

## *Original Paper*

# Research and Prospect of Big Data Audit Technology

Shicheng Li<sup>1</sup>

<sup>1</sup> School of Economics and Management, Dalian University, Dalian, Liaoning, China

Received: January 27, 2024

Accepted: March 08, 2024

Online Published: March 27, 2024

doi:10.22158/ibes.v6n2p92

URL: <http://dx.doi.org/10.22158/ibes.v6n2p92>

### ***Abstract***

*In the era of big data, technology is constantly innovated. For the audit industry, the combination of big data technology and audit should be the development trend of The Times. The joint application and combination of various technologies will promote the further development of audit. This paper mainly expounds some basic technologies of big data audit, explores and discusses the problems and corresponding solutions in the development process of big data audit in today's society, and looks forward to the trend of big data audit in the future.*

### ***Keywords***

*big data audit, big data technology, audit*

## **1. Research Background and Significance**

With the rapid development of information technology, big data has become an indispensable resource, and the audit field is no exception. It is difficult for auditors to deal with huge amounts of data when using traditional audit methods. Due to the continuous update of big data technology, the application of big data technology in the audit work has become an important direction of innovation and development in the audit field. In the early stage of the combination of big data technology and audit, big data audit mainly relies on simple data extraction and query tools for the preliminary processing and analysis of massive data. With the continuous improvement of big data audit technology, big data audit can adopt Hadoop, Spark and other distributed computing frameworks, through the use of data collection, processing and analysis technology, to realize the rapid and effective audit of massive data, so as to improve the audit efficiency and quality. In addition, the good development of big data audit not only helps to promote the green innovation and high-quality development of the audit industry, but also enables audit institutions to be able to quickly respond to market changes and provide better audit services for the audited units. Therefore, this paper expounds the application and development of big data audit in recent years from the perspective of literature review.

## 2. Related Concepts of Big Data Audit

In the era of developed science and technology, the technology and theory are constantly changing, and the audit mode is constantly changing. The traditional manual audit has been changed to big data audit, so the relevant theory and application of audit are updated combined with big data technology. Different scholars also have different definitions of big data audit based on different perspectives, which can be divided into technology, data characteristics, audit objects and other perspectives. First, based on a technical perspective, Qin (2014) defines the audit method, audit mode, audit results and other aspects of big data technology, big data can help auditors to collect data and analyze data in a certain period of time, the ability to efficiently provide available information to decision makers (Qin, 2014); next, based on the data feature perspective, Liu et al. (2016) believe that big data audit has the characteristics of mass, diversity and high value, Favto the goal of achieving the full sample, Promote the improvement of audit work quality (Liu, Niu, & Tang, 2016); last, Based on the perspective of the audit objects, Chen et al. (2017) proposed that the audit objects of big data include electronic data and computer information systems, thus becoming a brand-new way of the audit, make the audit work close to the continuous, dynamic, real-time goal, to achieve an ongoing audit of the (Chen, 2017). Therefore, the definition of big data audit can be summarized as the collection, processing and analysis of the data required by the audit work through big data technology, to obtain more objective and fair audit opinions, and to publish appropriate audit reports.

## 3. The Concept and Application of Big Data Qudit Technology

### 3.1 Data Acquisition Technology

#### 3.1.1 Network Crawler Technology

Traditional data collection methods rely strongly on auditors, but people's information processing ability is limited, prone to low efficiency and other problems. The web-crawler technology can be understood from the technical principles and practical steps. Based on the understanding of technical principles, web crawler technology is based on certain rules, auditors can not only automatically capture the required data from massive data, but also collect relevant audit information more comprehensively. After the web crawler technology determines the crawling content, it can automatically grab the required data on the web page, and the data can be preprocessed to a certain extent, such as removing blank data lines and so on. The audit data is collected based on practical steps. Firstly, the auditors have a preliminary understanding of the auditor through the collection of national policies and industry information; secondly, according to the industry environment and the specific situation of the management of the auditor, the crawler tool captures and collects data through keywords. (Chang & Li, 2023) According to the audit of Asia Pacific Industrial, due to the few external data related to the audited units, the importance level was mainly based on the information provided by the audited units, resulting in the audit failure (Wang, Luan, & Zhang, 2020). Therefore, it is very necessary to collect sufficient and high-quality audit data. The auditor should collect data efficiently

from the national policies, industry conditions, business environment and other parties, so as to improve the quality of audit data.

### 3.1.2 API Juggle

API (Application Programming Interface), namely the application programming interface, is an efficient way of data exchange that can realize data sharing between different systems. Due to the general improvement of the information degree of the auditor, the auditors need to collect electronic audit data from the information system of the auditor. Then, the auditors need to collect data through the API interface, which can reduce the work pressure and greatly improve the efficiency of the audit work. At present, all major firms will include the application programming interface (API) as an important measure to improve the speed and quality of data collection, which can not only promote the application of big data technology, but also reduce the dependence of auditors on the traditional audit process, so as to provide higher value audit services. For example, accounting firms successfully simplify the authentication process with the API interface, which not only provides continuous and safe electronic authentication service, but also contributes to the rapid and accurate acquisition of audit data from the auditor, thus improving the efficiency and quality of the authentication.

## 3.2 Data Processing Technique

### 3.2.1 Data Cleaning Technology

In the data processing technology, the data cleaning technology is the auditor's preliminary screening according to the collected data, selects the needed data from the massive data, and cleans the unneeded data. In this way, it can not only help auditors remove duplicate and wrong data, but also identify incomplete data, so as to supplement audit data in time, so as to ensure the accuracy of audit data, and also improve the efficiency of audit work. Before auditors operate the data cleaning technology, it is generally necessary to evaluate the original data obtained, select key points for cleaning, and retain the required data. Therefore, the data cleaning technology ensures the quality of audit data and the correctness of audit conclusions in the process of inspection, screening and cleaning.<sup>[6]</sup>

### 3.2.2 Data Conversion Technology

Data conversion technology can be simply understood as if it transforms one type of data into another. Audit data can be divided into structured data and unstructured data, among which unstructured data includes pictures, voice, text, etc. In this way, auditors cannot intuitively analyze the data, which can easily lead to low audit efficiency. However, one of the methods to solve this problem by data transformation technology is to transform unstructured data into structured data, and then intuitively presented through data analysis technology, which is convenient for auditors to understand the relationship and changes between the data through charts and other tools. However, the premise of data transformation is type compatibility, so that the transformed data can maintain its own accuracy. If you perform a specific task, the data is converted by writing a simple program.

### *3.3 Data Analysis Technology*

#### *3.3.1 Data Mining*

Auditors from the auditors internal and external data collection, through data cleaning and data conversion, can use data mining technology in the huge amounts of data, the hidden invisible relationship between data and data mining, found the meaningful relationship between data, also can intuitively feel the changing trend of data. In the process of data mining, clustering method, trend analysis method, structure analysis method, ratio analysis method, text mining method, multidimensional data analysis method can be used to further mine the hidden relationship between audit data. (Wang, Bao, Wang, etc., 2020) It can be seen that data mining technology facilitates auditors to find the hidden risk points of the auditors, and can provide sufficient audit evidence and insight into the various problems existing in the auditors, so as to reduce the risk of audit failure and improve the audit quality.

#### *3.3.2 Data Visualization Techniques*

The use of data visualization technology is through the visualization software, such as R language, SAS, visual analysis of data mining results, in the form of intuitive, understandable, such as bars, charts, label cloud, theme river, not only easy to understand the data structure, and can directly observe the abnormal phenomenon in the audit data. Therefore, auditors can use the visual analysis results, professional competence and their own judgment to gain insight into the existing problems and risk points of the auditor, so as to get appropriate audit opinions and form a more objective and fair audit conclusion. Mu Hongsheng, Nie Huiping proved in the study of water pollution audit in the audit process using data visualization technology, such as label cloud analysis technology, can unstructured data analysis, through the text vocabulary and word frequency statistics, drawing label cloud map, makes the auditors by observing font, deep color, can quickly understand the audit focus, also can find more audit clues. (Mu & Nie, 2023)

## **4. Application Problems and Solutions of Big Data Audit**

### *4.1 Data Security Issues*

With the continuous improvement of enterprise information degree, how to ensure the security of data in the process of data collection, processing and analysis has become an important challenge facing big data audit. First of all, in the process of data collection, the auditor collects the industry information and policy information about the auditor through network crawler technology, which may involve the illegal invasion of the information system; when the API interface collects the internal data of the auditor, it may encounter business confidential information. Because the data storage is basically stored on the Internet, it is difficult to guarantee the confidentiality of data, resulting in data leakage and irreparable economic losses to the auditor. Secondly, in the process of data collection, it is difficult to determine the authenticity and validity of the data due to the inconsistency of the system. Therefore, from the auditor's point of view. Want to solve the above problems, can start from two directions, one

of the audit information system defense ability, strengthen the ability of information processing: the data security monitoring, on the one hand, increase a variety of encrypted information transmission path, on the other hand time monitoring data abnormalities, to ensure the security of the audited confidential documents and complete. (CAI & Li, 2023)

#### *4.2 Data Quality Issues*

Data quality is directly related to the reliability and validity of audit results, and low-quality data may lead to audit failure. Audit quality may be problematic during the collection and processing of the data. First of all, in the process of data collection, the data sources are relatively extensive. If the data is originally tampered with, it will directly increase the audit risk. Secondly, in the process of data processing, on the one hand, due to auditors lack of experience and lack of judgment ability, lead to screening the audit data. On the other hand, in the process of conversion, auditors will be incompatible data through wrong coding processing to get the wrong audit data, the audit data quality is low. Therefore, block-chain technology can be combined with audit, which is because block-chain technology has decentralized and tamper-resistant attributes, which improves the quality and efficiency of big data audit forensics. In the audit process, the task of data conversion is delivered to professionals who know big data technology. Auditors review the converted data to ensure the quality of the data and ensure that the collected data is complete and effective.

#### *4.3 Talent Shortage*

In the audit industry, there is a serious shortage of compound talents who know both big data technology and audit professional knowledge, which leads to the failure to meet the requirements of high-quality development of the audit industry, making the audit work prone to the problems of low efficiency and low quality. Because the audit work needs to make use of the rich audit experience and good judgment ability, it is difficult for the big data technology majors to learn the professional audit knowledge. Similarly, for audit professionals, learning big data technology is very challenging. (Liu, 2022) But can't give up the training of audit talents, can start from two aspects, the first aspect, auditors can take "big data + audit" combination, and organize the auditors big data technology training, guide the thinking change, to cultivate compound talents with large data audit skills. Secondly, we should add the courses and class hours of big data technology, such as the application of web crawler software, visual analysis software, knowledge graph analysis software, etc. (Ma & Geng, 2023)

## **5. Future Development Trend of Big Data Qudit**

### *5.1 Combination of Artificial Intelligence and Big Data Audit*

Big data audit depends on auditors to know how to use big data technology to code, collect, process and analyze data, unable to realize the automation of the audit process, not only the audit efficiency is low, but also prone to operational mistakes. Therefore, it is necessary to introduce artificial intelligence. Because artificial intelligence can through machine learning, IPA technology to build the appropriate model, applied to the audit work, not only can realize the automation of the audit process, liberate the

auditors' hands, can also be advantageous to the auditors timely find potential problems in the audit process, and timely solve to reduce risk. In the audit industry, Deloitte has developed an audit robot named "Xiaoqinren", which realizes human-computer interaction and promotes the high-quality development of the audit industry.

### *5.2 The Combination of Block-Chain Technology and Big Data Audit*

Block-chain technology is a kind of decentralized, highly secure distributed ledger technology, has the characteristics of tamper-proof, ensure the security and authenticity of the audit data, to eliminate the risk of data is tampered with or forged, for a big data audit development provides a new direction, can effectively improve audit efficiency and ensure the accuracy of the audit. Block-chain technology application in the audit work mainly through the "block-chain + original invoice" deposit and traceability function, not only all data real-time login by time and place, provide complete audit traceability of context, can realize each transaction data on real-time chain and multiple validation, facilitate real-time audit, and block-chain technology can automatically perform many current audit procedures, simplify the audit process, ensure accuracy and credibility, reduce the audit risk. At the same time, block-chain technology has a smart contract function that can automate audit, and improve audit efficiency by reducing human intervention. (Lin, 2023)

### *5.3 The Combination of the Knowledge Graph and the Big Data Qudit*

Knowledge graph is a graph-based data structure, which can abstract entities, events, and concepts in the real world into nodes, and connect these nodes through relationships to form a huge knowledge network. In the audit work, the main method of using knowledge graph is to collect and gather all kinds of audit data, extract the entities, relationships and attributes in the data, use the knowledge graph to build the correlation between various business data, form the audit data relationship chain map, and assist the whole chain audit and cross-check audit of audit data. (Xu, Chen, Ge, etc., 2020) The combination of knowledge graph and big data audit can provide technical support for auditors to quickly form audit strategies and qualitative audit doubts. It can also realize visualization of complex business, assist to clarify business relations and data flow, and improve audit efficiency. In the audit practice, on the one hand, through the collection of relevant laws and regulations, systems to form a system database, combing the typical problems to form a problem database, and the digital key theme method database of audit supervision is established. On the other hand, the system knowledge map can be established to tie the audit data and into a normative format to achieve accurate positioning. At the same time, combined with the collected internal and external information of the audited units, a comprehensive map of suppliers is established for evaluation. In this way, it can not only promote the development of the audit industry, but also bring more comprehensive, intelligent and efficient audit services to enterprises.

## 6. Summary

With the rapid development of big data, in this paper, by examining the context and significance of the big data audit, the concept of big data audit is introduced; next, this paper expounds the big data audit technology, including data acquisition technology, data processing technology and data analysis technology; then, this paper describes the three problems in the application of big data technology are data security, data quality and talent shortage, and put forward the corresponding solution; finally, summarize the future development trend of big data audit, there are three main directions: the combination of artificial intelligence and big data audit, the combination of block-chain technology and big data audit, and the combination of knowledge graph and big data audit.

Big data can be conducive to the high-quality development of the audit industry, further promote the innovation of audit, and provide strong support. In the future, with the continuous development and improvement of big data technology, it is believed that the audit industry will usher in a broader development space and opportunities.

## Reference

- CAI, Y., & Li, Y. (2023). Impact of big data information technology on audit quality. *Journal of Hubei University of Economics (Humanities and Social Sciences edition)*, 20(09), 67-70. <https://doi.org/10.54691/bcpssh.v20i.2158>
- Chang, Q. J., & Li, L. L. (2023). Research on the application of web crawler technology in the acquisition and analysis of audit evidence—Take Kim Jong Da as an example. *Accounting Research*, 2023(07), 56-62.
- Chen, W. (2017). Big data audit: Current situation and Development. *Chinese CPA*, 2017(12), 77-81.
- Lin, Z. J. (2023). Research on the impact of blockchain technology on accounting audit. *Friends of Accounting*, 2023(20), 2-6.
- Liu, C. L. (2022). Analysis of the application of big data technology in the internal audit of railway enterprises. *Railway transportation and economy*, 44(05), 86-90+137.
- Liu, X., Niu, Y. F., & Tang, Z. H. (2016). Some thoughts on promoting the big data audit work. *Audit study*, 2016(05), 3-7.
- Ma, Y. Q., & Geng, G. Q. (2023). The dilemma and Countermeasures for the implementation of big data audit. *Financial management*, 2023(09), 64-66.
- Mu, H. S., & Nie, H. P. (2023). Water pollution audit study based on visualization technology. *Accounting and Communications*, 2023(17):125-130.
- Qin, R. S. (2014). Research on the impact of big data and cloud computing technology on audit. *Audit study*, 2014(06), 23-28.
- Wan, J. G., & Liu, R. Z. (2001). Audit data conversion technology (I). *China Audit*, 2001(04), 55-56.
- Wang, L., Bao, X., Wang, Y. etc. (2020). Audit data analysis and case application based on data mining algorithm. *Chinese CPA*, 2020(06), 101-105.

- Wang, Q. F., Luan, D., & Zhang, L. D. (2020). Research on the application of web crawler technology to obtain audit evidence—Take Asia-Pacific Industrial audit as an example. *Friends of the Accounting*, 2020(17), 131-136.
- Xu, C., & Chen, Y. (2021). Research on the application of Big Data Technology and Methods in audit supervision. *Quantitative, economic, technical and economic research*, 38(05), 135-153.
- Xu, C., Chen, Y., Ge, H. M. (2020). Research on audit technology based on Big data. *Journal of Electronics*, 48(05), 1003-1017.