Research on the Secure Medical Big Data Ecosystem Based on Hadoop

Xiangfeng Zhang  
Yellow River Conservancy technical institute

Yanmei Wang  (✉ wangyanmei@vip.henu.edu.cn )  
henan university

Research

Keywords: secure medical, big data, ecosystem, health data, Hadoop

DOI: https://doi.org/10.21203/rs.3.rs-45398/v1

License: ☑️ ⬅️ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

This paper designs and implements a secure medical big data ecosystem on top of the Hadoop big data platform. It is designed against the background of the increasingly serious trend of the current security medical big data ecosystem. Since traditional healing activities take place in medical institutions, patient users must travel to these institutions to learn about their treatment status. The personalized health information system designed for this purpose enables patient users to understand their treatment and rehabilitation status anytime and anywhere. The above is a consideration from the perspective of the patient user, from the perspective of personal health data. Because traditional medical health data is scattered throughout different independent medical institutions, and these databases are heterogeneous. As a distributed accounting technology for multi-party maintenance and backup information security, blockchain is a good breakthrough point for innovation in medical data sharing. The characteristics of blockchain without a central server make the system without a single point In case of failure, the stability of the system is well maintained. In this paper, the system realizes the personal health data centre on the Hadoop big data platform, and the original distributed data is stored and analysed centrally through the data synchronization module and the independent data acquisition system. Utilizing the advantages of the Hadoop big data platform, the personalized health information system for stroke has designed to provide personalized health management services for patients and facilitate the management of patients by medical staff.

1. Introduction

The development of medical information services is an international trend. With the rapid development of information technology, more and more domestic medical institutions are accelerating the implementation of information-based platforms based on overall construction. Such systems have improved the hospital's service level and core competitiveness. Information technology not only improves the efficiency of doctors' work, but also it allows doctors to have more time to see patients, improves patient satisfaction and trust, and establishes an invisible scientific image of medical institutions [1-3]. In the past few years, the US healthcare information industry has made great strides. Google works with the US Medical Centre to create an electronic medical record for hundreds of millions of communities that doctors can remotely monitor. Microsoft has also launched a new medical information service platform to help doctors, patients and families understand the latest real-time status of patients. A few years ago, Intel also launched a digital healthcare platform that helps doctors and patients build interaction through information technology. IBM has a lot of work in this area. In China, the “Twelfth Five-Year Plan” of the Ministry of Health clearly stated that health informationization is an important task for deepening medical reform. The Ministry of Health of China has initially established a health information technology roadmap, referred to as the “3521 Project”, which is to build a national, provincial and municipal health information platform, and strengthen public health, medical services, new rural cooperative medical insurance, and basic drug systems. With the aging of the population and the changes in diseases, the proportion of medical services for chronic diseases is rising, which seriously jeopardizes human life and
health. This makes personal health management more and more accepted and valued by people. Relevant guidance departments have clearly identified chronic diseases. Prevention and control information has incorporated into health information technology [4, 5]. How to use advanced information technology, patient-centred medical resources, achieve effective integration of information interconnection and resource sharing, innovate chronic disease health management system, improve chronic disease monitoring and information management, has become a key to chronic disease management and control in China.

At present, China's health care system has more focused on treatment, with a focus on medical plans and means to control personal life. Therefore, if there is a service system that can focus on the quality of life and lifestyle of individuals and meet the diverse medical needs of each individual, changing the allocation of medical resources tends to “just cure”, attach importance to the development of medical care, and despise the health of the population. Often, people need a warm, personal-centred health system. A very important aspect of this process is how to use information identification methods without notice, when early detection and early intervention, not only focus on those who have diagnosed, but also strengthen prevention efforts for high-risk groups. At the right time, we can have a much-targeted measure of intervention. In fact, health management has a trilogy: health management is the foundation and health assessment, and health intervention is the key. The World Health Organization survey found that lifestyles affect health by 5 percent. In this context, through health management, self-management is very valuable, can arouse everyone's health awareness, lifestyle changes, and improve people's health [6, 7]. However, how to manage the real-time control of personal health information, effectively prevent, and control chronic diseases still faces two challenges. On the one hand, how to mobilize participation with the patient has no obvious symptoms. On the other hand, for real prevention, doctors need to spend a lot of time and energy [8, 9]. How can doctors have so much time and how to have appropriate mechanisms to encourage doctors? These two problems have been plaguing the universality of the medical field for a long time.

This paper is a safe medical big data ecosystem for the establishment of personalized health information and services on the Hadoop platform to consider the above aspects of stroke, optimize the local medical resources structure, improve people's health care awareness, and promote the health of China's major social impacts [10, 11]. The system utilizes the massive data storage of the Hadoop big data platform, and the convenient scalability greatly reduces the difficulty of storage and system upgrade. It is based on the above considerations to establish a personalized security-oriented big data ecosystem on the Hadoop platform. The health information system has a significant role and social influence on the prevention and treatment of chronic diseases, optimizing the structure of regional medical resources, raising people's awareness of health care, and promoting the level of medical informationization in China. The system utilizes the massive data storage and scalability of the Hadoop big data platform, which greatly reduces the storage and upgrade maintenance costs of the system. With the current form of its popular mobile applications, patient users and doctors can stay in touch anytime, anywhere, which reduces the communication barriers that are inconvenient for medical communication and increases the level of medical digitalization. As the scale of personal health data continues to
expand, the computing power of the Hadoop cluster can solve the problem of large-scale concurrent access during peak traffic hours, ensuring the robustness of the system pair.

## 2. Methods

### 2.1 Construction needs of safety medical network

As the public health sector includes safe production, as well as the management of medical supply forests and modern logistics, the application of any technology is not separate; they must be holistic and systematic. Vaccines, blood products, including food anti-counterfeiting, and drug safety monitoring will all have included in the scope of technical monitoring through RFID technology. The application of IoT technology in the medical field, it is mainly in several aspects:

1. The community medical treatment is aimed at “difficult to see a doctor and expensive to see a doctor”. The medical reform has put forward many measures, including two-way referral to community medical care. In community medical care, we hope to form a complete network platform through the Internet of Things technology to achieve the entire region. I hope that 20% of the highly skilled health talent team and experts can be shared by 80% of the population. Specifically, to identify the patient's identity, which is implanted in the human body, and identify the patient's identity through labels and wristbands to carry out timely, accurate, appropriate, fair, inexpensive, and reasonable services to improve the service to the young and old.

2. The safety of medicine, blood, sanitation, consumables, medical equipment production, and distribution is also included in the scope of medical insurance. When a medical device is used up, it must be re-implanted into the body of other patients, which brings great safety risks to patients. I hope to provide good technical support through RFID technology and IoT technology.

3. Prosecution of drugs. It is hoped that the implementation of the technology will enable any drug to have a corresponding identification code, so that patients can get safe and effective drugs regardless of where they buy the drug.

4. The medical environment is very important. It is not only the safety of the patients themselves, but also the safety problems brought about by the social environment and the natural environment. This is also an important application. As a medical analysis of garbage work is also very important, which is also based on the current environmental protection including the proposed application aspects, although not the focus, but also one aspect.

5. The application of RFID to transposition security. The focus is on key aspects that may lead to safety. The management of blood collection and supply institutions, blood diseases caused by blood transfusion, including other forms of blood use, are also shocking. I hope that after donating blood to the public through such technology, I will post it. With the RFID tag, this blood can always know the safety of the blood itself, and can guarantee the safety of blood in real time.
6. The management of medical devices, the re-implantation of used medical devices into the human body is very unethical. It is now a market economy under reform. Other channels of such problems will be managed, but from the health industry, it is hoped that through technical means. Implement safety management of medical devices, including tracking and management of medical device packages and all other medical devices in the future.

7. In the basic functional function specification of the hospital information system issued by the Ministry of Health, the consumables before and after the operation have managed. We hope that the new technology can make the requirements of the Ministry of Health more effective and can remind the doctor or the relevant. The personnel are safe to the patient's medical equipment. The ten gauze that you took, the consumables have labelled, and ten gauze in the patient's abdominal cavity have scanned out after ten. We want to make medical devices through RFID technology [12-15]. Use management is more secure. The overall planning of the Ministry of Health is very important. The advent of the IoT medical era, through the application of all-round IoT technology, enables the integration of software applications, hardware promotion, and informationization of the entire health system. Today's application software is not only a problem of the software itself, but also organically connected with objects and people, so that all kinds of information can be collected in an all-round way, so that the status quo of medical treatment can be more thoroughly perceived. We can make such a goal.

8. According to the country to establish an innovative country, from the Ministry of Health, we also hope to provide an innovative and scientific modern service, which also has a significant relationship with the country's major exhibition modern service industry. We use the Internet of Things technology to achieve the traditional. The innovation of medical model and the realization of traditional medical informatization innovation are not the existing scattered and isolated applications. We hope to integrate and optimize the status quo of medical informatization and rapidly improve innovative technologies based on the existing work basis. Adapt to the needs of health reform, adapt to the needs of patients' medical and health services, and finally realize real-time intelligent and dynamic interoperability services, especially emphasizing intelligence. In the meeting including the Ministry itself, I said that informationization is not a form. We especially need to provide intelligent and automated information services to the masses through means of informationization.

9. Urban and rural medical care, the medical environment of the city is getting better and better. Although it is difficult to see a doctor and expensive to see a doctor, it is still more and more advanced from the medical means. Nevertheless, rural farmers’ friends, we hope that with such technology to make better use of our medical resources, so that our urban and rural services can be integrated, I also hope that everyone will focus on this; on the one hand, we hope to pass our rural medical care. At the same time, we also hope that the city's good medical resources will be transported to the countryside through our Internet of Things technology, so that everyone can enjoy fair medical services. By using the Internet of Things technology to build an electronic medical system, it will bring more convenience to the medical service field. To improve the modernization of medical services, we must not only improve the services of
high-end medical personnel for patients, but also improve the service capabilities of medical personnel through such means.

2.2 Key technologies of big data platform based on Hadoop

Hadoop is a distributed system infrastructure developed by the Apache Foundation. Users can develop distributed programs without having to understand the underlying details of the architecture, so that they can fully utilize the platform’s massive data storage and fast computing capabilities. Hadoop implements a distributed file system, referred to as HDFS [16]. HDFS has good features, high fault tolerance, and the need to use expensive hardware like other platforms. In addition, it provides a high-speed interface for application data access, suitable for applications with very large data sets. HDFS relaxes the requirements of the file system so that it can access the data in the file system in a stream form.

The core design of the Hadoop framework is HDFS and Map Reduce. HDFS mainly provides massive storage of data, and Map Reduce provides distributed computing services for data. The processing of data in Hadoop can be understood in a simple sentence: the data has processed by the Hadoop cluster to get the results. The processing flow is shown in figure 1.

The two core components of Hadoop are HDFS and MapReduce in figure 2. The HDFS function is the storage of massive data, and the role of MapReduce is the calculation of massive data. The data storage warehouse tool Hive and the distributed database system Hbase are also two important components of Hadoop. The full HDFS name is Hadoop Distributed File System, which is used to store files in a Hadoop cluster. The HDFS surface looks like a simple hierarchical file system with simple operations like creating, removing, moving, and more. However, the files stored in HDFS are divided into data blocks according to certain needs, and then many and many data blocks are placed in multiple slave nodes. This is where it is very different from traditional storage architectures. The user usually determines the size of the divided data block and the number of data blocks placed. The upper layer of DFS is MapReduce, which consists of JobTrackers and TaskTrackers. On HDFS, by default, large files have divided into equal parts. In the HDFS introduction document, this default value is set to 64M. In the diagram shown below, the file data1 has first divided into three parts, which are placed in three different machines [17, 18]. Map Reducel is a task for each component of Hadoop input, and then call Map to calculate. In the task, the system will process the input records one by one, and then the map will be key-value after processing. The form of the key-value pair will output the result. Hadoop will then pass the result of the previous step as input to Reduce by key. The output of the Reduce Task is the output of the entire job and is saved on HDFS.

2.3 Literature review

Developed countries such as Europe and the United States have begun to use information technology to integrate resources into existing medical systems. The main sign is that many medical institutions have gradually begun to develop and apply regional and even national large-scale integrated medical systems [16]. At present, the regional medical and health information construction in developed countries has developed to a new stage. Some countries in Europe and the United States have entered an aging society.
in the 1980s. In the face of the common problem of aging, some European countries do not encourage the construction of new old-age care institutions, and the pension models in developed countries are based on home-based care. Support community and family to provide high quality long-term care for the elderly. In order to provide high-quality services that satisfy the elderly, Switzerland, the Netherlands, Denmark, the United Kingdom and other countries have developed incentive programs to promote the cooperative operation of providers of different types of aged care services such as institutional pension, community pension, and home pension. Nowadays, the model of home-based care for the developed countries and related supporting systems has developed relatively mature. Shahriyar Rifat et al (2010) mentioned that mobile healthcare is an integration of mobile computing and medical monitoring. They believe that mobile devices are an integral part of our lives and can be seamless. Integrate healthcare services into our daily lives. Tang et al (2010), in conjunction with medical information systems and mobile communication technologies, established a telemedicine home care management system that enables long-term and sustainable health monitoring by transmitting multimedia information services.

Nanjing Mobile launched a series of “Healthy Patrons” project, which includes a variety of health terminal products with data transmission functions, which are used to measure physiological indicators such as blood pressure, blood sugar, ECG, body temperature, etc., and SIM cards built into each terminal product. The measurement data can have uploaded to the health management platform. The platform can automatically provide reminders and reports to users, and achieve multi-integrated service effects such as prevention, diagnosis, treatment, and rehabilitation. It is convenient for individuals, families, and doctors to dynamically control the health status of the elderly in real time and make corresponding reasonable measures in time [19]. Sichuan Mobile launched the “Platform. Terminal” smart care information platform mainly uses information technology to achieve effective connection between citizens’ pension needs and civil affairs agencies and community service organizations. The dedicated terminal sets the SOS emergency call button, which can locate the old man through LBS and GPRS in the emergency of the old man, and inform the 120 and relatives to help the first time [20]. The Health China 2020 Strategy Research Report released by the Health Planning Commission in 2012 clearly pointed out that in the next few years, the major projects of the seven major medical systems involving a total amount of 400 billion yuan will have launched. Among them, the national e-health system project budget of 6.1 billion yuan, including the implementation of the standardization of large-scale integrated hospital information system, the establishment of a national electronic health record and the creation of a regional medical information platform. This is the first time the government has clearly stated the scale of investment in the direction of medical informatization, indicating that the status of informatization in the reform of the medical system is constantly improving.

3. Experiment

3.1 Hadoop architecture

Hadoop uses a master-slave (Master, Slaver) architecture. In order to get Hadoop running smoothly in a complete cluster, a series of necessary background programs are essential. As shown in Figure 3.
Hadoop is a software framework that can handle massive amounts of data and is developed by the Apache Foundation. It is possible to satisfy a user to develop a distributed program without fully understanding the principle of "distributed". It is able to divide large chunks of programs into small work units, making full use of the ability of the cluster to make high-speed storage or calculations.

The following is a description of each role:

(1). NameNode

The NameNode is a daemon for HDFS in Hadoop and usually runs on a separate machine. It is mainly responsible for recording how the stored or calculated files are divided into data blocks required for fast processing, and the locations of the nodes to which the data blocks are divided are recorded. The NameNode decides whether to map the file to a copy block on the DataNode [21]. Although the main function of the NameNode is to manage the memory and I/O modules, the true I/O processing is independent of the NameNode. Only the metadata of the node location where the data block is stored has related to the NameNode. This is to avoid these. Unnecessary information reduces the processing speed of the server. However, the NameNode itself is a single node of the Hadoop cluster, and as long as there is a problem with the NameNode service, the entire system will be in crisis.

(2). DataNode

Compared to the singularity of the NameNode, the number of DataNodes is very large. The DataNode program runs on every slave server in the cluster. It is responsible for reading and writing the data block divided and allocated to it to the local system, that is, it is the address used to store the data block. When the user needs to use the data block of this block, the NameNode finds the corresponding location of the storage DataNode. The client then communicates directly with the daemon on the DataNode to process the required data blocks accordingly. The DataNode has presented in the form of a rack, and each rack connects all the data for the user through a switch, NameNode.

(3). Secondary NameNode

The Secondary NameNode can be seen according to its name. It is an alternate NameNode. It exists in a similar way to the NameNode, but it does not work for the NameNode. It is only used to periodically save the state of the metadata in order to monitor the working status of the NameNode. When an abnormality occurs in the NameNode, instead of completing the required functions, the error is largely avoided.

(4). JobTracker

Like the NameNode, there is only one JobTracker in each cluster, which acts as a linker and Hadoop. After the user has finished writing the code, submit it to the cluster, and the JobTracker will determine the file to be processed. There are several different tasks in the process. The job of the JobTracker is to assign nodes to these tasks. If there is a problem with the task, the JobTracker monitors it and the task is restarted.
(5). TaskTracker

The relationship between TaskTracker and JobTracker is similar to the composition of the DataNode and NameNode responsible for storing data, and is a master-slave structure. The JobTracker is located at the master node to manage the data analysis work, while the TaskTracker is responsible for managing the respective tasks of each slave node.

3.2 Building a secure medical big data ecosystem based on Hadoop

In this system, mainly to demonstrate the feasibility of the system, the deployed Hadoop cluster adopts a single-layer network topology, which will reduce many unnecessary problems in development. Of course, in the actual production environment, this is obviously not a suitable; most appropriate still requires a typical two-layer network topology. The network structure used by this system is as follows.

The Hadoop cluster of this system uses a master-slave architecture. From the perspective of scalability and performance, in a large-scale Hadoop cluster configuration, different component roles are assigned to different machines to avoid failure of a single machine and failure of the entire cluster. The cluster of this system is composed of 8 host computers, which uses the Ubuntu12.4STD operating system and uses Gigabit LAN to ensure the data transmission. The specific configuration of each component role is shown in table 1.

| CPU name | Program   | Process (Character)             |
|----------|-----------|---------------------------------|
| HIT02    | Hadoop, Hbase | Namenode                        |
|          |           | Tasktracker                     |
|          |           | Datanode                        |
| HIT03    | Hadoop, Hbase | Datanode                        |
| HIT04    | Hadoop, Hbase | Resource Manager                |
| HIT05    | Hadoop, Hbase | Namenode                        |
|          |           | Quorum Peer Main                |
|          |           | Node Manager                    |
|          |           | Datanode                        |
| HIT06    | Hadoop, Hbase | Namenode                        |
|          |           | Tasktracker                     |
|          |           | Quorum Peer Main                |
|          |           | Datanode                        |
The cluster uses a unified installation directory hadoop:/usr/local/hadoop, Hbase: /usr/local/hbase, Zookeeper: /usr/local/zookeeper. Figure 5 shows the overall design of the distribution Hadoop. Among them HIT02 and HIT03 are Name Node (can change Name Node, high availability), Hmaster, HIT04 and HIT05 are Resource Manager, (HA mode, high availability mode requires at least 2 RM), HIT06-HIT10 is Data Node, Zookeeper. In addition to HIT04 and HIT05, all Hbase is installed. The basic operation commands are shown in the table below.

### 3.3 Safety Medical Big Data Ecosystem Construction Process

The security medical big data ecosystem of this paper is improved because of the traditional information retrieval system. It utilizes the concept of knowledge map proposed by Google in 2012 and expands the search keywords. This paper realizes the construction of the stroke knowledge map. The knowledge map of the system is built on the Wikipedia database, which guarantees the quality of the map.

First, we choose the knowledge vocabulary collection related to the security medical big data ecosystem as the most program input. According to this knowledge vocabulary collection, we find the corresponding page collection, and then grab these pages to analyse and judge the link information in the page. Whether the link information is a knowledge term related to the secure medical big data ecosystem, and if so, the knowledge term is added to the term set, and the link relationship between them is saved to the relationship set for the system. Knowledge Graphs limit the path length of nodes and starting nodes to facilitate experiments. Otherwise, the constructed knowledge map will be a very large collection. Finally, an undirected graph is formed according to the set of knowledge terms and the set of relationships, which is the stroke knowledge map. The construction flow chart of the safety medical big data ecosystem knowledge is shown in the figure below.

### 4. Results

The research and analysis of big data storage under the Hadoop platform needs to complete the loading of virtual machines, the installation of Hadoop and platform construction, the application of some file management instructions of HDFS and the problems of Hadoop storage. The first thing to solve is the installation of Linux. I have consulted a lot of materials and articles. I have experienced many failures, and I have slowly understood many problems that I have not succeeded in. Finally, I have an idea of installing a virtual machine system. CentOS7.0 was successfully installed and its basic configuration and installation of some basic tools were completed. The high efficiency of Hadoop-based secure medical big data ecosystem is determined by its own characteristics—fast information transmission, transparent information and information. Because information reduces transaction costs, coordinated costs, and execution transaction costs, it increases transaction efficiency and increases patient value. Improving the efficiency of medical business is one of the important goals of medical digitalization. With the help of electronic information mobilization means and information mode, such as automatic medicine
dispensing machines and logistics lines created by modern information technology, medical services such as automatic medicine dispensing machines and logistics lines can be used. Rapidly improve medical efficiency and make labor-intensive business transition to high efficiency and precision. The increase in transaction efficiency leads to a reduction in costs and thus more value. At the same time, more information or improved information reduces the cost of seeking medical consultations for patients. Through the remote video clinic of the upper hospital and the logistics distribution of the medical platform, third-party payment and other background management, patients can find superiors in the nearest community hospital. Hospital doctors have simplified medical procedures, effectively reduced health care costs, saved time and reduced medical costs. Figure 9 shows the Test result between Hadoop and Traditional method.

China has not formed an effective technology construction mode and business operation mode in the construction of information platform services. In terms of technology, standardization management of information has high requirements for information transmission standards and exchange standards, and the state has no system. Advance standards, comprehensive evaluation of system platform construction, including supervision and evaluation of institutions, stay in the line and framework stage; in the service operation, there is a lack of completed and mature standardized management system, and no effective service model is formed. The Safety Medical Big Data Ecosystem is an information platform for all medical and health departments in the society. It involves the special group of the elderly, the health administrative department, the medical service organization, the disease prevention, and control center, the social medical security and other related departments. Multi-sectoral integration and the management of the elderly in China is still multi-sectoral management, the government functions are not clear, resulting in uneven pension services, unbalanced demand for old-age care, and low resource utilization, which seriously affects the development of services for the elderly. The government is required to form a complete and mature standardized management mechanism and manage it in an integrated manner to integrate health care resources.

Medical expenses are the largest expenditure in addition to basic living. China's medical insurance is limited to a single business operation, and residents' medical insurance has not yet involved regional collaborative services. The medical care system has not yet established, and the medical insurance costs for inter-patient diagnosis cannot have settled. Many local telemedicine services are not included in the scope of medical insurance reimbursement. Therefore, the elderly are more willing to go to the first-level hospital or the secondary hospital to set up a medical insurance hospital for treatment.

The development of informatization in China has only begun in recent years. There is a serious information bias. In the city and large hospitals, there is still an initial construction of informatization, and the level of informatization of grassroots health institutions is not widely applied, resulting in the inversion of the hierarchical structure [21]. It is because the awareness of informatization construction of grassroots health institutions is weak, and the lack of introduction of informationized talents is lacking. Secondly, the level of informatization application is low. The health informationization construction of primary medical institutions has not yet covered comprehensively. The treatment of basic diseases has
not involved informatization. Informatization applications are still at a lower level, lacking the collection and accuracy of data, leading to the informatization construction in the community.

In the development of the safety medical big data ecosystem, the role of government departments is the most critical, and it is the main bearer of medical services. Therefore, the government must scientifically develop the old-age service system, conscientiously implement the multi-level and diversified service needs of the elderly, and adhere to the development of China's elderly services in the direction of intelligence and information. The health management department should avoid the information system project in the construction process, because the standards are inconsistent, which are leading to the original intention of leaving the design, meaningless waste of the sword. At the same time, the government should give full play to the guiding role and break the informationization of medical resources. In addition to the horizontal management of health services, it pays attention to the vertical management of health services, coordinates the interests of various departments, and cooperates regionally to form a scientific system. Integrate more old-age resources, strengthen overall coordination, study and formulate a unified, specific and standardized elderly care service process, give full play to the government's binding role, and establish a benign interaction mechanism between government, society and market. The design and grassroots advancement have combined at the same time, and the government should form a unified payment system and provide financial support for the aged care service.

Because of the weakening of physical function, the elderly have a high incidence of illness in the elderly, frequent frequency of medical treatment, medical expenses have characterized by “rigid” rise, and the burden of medical expenses is heavy. The government should adjust the medical insurance policy and increase the expenditure on medical insurance for the elderly [22]. The government should also encourage the development of medical commercial insurance, guide residents to make more scientific and rational medical financial planning, broaden the source of medical insurance funds for the whole society, and mobilize social resources.

Whether it is a traditional nursing home that has already started operating in the early days, or a large capital force that is planning to build an industrial platform, they are in urgent need of a comprehensive and efficient information system. Primary medical institutions are the health institutions closest to the elderly. The informationization of grassroots medical institutions directly affects the amount of medical data provided by the health care big data ecosystem. Therefore, it is recommended to increase capital investment in grassroots medical institutions, strengthen information construction, and introduce informationization.

The advantages of the medical information health platform should have popularized for the elderly, and the elderly can experience the convenient service brought by the medical information platform. The elderly can understand their physiological parameters at home and operate easily. Through platform remote consultation, the space distance is greatly shortened, online consultation and diagnosis services for elderly people and doctors have provided, and the self-health management of the elderly, especially
elderly patients with chronic diseases, and the transition from public management to personal management are guaranteed to a certain extent. The efficient service quality of the elderly and the improvement of the satisfaction and participation of the elderly in medical treatment can stimulate the enthusiasm of the elderly to use the platform.

5. Discussion

In general, with the continuous deepening of China's medical and health system reform, medical and health informationization has become a booster for the development of health care, especially in the face of a traditional service such as old-age services; there is a lot of space for medical and health informationization. Integrating the medical resources of individuals, hospitals, and society through the secure medical big data ecosystem has improved the accessibility of older people to medical services. Reducing medical costs is bound to be the development direction of the health care industry. This paper designs and implements a secure medical big data ecosystem on top of the Hadoop big data platform. It is designed against the background of the increasingly serious trend of the current security medical big data ecosystem. The system mainly designs personal health data center, personalized information recommendation subsystem, and other modules. Utilizing the advantages of the Hadoop big data platform, the personalized health information service has designed to provide personalized health management services for patient users, while providing convenience for medical staff to manage patients.

List Of Abbreviations

(IBM) International Business Machine

Declarations

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

declares that he has no conflict of interest.

* Research involving human participants and/or animals

Ethical approval:
This article does not contain any studies with human participants or animals performed by any of the authors.

* Informed consent

All authors agree to submit this version and claim that no part of this manuscript has been published or submitted elsewhere.

**Funding**

Yellow River Conservancy Technical Institute innovation fund, KJ18027

**Authors' contributions**

Xiangfeng Zhang is responsible for the collection of experimental data, while Yanmei Wang is responsible for the simulation of experiments and the writing of papers

**Acknowledgements**

Yellow River Conservancy Technical Institute innovation fund, KJ18027

**References**

1. HASHEM, Ibrahim Abaker Targio, et al. The rise of “big data” on cloud computing: Review and open research issues. Information Systems, 2015, 47: 98-115.

2. LUO, Jake, et al. Big data application in biomedical research and health care: a literature review. Biomedical informatics insights, 2016, 8: BII. S31559.

3. YE, Ying; LAU, Kwok Hung. Designing a demand chain management framework under dynamic uncertainty: An exploratory study of the Chinese fashion apparel industry. Asia Pacific Journal of Marketing and Logistics, 2018, 30.1: 198-234.

4. SAGIROGLU, Seref; SINANC, Duygu. Big data: A review. In: Collaboration Technologies and Systems (CTS), 2013 International Conference on. IEEE, 2013. p. 42-47.

5. FIERRO HERNANDEZ, Daniel; HADUD, Abubaker. Value creation via supply chain risk management in global fashion organizations outsourcing production to China. Journal of Global Operations and Strategic Sourcing, 2018, 11.2: 250-272.

6. EICKER, Themari; CILLIERS, Jacoba O. Equipping small business retailers to manage logistical supply chain drivers: A theoretical guideline. Journal of Transport and Supply Chain Management, 2017, 11.1: 1-12.
7. DEMCHENKO, Yuri; DE LAAT, Cees; MEMBREY, Peter. Defining architecture components of the Big Data Ecosystem. In: Collaboration Technologies and Systems (CTS), 2014 International Conference on. IEEE, 2014. p. 104-112.

8. RAGHUPATHI, Wullianallur; RAGHUPATHI, Viju. Big data analytics in healthcare: promise and potential. Health information science and systems, 2014, 2.1: 3.

9. LV, Zhihan, et al. Next-generation big data analytics: State of the art, challenges, and future research topics. IEEE Transactions on Industrial Informatics, 2017, 13.4: 1891-1899.

10. CARDENAS, Alvaro A.; MANADHATA, Pratyusa K.; RAJAN, Sreeranga P. Big data analytics for security. IEEE Security & Privacy, 2013, 11.6: 74-76.

11. DEMCHENKO, Yuri; NGO, Canh; MEMBREY, Peter. Architecture framework and components for the big data ecosystem. Journal of System and Network Engineering, 2013, 1-31.

12. RIGGINS, Frederick J.; WAMBA, Samuel Fosso. Research directions on the adoption, usage, and impact of the internet of things through the use of big data analytics. In: System Sciences (HICSS), 2015 48th Hawaii International Conference on. IEEE, 2015. p. 1531-1540.

13. AHMED, Eajaz, et al. The role of big data analytics in Internet of Things. Computer Networks, 2017, 129: 459-471.

14. VATSALAN, Dinusha, et al. Privacy-preserving record linkage for big data: Current approaches and research challenges. In: Handbook of Big Data Technologies. Springer, Cham, 2017. p. 851-895.

15. RATHORE, M. Mazhar, et al. Urban planning and building smart cities based on the internet of things using big data analytics. Computer Networks, 2016, 101: 63-80.

16. THOTA, Chandu, et al. Centralized fog computing security platform for IoT and cloud in healthcare system. In: Exploring the convergence of big data and the internet of things. IGI Global, 2018. p. 141-154.

17. SHARIF, Ather, et al. Current security threats and prevention measures relating to cloud services, Hadoop concurrent processing, and big data. In: Big Data (Big Data), 2015 IEEE International Conference on. IEEE, 2015. p. 1865-1870.

18. GUPTA, Maanak, et al. Multi-layer authorization framework for a representative Hadoop ecosystem deployment. In: Proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies. ACM, 2017. p. 183-190.

19. YANG, Ji-Jiang, et al. Emerging information technologies for enhanced healthcare. Computers in Industry, 2015, 69: 3-11.

20. GUPTA, Maanak; PATWA, Farhan; SANDHU, Ravi. An Attribute-Based Access Control Model for Secure Big Data Processing in Hadoop Ecosystem. In: Proceedings of the Third ACM Workshop on Attribute-Based Access Control. ACM, 2018. p. 13-24.

21. AHMAD, Awais, et al. Multilevel data processing using parallel algorithms for analyzing big data in high-performance computing. International Journal of Parallel Programming, 2017, 1-20.
AHMAD, Awais, et al. Toward modeling and optimization of features selection in Big Data based social Internet of Things. Future Generation Computer Systems, 2018, 82: 715-726.

Figures

Figure 1

The processing flow framework
Figure 2

MapReduce framework of Hadoop

Figure 3

The master-slave architecture of Hadoop
Figure 4

The network structure used by this system
Figure 5

Overall design of the distribution Hadoop
Figure 6

Command sets of the Hadoop management system
Figure 7

Overall dataflow of Hadoop
**Figure 8**

The construction flow chart of the safety medical big data ecosystem

**Figure 9**

Test result between Hadoop and Traditional method