



Design of Medical Insurance Fund Audit Platform Based on Big Data Analysis

Xiaolan ZHANG^{a,1}

^aBeijing Jiaotong University, Beijing, China

Abstract. The medical insurance fund is an important guarantee for people's health. It is of great significance for the country to maintain the people's livelihood by timely discovering illegal acts and avoiding fund losses. Due to the characteristics of high dimension, large quantity and complex data type of medical insurance data, traditional audit methods have been unable to be fully applicable. Therefore, aiming at the massive medical insurance data, this paper takes the HDFS distributed storage system based on Hadoop as the basic system architecture, makes full use of the data analysis method in big data processing, designs and develops the medical insurance fund audit platform for the majority of medical insurance audit professionals, with a clear and friendly analysis process. This program can help to improve the accuracy and efficiency of medical insurance fund audit work, accelerate the transformation of results of big data analysis technology, and promote Chinese progress of medical insurance fund audit work.

Keywords: Medical Insurance Fund Audit; Big Data Analysis; Distributed Storage; Audit Platform

1 Introduction

With the rapid development of information technology and economic level, in order to ensure the rational use of national resources, the state is gradually improving the digital and scientific management of national public resources. Among them, the audit management of medical insurance fund is the primary task of medical insurance reform and development. Due to the large number of insured people in our country, the medical insurance data involved in the audit has the characteristics of high dimension, large quantity and complex structure, so the traditional audit method can't be fully applied. With the increasing popularity of national medical insurance in China, the medical insurance audit platform needs to process more and more data. The amount of data in many hospitals has risen from TB level to PB level, and these medical data are stored in different insured units, insured departments and computers. Therefore, it is required that the medical insurance fund audit platform can accurately deal with complex and massive data, and can deal with various types of data. In this regard, many scholars put forward the use of big data processing technology to solve the above problems.

¹ Corresponding Author, Zhang Xiaolan, Beijing Jiaotong University, Beijing, China; Email: zhangxiaolan995zz@163.com

Therefore, based on the Hadoop distributed system infrastructure, this paper aims to rationally utilize and optimize existing 5T data analysis algorithms for massive medical insurance data, and design and develop a medical insurance fund audit platform for medical insurance audit practitioners with a clear and friendly user management interface and data analysis process to promote the progress of our medical insurance fund audit.

2 Research Status

In order to better realize the management and analysis of medical insurance fund data, there have been many researches on the construction of medical insurance fund audit platform. Wu Hengliang [1] proposed a "top-down" implementation path for big data audit construction from the three dimensions of technology, management and mode, including five aspects: big data culture construction, overall planning, team building, platform construction, process design and problem breakthrough. Liu Guocheng et al. [2] simplified the big data audit platform into complexity, and proposed a theory that the big data audit platform can be divided into four service layer sub-platforms: acquisition, pre-processing, analysis and visualization, aiming to improve the intelligence of the big data audit platform. Lai Lizhen [3] built a "cloud audit" platform system in order to fully optimize and utilize audit resources and better promote information interaction and sharing. Based on cloud computing, key elements such as ETL, data warehouse, online analytical processing, data mining, intelligent engine and SOA service were integrated to improve audit efficiency. In addition, Ye Miaomiao [4] et al. also put forward the idea of using multi-layer big data to build an audit platform and realize the integration of audit data management sub-platform, audit data analysis sub-platform and big data design sub-platform. Helen Brown-Liburd [5], M Cao [6] and Chen Wei [7] all discussed the opportunities and challenges faced by the medical insurance fund audit platform under the new big data environment, and their research results provided a theoretical basis for improving the audit of the medical insurance fund audit platform.

Through the above introduction, it can be found that there are still the following problems in the existing medical insurance audit platform: Few of the existing researches are involved in the practical application; The existing research has some challenges to the popularization of the platform; The existing research requires auditors to be highly professional in computer, which is difficult to operate, and the data analysis process is complicated, which increases the difficulty of platform operation.

3 Platform Design and Implementation

3.1 Data Collection and Processing

Data is the most basic and important part of the medical insurance fund audit platform. Due to the large population base and many kinds of diseases in China, the data types required for medical insurance audit are complex, mainly including inpatient compensation data and outpatient data. All the data used in this study were desensitized, mainly

from medical insurance centers, designated medical institutions and banks in two county-level areas:

1) Relevant data of the medical insurance center: unit network of medical institutions, composition of the number of insured, name of designated medical institutions, payment status, medical expenses, composition of medical institution expenses, name of insured diseases, settlement of drug costs, settlement of hospital costs, attending doctors, etc.

2) Some relevant data of designated medical institutions: insured medical institution unit, insured user name, personal account information, treatment and medical expenses, medical institution fee composition, drug out-of-pocket expenses, hospital expense settlement, etc.

3) Relevant bank data: accounts of insured units, accounts of insured individuals, settlement of medical structures, settlement of medical insurance centers, etc.

Data processing refers to the modular processing of the original data collected, the screening of contradictory data, the cleaning of redundant data, and the supplement and adjustment of the data collected with incomplete information, and the batch cleaning of the data, so that the subsequent auditors can carry out more accurate data analysis.

3.2 Platform Architecture Design

The medical insurance fund audit platform deeply studies the business needs of medical insurance fund audit and the new features of the current medical insurance fund audit business in China, uses existing big data analysis algorithms and analysis technologies, builds a medical insurance fund audit platform based on big data analysis technology, and uses multidimensional analysis, query statistics Data analysis algorithms such as decision charts are used to analyze and process data. On the overall platform construction, platform design principles such as ease of use, efficiency, and scalability are also implemented, and a medical insurance fund audit platform that is easy for users to operate and system expansion is realized.

This section mainly introduces the technical architecture of the platform from the technical level based on the principle of platform architecture, and further introduces the technologies used by the platform at all levels. First of all, in terms of data storage, the medical insurance fund audit platform uses Hadoop distributed file storage system. The data computing layer mainly uses Spark distributed parallel computing to clean and process data. Secondly, the system framework is improved according to the actual operational requirements of the audit business. The technical architecture of the medical insurance fund audit platform is shown in Figure 1.

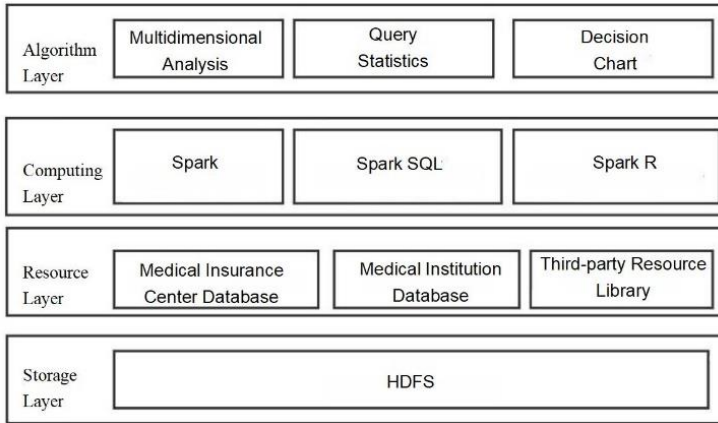


Fig. 1. Technical Architecture Diagram of Medical Insurance Fund Audit Platform.

3.3 Platform Database Design

The data storage system adopted by the medical insurance fund audit platform is MySQL and Hadoop HDFS storage system. In HDFS, Excel files are used for storage, and the storage is mainly original data. The structured data stored in MySQL is mainly used for audit query, statistical analysis and other work in the audit work.

The first is the design of the conceptual structure of structured data storage and the design of the specific contents in the database table according to the conceptual structure. Next, we will use the E-R diagram of the platform to represent the entities in the database and their attribute information, as shown in Figure 2.

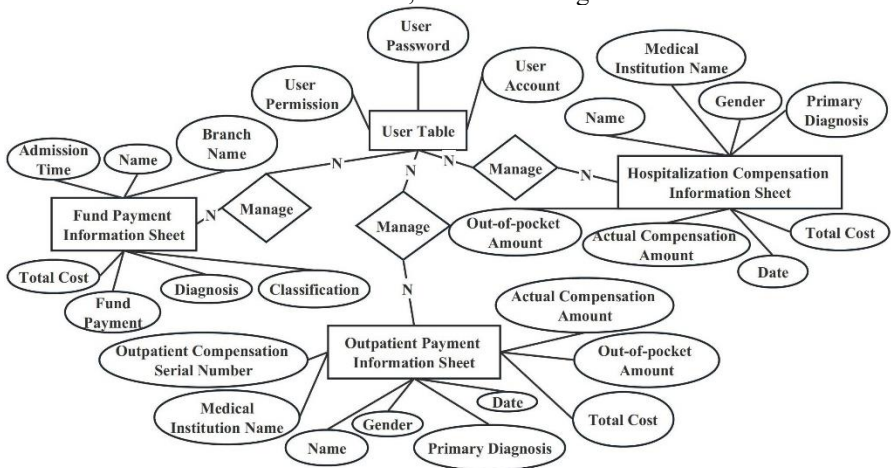


Fig. 2. E-R Model Diagram of Medical Insurance Fund Audit Platform.

In Figure 2, the rectangle represents the entities of the system, namely various information tables. The ellipse represents the attribute information of the entity, that is, the

attributes in various information tables. The diamond represents the relationship between entities. Entity relationships in this system are many to many. As can be seen from the figure, the entities in the medical insurance fund audit platform mainly include user information, fund payment information, outpatient payment information and inpatient compensation information. The attribute content of user information mainly includes user account, password and authority. The attribute content of fund payment information mainly includes outlet name, name, admission time, total expenses, fund payment, etc, The attribute content of inpatient compensation information mainly includes the name of the medical institution, the main diagnosis, the actual compensation amount, and the self-paid amount after the rural cooperative medical system compensation. The attribute content of outpatient payment information mainly includes the outpatient compensation serial number, the name of the medical institution, and the self-paid amount.

Unstructured data storage mainly refers to the storage of data files, such as Word files, PDF files, etc. HDFS distributed storage system is used in this platform to manage unstructured data, and the main file format stored is Excel files. In the storage process of HDFS, more than one node is used to store data. Data can be stored in multiple nodes without mutual influence between nodes to prevent data loss caused by a node failure, so as to improve the reliability and security of the system. The specific storage process is shown in Figure 3.

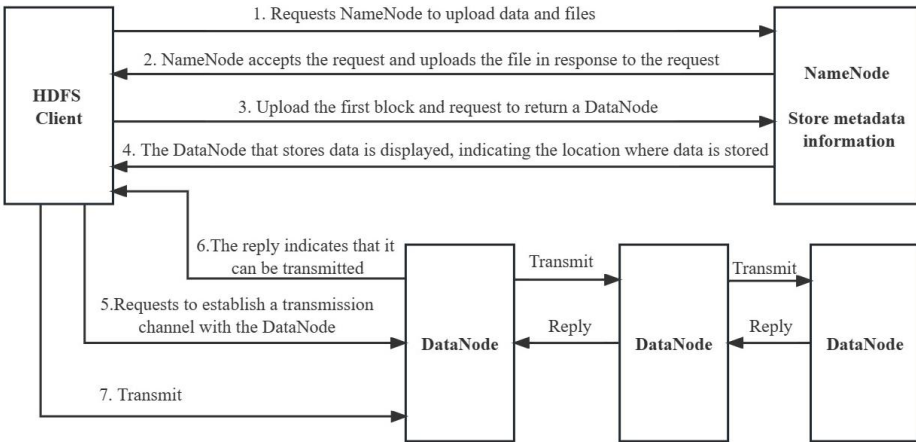


Fig. 3. HDFS Data Storage Process Diagram.

3.4 Function Module Design and Implementation

From the above demand analysis, we can see that the main function modules of this platform are user login, medical insurance fund data query and data statistical analysis. The following will mainly describe the design and implementation of data query and data statistical analysis from three aspects of function overview, technical analysis and implementation process through flow chart, sequence diagram, code design and other ways.

- Design and implementation of audit data query function

The audit data query function is one of the main application functions of the medical insurance fund audit platform, which facilitates users to quickly and accurately obtain the required data.

In the medical insurance fund audit platform, the data is mainly divided into three categories according to the payment type: fund payment data, outpatient payment data, and inpatient compensation data. Then the data query function can be realized in each type of data. The specific processing flow chart of the system data query function is shown in Figure 4.

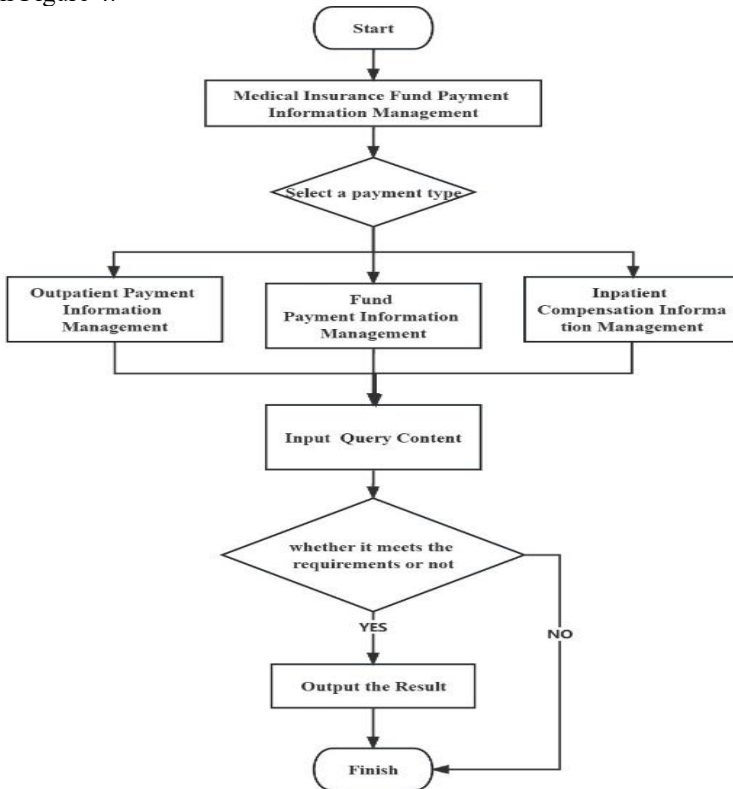


Fig. 4. Flow Chart of Audit Data Query Function.

The audit data query function adopts the front end and rear end separation to improve the operation efficiency. The front end uses the Java programming language. The Spark SQL distributed memory data query is combined with the Spring Boot framework and implemented through Eclipse. According to the query conditions selected and entered by the user, the back-end database queries and displays the results in the front end. After the front-end user clicks the query button, the front end actually makes a query request to the back end, and then the database queries. The final front end page displays the results.

- Design and implementation of audit data statistics and analysis function

The audit data statistics and analysis function is one of the core application functions of the medical insurance fund audit platform, which aims to achieve the statistics and analysis of medical insurance data and provide clear and definite data statistics and analysis results for auditors.

In the medical insurance fund audit platform, multi-dimensional data analysis of fund payment data, outpatient payment data and inpatient compensation data is mainly realized. Two dimensional analyses are realized for each data type, and the final data statistical analysis results are displayed in a visual chart, which helps auditors to make the next audit decision. See Figure 5 for the detailed processing flow chart of audit data statistics and analysis function.

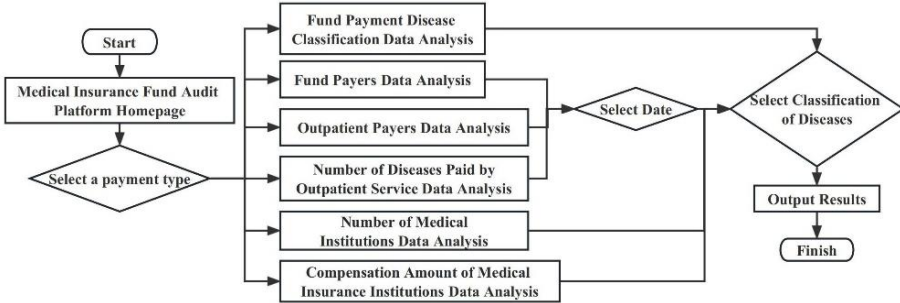


Fig. 5. Flow Chart of Audit Data Statistical Analysis Function.

The statistical analysis function of audit data is implemented by combining Spark distributed parallel algorithm and Spark R. The main process is to first use Spark technology to read the cleaned data from Hadoop HDFS, then extract some of the data that needs statistical analysis according to different audit requirements, store them in the Data Frame in Spark R, and then go to the Data Frame for the next step of data analysis, The final data analysis results are written into MySQL, and the E-charts database is used to visualize the data results. The specific technical process used in the statistical analysis process is shown in Figure 6.

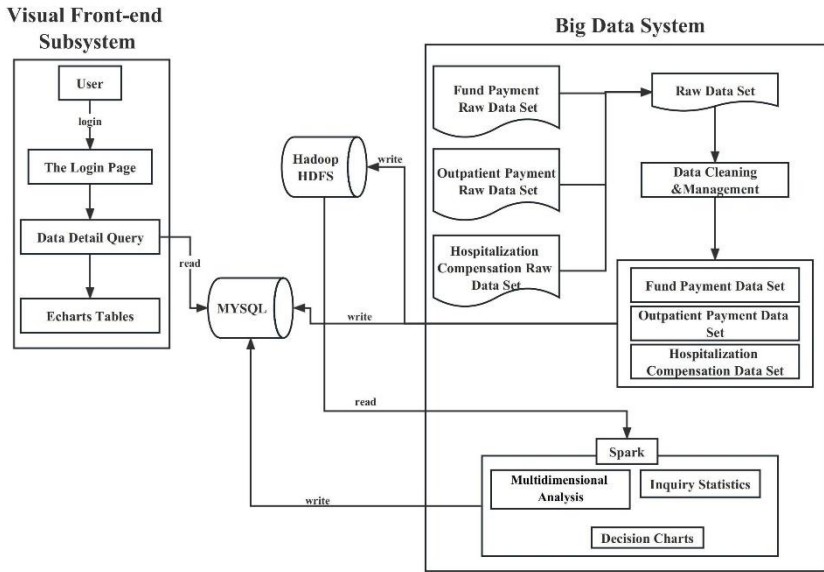


Fig. 6. Schematic Diagram of Technical Analysis Process of Audit Data Statistical Analysis Function.

In Spark distributed parallel computing, users submit data analysis requests to DataAnalyseController, and DataAnalyseController sends analysis conditions and requests to DataAnalyseService. Then SiteBaseDataService and SiteBaseData are used to determine the data analysis and implementation methods, retrieve data from Hadoop, and finally return the survey results to the client. See Figure 7 for the implementation process of Spark distributed parallel computing for specific data statistical analysis functions.

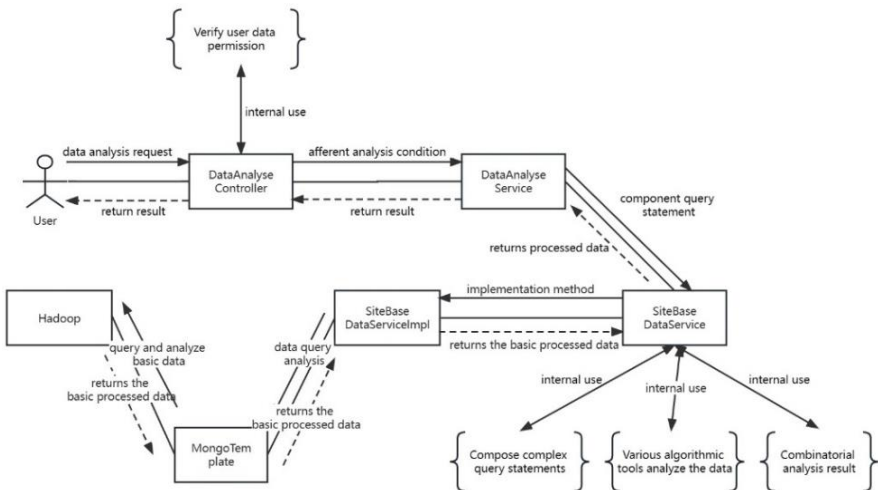


Fig. 7. Use Case Diagram of Spark Statistical Analysis Technology.

4 Platform Test and Operation Effect

After the basic business functions have been realized, the next step for the medical insurance fund audit platform is to test the platform. The test content is mainly divided into two aspects: function test and performance test to determine whether the platform meets the business and platform requirements.

4.1 Function Test

The function test aims to ensure that all functional modules of the platform can be executed normally, and the processing results are error free, which can support the audit work of auditors. The main method of functional testing is the black box method, which treats the entire software program as a black box, regardless of the specific implementation process of the platform. It focuses on testing whether the business functions of the platform meet the goals determined in the requirements analysis phase, and records the problems that occur in the testing process.

4.2 Performance Test

The medical insurance fund audit platform has passed the function test and performance test. The specific test conclusions are summarized as follows:

- Functional test

The final pass rate of test cases is 100%, and all meet the end criteria of the test, and all pass the functional test.

- Performance test

According to the performance test results of the medical insurance fund audit platform, the response time and various monitoring indicators of the platform meet the business needs, with an average response time of 2 seconds, effectively meeting the business needs.

5 Conclusion

On the basis of fully investigating the relevant work of medical insurance fund audit and the research status at home and abroad, this paper uses big data analysis technology, fully integrates the needs of current medical insurance fund audit business, develops and designs an easy-to-use and professional web-based medical insurance fund audit platform to clean. As the main function module, query and analysis effectively improve the efficiency of medical insurance fund audit and relieve the pressure of medical insurance auditors in China.

This paper focuses on the development process of this platform, which mainly goes through four stages: requirements analysis, system design, system implementation and system testing. The main implementation results of each stage are described as follows:

- 1) In the demand analysis phase, we realized in-depth research and learning on the audit business of medical insurance funds, understood the current research status at

home and abroad and the needs of the majority of medical insurance auditors, and determined the overall functional and non-functional requirements of the platform.

2) In the system analysis and implementation phase, the overall design principles of the medical insurance fund audit platform were elaborated in detail, and the design and implementation process of the main functions of the platform and the design and implementation process of the platform database were emphatically introduced, and the results of the platform development and design were displayed.

3) In the system testing phase, we fully tested the medical insurance fund audit platform, tested the internal architecture of the platform with the white box method, and tested the functions and performance of the platform with the black box method. Finally, according to the test results, the platform has passed all tests.

References

1. Wu Hengliang, Yu Benhai, Zhai Yuxiang, Zhang Weiwei, Research on the implementation path of medical insurance fraud audit in the big data environment [J]. *Finance and Accounting Monthly*, 2019 (24). 105-110
2. Liu Guocheng, Wang Huijin, Research on the Construction of Big Data Audit Platform [J]. *Audit Research*, 2017 (06). 36-41
3. Lai Lizhen, Construction of cloud audit platform system in the context of big data [J]. *Finance and Accounting*, 2017 (03). 43-45
4. Ye Miaomiao, Xu Sixing, Construction and implementation method of big data audit platform [J]. *Financial Supervision*, 2020 (19). 99-104
5. Helen Brown-Liburd, Miklos A, Vasarhelyi. Big data and audit evidence [J]. *Journal of Emerging Technologies in Accounting*, 2015(12).1-16.
6. M Cao, R Chychyla, T Stewart. Big data analytics in financial statement audits [J]. *Accounting Horizons*, 2015(29).423-429.
7. Chen Wei, Smielauskas Wally. Electronic data audit in big data environment: opportunities, challenges and methods [J]. *Computer Science*, 2016 (43). 8-13

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

