Research on intelligent medical big data system based on Hadoop and blockchain

Sudhir Allam, *Sr. Data Scientist, Department of Information Technology, USA*

Abstract— *This paper examines the concept of Hadoop and blockchain in developing intelligent medical systems. Health knowledge systems are growing in popularity across the world. As digital technology develops rapidly, many local medical organizations are expanding the introduction of information-based, comprehensive systems. Such programs also increased the quality of operation and central efficiency of the hospital [1]. IT not only increases the quality of the clinical practice, but it also encourages physicians to have even more time to treat patients, to increase patient comfort and confidence, and to develop an invisible empirical glimpse of medical institutions [1]. This is a really interesting moment for healthcare and IT (IT). Thanks to advances in genetic testing and the development of precise medication, healthcare has a groundbreaking strategy for preventive care that integrates the genetic structure, lifestyle, and environmental conditions of the particular patient. At the same time, IT progress has created vast health information databases, provided resources for tracking health data, and involved people in their health. Combining these advances in healthcare will encourage transformative changes in health IT. Blockchain is a successful game changer for advancement in medical data storage because it is a collaborative reporting technology with multi-party maintenance and data backup security [1,2]. This paper creates the medical records center on the large data network Hadoop and securely stores and analyzes the initial dispersed data via the sophisticated processing module and the autonomous data acquisition framework. The customized hospital information system for chronic conditions like stroke utilizes the benefits of the Hadoop Big Data framework to provide patients with personalized health care systems and to promote patient management by medical professionals.*

Keywords: *Blockchain; Hadoop; Data Sharing; MapReduce; HDFS; Electronic Health Records*

I. INTRODUCTION

Software engineers and data analysts work with and analyze large amounts of data to quickly uncover valuable insights and create better technologies for clinical systems and patient care. In healthcare and clinical science, secondary usage of medical large data is increasingly widespread. Currently, the U.S health care system focuses more on recovery, focussing on emergency plans and ways of controlling personal care. Thus the, as the scheme of services is designed to reflect on people's quality of life to lifestyles and satisfy individuals' various medical requirements, modifying the distribution of medical resources seems to "only treat," emphasizing medical growth and people’s wellbeing [2]. People frequently need a welcoming, person-centered health care system. A significant part of this process is how to utilize knowledge recognition approaches without warning, not only focusing on early warning and early response on diagnoses but also strengthening preventive measures in high-risk populations. Comprehending the rationale underlying medical big data highlights trends in information technology in hospitals is essential in the development and expansion of healthcare information systems. Hadoop and blockchain receive publicity due to their position in clinical decision-making [2]. Electronics, computerization, and hospital intelligence will increase the performance of the medical system efficiently. Massive digital medical data help intelligent auxiliary diagnostic algorithms and also face new problems for dynamic, multi-source data collection and access. Patient information systems are an international trend. With the accelerated advancement in information technology, an increasing number of domestic primary healthcare centers are rushing to introduce information-driven technologies. Such programs also increased the quality of care and strategic productivity of the hospital. Blockchain is a technology to solve today's interoperability problems in health IT systems and to be the technological norm for the safe sharing of electronic health data among people, health providers, health services, and medical researchers [3]. This paper examines how Hadoop and blockchain work in the medical big data analysis environment. It examines how the medical sector can intelligently process medical data and reveal certain characteristics of consumer behavior in the hospital information system.

II. RESEARCH PROBLEM

The main problem that this explorative study aims at solving is to understand how Hadoop and blockchain are significant in developing smart systems to support the medical field. The proportion of care facilities for chronic illnesses is increasing as the population ages and ailments evolve, putting human life and wellbeing in jeopardy. This increases public acceptance and value in individual health management. Clearly defined chronic illnesses became the relevant guidance agencies. Data prevention and monitoring have been integrated into information technology for health [4, 5]. The utilization of modern information technologies and patient-centered care tools, successful application of the information and resource exchange process, innovative chronic disease management systems, and the improvement of chronic disease surveillance and information management have been a gateway to the assessment and tracking of chronic diseases. The World Health Organization study showed that 5 percent of lifestyles impact health. In this sense, self-help is extremely useful in terms of health management, since it can raise health consciousness, encourage behavioral improvements, and increase people's health [5]. However, two problems remain in managing confidential health records in real-time, successfully preventing and controlling chronic diseases. On the one side, there are no clear signs of mobilizing patient involvement. However, for actual prevention, physicians need to invest a lot of time and resources.

III. LITERATURE REVIEW

A. How Hadoop works in a medical system

The use of all technologies must be multidimensional and systematic since the public health field requires healthy output, the provision of medical supplies, and advanced logistics. The scientific auditing of RFID technologies would cover the vaccines, blood supplies, particularly anti-counterfeiting foodstuffs, and pharmacovigilance. Hadoop has been built by the Apache Foundation for a distributed framework architecture. Without the fundamental specifics of the system being understood, users may build distributed programs to make good use of the vast data storage resources and fast computational power. The distributed file system known as HDFS is implemented by Hadoop [6]. HDFS has positive characteristics, strong tolerance for failures, and the need to employ costly hardware like other frameworks. Furthermore, it offers a high-speed platform for systems and very broad data sets for data entry. HDFS stabilizes file system specifications to view file system data in stream type [7]. HDFS and Map Reduce are the central architecture of the Hadoop framework. HDFS primarily offers large data processing, and Map Reduce offers distributed data computing capabilities. Hadoop data analysis can be interpreted in a short statement: the data are analyzed to meet customer requirements by the Hadoop cluster.

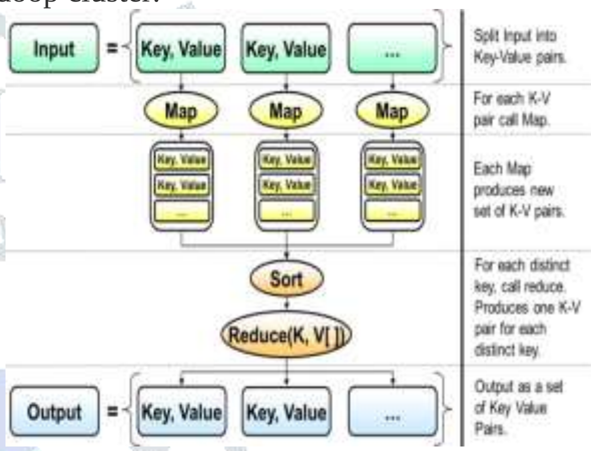

Fig i: How Hadoop processes information in a medical system

The HDFS platform is mass storage data and MapReduce's function is to calculate huge data. Additionally, two essential elements of Hadoop are the data management storage tool Hive and the distributed database framework Hbase. HDFS stands for Hadoop Distributed File System. It is used to store data in a Hadoop cluster. The HDFS area looks like a basic file system, with easy functions such as development, deletion, movement, and more [6]. Nevertheless, HDFS files are broken down into data blocks based on some requirements, and several and several data blocks are located in several slave nodes. It differs greatly from standard storage structures [7]. Typically, the size and amount of data blocks are determined by the individual. The upper layer of DFS is MapReduce, consisting of JobTrackers [7,8]. The option for HDFS is to split big files into equal bits. The reduction of the map is a job for each Hadoop input component and the calculation of the map. In this job, the machine processes the entry records one by one, and after processing, the chart becomes the main value [9]. This results from the shape of the main value pair. The product of the previous stage is then passed to Hadoop as a main Reduced input. The Reduce Task output is the complete job output and it is HDFS saved.

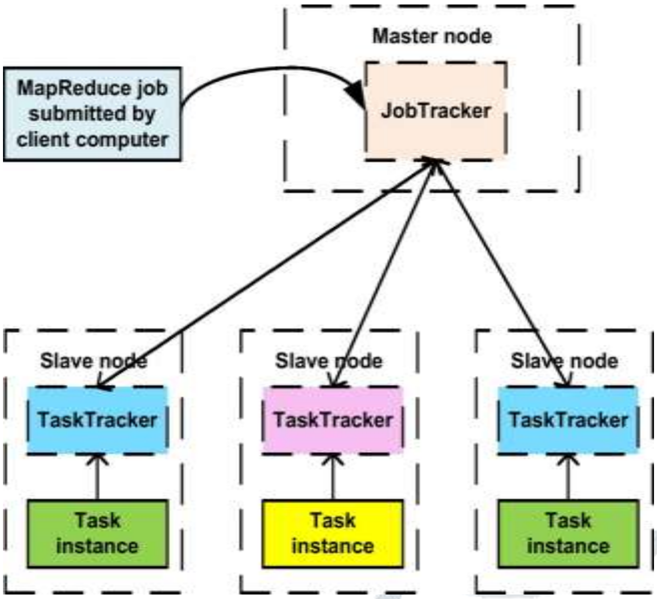

Fig ii: MapReduce in Hadoop

The NameNode is an HDFS daemon in Hadoop which normally works on an individual system. The key responsibility is to document how the processed or computed files are split into the required data blocks for rapid processing and to record the position of the nodes to which the data blocks are split [10]. The NameNode determines whether the file is to be mapped into the DataNode copy block [10]. Although NameNode mostly supports the management of memory and I/O units, the actual I/O processing does not include NameNode. The NameNode is only linked to the metadata of the node position in which the database is located. This can be avoided. Unnecessary data limits the server processing speed. However, the NameNode itself is one Hadoop cluster node, and the whole machine is in trouble, as long as the NameNode operation is in difficulty. The number of DataNodes is very high in comparison to the singularity of the NameNode. On each slave server in the cluster, the DataNode software is operating [11]. The data block is separated and assigned to the local machine, i.e. the domain used to hold the data block, by reading and writing. The NameNode determines the corresponding position of the DataNode [11] storage when the consumer wants to access the block. To process necessary data blocks inline, the client directly interacts with the daemon in the DataNode. The DataNode is a rack and each rack uses a switch, called NameNode, to bind all data for the user.

B. Hadoop based medical system

Hadoop is a distributed computing platform created by the Apache Foundation that can handle large volumes of data [12]. It is necessary to fulfill a user's need to create a distributed application without completely comprehending the concept of "distributed." It will split big pieces of programs, using the cluster's capacity to render high-speed storage or measurements, in limited working units. The example of Hadoop's implementation is a structured and digital mechanism in which a central decision-making server is used to communicate with the doctor, the pharmaceutical stores, and the medical manufacturing company. This database allows one to analyze various cases such as regional wealth sales, regional wealth measurement, etc [12]. This is a useful tool. Thus, with DSS focused on distributed environments our concept is to resolve issues. The mechanism suggested comprises four systems: a doctor node, a pharmacy enterprise node, a medical store node, and a centralized judgment node. One, a centralized decision node, would be the master and the other three are slaves of the master (name node & data node) (Only Data nodes). The SQL patient knowledge table from any area will be used in the doctor node. The table includes details such as Doctor ID, patients' data, diagnosis information, and prescription medications, as well as the year. The Pharmaceutical Node is supported by details on medicines manufactured by various pharmaceutical firms, the number of such medicines manufactured and the price set by the pharmaceutical corporation, which would also be a SQL chart. [12] Similarly, a medical node stores medicines stock record. The centralized node will recover the details for the analytical component from these SQL tables and will store it at the centralized nodes in the Hbase table. Here, we're using Apache Phoenix at the top of Hbase, Hbase is NoSQL, we can get the data from the Hbase table using SQL queries using apache phoenix, and Apache phoenix has faster results than Hbase for regular transactions [13,14]. There will be several MapReduce systems on various columns in the centralized base table that will get different analytical outcomes. The below are the results: A. Proportion to the use by a single practitioner of the Generic drug. b. Regional criteria for wise medication. c. Disease analysis over a specified period in a given area. For instance, if we want to know how many patients were diagnosed for swine flu on 1 January then, for two columns, i.e. swine flu names and dates i.e. 1 January, we can then submit the MapReduce software to the

HBase table and fetch those records that meet the requirements and store them in Hadoop based medical system a new table [14].

The concept of using blockchain technologies in healthcare applications is very productive. The key advantages are that health records are saved on blockchains. It may be analyzed but held private and the data contribution and utilization are compensated for by the layer. Blockchains are integrated into a structure of centralization (high efficiency, low cost), and achieving full detail and transparency [14]. Programmable smart contract (no burdened evolutionary model) and the open world's anonymity (protection of Privacy) are some of the capabilities of blockchain. Due to these features, blockchain technology has very important key application advantages in the area of medical information. De-centralized distributed infrastructure will also save many intermediary costs. Time characteristics that do not tamper will address data monitoring and information protection issues [14]. The security of the new health information management authentication can be resolved by a secure confidence mechanism. Flexible special programmable aids clinics to create expanded applications.

C. *Medical health blockchain*

i.Practical application

The implementation of a structured and fully open infrastructure for medical records is the main feature of blockchain systems [14]. It ensures that requirements for enforcement are quickly traced. The use and storage of distributed accounts Improve surgical facilities' abilities to tolerate and fix errors. Providers in pharmacy and medical services are capable of building a sustainable framework without the expense of converting processes and the checking of prescriptions.

ii. Personal health record storage on the blockchain

Blockchains can be used for the storage and management of personal health information. For example, the vast medical electronic record system (EMR). Cana uses technologies for block chaining and privacy (private key access only). The benefit of digital encryption not name coding. Digital properties may be coded as personal health documents and digital currencies in the blockchain. Individually, hospitals, clinics, insurance providers, and other individuals may have the requisite health information through private keys [15]. The E-Record service (EMR) may also encourage general format in the blockchain. This dilemma should be solved, despite having switched to the electronic medical record method so such major suppliers of services [15]. They still differ widely and cannot be exchanged or interoperable. The blockchain will provide EMR across the nation with a common repository and exchange format. Personal data loss of the patients, medical data leakage, inconsistency of medical records. The trouble with the cross-regional referral of hospitals and the division of clinical data are the key problems in this area. The principal cause of such problems is the lack of standardization and the rigorous, full, and coordinated EMR [16]]. This results in productivity in hospitals and the whole care sector with medical and administrative staff. The blockchain records book content, ensuring the authenticity of all data by a consistent algorithm. For instance, whether the Blood Type is one of the medical documents. But blood type 0 from other medical facilities is registered in the "data block." The details would not be registered in the blockchain, and the machine would not be able to balance the information [17]. This protects the records on the patient's medical report. Therefore, any time patient visits a new medical facility it is not essential to record medical history.

IV. THE FUTURE IN THE UNITED STATES

The U.S. healthcare information sector has seen tremendous strides in recent years. Google is collaborating with the US Medical Center to develop an integrated medical database for hundreds of millions of people that physicians can monitor remotely. Hadoop and Blockchain technologies have the potential to transform health knowledge sharing by allowing for greater data privacy, better patient treatment, more efficient hospitals, and more rigorous medical science. Despite the positive aspects, many underlying problems must be addressed before widespread deployment can be carried out safely and successfully. Healthcare organizations must evaluate blockchain in the light of their requirements, much as they must for other new technologies, and provide clinicians with the expertise to efficiently leverage these resources [18]. Additionally, Microsoft introduced a new medical information service portal to assist physicians, staff, and relatives in determining a patient's current real-time condition. Intel has introduced a digital health care portal a few years back that aids doctors and patients in establishing contact through information technology. This is an environment in which IBM has done a lot of studies.

V. ECONOMIC BENEFITS TO THE UNITED STATES

Health operations can be optimized with Hadoop and blockchain medical technologies, resulting in lower costs, manpower, and failures compared to the new scheme. The United States has continued to bring services into modern care facilities through the application of information technology. The key indicator is that many medical facilities have increasingly started to establish and implement large-scale interconnected medical services on a local and even national scale. Since it stores data in a cluster of commodity hardware, Hadoop is a cost-effective approach. Since commodity hardware is inexpensive, the expense of connecting nodes to the framework is not prohibitively large. Since the amount of redundant data has reduced, fewer machines are needed to store the data.

VI. CONCLUSION

The significance of Hadoop and blockchain in the development of smart systems for medical diagnostics was investigated in this study. Hadoop-based analytics for the medical sector would enable doctors to make decisions relevant to disease diagnosis at the regional level as soon as possible, allowing the regulatory body to take appropriate and decisive steps and save lives. Health stores and pharmacy firms would benefit from this framework because it would enable them to maintain track of their inventory. Blockchain technology solves interoperability issues, is built on open principles, which offers a decentralized distributed view of health records. It is expected to gain mainstream adoption and implementation in all industries. The benefits of blockchain as outlined in this paper can bring together millions of people, health care professionals, health care agencies, and medical practitioners to exchange large quantities of genomic, diet, lifestyle, environmental, and medical information while maintaining assured security and privacy. The collection, storing, and dissemination of this information would provide a practical basis for the progress of medical science and precision medicine, as well as aid in the identification and development of innovative approaches to cure and avoid disease. It would also be used to see how mobile technologies would help people become more involved in their medical services, which will lead to better wellbeing and preventive medicine. Hadoop and Blockchain technology is certainly placed in the healthcare IT environment, and the ONC could seriously suggest using blockchain for the development of accurate medicine to base their integration plan.

References

1. Y. Hau and M. Chang, "A Quantitative and Qualitative Review on the Main Research Streams Regarding Blockchain Technology in Healthcare", *Healthcare*, vol. 9, no. 3, p. 247, 2021.
2. A. Hussein, "Using Hadoop Technology to Overcome Big Data Problems by Choosing Proposed Cost-efficient Scheduler Algorithm for Heterogeneous Hadoop System (BD3)", *Journal of Scientific Research and Reports*, pp. 58-84, 2020.
3. Q. Jia, "Research on the medical system based on blockchain technology", *Medicine*, vol. 100, no. 16, p. e25625, 2021.
4. K. Surendran, "Student Academic Management System Using Blockchain Technology", *Journal of Advanced Research in Dynamical and Control Systems*, vol. 12, no. 3, pp. 1410-1415, 2020.
5. S. Khanra, A. Dhir, A. Islam, and M. Mäntymäki, "Big data analytics in healthcare: a systematic literature review", *Enterprise Information Systems*, vol. 14, no. 7, pp. 878-912, 2020.
6. F. Ouatik, M. Erritali and M. Jourhmane, "Student orientation using machine learning under MapReduce with Hadoop", *Journal of Ubiquitous Systems & Pervasive Networks*, vol. 13, no. 1, pp. 21-26, 2020.
7. H. Abu-Alsaad, "Retailing Analysis Using Hadoop and Apache Hive", *International Journal of simulation: systems, science & technology*, 2020.
8. U. Dr. K and D. Bhavanam, "Usage of HIVE Tool in Hadoop ECO System with Loading Data and User-Defined Functions", *International Journal of Psychosocial Rehabilitation*, vol. 24, no. 04, pp. 1058-1062, 2020.
9. X. Zhang and Y. Wang, "Research on intelligent medical big data system based on Hadoop and blockchain", *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, 2021.
10. M. Deng, "Robust human gesture recognition by leveraging multi-scale feature fusion", *Signal Processing: Image Communication*, vol. 83, p. 115768, 2020.
11. J. Eijsink, A. Fabian, J. Vervoort, C. Boersma and M. Postma, "PMU102 Western-World Hospitalization Implementation Examples of VALUE-Based Healthcare in Expensive

Intervention Settings: A Systematic Review", *Value in Health*, vol. 23, p. S620, 2020.

12. Q. Zeng, Q. Wang, L. Zhang, and X. Xu, "Comparison of the Measurement of Long-Term Care Costs between China and Other Countries: A Systematic Review of the Last Decade", *Healthcare*, vol. 8, no. 2, p. 117, 2020.
13. H. Liang, J. Zou, K. Zuo and M. Khan, "An improved genetic algorithm optimization fuzzy controller applied to the wellhead backpressure control system", *Mechanical Systems and Signal Processing*, vol. 142, p. 106708, 2020.
14. P. Kakarlapudi and Q. Mahmoud, "A Systematic Review of Blockchain for Consent Management", *Healthcare*, vol. 9, no. 2, p. 137, 2021.
15. M. Hoffman, L. Ibáñez and E. Simperl, "Toward a Formal Scholarly Understanding of Blockchain-Mediated Decentralization: A Systematic Review and a Framework", *Frontiers in Blockchain*, vol. 3, 2020.
16. E. Chukwu and L. Garg, "A Systematic Review of Blockchain in Healthcare: Frameworks, Prototypes, and Implementations", *IEEE Access*, vol. 8, pp. 21196-21214, 2020.
17. [C. Agbo, Q. Mahmoud and J. Eklund, "Blockchain Technology in Healthcare: A Systematic Review", *Healthcare*, vol. 7, no. 2, p. 56, 2019.
18. X. Yaya and Z. Bi-Geng, "Research on Medical Image Storage and Retrieval System Based on Hadoop", *Journal of Physics: Conference Series*, vol. 1544, p. 012119, 2020.