

Research on Security Auditing Technology of Chemical Enterprises for Big Data

Zhou Fuping, Fang Mingyong

Guangzhou College of Technology and Business, Guangdong, China

Keywords: big data; security auditing technology; chemical enterprises

Abstract: The emergence of the continuous progress of information technology and data technology, prompting the internal audit of our country whether in theory or concrete operation practice have great progress, at the same time the use of big data technology can also promote the development of national economy. Then, based on the reference of domestic and foreign literatures, this paper gives the framework and content of the research, introduces the research ideas and methods, innovation and deficiencies. Next, the status of the development of internal audit in the big data environment is described. In recent years, some of the advanced Internet companies and large state-owned enterprises have recognized the importance of big data technology, and actively explore the use of information technology, and has been applied to the internal audit to the large data environment under development related to the work of internal audit theory research provides a good source of information. On this basis, this paper analyzes the concept and characteristics of big data points out the difficulties faced by the development of internal audit in the big data environment. and puts forward the corresponding improvement measures. How under the environment of big data further development of internal audit. Making big data of internal audit has become the leading force of audit technology in our country. Play its role with respect to other auditing methods and incomparable. has already become the industry to discuss a new topic. I hope these suggestions and opinions can contribute to the healthy and orderly development of our country's economy.

1. Introduction

In recent years, the explosive growth of information technology, the rapid development of large data. in all areas of society have formed a certain influence. The society of all walks of life and people's daily life are gradually applied to large data technology. In large data environment, more and more enterprises use big data technology development of internal audit, which to the development of enterprises brought some opportunities, but also can not ignore the with the technology of the data to the data and information security of a series of problems, which has become the problem of large data environment development some of internal audit can not be ignored. This paper first introduces the background and significance of the subject. Western information audit direction has been developed to a certain stage of maturity, the application of large data technology in the internal audit work has become an important breakthrough in the development of the western developed countries. In our country with the arrival of the era of big data, and gradually big data internal auditing theory continues to advance, scholars has in terms of content, method and technology made the induction and the summary, and put forward their own proposals.

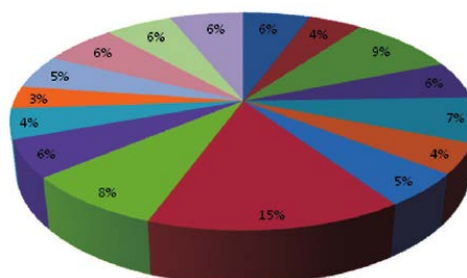


Figure. 1 Chemical enterprise process security audit factor scores.

Security audit is the audit institutions and their staff according to national laws and regulations and other requirements of the audited enterprise safety responsibility to perform evaluation and verification, to promote the enterprise to fulfill the main responsibility for production safety, improve production safety, and to make it comply with the safety requirements of the development of an independent oversight activities “}1. At present, many fields of safety audit have been researched and applied in many fields, including information field, road area and aviation field. However, there are few researches and applications about process safety audit in chemical enterprises.



Figure. 2 Data access mode

Recently, the rapid development of information technology has been bringing an unprecedented revolution to human life. Among all the new information technologies, cloud computing and big data are typical representatives, there is a close relationship between them: the framework of cloud computing provides the basic platform of information storage and data mining process for big data, big data technology promotes the rapid development and improvement of cloud computing. However, the security problem involved in each phase of the whole chain of big data based on cloud computing becomes more and more highlighted. In this paper, we do the further researches on the secure storage mechanisms for big data based on cloud computing from three aspects, which are the integrity auditing of big data, the secure retrieve of big data, and the secure sharing of big data.

2. The big data technology

As an application system that needs to collect, analyze, and present massive data, the security audit system naturally has a strong demand for a technology capable of processing massive amounts of heterogeneous data with high performance. The emergence and development of big data-related technologies has provided new technical support for adapting to the needs of network development. The use of big data-related technologies to transform or develop new security auditing systems, compared to traditional security auditing systems, can be significantly improved in the following areas:

- (1) The performance bottleneck of the system in acquisition, retrieval, analysis and storage will be broken;
- (2) It can better deal with structured data and unstructured data at the same time;
- (3) It is possible to make use of the relevant algorithms and models of big data analysis to perform more extensive and in-depth analysis of historical data, and to dig out more valuable information for users from the massive data;
- (4) As the system is deployed on cheaper hardware devices, the procurement and maintenance costs are even lower, and the level of expansion can be flexibly implemented according to the development of user network applications, effectively reducing the total cost of ownership of the system and increasing flexibility. Better protect user investment.

Table 1 Properties of raw materials

Primary Key	Column A1	Column A2	Column B1	Column B2
Id-1	ValueA 11	Null	Null	Null
Id-2	ValueA 12	Null	Null	null
Id-2	ValueA 13	Null	Null	Null

Security audit is an important feature for database security. It is widely used in relational databases. For example, Oracle 9i and above versions provide users with audit. The audit module can monitor and record users' operations on the database. With the rapid development of Internet, the size of data grows every day, which makes relational databases unable to process big data. Non-relational databases have dramatically risen in popularity. These databases are commonly known as data base. Big databases provide us with highly available and highly scalable data storage solutions.

But these databases are not initially designed with considering security as an important feature. Consumers have to protect these databases themselves by using third party tools and services. Data security is probably one of the most difficult challenges faced by big databases. Therefore, it is necessary to design a security audit system for big databases.

The first stage will go through the exploration phase. It includes research and exploration in theory and exploration in practice. In the research work, on the one hand, it is necessary to strengthen the research on the theoretical field, start with the cross-domain analysis of big data technology and computer means, explore the similarities and development path, and form a complete theoretical system research on big data. On the other hand, we must also pay attention to the exploration of practice. We should gradually apply big data technology in real internal audit work. We must combine the experiences of practitioners and actual cases to preliminarily summarize the internal audit work plan that is in line with the development of big data environment.

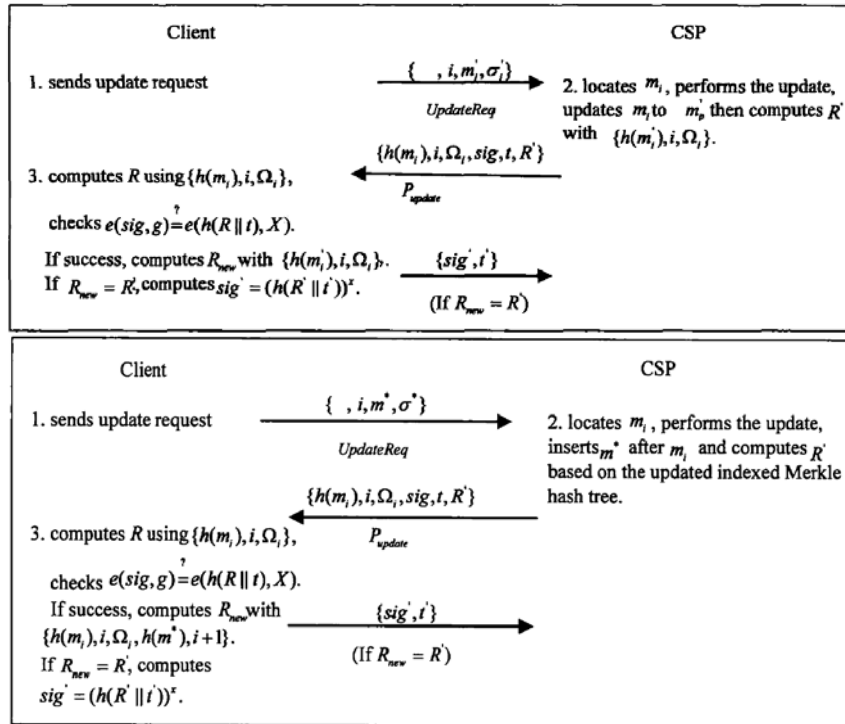


Figure. 3 Data change and insertion process

The second stage has to go through the trial stage. Will be formed in the exploration of theory and practice achievements of application in the field of internal audit, the technology and application of big data in some of the distinctive characteristics of the subsystem in the first trial, and before the audit results compared to the analysis of the development, and the lack of experience to make up the defect extraction, correction of unfavorable factors in the application of this post, to reach the mature stage gradually.

The third stage reached the stage of promotion. It is possible to apply the big data internal auditing platform initially applied in the second stage to the entire internal auditing system and connect each related subsystem with the internal auditing platform. In this stage, facing the overall promotion of big data in internal audit, we must pay more attention to the efficiency and quality of data use, making its application more recognized by enterprises and institutions. At the same time, in the process of promotion, computer technology and software means also need to be continuously integrated with the concept of big data, repeatedly correcting the resistance that may be encountered in the convergence of organizational systems, human and material resources, and achieve a balanced effect.

3. Design of security audit system based on big data technology

Security audit system based on big data technology is based on the traditional security audit

products, using a variety of technical architecture and implementation of big data, with the traditional security audit function and support the massive data in real time analysis and historical data collection, data mining and security audit system can be extended according to the level of.

The security audit system based on big data technology needs huge amount of data and lots of unstructured data. Therefore, in every process, data processing is different from traditional audit products. The following is a brief introduction to how to use large data related technologies to solve these problems.

3.1 Audit data collection and storage of large amount of data

When the amount of audit data to be collected is large, the auditing system's acquisition performance will also face greater challenges, which may result in data loss, process congestion, slow system response, or stop responding. Therefore, in a large data volume environment, a distributed acquisition probe deployment method may be adopted, and one or more data acquisition probes may be deployed according to the actual amount of data to be collected. After obtaining the data, the probe may use the TCP protocol and analysis. The platform establishes a connection and can compress and encrypt the data to be transmitted, ensuring the performance of the data collection end and reducing the occupation of the network bandwidth. The data redundancy module is not required, but if there is not enough dimension information in the log data, or if you need to increase the dimension frequently, you need to define the data redundancy module. Redundancy dimension definer defines the dimension information and source (database, file, memory, etc.) that need to be redundant, and specify the extension mode to write information in the data log. In the data storage, some relatively fast memory NoSQL can be used to redundance the original data, and as many nodes as possible can be used for parallel redundancy; or it is also possible to execute a batch Map in Hadoop to convert the data format.

Key codes:

```
Configuration conf=HBaseConfiguration.createU;  
HTable table=new HTable(conf, "user");  
Put put=new Put(Bytes.toBytes("rk001"));  
put.add(Bytes.toBytes("fam"),Bytes.toBytes(" coll"),Bytes.toBytes("valuel"));  
put.add(Bytes.toBytes("fam"),Bytes.toBytes("co12"), Bytes.toBytes("value2"));  
  
table.put(put);  
  
table.close();
```

3.2 Data normalization and correlation analysis

After collecting audit data, we first need to classify these massive data, normalize them according to certain standards, and do some simple cleaning and preprocessing of data. These are consistent with the process of event handling taken by traditional audit products. The difference is that the amount of data to be processed is very large. The amount of data may reach 5 million EPS (Events Per Second, the number of events per second), and the peak value may reach 10 million EPS. This performance requirement is not able to be done by traditional audit products and must depend on the ability of large data clusters to process concurrency. The general use of traditional products in memory database real-time correlation analysis method, due to the stand-alone version of memory resources and the efficiency of the SQL statement event when the quantity is large, the concurrency rule engine and processing capacity are greatly decreased, resulting in the rule engine can not detect the abnormal. The distributed computing framework based on big data cluster is combined with the complex event processing flow based on big data cluster as a real-time rule analysis engine. It can run multiple rules efficiently and detect abnormal events in real time. Storm+Esper can be used in concrete implementation. Storm is very suitable for real-time statistical calculation of mass data, and can quickly feedback the results of statistics. The Storm framework uses a strict and efficient event processing process to ensure the accuracy of the data in the operation, and provides a variety of real-time statistical interfaces for use. We use Storm's memory data iterative computing framework to do association analysis operation, and use the complex event

processing function implemented by Esper as a real-time correlation analysis engine, which can improve the real-time and accuracy of system association analysis.

Key codes:

```
Configuration conf=HBaseConfiguration.createU;
HTable table=new HTable(conf, "user");
Get get=new Get(Bytes.toBytes("rk001"));
get.addColumn(Bytes.toBytes("fam"), Bytes.toBytes("col1"));
Result result=table.get(get);
byte[] value=result.getValue(Bytes.toBytes("fam"), Bytes.toBytes("col1"));
System.out.println("value:" + Bytes.toString(value));
table.close();
```

3.3 Statistical analysis of historical data

Off-line statistical analysis of the massive data stored in the cluster is another important problem that the audit system needs to solve. For most applications where the feedback time requirements are less stringent, such as offline statistical analysis, machine learning, reverse index calculation of the search engine, and calculation by the recommendation engine, offline analysis should be used. In the face of massive amounts of data, the ETL tools of traditional audit products often fail. The main reason is that the overhead of data format conversion is too large, and it cannot meet the processing requirements of massive data in terms of performance. The amount of computation of this data and the scope of data retrieval is that traditional audit products cannot be implemented based on relational databases. The off-line statistical analysis function of the audit system based on big data technology mainly uses distributed computing clusters to perform general analysis and classification and aggregation of the mass data stored therein, which can satisfy most common analysis needs. The Hive+Hbase framework is deployed on top of Hadoop. Hive and Hbase have different special features.

Key codes:

```
Configuration conf=HBaseConfiguration.createU;
HTable table=new HTable(conf, "user");
Delete del=new Delete(Bytes.toBytes("rk001"));
del.deleteColumn(Bytes.toBytes("fam"), Bytes.toBytes("col2"));
table.delete(del);
table.close();
```

Hive is high-latency, structured, and analysis-oriented. Hbase is low-latency, unstructured, and program-oriented. Hive data warehouse is highly delayed on Hadoop. Hive's Hbase integration is to use Hbase features, use big database to simplify the writing of Map/Reduce tasks, and use Hive and Hbase to complement each other to speed up the calculation of event analysis results. After the command is parsed into Map-Reduce by the core module and submitted to the Hadoop cluster, the report is generated for display in the report center, so that the data stored on the HDFS is analyzed offline.

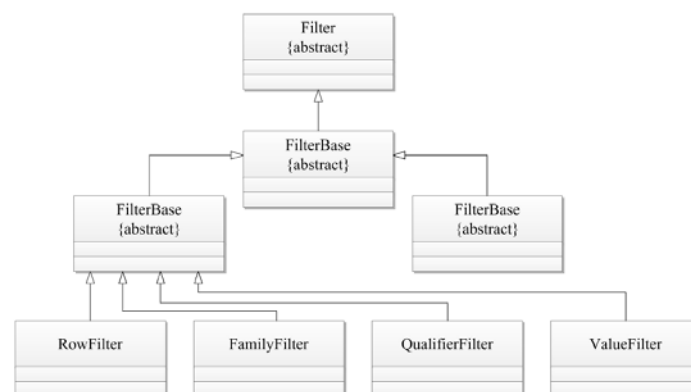


Figure. 4 Filter hierarchy.

3.4 Data mining

Data mining is used to analyze historical data, but with the previous statistics and analysis process is different, the data mining of large data environment is not what the predefined theme, mainly in the existing data of various algorithms based on the above calculation, so as to predict the effect, and further analysis high level data demand. At the same time, a variety of mining algorithms can be used to achieve a certain degree of model mining or the operation of the established audit analysis model. The methods and contents of these data mining are also the lack of traditional audit products.

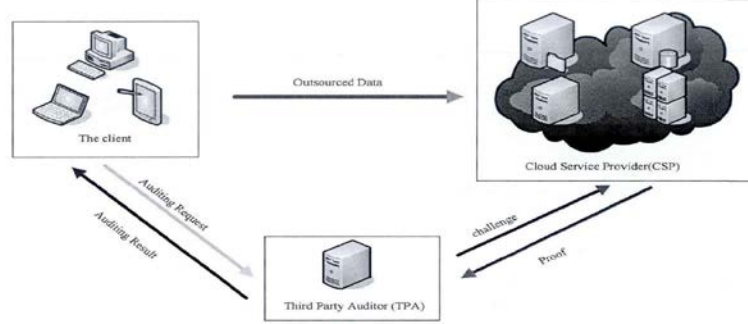


Figure. 5 System model

Through data mining technology, some more hidden network attacks, illegal access and misallocation of system configuration can be found. At present, there are many open data mining algorithms, which have no difficulty in technology implementation, but the algorithm and information security industry model still need a long time to train and debug.

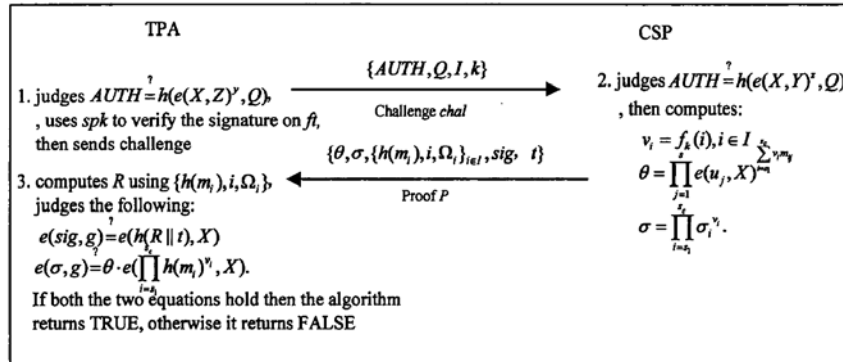


Figure. 6 Integrity audit process

All walks of life have different analytical models for big data applications. Faced with different development environments, model building should fully consider their own characteristics. In general, there are three types of commonly used internal audit data analysis: query-type analysis, multi-dimensional analysis, and mining-type analysis. The content that can be covered by big data is multi-faceted and multi-faceted. Therefore, when solving internal audit questions, the requirements for industry personnel are higher. In addition to data analysis experts, there may be more social areas, such as ecology, geography, mathematics, and statistics. Many other talents participated. In addition to the need for an audit function, the big data internal audit analysis model is more important in forecasting functions, looking for patterns between data and using models to analyze the magnitude of event risks. The auditors used their own auditing experience to make judgments and forecasts in their past work. Today, they rely on audit data to judge. This change is the biggest contribution of the internal audit analysis platform of big data.

4. Conclusion

According to the characteristics of big data, we design general audit rules. Audit rules are divided into two parts, namely audit level and audit condition. Audit level contains coarse-grained

audit, fine-grained audit and data audit. Audit condition contains table condition, condition and data condition. Administrator can select appropriate security level according to the safety requirements.

In this paper, we analyze general architecture and security mechanism of one of the most popular big databases. In our security audit system, according to the working location of coprocessor, we define audit aspect by combination. We define a series of hook functions in this aspect. The results show that the security audit system works well. The system also support combination of audit level and audit condition, which reflects the flexibility of the audit policy.

Acknowledgements

2018 Guangdong Province Higher Education Teaching Reform Project “Managerial & Accounting Talents Pattern Research based on Big Accounting Perspective”, 2018 Guangzhou College of Technology and Business Quality Engineering Project(ZL20181131)2019 Guangzhou College of Technology and Business” Practical Research on Construction of Accounting Experimental Training Room and Training of Talents”.

References

- [1] Uke N J, Uke N J. Efficient Security Framework for Sensitive Data Sharing and Privacy Preserving on Big-Data and Cloud Platforms[C]// International Conference on Internet of Things and Cloud Computing. ACM, 2016, pp. 38-42.
- [2] Li Z, Shao X, Duan S, et al. The Energy Audit and Energy-saving Benefits Analysis for One Chemical Enterprise[J]. Guangdong Chemical Industry, 2016, pp. 45-65.
- [3] Yang G, Liu X, Liu Z. The Example of Cleaner Production Audit in a Phosphorus Chemical Enterprise [J]. Shandong Chemical Industry, 2015, pp. 15-19.
- [4] Wang X, Zhang J, Wang M, et al. CDCAS: A Novel Cloud Data Center Security Auditing System[C]// IEEE International Conference on Services Computing. IEEE, 2014, pp. 605-612.
- [5] Alomari S, Alghamdi M, Alotaibi F S. Enhancing Big Data Auditing[J]. 2018, 11(1):90.
- [6] Zhang Y, Wang F, Fang Y. The Information Construction of Enterprises' Archives in Big Data Era[J]. Journal of Beijing University of Chemical Technology, 2014, pp. 007.
- [7] Gowrigolla B, Sivaji S, Masillamani M R. Design and auditing of Cloud computing security[C]// International Conference on Information and Automation for Sustainability. IEEE, 2010, pp. 292-297.
- [8] Smith P. Real-time auditing of covert data marks: US, US 20070091377 A1[P], 2007.
- [9] Mahzan N. IT Auditing Activities of Public Sector Auditors[J]. Canadian Journal of Chemical Engineering, 2011, 5(5), pp. 1551-1563.
- [10] Gaganis C, Zopounidis C. Multicriteria decision support methodologies for auditing decisions, Multicriteria Decision Support Methodologies for Auditing Decisions[M]// Encyclopedia of Optimization. Springer US, 2009, pp. 2362-2372.