

# *Research on Data Auditing Mode in Big Data Environment*

**Yingli Wang**

*Guangzhou Huali Science and Technology Vocational College, Guangzhou 511325, China.  
105624725@qq.com.*

**Keywords:** Cloud computing; Audit big data; Aggregation and decentralization; Audit data center; Audit hybrid cloud platform

**Abstract:** In the big data environment, the data center is a resource pool for the collection, processing, storage and sharing management of audit data. It is an important platform for data aggregation and decentralization. However, at present, the data centers in China are too scattered and regional, and it is difficult to implement auditing. Data integration and sharing between agencies and ministries is not conducive to the smooth progress of audit data aggregation and decentralization in the big data environment. In order to better play the "immune system" function and realize the modernization of national governance, this paper takes the data flow as the main research perspective, and proposes that the national audit should conform to the development of information technology, adopt advanced technology, integrate the national data center, and coordinate the central and provincial governments. The city integrates auditing big data resource pools in an all-round way, and builds a cloud platform for auditing big data with high-efficiency data processing technology--"cloud computing" as a carrier to improve the aggregation and dispersion efficiency of audit data streams.

## **1. Introduction**

Big data emerges as an important technological change when the new generation of information technology is integrated and innovated. Big data technology will also have an important impact on audit development. The "Opinions of the State Council on Strengthening the Audit Work" and the "Framework Opinions on Improving Some Major Issues in the Audit System" all propose ways to explore the use of big data technology in audit practice. In the traditional environment, the research of data-based audit mode is relatively clear, and the data-based audit mode in the big data environment remains to be studied. The data-based audit mode is an audit forensic model adopted after the audit object is transformed from the traditional paper account to the electronic data under the condition of acclimatization. In recent years, the information environment has continued to change, and audit targets have been moving toward big data. Faced with this change, the research of data-based audit mode should also be developed and promoted. Analyzing and summarizing the data-based auditing model in a big data environment will help improve auditing capabilities and help auditors better perform big data auditing.

## 2. Review of related theories

### 2.1 Auditing Big Data

The big data of national audit, from the source point of view, it is a cross-domain data model based on audit objects. The collection of audit data is based on a comprehensive analysis of the data in the information system of the unit under review [1-2]. From the perspective of data structure, the data in the big data environment presents a variety of features such as unstructured and semi-structured. These data seem to be chaotic, but the essence is rich in resources, and the key wealth of auditing big data should be tapped. To contribute to the innovation of China's audit acclimatization. It has a large amount of data, diversified data composition, high data processing time, and large data potential value.

### 2.2 Cloud Computing

The idea of the cloud computing center is to uniformly manage and schedule computer resources, and build it into a computing resource pool that can realize on-demand services through the network. Users are more focused on applications and conveniently access computing and storage anytime, anywhere according to specific business needs [3]. Resources to effectively achieve work goals. Simply put, in the cloud computing market, customers spend money on computing, storage, and services rather than computers, storage, and servers. Some scholars predict that the future cloud computing market is just like the hydro power market we consume, and it can be realized anytime and anywhere as long as there is demand.

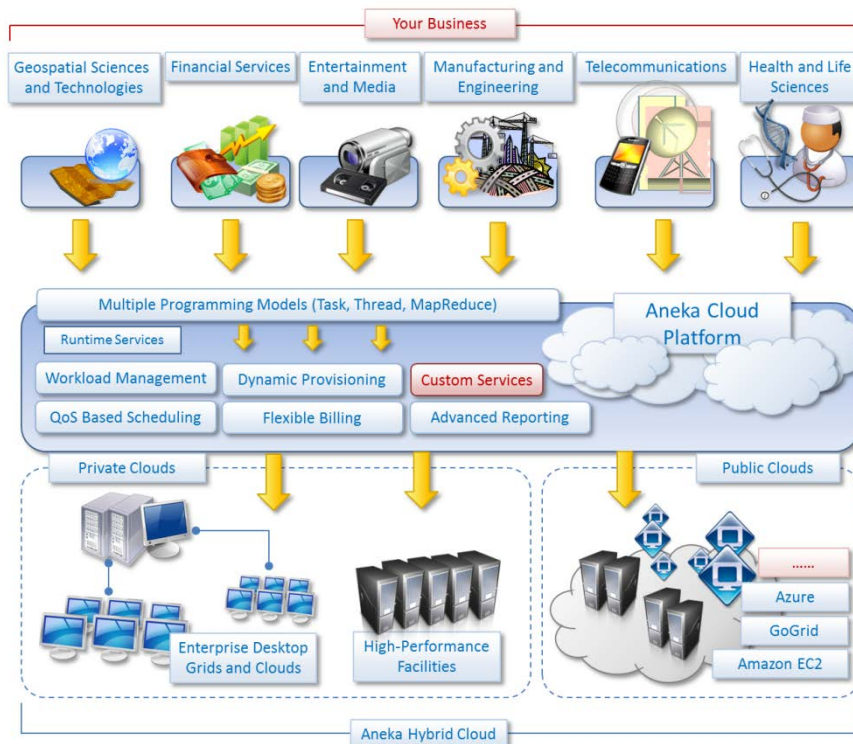


Fig.1 Cloud computing model

### 2.3 Data Mining

Data mining refers to the extraction of implicit, prior unknown and potentially useful knowledge

from a large amount of data. Using the advantages of data mining technology, it can find some knowledge beyond the audit experience, which is a useful supplement to the current computer audit method. To put it simply, it is the process of extracting knowledge and information that cannot be discovered but hides value from massive, random, fuzzy, incomplete, and noisy data [4].

Data mining is a specific step in the knowledge discovery process, which includes many data processing stages. In general, it can be roughly divided into three stages: data preparation, data mining, and result analysis. The data preparation phase specifically includes data collection, selection, processing, and conversion. In short, the fundamental task of data mining is to perform format conversion, cleaning and other processing on the data from the vast raw data through data selection and data collection. Finally, the formatted data set is used. The foundation makes data mining and model building for the target number, and draws knowledge and patterns, and makes an easy-to-understand analysis of the knot dance. Figure 2 shows the general process of data mining.

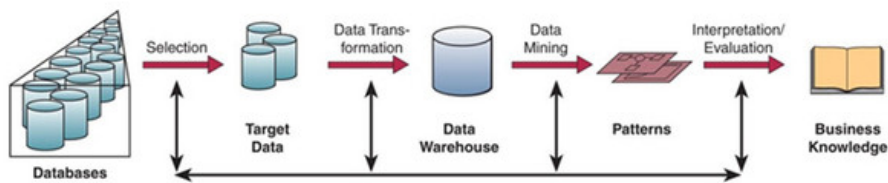


Fig.2 Data mining process

### 3. Path Design of Data Auditing Mode in Big Data Environment

Therefore, under the impact of big data, in order to ensure the smooth and orderly operation of the audit big data stream, the first task is to organize and upgrade the original data center. The basic data resources accumulated during the construction of the first and second phases of the gold phase will be integrated and planned, the construction of the audit big data resource pool will be carried out, and the cloud computing with powerful data processing capability will be used as the infrastructure to build the audit big data platform. Through the platform, the audit data resources stored by the auditing organizations of each locality are aggregated, and the platform is used for unified and decentralized management to realize the collaborative promotion and data sharing and separability among government departments, and to open up information channels of internal auditing organs and external departments of the national audit to form vertical And an information sharing network that runs through the horizontal line. It can be said that the audit big data platform is the basic guarantee for the auditing of big data stream aggregation and dispersion, and the latter is the core idea of the former.

In the design of the audit big data platform in the context of national governance, this paper focuses on the data flow aggregation and decentralization, selects the data center as the entry point, and proposes the theoretical idea of the construction of the audit big data platform based on the current status of the center and the challenges faced. Firstly, considering the impact of current information technology on audit data flow from the perspective of national governance and resource allocation, the data center is selected as the representative, and the integration of audit big data resource pool is proposed, and the resource pool framework is discussed from the two directions of aggregation and decentralization [5]. Cloud computing build platform. The platform application data mining and visualization technology conducts audit practice research, and analyzes the data aggregation application of new technology in networked audit, and the data decentralized security supervision system for cloud security.

### 3.1 Data mining-based audit data aggregation model

According to the technical method of data mining (the principle of data mining in the auditing business), this paper gives the working model of data mining technology in the audit big data cloud platform, as shown in Figure 3.

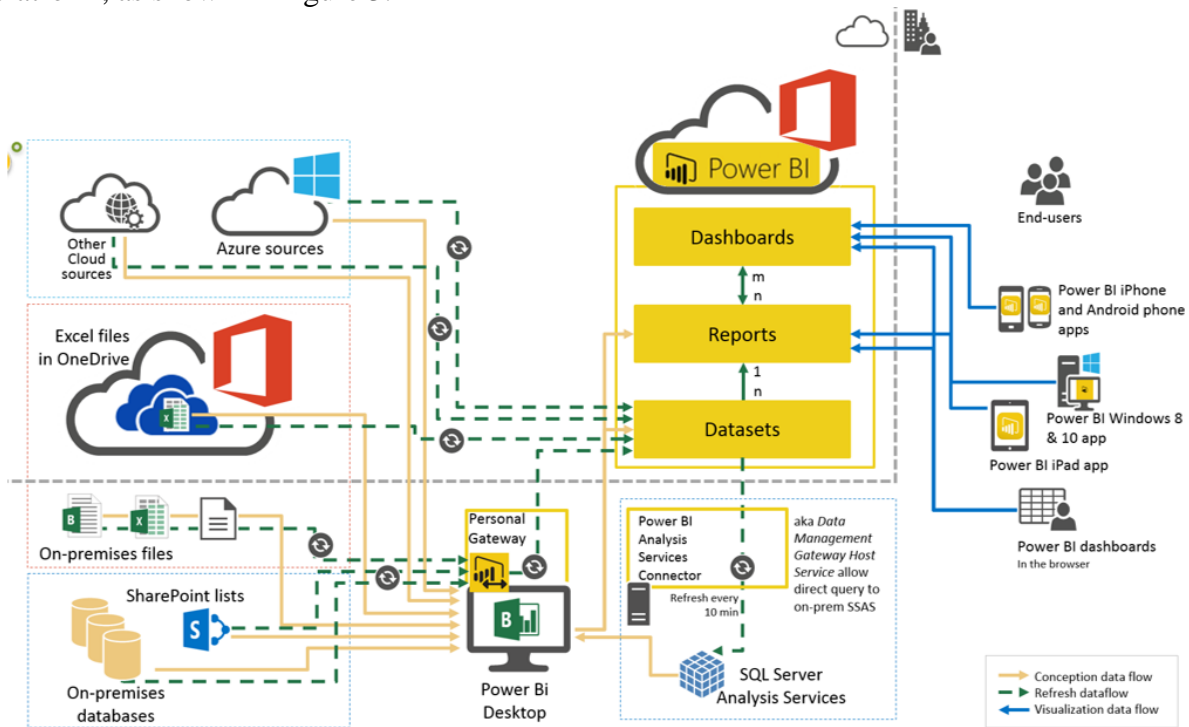


Fig.3 Working model of data mining technology in auditing big data aggregation

The auditor sends a data collection request to the cloud according to the audit task, and the cloud platform application data mining technology processes the audit task data request. The data processing process in the middle-dashed box in the figure is the data mining framework, and the whole process is completed on the cloud platform.

(1) Collect original audit data. The cloud collects the original data of the audited unit according to the audit task target, and imports the collected data into the audit database in the cloud platform infrastructure layer IaaS.

(2) Preachment. The cloud needs to be cleaned and pre-production for the collected data. It includes unifying data in different formats, sorting into a format recognizable by the data mining algorithm, and filtering data that is not related to the audit task. The processed data is the processing object of data mining technology.

(3) Mining analysis. According to the industry nature, business characteristics and data patterns of the audited units provided by the auditors, the cloud adopts different data mining algorithms, such as association rule analysis discovery, sequence pattern mining, exception data patterns or cluster analysis. Appropriate algorithms are used to obtain patterns implicit in the data, and these patterns are compared to patterns in the rule base to identify suspicious data. At the same time, the new data schema is returned to the association rule base for later similar audit service calls.

(4) Audit processing. For the results of (3) analysis, the auditor calls the expert experience library of the platform service layer PaaS, the classic case database, and the hooking model for comprehensive analysis.

(5) Collect abnormal data. Collect the abnormal data in (3) and (4) for the auditor to analyze and

verify the suspected data in (6).

(6) Audit analysis. The data analysis results and abnormal data will be excavated. The distributed interface of the cloud platform is transmitted to the auditing end, and the auditor conducts further investigations, draws conclusions, and makes recommendations and opinions [6].

### **3.2 Data flow aggregation process under networked audit**

(1) Data collection. Refers to the electronic data collection of the audited entity by the audit department. Under the current information conditions, the process is generally implemented by means of “data acquisition front-end machine”, which is arranged by the auditing authority in the database server of the audited unit, and realizes the audit under the networked audit mode through the data acquisition software in the front-end machine. Data collection work.

(2) Data transmission. The collected electronic audit data is transmitted to the audit department through the network. The transmission between the two depends on the network connection, and the connection can be made by a dedicated line or by means of a wide area network. For the geographical distance between the audit unit and the audited entity, the former form is adopted, otherwise the latter is adopted.

(3) Data storage. The collected electronic data is stored in the data storage system of the audit terminal. The amount of data storage and security depends on the physical conditions of the audit department's system configuration. In the case of increasing data volume, the audit terminal must provide enough servers to ensure data storage.

(4) Data analysis. Perform calculation and analysis on the data to provide audit trails.

## **4. Exploration and effect evaluation of data-based audit mode under big data environment**

### **4.1 Exploration of Data Auditing Mode in Big Data Environment**

The data-based auditing model in the big data environment has changed. First, data collection has become more diverse. The Provincial Audit Office collects the provincial audit data volume 8 T B and the city and county audit data volume by constructing the financial database of the audited entity, the database of auditing results, the third-party information database, and the Internet public opinion database. TB, the source of accumulated big data audits. The second is to pay more attention to data quality inspection and data processing [7]. Collecting data is not an end in order to use the data and to address issues such as data consistency, data conversion, and data filtering. Third, the data analysis method is more in-depth, through the business data and financial data, unit data and industry data, and cross-industry, cross-domain data comprehensive comparison analysis, research and development of audit results analysis system, the structure of unstructured text data Process and analyze the results of the audit. Fourth, the audit process can be optimized. According to the frequency and audit results of the audited units in the past, the auditing of the audited units is summarized to provide a basis for the preparation of the annual project plan. Big data analysis is no longer limited to specific audit projects and can be applied to all aspects of the audit process. The improved data-based auditing model has achieved outstanding audit results. The budget execution audit can only audit more than 10 units a year, and can basically achieve full coverage of provincial budget units. In the audit of public housing in 2004, 2,272 suspected problems were found, and the amount of violations of various disciplines was investigated and dealt with. With an investment of 1.2 billion yuan, the audit results have improved significantly. At the same time, the improved data-based audit model has created conditions and opportunities for the use of audit results [8].



## 4.2 Evaluation of the effectiveness of the data audit mode in the big data environment

As an important method of performance evaluation, the Balanced Scorecard can be used to evaluate the effectiveness of the data-based audit mode in a big data environment. For enterprises, the Balanced Scorecard is a performance management system that implements the company's strategy into operational metrics and target values from four perspectives: finance, customer, internal operation, learning and growth. For the audit, the above-mentioned perspective can be converted into “effects”, that is, the task of maximizing the budget and capital efficiency while completing the audit task”, that is, improving the quality of the report and enhancing the function of serving the audited entity, “efficiency” means improving the internal Operate and optimize the audit business process, “innovation” is to create a learning and growth environment and then innovative ways. By setting evaluation indicators from the aspects of effect, service, efficiency and innovation, based on the accumulation of test data, it can effectively evaluate the big data environment. The effect of the data audit mode.

## 5. Conclusion

This paper analyzes the current status of the audit data center accumulated in the existing dramatization process, and analyzes the limitations of its existing data storage facilities, such as large input cost, limited data storage, low data sharing, and lack of data value mining. It is difficult to meet the modernization of national governance. New requirements and national audit “immune function” requirements. It is urgent to coordinate and integrate distributed data resources to form an audit big data resource pool. In view of the advantages of cloud technology for data processing, such as high efficiency, high value and low cost, the cloud platform is proposed as a resource pool. Audit big data as the core platform. And from the perspective of audit data flow, we design an audit big data platform with data aggregation specification system, data mining technology, data dispersion platform, visualization technology and data security supervision system. Finally, suggestions for platform development are proposed, including the choice of hybrid cloud as the platform infrastructure model, standardizing data security management, ensuring continuous and efficient service of cloud services, and strengthening auditor capacity development.

## References

- [1] Cheng Ping, Bai Wei, et al. *Big Data Auditing Based on COBIT Standard in Cloud Accounting Environment*. *Friends of Accounting*, Vol. 4(2016) No.12, p. 125-128.
- [2] Yang Kaiqi. *Discussion on the Characteristics and Realization of Big Data Audit—Taking the Audit Commission's Audit Results of 2012 PetroChina as an Example*. *Finance and Economics (Academic Edition)*, Vol. 15(2015) No.12, p. 315-318.
- [3] Cheng Ping, Bai Wei. *Research on Big Data Audit Based on Financial Shared Service Model*. *Chinese Certified Public Accountant*, Vol.5(2016) No.18, p. 84-87.
- [4] Huang Jianghai. *Analysis of the Basic Characteristics of Big Data Auditing Evidence*. *Audit Monthly*, Vol. 8(2016) No.19, p. 6-9.
- [5] Chen Wei, Wu Zheng, Liu Hai. *Research and Implementation of Big Data Audit Method Based on Benford's Law*. *Chinese Certified Public Accountant*, Vol. 9(2017) No.33, p. 80-84.
- [6] Fu Xin. *On the Risks and Opportunities Brought by Big Data to Audit Work*. *China Market*, Vol. 1(2016) No.12, p. 44-47.
- [7] Jia Jia, Liu Jianli. *Thinking on Auditing Work Mode under Big Data Based on Audit Independence*, Vol. 4(2016) No.12, p. 61-67.
- [8] Tang Lin, Fu Dajie. *Connotation Characteristics, Realistic Dilemma and Development Countermeasures of Big Data Auditing*. *Western Finance and Accounting*, Vol. 7(2017) No.17, p. 70-73.