

Analysis of Data Audit Mode in Big Data Environment

Yan Li

Lingnan Normal University, Zhanjiang, China

Abstract:

As the latest product of the new generation of information technology innovation and integration, big data has had a great impact on many fields. Combined with the research status of big data audit at home and abroad and based on the characteristics of big data environment, this paper proposes the logical workflow and technical methods of data-based audit, and puts forward suggestions on the construction of big data audit platform, information sharing, talent training and data security, providing ideas for the research and practice of big data audit mode.

Keywords: *Big data, Data audit mode, Technology method*

I. INTRODUCTION

Big data, artificial intelligence, cloud computing and other digital technologies have pushed the world into the era of digital economy. At the Us Government Audit Forum (2013), T.M. Persons proposed to pay attention to the impact of big data trend on audit. The 2014 annual meeting of American Accounting Institute (AAA) introduced the discussion on the application of big data analysis technology in the field of audit, and the international Big Four also began to actively promote the application of data analysis technology in the audit practice. In December 2021, the Ministry of Finance issued the outline of the 14th Five-year Plan for Accounting Reform and Development, which proposed that "actively promote the digital transformation of audit work and encourage accounting firms to use digital audit technology in accordance with laws and regulations". In June 2021, the National Audit Office issued the "14th Five-Year Plan for The Development of National Audit Work", pointing out that "strengthening the innovation of audit technology and methods, making full use of modern information technology to carry out audit, and improving the quality and efficiency of audit".

Applying big data technology to audit work, realizing the development from traditional accounting audit to computer aided audit and then to big data audit is the general trend of future audit development. Big data audit has become an important way for audit departments at all levels to improve audit efficiency, ensure audit quality and promote audit work digital transformation.

II. LITERATURE REVIEW

2.1 Research on the Relationship between Big Data and Audit

T.M. Persons (2013) proposed in the Us Government Audit Forum that auditing should consider the impact of the development trend of big data. Costonis (2013) pointed out that audit data under big data also includes unstructured data, and auditors should master its nature and data selection [1]. Andy Lymer (2014) studied the r&d of audit technology and the input-output ratio of audit work and found that the r&d of big data technology would further improve audit efficiency [2]. Wang (2015) pointed out that it is very important to use big data technology for audit work, but accounting firms do not pay enough attention to it and lack application experience [3]. Kitchin R et al. (2016), by comparing the characteristics of small data and big data in different fields, found that big data has more key characteristics of high frequency and completeness. Therefore, in the era of big data, the audit forensics environment will undergo significant changes [4]. Domestic academia and practice of the great data audit research began in 2013. Liu Bixiang (2013) analyzed the opportunities and challenges posed by big data to computer audit and proposed to use big data to promote computer audit. It also proposes to solve the problems of ensuring data quality, establishing a public data analysis platform, strengthening the training of auditors and establishing a reasonable audit result evaluation system [5]. Qin Rongsheng (2014) pointed out that big data technology promoted the comprehensive application of audit results and the development of efficient data audit. Some feasible suggestions are put forward on how to develop data-based audit [6]. Lu Qingfang and Liang Zihui (2015) pointed out that the use of big data can reduce accounting fraud and audit risk caused by CPA's reliance on error of experience judgment, thus improving the efficiency and effect of audit [7]. Long Ziwu and Wang Yunpeng (2016) point out that big data auditing can reduce audit risks of accounting firms and improve the quality of audit conclusions, but also put forward higher requirements on the quality and technology of accounting firms' staff and audit data security [8].

2.2 Research on Data Audit Technology

Chen (2011) proposed to change the way of obtaining audit evidence by collecting and analyzing the financial data of enterprises through remote networking [9]. Overpeck J T (2014) proposed to integrate Python into the audit work and solve the big data problems with its rich extended library, and made a specific analysis of Python module functions [10]. Brown-liburd (2015) believed that CPA's big data processing ability should be emphasized, and analyzed how to train auditors in big data auditing [11]. Sookhak (2017) provides a remote data audit solution that identifies the integrity of data stored in the cloud [12]. Many domestic scholars have studied the development and application of data-based audit technology. Lu Jinsong (2014) studied how to make use of data analysis technology in financial audit under big data, and proposed to focus on the analysis of policy implementation, capital investment, risk control and prevention, so as to achieve continuous tracking and full coverage of financial audit [13]. Cheng Cheng and Li Rui (2016) analyzed how to take electronic data forensics directly as audit evidence in the big data environment and how to prevent the risks of electronic data audit forensics [14]. Chen Wei and Sun Mengdie (2018) put forward the working principle of using web crawler technology to conduct big data

audit and verified it through case analysis. The technical methods of their research are conducive to the development of big data audit in the future [15]. Chen Wei (2019) took financial audit as an example and analyzed case data with commonly used visualization technologies such as scatter chart, bar chart, broken line chart, histogram, bubble chart and violin chart, providing a practical basis for accounting firms to develop big data audit tools or other software tools in the future [16]. The research group of theoretical Research Institute of Shanghai Special Commissioner office of National Audit Office (2020) applied Python language and Neo4j graph database technology into the policy implementation tracking audit, combined big data technology with the practice of policy implementation tracking audit, found the focus of data-based audit, and proposed improvement measures based on big data technology [17].

Based on the above research, accounting firms in China are in the initial stage of big data technology application. The research on big data technology mainly focuses on the advantages of big data technology. Although some government auditing has applied big data technology to practical research, it provides limited help to the auditing work of accounting firms. Therefore, it is very important to conduct theoretical and practical research on the implementation of big data technology in the auditing work of accounting firms.

III. DESIGN OF DATA AUDIT MODE IN BIG DATA ENVIRONMENT

3.1 Data Audit Workflow

Today, most accounting firms still choose traditional auditing methods. In the process of discovering audit evidence, due to the limitation of actual conditions, only part of the information of the audited company can be collected, thus affecting the accuracy of audit evidence. Big data audit data includes unstructured and structured data. The quantity is large and difficult to deal with, so auditors cannot effectively deal with and analyze them. In addition, the way of generating audit report is relatively simple, which cannot be effectively applied in similar audit projects and cannot continuously provide audit suggestions for enterprises. Therefore, in order to improve the defects of traditional audit methods, the audit process framework in the big data environment is constructed, as shown in Fig 1 below.

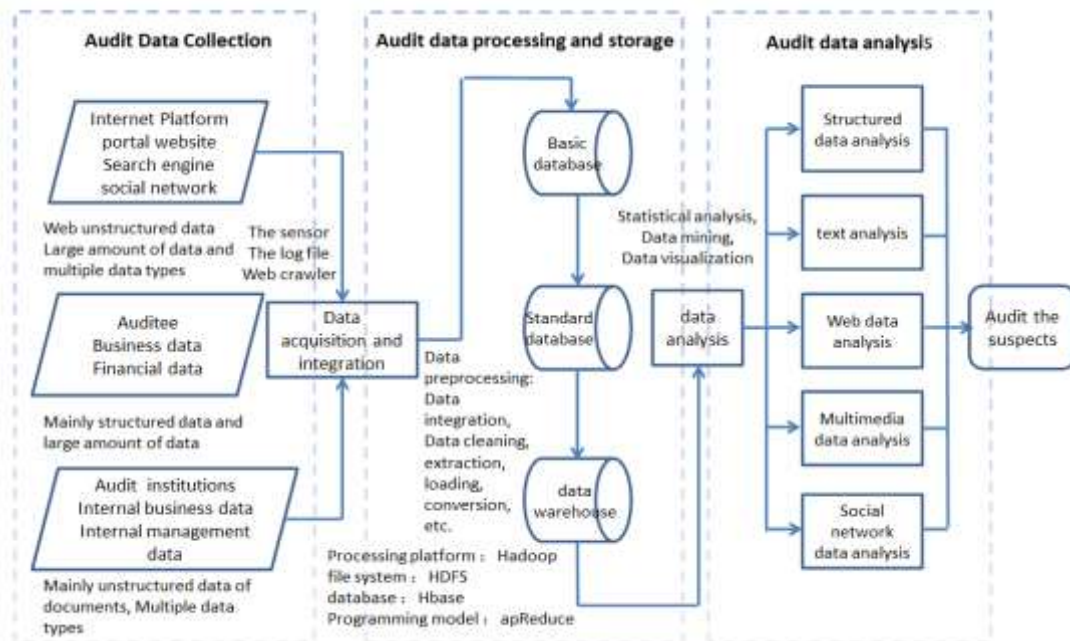


Fig 1: data audit logic workflow

3.1.1 Auditing the collection phase of big data

"Big data" refers to the collection of massive electronic data, such as structured data and unstructured data, whose contents cannot be captured, managed and processed by conventional software tools within a certain period of time. This concept indicates that new data processing modes need to be developed to stimulate the information that can bring useful value to decision-making. For auditing, big data is mainly divided into internal data and external data, among which internal data refers to the data generated by the continuous operation of enterprises, while external data mainly refers to the Internet and industry information in the form of documents or images. It is generally convenient to collect data inside the enterprise. For example, the accounting books of the current year and the previous year of the enterprise can be obtained directly from the server of the finance department. The establishment of financial sharing center will further realize the electronization of enterprise financial and business data resources. Its working principle is that each business department and each branch structure will scan the original vouchers, and the results are summarized to the sharing center for centralized processing, which is also very convenient for the collection and reference of audit evidence. External data mainly come from the statistical data of major organizations and industry associations, such as the Statistics Bureau, China Securities Regulatory Commission, and other official websites as well as major media news websites. This data can be accessed, viewed, and downloaded by itself, and some can be crawled using Python crawlers. By using crawler, we can obtain a large amount of value data, so as to obtain information that cannot be obtained in perceptual cognition.

3.1.2 Auditing big data processing and storage

The data used by auditors are large in scale, diverse in type and different in structure. Generally, the data taken cannot be used directly. The data needs to be cleaned first. Cleaning is to check the quality of data, discover and correct existing errors. The main purpose of this process is to check the consistency of data and filter out those duplicate, contradictory and empty data, thus reducing the deviation of audit work. Using Apache Drill, Hadoop and other big data processing technology, after extraction, loading, conversion and other steps, to obtain the audit required data. For structured data such as financial data, it is easy to collect and analyze, and generally does not need the pre-processing stage. Auditors can directly import it into the audit software.

The performance of computing and storage devices cannot meet the development trend of big data through simple upgrades. In particular, big data technology must be used to process and store unstructured data. The big data audit platform can use the Hadoop massive data processing platform, HDFS distributed file system, Hbase database, and MapReduce programming model as the core, and realize the storage and management of big data audit. Hadoop Distributed File System (HDFS) technology can be deployed and constructed on inexpensive server hardware. It has the advantages of high fault tolerance and high scalability, and is easy to install and deploy. The NoSQL database technology has become a standard for processing big data to meet diversified and large data requirements. Hbase column storage databases integrate existing relational databases to facilitate batch data processing and real-time query. The MapReduce distributed programming model is a simple but powerful technology that distributes large-scale computing tasks among clusters of inexpensive server hardware, allowing parallel applications to be developed without understanding the underlying details of a distributed system.

3.1.3 Auditing the big data analysis phase

In the context of big data, there are three principles for big data analysis, which is, not sampling for all, not absolute accuracy for efficiency, and not causality for correlation [18]. Statistical analysis, data mining and data visualization can be more used in big data analysis to expand the capability of audit data analysis. Different types of data analysis require different analysis methods. Structured data analysis mainly adopts data mining and statistical analysis. Text analysis mainly adopts the methods of text expression, natural language processing, information extraction, summary, classification and clustering, question answering system and opinion mining, etc. Webpage analysis mainly adopts webpage content mining, webpage structure mining and webpage usage mining methods. Multimedia analysis mainly uses abstract, annotation, index retrieval, and recommendation and event detection methods. Social network analysis mainly uses link prediction, community discovery, social network evolution, impact analysis, keyword search, classification and clustering, and transfer learning methods [19]. Webpage analysis mainly adopts webpage content mining, webpage structure mining and webpage usage mining methods. Through the above big data analysis methods, audit data analysis can conduct in-depth mining of all kinds of collected data, and achieve audit data analysis results that are difficult to obtain in the past. Data auditing will play a big role in this phase.

3.2 Technical Methods of Data Audit

With the explosive development of information data, auditors need to make use of new technical methods for the audit process under big data. The following describes three technical methods used in big data auditing. First, web crawler technology can be applied in the stage of auditing data collection or processing, which can automatically capture or screen out the required data. Second, natural language processing risk assessment can be used in the data analysis stage to intelligently process the audio data obtained from the audited units. Third, visualization technology can be applied in the data analysis and display stage to analyze all the data as a whole and display them intuitively.

3.2.1 Web crawler technology

Now the most common way to obtain external data is to use the web crawler technology. The technology is designed by the programmer to crawl the web information that needs to be obtained, analyze and filter the web information automatically, and obtain the content and retrieval method of web information, improve the efficiency of obtaining web information.

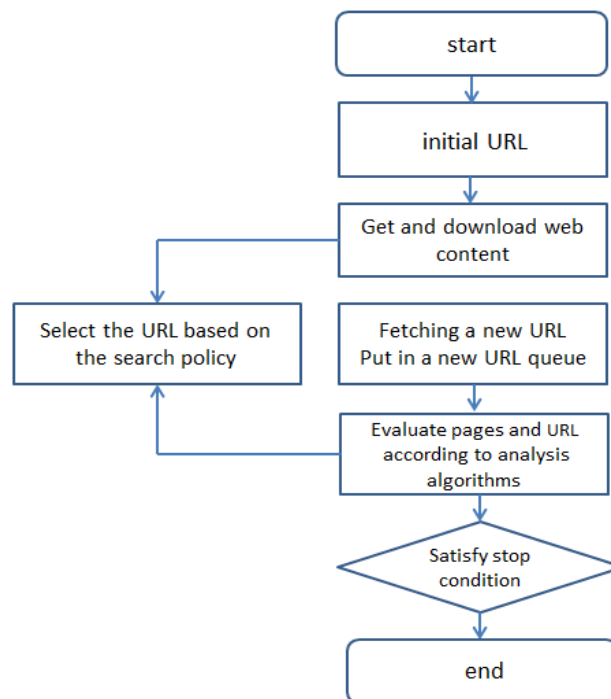


Fig 2: big data acquisition process based on web crawler technology

Web crawler technology can be used to purposefully grab information according to the required information, locate the target in the website related to the theme of the page and then start to crawl. Fig 2 shows the big data acquisition process based on web crawler technology. The first is to determine the network information to crawl. On the basis of meeting the conditions, determine the target website and analyze the corresponding web pages. The second is to capture the relevant information, using the web

crawler technology, to capture the required network data. There are two commonly used web crawler methods at present. One is the Request module based on Python language, which inputs relevant codes according to the required data. The software will automatically grab the information on the website according to the codes. This method requires users to write codes to screen out useless information while crawling information, but it requires high computer capability of users. The other is through excel to crawl, suitable for the weak computer ability of the people to use, through the excel data in the website can be directly operated. This method is simple and quick to operate, and users do not need to have a high level of computer power, but it has a certain impact on the accuracy of the information collected. Finally, the results are exported by software. Interprets the code that crawl the site information and saves it when it finds a site that matches the user's needs. And when there is a new url, the software will put the new url into the new capture team, until capture to meet the required conditions of all information. Data results can be saved in Text text or Excel spreadsheet for easy viewing.

3.2.2 Natural language processing

Auditors obtain further audit clues through interviews with relevant personnel. When talking with multiple employees on the same question and finding inconsistent or contradictory answers, it is easier to obtain audit clues. Compared with manual analysis, analysis using natural language processing technology is more efficient and effective. Intelligent analysis can be performed through natural language processing. The specific steps are shown in Fig 3. First of all, the conversation content with the relevant personnel of the audited unit will be recorded into audio files. Second, the computer automatically transfers audio files into text information. Finally, through the automatic analysis of natural language processing technology, the important information is obtained and the audit conclusion is formed.



Fig 3: analysis of interview information based on natural language processing

When natural language techniques are used in internal control tests, it is possible to find out whether the organization being audited is strictly following the internal control system. For example, to determine whether each business has been authorized, the authorized person information can be analyzed through the computer, and the authorized person of each business is automatically checked with the requirements in the internal control manual to check whether the two are consistent. When a discrepancy is found, it is automatically treated as an internal control failure. At the same time, through natural language processing technology, audit reports can be formed automatically. Natural language processing technology can automatically obtain the key information of the audit paper, thus forming the audit report.

3.2.3 Visualization technology

At present, among the large data audit application technologies, data visualization analysis technology is a commonly used data analysis means in many industries, and its development degree is relatively mature. Commonly used visualization techniques include bar chart, line chart, and violin chart, scatter chart, bubble chart and label cloud. Scatter diagram can be used to analyze the degree of data dispersion to find audit doubts; the bubble chart can intuitively discover the relationship between three or four variables; the tag cloud can collect all the text information on the website to process, analyze the key words. In order to achieve the optimal effect of audit data analysis, in the application of visualization technology, certified public accountants should specifically analyze in a variety of different situations to select the corresponding visualization method for analysis.

IV. SUGGESTIONS ON PROMOTING THE APPLICATION OF DATA-BASED AUDIT

4.1 Build Big Data Audit Platform and Information Sharing

The application of big data and cloud computing technology in audit is not a matter of a short time, and the platform construction is particularly important. The big data audit and analysis platform will greatly promote the application of big data in audit. The role of big data audit and analysis platform and cloud audit platform is to provide media and channels for data acquisition and analysis, and to realize remote storage and mobile computing. Big data audit and analysis platform can realize its functions through IaaS, PaaS and SaaS three service modes, and build a cross-industry and cross-field industry cloud audit platform.

The big data analysis models of different industries are different. The audit industry needs to research and develop the big data analysis models and software according to its own industry characteristics. Audit data analysis includes query, analysis and mining. Due to the wide range of big data content, specific audit practices may require the technical support of experts in the fields of geography, mathematics, statistics and social networks.

4.2 Attach Importance to Personnel Training and Data Security.

The concept of Industry 4.0 and intelligent manufacturing is proposed to integrate emerging technologies such as the Internet of Things and big data with the design, production, management and service of traditional manufacturing activities. The gradual development of accounting and auditing informatization will put forward higher requirements for accountants and auditors. The emergence of new technologies requires new talents, so it is necessary to strengthen the construction of big data audit talent team. As a link between students and enterprises, the school should actively explore the reform of the training mode of audit informatization talents, so as to supply a group of talents with professional ability suitable for practice. At the same time, audit practitioners should also put forward higher requirements for themselves in their work and continue to learn. Enterprise data security is the privacy of the enterprise.

Only in the case of privacy security, big data analysis can be benign development. Most companies consider it more secure to store and manage data themselves than to have a third party. At present, the cloud service development speed is very fast, related applications are more and more. If the providers of cloud services cannot effectively solve the security problems of the services they provide, then the whole system of cloud computing will be full of risks, and the consequences will be unimaginable.

4.3 Accelerate the Standardization and Construction of Relevant Laws and Regulations and Auditing

Current laws, regulations and auditing standards do not have clear provisions on the application of big data and cloud computing in auditing, and there is no relevant legal basis for the application of big data technology in auditing. In order to truly carry out the application of big data audit in the field of audit, it needs explicit support in terms of laws and regulations. Data collection and storage behavior, data analysis results and the legal status of relevant electronic evidence are all urgent problems to be solved in the application of big data technology in auditing. There is a variety of financial software on the market at present, although they are satisfied with the current accounting standards, but there are still some differences. Comparing and analyzing the financial data of enterprises using different accounting software, it is inevitable that there will be no corresponding situation. Therefore, it is necessary for the country to implement a standard architecture, so as to achieve consistency in the data structure and lay a good foundation for the future application of big data technology.

V.CONCLUSION

The rapid development of big data information technology promotes the audit work of accounting firms to a new stage. The information of the auditee is more and more digitized, and it is more and more convenient for the accounting firm to obtain the information of the auditee. The big data environment affects the audit work content, audit work flow and audit methods of accounting firms. Based on the characteristics of big data environment, this paper expounds the logical workflow and technical methods of data audit in accounting firms, and puts forward suggestions on the construction of big data audit platform, information sharing, attention to talent training and data security, so as to provide ideas for the research and practice of big data audit mode in the big data environment.

REFERENCE

- [1] COSTONIS M (2013). Tackling Big Data .Bests Review 7: 2-40
- [2] Andy Lymer (2014). Global technology audit guide (GTAG): Data analysis technologies. The Institute of Internal Auditors
- [3] Wang T, Cuthbertson R (2014). Eight issues on audit data analytics we would like to see researched. Journal of Information Systems 1:155-162
- [4] Kitchin R and McArdle G (2016). What Make Big Data? Exploring the Ontological Characteristics of 26 Datasets. Big Data & Society 1: 67-82
- [5] Liu Bixiang (2016). How to Use big Data to promote computer Audit. Science and Technology Information 21: 107-125

- [6] Qin Rrongsheng (2014). Research on the impact of big data and cloud computing technology on audit. *Audit Research* 6: 23-30
- [7] Lu Qingfang, Liang Zihui (2015). Research on the impact of big data on risk-oriented audit. *Journal of Henan Normal University* 2: 55-58
- [8] Long Ziwu, Wang Yunpeng (2016). Research on the impact of big data era on CPA audit risk and audit quality. *Friends of Accounting* 8: 112-114
- [9] Wei Chen et al (2012). Influence factors analysis of online auditing performance assessment. *Kybernetes* 6:587-598
- [10] Overpeck J T, Meehl G A, Beny S, et al (2014) .Dealing with data. *Science* 6: 639-806
- [11] Helen Brown Liburd and Hussein Issa and Danielle Lombardi (2015). Behavioral Implications of Big Data's Impact on Audit Judgment and Decision Making and Future Research Directions .*Accounting Horizons* 2:423-429
- [12] Mehdi Sookhak et al (2017). WITHDRAWN: Dynamic remote data auditing for securing big data storage in cloud computing. *Information Sciences* 3:101-116
- [13] Lv Jinsong, Wang Zhong (2016). Data Analysis in financial Audit. *Audit Research* 5: 26-31.
- [14] Cheng Cheng, Li Rui (2016). Research on electronic data audit forensics mode. *Audit Research* 5: 8-13.
- [15] Chen Wei, SUN Mengdie (2018). Research on Big Data Audit Method based on Web crawler Technology. *China Certified Accounting* 7: 76-80.
- [16] Chen Wei (2019). Case study of big data audit based on visualization analysis Technology. *China CERTIFIED Public Accountants* 6: 61-64.
- [17] Research Group of Theoretical Research Institute of Shanghai Commissioner's Office of National Audit Office (2020). Research on the application of big data technology in tracking audit of the implementation of major national policies and measures. *Audit Research* 2: 14-21
- [18] Victor Mayer-Schonberg (2012). *The Era of Big Data*. Zhejiang People's Publishing House
- [19] Li Xuelong, Gong Haigang (2015). Overview of big data systems. *Science China Information Science* 1: 1-44.