BIG DATA E-HEALTH RECORDS MANAGEMENT ON CLOUD USING MONGODB

[1]Muralidharan G, [2]Nandha Kumar T, [3]Mrs. Suguna M

[1]Department Of CSE,Kumaraguru College of Technology,Coimbatore
[2] Department Of CSE,Kumaraguru College of Technology,Coimbatore
[3] Department Of CSE,Kumaraguru College of Technology,Coimbatore

Abstract- *Electronic Medical Records are digital version of a patient's health records. These records are very sensitive and critical in analysing the patient's medical history. The project proposes a new design for large-scale Patient Health Record Maintenance System Using Cloud Storage. The proposed system can be used to protect different data content types, including videos, images, audio clips, and text and report document files. A "NoSQL" database can be well suited for such voluminous and unstructured data. With the use of Mongo DB, a NoSQL database, storage can be provided for various unstructured data gathered from various different sources. By storing the EHRs in cloud the proposed system provides high security, scalability and rapid and quick access to the files. The data is stored in Hadoop Distributed File System (HDFS) which is very useful framework for handling Big Data like EHR. The EHRs consist of large amount of various types of data, both structured and unstructured, coming from different sources which is a form of Big Data. The Project is mainly for managing patient information in a centralized manner and to provide retrieval of data from anywhere. The system also provides patients with access to their information so that the patients can enter their login credentials and they can easily get their information. The system will successfully manage the medical records and minimize the workload of the stakeholders.*

Key Words- *Cloud computing, MongoDB, Electronic Health Record (EHR), Big Data*

1. INTRODUCTION

Health care data are valuable resources that should be managed efficiently for maintaining the health history of an individual. An Electronic Health Record (EHR) is a digital version of a patient's health record. EHRs consist of dynamic information that varies with time. It includes patient's medical history, diagnosis, medications, treatments undertaken, laboratory scan results etc. But the electronic, management and retrieval of health data is complex, voluminous, distributed, dynamic, unstructured and heterogeneous. The data gathered come from various sources and does not have a specific structure associated with it. Storing such data in traditional database has many setbacks like high cost, less scalability, reliability etc.

The volume of health care data increases with time. Health care data is a form of Big Data. By using a distributed file system like HDFS data can be stored efficiently and it provides powerful processing capabilities with high security and scalability. Hadoop uses Map Reduce framework which can be used to write applications to process huge amounts of data, in parallel, on large clusters of commodity hardware in a reliable manner.

A "NoSQL" database can be well suited for such voluminous and unstructured data. "NoSQL" stands for "Not only SQL" and most of the NoSQL databases are open-source thus providing a cost free resource. By storing the EHRs in cloud this system provides high security, scalability and rapid and quick access to the files. This project mainly focuses on providing a EHR management tool with data

stored on cloud using Mongo DB to provide security, rapid access and scalability for the voluminous health records.

2. LITERATURE REVIEW

Mehmet Zahid Ercan et al (2014) analysed the use of NoSQL databases on EHR systems. They propose [1] that The structure and size of the healthcare data have changed considerably over time. Recent literature shows that the emerging NoSQL databases have significant advantages such as easy and automatic scaling, better performance and high availability which address the limitations of relational databases in distributed healthcare systems. The key benefits of of NoSQL as proposed by the paper are , a. Information sharing function of EHR in improving healthcare outcomes, b. Support continuity of care, efficient, consistent and universal sharing of health information, c. Support for the development of health policies, medical education and advanced research. This paper compares the performance of SQL and NoSQL databases in handling health records. Considering the importance and sensitivity of EHR data, NoSQL databases have significant potential to lead to better EHR applications in terms of scaling, flexibility and high availability.

Olutayo Boyinbode et al (2015) studied the utilization of modern information technology in the delivery of healthcare is to enhance the availability and reliability of improved healthcare services to patients at a reduced cost. They proposed 'CloudeMR: A Cloud Based Electronic Medical Record System' [2]. The healthcare industry in Nigeria has not been able to tap into the full potential of modern information technology to improve on healthcare delivery. Accesses to patients' longitudinal records are often times difficult and cumbersome. This study differs to every other works considered by taking into account all the lapses to design a robust, yet simple and effective electronic medical record (EMR) system that will ensure security, privacy, reliability and availability of patient's records.

This paper proposes and implements a cloud-based electronic medical record (CloudeMR) [2] system to improve the delivery of healthcare system in the rural communities of Nigeria. It implemented a simple and efficient cloud based medical records system in Nigeria. It is understood that cloud computing can significantly improve the performance and reduces the cost of healthcare delivery through reduced access bottlenecks.

Sophia Kingsley et al (2015) portray the concept and benefits of big data PHR (Public Health Record) on cloud in detail in their paper 'Big Data Personal Health Records Management and Analysis on Cloud Using No SQL-Mongo DB' [3]. To ensure the availability of individual and ease and fast retrieval of data on PHR cloud, the use of Mongo DB-NoSQL database as a service is essential. The platform such as Mongo DB is suitable for storing PHR data due to its scalability and flexibility in structural format for storage. Mongo DB stores structured data as JSON-like heterogeneous documents with dynamic schemas. Mongo DB has the capability to scale horizontally through sharding. It proposes a PHR system that uses MVC framework and Entity framework to keep maintainability. PHR mainly consists of three parts. The model part includes two main parts which are the security model and the data model. The middle part of the PHR is controller that handles all data logic and user requests. The paper also proposed the use of Stemming algorithm for analysis purposes. Stemming programming was used to filter the sensitive information concerning the patient from the normal information which may be hosted on public cloud.

S. Naisha Sultana et al (2014) propsed 'Cloud-based development of Smart and Connected data in Healthcare application' [3]. This paper describes the design of an Information Integration and Informatics framework that allows storing, integrating and analyzing healthcare data in the cloud. The Information Integration and Informatics

framework allow the development of advanced healthcare application with data integrated in difference database. Application developers can quickly develop healthcare

applications by not thinking about the data management in cloud and cloud infrastructure management deployment configuration which are taken care by the Information Integration and Informatics framework. The proposed system used Aneka Cloud Platform for the development.

Jyotsna Talreja Wassan et al (2014) proposed a sample modelling of general framework [5], based on extracting patterns and generating recommendations is reviewed for maintaining and accessing EHRs, which are form of Big Data. It proposes how data storage with Big Data stores and analytics with Map Reduce paradigm, may be performed on simulated health data. Big Data paradigm may help greatly in generating new knowledge in health care industry by analyzing unstructured and schema less data efficiently. Also it is useful in sharing information across hospitals and various stakeholders via distributed computing. The layer-wise modular framework for accessing EHRs is suggested in this paper to cope up with the complexities involved in dealing with large scale digitized health data. The paper reflects that mining components and recommendation engines working above them could be useful in effective utilization of EHRs for various stakeholders with the growing number of recorded features in EHRs. The framework gathers information about consumers of health data from EHRs stored in a scalable fashion and tries to calculate similarities between different concepts to produce recommendations. The Big Data platforms available in today's world can effectively be used for storing health data and analyzing it with Map Reduce paradigms.

3. PROPOSED WORK

Most of the EHR systems implemented on the cloud, store data using the standard relational databases like SQL and ORACLE. This model has many restrictions and do not perform well when working with large voluminous data. Those are also proprietary software and require high cost. The standard SQL based system can only handle structured data while most of the electronic health records are unstructured and dynamic. This leads to many problems like scaling, availability, processing time and security.

The main objective of the proposed system is to provide a centralized and distributed system for managing electronic health records of patient. The system helps to improve security of e-health records stored in centralized database. The proposed system provides the patient with a report about their health. By using a distributed file system like Hadoop File sharing System (HDFS) the large data sets can be stored reliably and can be streamed quickly to the application. The proposed system uses Mongo DB, a open-source NoSQL database to handle unstructured and voluminous data gathered from various different sources. The proposed system also provides security by encrypting the files using the files in the cloud.

The proposed system provides various advantages over the existing system like improved security, centralized and quick access, easy maintenance, high accuracy, high scalability and improved reliability.

Data Flow Diagrams of Proposed system
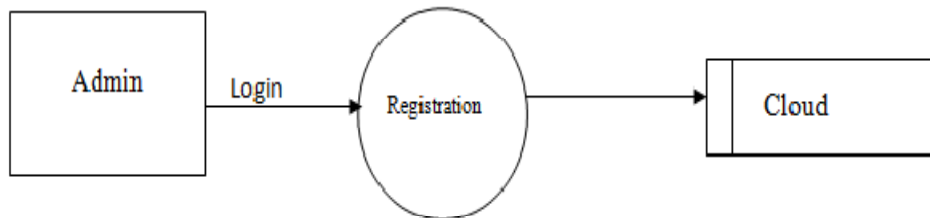
LEVEL 0

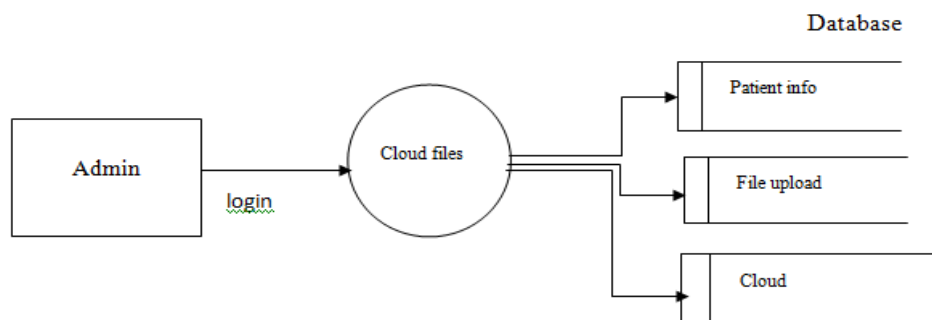Fig.3.1. Level 0 DFD

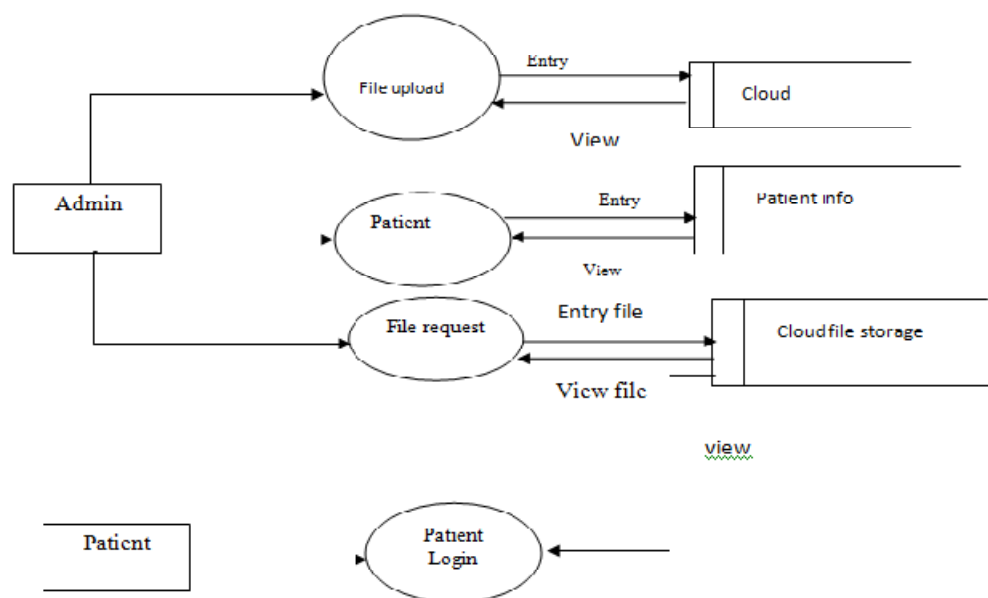## LEVEL 1

Fig.3.2. Level 1 DFD

LEVEL 2

Fig.3.3. Level 2 DFD

The proposed system has various modules for doctor and patient registration, uploading EHR Files, Viewing Patient Files, Viewing Cloud Files, Secret Key Request & Response, Downloading Files , Patients Requesting Files and Report generation.

4. CONCLUSION

With the vast and rapid development of various health care techniques and so that of related health care data, the existing systems developed with standard SQL based data model has several disadvantages such as less scalability, high cost and inability to handle unstructured and Big Data related to Electronic Health Records. The shortcoming of the existing models is overcome with the proposed project. By distributing the Big Data file over multiple cloud servers there is a great improvement in efficiency, reliability, scalability and data security. It uses Mongo DB, an open-source NoSQL document oriented database, to handle all the unstructured data coming from different sources. The Mongo DB also provides high scalability through automatic sharding thus enabling quick and rapid access. By storing all these large amount of data in cloud, this system provides a centralized access to the database and improves the security and efficiency of the E-Health Records management system.

REFERENCES

[1] Mehmet Zahid Ercan and Michael Lane , ' Evaluation of NoSQL databases for EHR systems', 25th Australasian Conference on Information Systems, December 2014.

[2] Olutayo Boyinbode and Gbenga Toriola , ' CloudeMR: A Cloud Based Electronic Medical Record System', International Journal of Hybrid Information Technology Vol.8, No.4 (2015), pp. 201-212 , 2015.

[3] Sophia Kingsley Okore and Edwin Sweety Bakyarani , 'Big Data Personal Health Records Management and

Analysis on Cloud Using No SQL-Mongo DB', International Journal of Computer Trends and Technology (IJCTT) – volume 28 Number 2, October 2015.

[4] S. Naisha Sultana, Gandikota Ramu and Prof. B. Eswara Reddy, 'Cloud-Based Development Of Smart And Connected Data In Healthcare Application', International Journal of Distributed and Parallel Systems (IJDPS) Vol.5, No.6, November 2014.

[5] Jyotsna Talreja Wassan, ' Modeling Stack Framework for Accessing Electronic Health Records with Big Data Needs', International Journal of Computer Applications (0975 – 8887) Volume 106 – No.1, November 2014.