Big Data Security Storage Based on Hybrid Large-scale Database

Xuemei Peng¹,*, Jie Hu¹
¹Nanchang Institute of Technology, Nanchang 330044, Jiangxi, China
*Corresponding author E-mail: pxmei712@nit.edu.cn

Abstract. With the development and progress of network technology, hidden dangers of network security have also become a threat to network storage systems. This research mainly discusses the secure storage of big data based on hybrid large-scale databases. The user sends an authentication request to the SSP, and the SSP verifies the user's time stamp to ensure that it is not a replayed message, and checks whether the user ID is legal. The virtualization layer logically divides the hardware devices at the facility layer through virtualization technology, provides unified services to the outside world, and supports the operation of the platform layer; the platform layer builds a file system and database to provide an operating environment for the operation layer; the operation layer provides users with a big data environment resource access behavior and platform traffic are reasonably allocated and managed, platform business deployment, etc.; the application layer provides users with big data computing services. During the file download process, the big data storage module is relatively fast (300-600k/s), while the speed of Seccloud and invisible cloud is relatively slow (200-400k/s). This research helps to improve the data access efficiency of the big data secure storage module.

Keywords: Big Data, Hybrid Large-Scale Database, Secure Storage, Computing Service

1. Introduction
With the rapid development of big data computing technology, software technology and architecture have undergone significant changes in the big data environment, and users have begun to migrate systems and data to the big data environment to meet the on-demand resource acquisition, load balancing and disaster tolerance capabilities and other needs. However, big data platforms are not safe. Due to their openness, big data platforms face various threats such as account hijacking, data leakage, and permanent data loss.

Big data services provide external APIs to facilitate the IT team to manage and call big data resources. At the same time, the risks they face are also greatly increased. Therefore, how to ensure the security of data storage on big data platforms is an urgent problem that needs to be solved at...
present\textsuperscript{[1-2]}\. Big data secure storage is the product of the extension and derivative development of big data computing concept. With the development of big data computing, it continues to mature, and it is an indispensable part of big data computing\textsuperscript{[3-4]}. Because big data computing often has to deal with large-scale data sets, in order to manage data safely and effectively, big data security storage is bound to be used\textsuperscript{[5-6]}\. The big data secure storage system is the most important basic component of big data computing\textsuperscript{[7-8]}\. However, the current big data security storage system is no longer just an auxiliary system for big data calculation, but access storage resources through any network device\textsuperscript{[9-10]}\. 

There are two types of deficiencies in big data service providers: one is that privacy is leaked due to low attention to private data; the other is that the system does not guarantee complete reliability and stability. In terms of the secure transmission of big data, first of all, a security technology for preprocessing data during big data collection is proposed, that is, the data from each terminal to the central station needs to be authenticated to ensure that the collection of big data is collected during the collection process. The received data are all valid data, which greatly reduces the difficulty of analyzing the collected data, and at the same time ensures that a large amount of illegal data will not impact the communication management site, reduce the storage equipment overhead of the communication management site, and reduce the equipment burden of the communication management site.

2. Big Data Secure Storage

2.1 Data Secure Transmission

In order to ensure that the XML configuration file is not tampered with during transmission, and to verify the identity of the configurator for S and ensure the security of the XML encrypted configuration file stored on the platform, the identity of the configurator needs to be signed. When the alliance’s business data security puts forward different requirements and needs to modify the configuration file, the platform needs to verify the information of the configurator. Based on the above security needs, the objects involved are defined as follows. Use $Ab(Cig_{file}(L_i))$ to represent the digital digest of the XML encrypted configuration file, use $Ae(Cig_{file}(L_i))$ to represent the encrypted XML file digest, and $Ab User$ to represent the digital digest of the configurator’s information.

$$Ab(Cig_{file}(L_i)) = DA(Cig_{file}(L_i))$$

$$Ae(Cig_{file}(L_i)) = DM(Ab(Cig_{file}(L_i)), SK(C_v))$$

$$Ab User = DA(UserInfo)$$

Represents the summary and encryption of the encrypted configuration file, and the process of summarizing the configuration information.

2.2 Data Security Access

In order to ensure that the security access control module has minimal intrusion to other modules and achieves loose coupling, all interfaces that need to control access are marked with annotations, and their specific permission registration information is read from the annotations through reflection during the system scan stage and organized into kv format. Stored in the cache, in order to ensure that the impact of the module on the efficiency of the user's use of the system access interface is as small as possible, the specific permission information is stored in the local cache, and at the same time, a copy is stored in redis and persisted in the way of AOF for single Read the information directly after the server is down and restart to avoid the overhead caused by repeated scanning and registration. In order not to invade the existing code, the dynamic proxy is used to complete the function addition of the
interface that needs access control, and the aop interception is used to complete the authorization check every time the interface is called externally or internally, so that the premise of not affecting the code of other modules is realized completed the security access control function. The user’s direct experience of file service performance is reflected in the page response time. The shorter the page response time, the better the service system performance; the higher the throughput rate, the greater the number of concurrent users that the web server can withstand, and the service system better the performance.

3. Big Data Secure Storage Experiment

3.1 Public Key-Based Scheme
First, the user sends an authentication request to the SSP. The SSP verifies the user's time stamp ts to ensure that it is not a replayed message, and checks whether the user ID is legal. After verification, the SSP generates a random token and sends it to the user. The user also needs to verify the SSP time stamp, and then use token to send operation request to SN. All messages use signatures to ensure integrity and non-repudiation.

3.2 Data Protection Module
The facility layer provides basic hardware equipment for the system, including network, storage, and server hosts; the virtualization layer logically divides the hardware equipment at the facility layer through virtualization technology, provides unified services to the outside world, and supports the operation of the platform layer; the platform layer builds files system and database provide the operating environment for the operating layer; the operating layer reasonably allocates and manages the user's resource access behavior and platform traffic in the big data environment, and performs platform business deployment; the application layer provides users with big data computing services. Application server parameters are shown in Table 1.

| Name             | Parameter                        |
|------------------|----------------------------------|
| Device model     | ThinkPad E480 (20KNA038CD)       |
| operating system | CentOS Linux 6.7 x86 64          |
| CPU              | i5-7200U 2 5GHz                  |
| RAM              | 8GB 2133MHz L PDDR3             |
| Graphics card    | AMD Radeon RX5502GB GDDR5       |

4. Big Data Secure Storage Analysis

4.1 Performance Comparative Analysis
The performance comparison analysis results are shown in Table 2. The safety protection module is designed with three important parameters of emergency destruction means. One is that when the security protection module detects that a security event occurs, it generates a destruction order according to the security strategy and destroys the key data stored in the protection module. The second is to set a quick destruction switch on the equipment shell to provide the destruction of important parameters on the module in an emergency; the third is to set up an organic shell disassembly sensing device on the equipment shell, which acts on the destruction circuit on the module. When the attacker disassembles the shell, the security protection module passively executes the security policy destruction. Through the design of database storage, the scheme of data file storage and key management on the big data storage server is given. Next, a high-availability solution is given for the security and integrity of the data in the Oracle database, so as to ensure that the data uploaded
by the data owner on the cloud server side will not make users unable to access and download the data normally due to server failure. At the same time, remote disaster recovery and backup technologies also ensure that the uploaded data will not be lost. This article adopts the encryption method combining CP-ABE and IBE. Now we compare this scheme with the ABE scheme of Bethencourt, Sahai and Waters, named BSW_ABE. Compared with the data sharing scheme of KP-ABE, the four encryption and decryption processes are respectively compared. The time spent by the function is shown in Table 1 (unit ms), and the test data is 10MB. It can be seen from Table 2 that the time difference between the four functions of the proposed scheme and the BSW-ABE scheme is small, which proves that the scheme does not increase additional system overhead on the basis of enhanced security. Compared with the KP-ABE scheme, the encryption and decryption phase consumes less time, so this paper chooses the CP-ABE and IBE scheme to combine better.

Table 2. Performance comparison analysis results

| Scheme    | Setup  | Keygen | Enc  | Dec  |
|-----------|--------|--------|------|------|
| KP-ABE    | 487    | 370    | 712  | 203  |
| BSW-ABE   | 493    | 354    | 703  | 214  |
| Average   | 495    | 361    | 714  | 209  |
| This article scheme | 491.7  | 361.7  | 709.7| 208.7|
| Average   | 490    | 367    | 619  | 147  |
| KP-ABE    | 49 6   | 381    | 604  | 153  |

4.2 Comparative Analysis of file Upload Speed

The comparative analysis result of file upload speed is shown in Figure 1. Here is a record of the time spent uploading and downloading files of different file sizes under the same bandwidth environment. Here we can see that the upload speed of the big data storage module is the fastest, while the speed of our Seccloud and invisible cloud It is roughly equal to 400k/s. Considering that the file needs to be encrypted in blocks, this process will take time, and part of the time is sacrificed in exchange for data security. Such a price is acceptable. File download is the same. Big data storage module is faster (300-600k/s), Seccloud and invisible cloud speed are slower (200-400k/s). System deployment is achieved through the establishment of web service platform, big data storage server platform and backup platform. The system is tested in terms of function and performance. The function test is basically normal. In terms of performance, the speed of encryption and decryption algorithms is first tested, and then compared with the current popular comparing cloud storage products, although there is still a gap in upload and download speeds, we believe that it is acceptable to sacrifice a little speed for data security. The high-availability architecture can ensure the data integrity of the big data server, and greatly reduce the downtime of the data server, and improve the trust and goodwill of users. Therefore, the big data storage module in this article protects the privacy of data and the high-availability architecture for data integrity protection has high practical value. This paper proposes an effective and scalable big data sharing encryption service, which realizes complete privacy protection, resistance to collusion attacks and data confidentiality. Under the general bilinear group under the random language model, the semantic security of the proposed scheme is proved, and the fine-grained access and backward security are realized. All extended operations, including user withdrawal, can only affect the current file or user, without involving key updates, reducing system overhead.
4.3 Server Access Delay

The server access delay is shown in Figure 2. As for the network latency test under concurrent requests, because there is a gap between the hardware we use and the big data storage module, invisible cloud and other listed products, and the website deployment does not use load balancing technology, there is no comparability, we will test Under Seccloud under different concurrent requests, the corresponding delay. Taking into account that our system did not use load balancing technology when deploying, the Apache server has limited concurrent support requests. Here we set the maximum number of requests to 200, and the number of concurrent tests is performed at three levels of 10, 20, and 30. When the concurrency level is the same, as the number of requests increases, the network delay will gradually increase. When the number of requests reaches the peak, the network delay reaches the peak. When the number of requests is equal, as the level of concurrent requests increases, the network delay will increase accordingly. This network delay result is affected by the hardware and Apache server. When deploying the web server, choose higher-configuration hardware equipment, and consider the server-side load balancing and redundancy technology. This will greatly increase the number of concurrent requests, and it will Reduce network latency.
5. Conclusion
As the big data transmission and storage security technology, the security technology proposed in this paper needs the support of a large amount of hardware environment. There are problems of high cost and long construction period. Although the authentication management module is small, it needs to be configured in each user terminal device. The next step is to follow the existing research direction. In the development of user terminal equipment, the security authentication function will be integrated in the user terminal equipment to reduce user usage costs.

The design of the security protection module is not perfect. The current design of the security protection module is focused on physical protection. In the next step, password protection and software protection will be added to the design of the security protection module. The dual-system hot backup technology solves the problem of data packet loss. It can be used not only for data storage equipment, but also for data processing equipment. In the future, the technology can be promoted to other fields.

Acknowledgments
Science and technology research project of Jiangxi Provincial Department of education in 2019（Fund number: GJJ191012）

References
[1] Wang F, Chen L, Liu S, et al. Studying safe storage time of orange peel (Citrus reticulata) using high-throughput sequencing and conventional pure culture[J]. Food Science & Nutrition, 2018, 6(8):2545-2552.
[2] Simonetti, Joseph A, et al. Evaluation of a community-based safe firearm and ammunition storage intervention[J]. Injury prevention: journal of the International Society for Child and Adolescent Injury Prevention, 2018, 24(3):218-223.
[3] Armstrong K. The safe storage and management of vaccines[J]. Practice Nursing, 2018, 29(4):167-170.
[4] Toshinsky G, Grigoriev S, Dedul A, et al. Safe Controlled Storage of SVBR-100 Spent Nuclear Fuel in the Extended-Range Future[J]. World Journal of Nuclear Science and Technology, 2019, 09(3):127-139.
[5] Shevtsov S A, Kargashilov D V, Shutkin A N. Fire and Explosion Safe Technology of Storage and Regasification of Liquefied Petroleum Gas[J]. Chemical and Petroleum Engineering, 2018, 54(1-2):38-40.
[6] Sondej M, Ratnayake C, Micha Wójcik. Economical and Safe Method of Granular Material Storage in Silos in Offshore Port Terminals[J]. Polish Maritime Research, 2018, 25(3):62-68.
[7] Capa R B, Jan Odom-Forren R C F. 4APromotion of Safe Use, Storage, and Disposal of Opioids in the Ambulatory Surgery Setting[J]. Pain Management Nursing, 2020, 21(2):211-212.
[8] Jie X, Gonzalez-Cortes S, Xiao T, et al. The decarbonisation of petroleum and other fossil hydrocarbon fuels for the facile production and safe storage of hydrogen[J]. Energy & Environmental Science, 2019, 12(1):238-249.
[9] Violoano, Pina, Bonne, et al. Prevention of firearm injuries with gun safety devices and safe storage: An Eastern Association for the Surgery of Trauma Systematic Review[J]. The journal of trauma and acute care surgery, 2018, 84(6):1003-1011.
[10] Hao J T, Sun Y, Luo H. A safe and efficient storage scheme based on blockchain and IPFs for agricultural products tracking[J]. Journal of Computers (Taiwan), 2018, 29(6):158-167.