May - June 2020
ISSN: 0193-4120 Page No. 5194-5198

Novel Design using Hadoop Technique for Managing and Integration of Clinical Data

Krishnamoorthy R[1], Dineshkumar T[2], Syed Mazhar A[3], Thirumalai M[4]

[1,2,3,4]Department of ECE
[1]Sree Sastha Institute of Engineering and Technology, Chennai, India
[2]Kongunadu College of Engineering and Technology, Tamil Nadu, India
[3]Dhanalakshmi Srinivasan Institute of Technology, Tamil Nadu, India
[4]Saveetha Engineering College, Chennai, India
[1]krishnamoorthy.mkr@gmail.com, [2]dineshrajumanibe@gmail.com, [3]syedmazhartn@gmail.com, [4]thirumalai3788@gmail.com

Article Info
Volume 83
Page Number: 5194-5198
Publication Issue:
May - June 2020

Abstract

Conventional methods and tools for processing the huge amount of data in medical field does not meet the need hence these challenges of big data of medical field has to be resolved in order to realize the big data's promise. The process of integration is aimed as it is a critical problem. This paper proposes the integration of clinical data with the system to manage the data based on the platform of Hadoop. The system is connected with various sources of heterogeneous source of data namely LIC, EMR and PACS. Based on the sharing of data, the history of the patient's medical record can be retrieved in order to understand different information about the patient's health issue. Information system in medical field collect data and high storage is required for saving medical history of patients. Extraction of useful information can be done by PHR or EMR for providing prescription for the patients while the same data can provide support for doing research through statistics of the results and data analysis. But data storage for massive health care is not possible through the framework of RDBMS. Hence, the proposed method implements Hadoop framework for management of data in massive health care which archives scalability, better performance and tolerance of faults.

Article History
Article Received: *19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *16 May 2020*

Keywords: *Heterogeneous Source, Hadoop Platform, Big Data, Data Management.*

1. Introduction

In the current era, data generated by the industrial applications has increased reaching several Terabyte and Petabyte as incorporation of computer technology has increased the size of data. This is the reason for the application systems for being expanded to operate at high speed. Hence the explosion of information is evolved [1] such that a change is leaded. Explosion of data have occurred in the medical field too in this current era of big data. Additionally, big prospects are shown in the applications of clinical data comparable to the efficiency of research, monitoring of remote data and preventing chronic disease by research of drugs with personalization of the medicine. Integration and management of big data of medical field is made critically for using the technology of the big data with the data's potential value to be mined [2].

The framework of Hadoop is implemented based on cloud computing meets the big data's requirements which can also solve various problems in relation with big data. Hence this framework can be utilized by many institutions and organizations. But the integration of big data is a challenging process which is not easier to handle. As data of different varieties are involved with

May - June 2020
ISSN: 0193-4120 Page No. 5194-5198

data sources being heterogeneous and distributed widely which is due to updating operation of data frequently with higher demands in real time with the standards that are lacking. This problem can be solved by the proposed work in this paper which avoids the information isolation problem thereby improving the sharing of data. Based on this method of management of data, full advantage is taken for the resource of data that has huge amount of information in the medical industries [3]. Therefore, an effective method is proposed in this paper for integrating the big data in the medical field with utmost security which is managed in such a way that the person who proceeds with the further treatment gets the correct data about the patient. Data analysis is done by two different methods namely Hive and Map reduce. Performance of the proposed technique is improved greatly which are shown by uploading of data, querying of data and analysis of data as the experimental results.

2. Related Work

The system of clinical data management and integration consists of two modules primarily namely integration of data and management of data.

A. Integration of Data

Integration of data is the first module in the architecture of the integration of clinical data which is based on the platform of Hadoop. This Hadoop platform consists of three major parts [4]. They are the system of clinical information, processing engine based on the rule message and the platform of Hadoop as shown in fig.1. Firstly, the platform of Hadoop is framed by the cluster of Hadoop, the cluster of HBase and the Zookeeper. In this system Map Reduce and HDFS provides the function of distributed basic storing process and the processing of the function.

The distributed applications are coordinated using the Zookeeper which can also manage the configurations related to it. One of the databases which is non-relational is used namely HBase which is oriented column wise with the database NoSQL. This database HBase has various advantages than the databases that are used traditionally for the management of the data relationally in storage of data and operation. Hence adoption of HBase is done in the model of data storage as it will be able to meet the demands the current era such as huge storage, higher concurrency and higher scalability of the system [5].

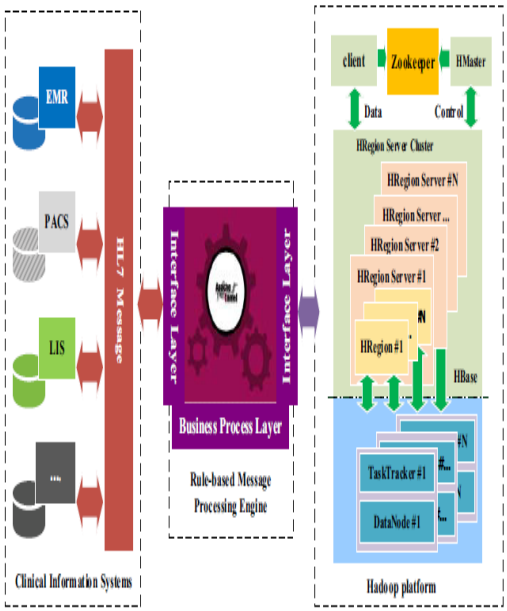

Figure 1: Framework of System for Integrating Clinical Data

Second module in the system is the processing engine based on the rule message which includes the interfacing layer and the processing layer for the business which is devised on the framework of camel integration of Apache.

The library that is useful for processing the events helping the messages from various sources to be processed [6]. The pattern of the work between the consumers that are competing for achieving the highly effective ability for the process of processing that are parallel connection which avoids the failure of single point as the interfacing layer is designed for the Camel does the interaction with the system of clinical information and the platform of Hadoop. Selection of the corresponding component is done for connecting with the applications of the external environment for delivering of the message. The best practice is primarily used by the layer of business process which emerges from the patterns of the integration of enterprise for devising the rules of routing and steps of processing.

Standardization of system adoption is done for the data to produce the system for the clinical information such as PACS, EMR and LIS. The system of HL7 is a standard that is accepted internationally for the exchange of medical information which can be utilized for the transfer of specific electronic data where the information can be administrative data, financial data and clinical data. The observation is described for treating the patients using the model which of basic class of HL7 as it gives the entire information and description about the transactions of the medical data [7-9]. Hence the HL7 message is utilized of 2.4 versions. Standardization of the

May - June 2020
ISSN: 0193-4120 Page No. 5194-5198

ADT class is done by the basic information about the patient with the information about the visit.

B. Management of Data

The characteristics of the big data are significant such as the velocity, volume, veracity and variety [10]. The platform that is used traditionally has its own limits hence cannot meet the challenges of the current era. But the integration of big data leads the traditional platform to be extended to the platform of Hadoop such that it can enable the heterogeneous data of the environment to run for providing the efficient data for managing the reliable data services [11]. The architecture for managing the data is as shown in Fig.2 which consists of five modules.

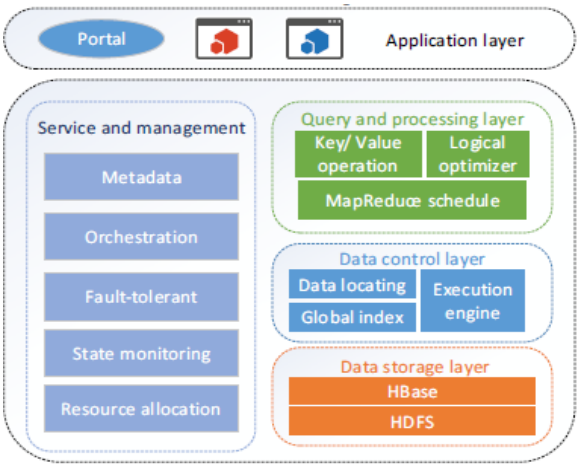

Figure 2: Framework for the Management of Clinical Data

i) Application layer

This layer is responsible for the reception of request which is sent by the users which in turn sent to the modules about the task for querying and handling of the data to be processed. Ultimately the results are presented to the users through the portal used for managing of the data [12].

ii) Layer for Querying and Processing

Based on the type of information in the query and distribution of data this layer optimizes parsing the request such that data is generated based on the scheduling and generation of the executive Map Reduce.

iii) Layer for Controlling Data

Information from the metadata is used with the global index utilized for locating the data for delivering the schedule of the Map Reduce for executing the engine in order to perform the necessary query of the system [13].

iv) Layer for Storing of Data

This layer is responsible for storing the actual data with the index design of the data and for managing the buffer with the file of the journal [14]. Also adoption of the structure for the master slave process is done such that storage of data in each of the node achieves improved efficiency of the system.

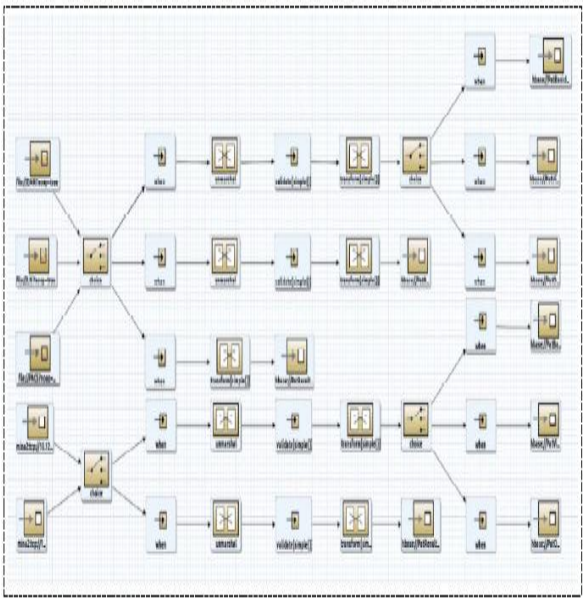

Figure 3: Flow Diagram of the Engine with the Message Processing

v) Management of Data and Service

This module is responsible for the management of metadata with the service of orchestration, management of fault tolerance, monitoring of the state and service of allocation of service [15]. Hence consistency can be ensured between the schema and the metadata along with the management with coordination of the system.

3. Implementation of the Proposed Method for Integration of Data and Management

The system is totally built on the platform of Hadoop which uses the engine of processing the message designed by the framework of the Apache Camel for implementing the integration of data. A group is created of workflows for integrating the LIS, EMR and PACS as a single system. The model is simplified with the workflows with the transactions of the medical field of the hospital. Implementation of the integration of the data is done as shown in Fig.3.

The processing of message with the interfacing layer is created firstly for the endpoint to be connected with the source of external data. Parsing of the business layer is done secondly which validates the messages of the HL7 which is delivered to the interfacing layer. Then the message is sent to the endpoint which is the target of the system which is pre-defined through the workflow containing the rules of routing of the system with the transformation of the message to be transferred in the system with the data source is considered externally with the rules of the routing.

Hence the system of clinical information interacts with the processing of message engine through the message of HL7 which is then standardized as the delivery of the data is done for the HBase for the data storage. Based on the sharing of data, management of the

May - June 2020
ISSN: 0193-4120 Page No. 5194-5198

information is done for the system of clinical data management. The data can also be retrieved and the medical records can be queried about the history of the patient as shown in Fig.4.

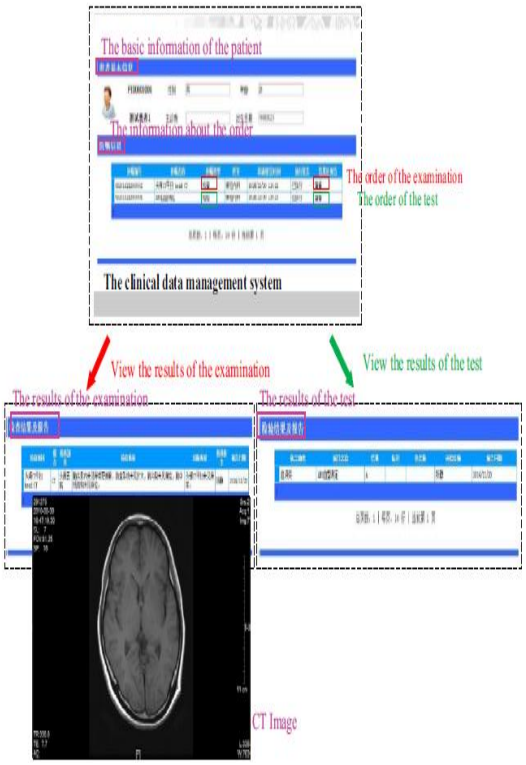

Figure 4: System of Querying the Medical Record of the Patient

4. Results and Discussion

The proposed technique is validated by analysis of data management of massive health care based on Hadoop platform which is designed and implemented for various testing scenarios. These scenarios include uploading of data, taking data statistics and querying a data. Operation of data uploading and querying of data is done by Hbase whereas statistics of data is implemented two times on Hive and Map Reduce respectively. Simulation is done using the platform of health care management. This platform consist of several patients health indexes such as heart rate, blood pressure, body temperature etc. Storage of data is done in Hbase. Simulation is performed on large number of patients such that from every patient collection of health indexes is done atleast five times in a day.

Finally 500 million data records are obtained from several patients which involves 950Giga bytes of data that has to be recorded hence 8 server that are identical to each other are involved with CPU with 8 cores with higher hard disk for extra storage. App server is used for serving all the servers involved in the implementation process. Network topology of the proposed system is as shown in figure.5.

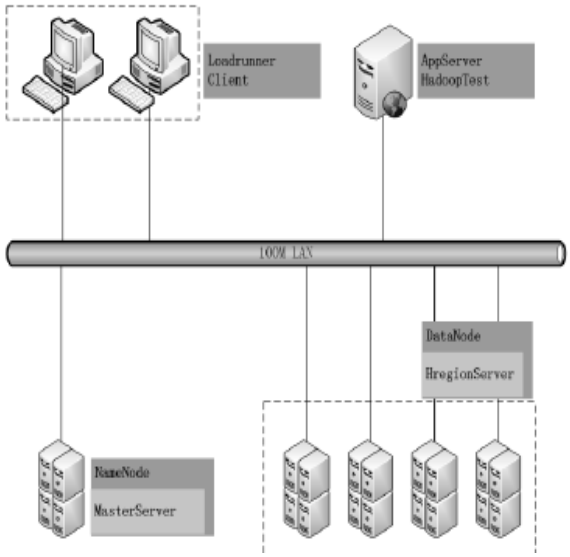

Figure 5: Network Topology of Proposed System

The network topology of the proposed system is designed such that data collection from all the servers is done in a smart way which is then provided to the Hadoop platform. Even though the amount of data is high that has to be managed, efficient method is implemented such that delay is considerably less than the conventional methods. Data from 90 patients with their medical history is uploaded into the system which is stored in the data base. Later query was raised on blood sugar level of a particular patient on a particular day, then in randomly chosen five days where it the level has exceeded normal level. Finally, query was raised on list of patients with female gender in a particular region with diabetes where system gave exact outputs which improves the data management of the patients in an efficient way.

5. Conclusion

This paper has focused primarily on meeting the demands of data integration and management of big data of medical field. Hence this paper designs the method and implements the integration of the clinical data with the system management of the data is done based on the platform of Hadoop. Firstly, the system is built for processing the message engine which connects the system of clinical information and the platform of Hadoop. Then the medical data about the patient is delivered in a standardized way to each of the user. The goal is achieved by this framework of data sharing and management of the clinical data which inquiries about the medical record of the patient. This system possesses features that are significant with the capability of parallel processing which has higher performance, huge storage of data, sharing of medical information in a standardized method, faster retrieval of data with querying in a reliable

May - June 2020
ISSN: 0193-4120 Page No. 5194-5198

service of data management. Finally, it also has provided solution for some of the problems for the demerits that are exist in the integration of the big data thereby providing an effective solution for managing of the big data which can efficiently integrate the big data of the medical field of the system.

References

[1] X. L. Dong, D. Srivastava, Big data integration, Data Engineering (ICDE), 2013 IEEE 29th International Conference on. IEEE, April 2013, pp.1245-1248.

[2] I. Merelli, H. Pérez-Sánchez, S. Gesing, D. D'Agostino, "Managing, analysing, and integrating big data in medical bioinformatics: open problems and future perspectives," J. BioMed research international, 2014.

[3] A. Ma'ayan, A. D. Rouillard, N. R. Clark, Z. Wang, Q. Duan, Y. Kou, "Lean Big Data integration in systems biology and systems pharmacology," J. Trends in pharmacological sciences, vol.35, pp. 450-460, 2014

[4] V. Madisetti, A. Bahga, "Cloud-Based Information Integration & Informatics Framework for Healthcare Applications," J. Computer, 2013.

[5] R. C. Taylor, "An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics," J. BMC bioinformatics, 2010.

[6] C. Ibsen, J. Anstey, Camel in action, Manning Publications Co., 2010.

[7] B. Blazona, M. Koncar, "HL7 and DICOM based integration of radiology departments with healthcare enterprise information systems," J. International journal of medical informatics, pp.425-432, 2007.

[8] Y. Genovese and S. Prentice, "Pattern-based strategy: getting value from big data," Gartner Special Report G00214032, 2011.

[9] A. Bahga, V. K. Madisetti, "A cloud-based approach for interoperable electronic health records (EHRs)," J. Biomedical and Health Informatics, vol. 17, pp.894-906, 2013

[10] Wei Liu, and Dedao Gu, "Research on construction of smart medical system based on the social security card", 2011 International Conferenceon Electronic and Mechanical Engineering and Information Technology(EMEIT), pp. 4697–4700, 2011.

[11] Xiaolin Lu, "Design and implementation of cooperative distributed dental medical information system", Proceedings of the Ninth International Conference on Computer Supported Cooperative Work in Design, vol 2, pp. 799–803, 2005.

[12] Boqiang Liu, Xiaomei Li, Zhongguo Liu, et al. "Design and implementation of information exchange between HIS and PACS based on HL7 standard", International Conference on Information Technology and Applications in Biomedicine, pp.552–555, 2008.

[13] Yuwen Shulim, Yang Xiaoping, Li Huiling, "Research on the EMR Storage Model", International Forum on Computer Science-Technology and Applications, pp.222–226, 2009.

[14] Faro, A., Giordano, D., Kavasidis, I., Spampinato, C., "A web 2.0 telemedicine system integrating TV-centric services and Personal Health Records", 10th IEEE International Conference on Information Technology and Applications in Biomedicine (ITAB), pp.1–4, 2010.

[15] Hudson, D.L. and Cohen, M.E., "Technologies for patient-centered healthcare", Life Science Systems and Applications Workshop, pp.148–151, 2007.