine and Education (ITME), 2011.