Construction and Research of Digital Archives Cloud Platform Based on Big Data Management

Yazhi Sun*
Heilongjiang International University, Heilongjiang Harbin, 150076, China

*Corresponding author: 19742@sina.com

Abstract. Cloud computing has the characteristics of super computing power, low cost and high security, as well as providing powerful data storage and network services. In the commercial field, it has begun to take shape. This paper attempts to introduce cloud computing technology into the construction of digital archives, analyzes the possibility and reality of the application of cloud computing technology services in the service and management of digital archives from the common point of view of cloud computing services and digital archives, and puts forward the corresponding ways of cloud computing application in digital archives. This paper analyzes and integrates all kinds of risks that digital archives may face under the cloud computing environment, and establishes the security evaluation index system including the security elements of information system and cloud computing technology. In this paper, fuzzy comprehensive evaluation method is used to evaluate the mathematical model of cloud digital archives security evaluation system, and a set of security evaluation system suitable for cloud digital archives is proposed.

Keywords: Cloud computing, digital archives, information system security, fuzzy comprehensive evaluation

1. Introduction
As a virtual form and function extension of traditional archives, digital archives meet the needs of people in the information age for filing [1-2], long-term preservation and development and utilization of archives information [3]. As the inheritance and development of distributed computing, parallel computing and grid computing technology, cloud computing has unique advantages in the analysis, processing and storage of massive data. Using this advantage to build digital archives based on cloud computing technology is beneficial to improve the work efficiency and service level of digital archives [4-5]. However, the cloud computing architecture is more complex, which has the characteristics of open environment, on-demand service and multi tenant, which brings significant threat to the information system security of cloud computing technology [6]. To determine the source of security threats, evaluate the importance of security risks, and ensure the security of digital archives resources in the cloud environment is the key to digital archives using cloud computing technology. It is of great significance to evaluate the security performance of cloud digital archives, to maintain the security of digital archives resources, to improve the service ability of cloud digital archives, and to enrich the theory and practice of Archival Science in the system security evaluation under the cloud computing environment.
environment [7-8].

2. Cloud Computing Technology and Digital Archives Security

2.1 Security Issues of Cloud Computing Technology
Cloud computing technology is a new IT architecture that provides computing resources on demand through the division of service levels. The innovation points are mainly reflected in the service mode level, and its commercial value is realized through the core operation characteristics such as resource leasing, application hosting and service outsourcing. Limited by its structural characteristics, cloud computing brings convenience to organizations and individuals, but also inevitably brings security risks from the computer network environment, thus posing a threat to the security of archival information resources. "Gartner has proposed seven security risks of cloud computing technology, which are access risk of privileged users, compliance risk, uncertainty of data location, shared storage data risk, data recovery risk, investigation support (data tracking function) risk and long-term development risk. [9]" Among the seven cloud computing security risks, the long-term development risk and the compliance risk can be regarded as the internal and external environmental factors that affect the operation and development of cloud computing technology, while the other four risks directly point to data security issues. Among them, the uncertain risk of data location indicates that it is difficult for data owners to master the specific physical address of data and directly monitor the security status of data under the separation of ownership and control of data. The storage risk of internal data and secret data is constantly amplified. It can be seen that most of the risks contained in cloud computing technology threaten data security, which is a problem that digital archives, which contain a large number of digital archives resources, must focus on and solve when using cloud computing technology.

2.2 Security of Digital Archives
Compared with the common information management system, digital archives has a strong professional in the aspects of information storage and utilization. Digital archives are original recorded information, whose voucher value and historical value are not possessed by other types of information. Digital archives information must be true, unmodified, properly preserved, long-term readable, and complete descriptive information. This requires the digital archives from the content and carrier of digital archives to provide a wide, long-term, deep-seated security. At the same time, in the construction of Digital Archives in cloud environment, the security assessment of information system is conducive to obtain the security risks of digital archives system in advance, and then put forward the corresponding solutions, which is of great significance for improving the security of Digital Archives in cloud environment, but the domestic research in this area is still relatively lacking.

3. Construction of Security Evaluation Index System For Cloud Digital Archives
The security evaluation index system is the basic content of the whole cloud digital archives security evaluation system. The rationality of the index system will directly determine the value of the security assessment. At present, there is no scientific and reasonable index system for the security evaluation of cloud digital archives at home and abroad. The security evaluation of digital archives mostly refers to the existing information system security evaluation system. In this paper, from the perspective of archival science, the existing archives information system security assessment system is used for reference, and the international information security assessment standards represented by ISO27001-2013 international standards for information security management and enterprise IT governance framework are referred, and the physical, technical, regulatory, system, management, system and other factors affecting the security of cloud digital archives are comprehensively considered, and the security assessment index system of cloud digital archives is established.
3.1 Information System Security Evaluation Index

(1) Physical device security. It mainly focuses on the safety performance of the computer and its related hardware equipment, such as whether the storage location of the equipment is appropriate, whether there is a strong magnetic field and strong noise area around the equipment, the safety of the power supply and cable of the equipment, whether there are fire prevention, waterproof, anti-theft and lightning protection measures, the conditions and requirements of asset transfer, and the data security when the equipment is scrapped or reused.

(2) Operating system and platform security. The main goal of this layer is to prevent unauthorized access to the system and application platform, including restricting the access to source code, information access control, security management of login and password system, and the robustness and stability of the operating system itself, as well as preventing system leakage and backdoor.

(3) Database system and software security. The database system and application software are related to the daily operation of digital archives, and their security importance is unnecessary to elaborate. Digital archives database system should have high security performance, widely used in related industries at home and abroad, with structured and unstructured data processing capabilities, which should be paid attention to in the early stage of database development. In daily use, the relevant monitoring software should be used to scan and monitor the database regularly, so as to discover and correct the possible security loopholes in the database system.

3.2 Evaluation Index of Digital Archives Security

(1) Physical security of digital archives. The security and integrity of Physical Archives is the basis of digital archives work. When building cloud digital archives and archives digitization, business departments must strictly follow the existing laws and regulations and work procedures to ensure the integrity and authenticity of archives entity digitization work.

(2) File data security. After digitalization, regular migration, upgrading, transmission, development and utilization, the archives should ensure its integrity, readability and authenticity, ensure that the value does not change, the content is not distorted, and has a reasonable metadata structure and complete information description. It includes "regular evaluation of the content and formal characteristics of metadata, so as to ensure that digital archives resources are always in a safe, readable and complete state".

(3) Internal business security. There are many kinds of file carriers and contents. The different standards and huge number of file sets increase the work difficulty of archives in the process of filing and handing over. It is easy to make mistakes in work, resulting in missing, wrong, missing and blank digital content of archives.

(4) Outsourcing security. The commercial companies in charge of outsourcing work often adhere to the principle of the best interests of the company. The outsourcing work includes risks brought by the differences between the company and the archives, such as "demand expression risk, adverse selection risk, technical environment risk and post operation and maintenance risk".

4. Security Evaluation of Cloud Digital Archives Based on Fuzzy Comprehensive Evaluation Method

On the basis of the evaluation index system, we need to draw a qualitative conclusion, and then grasp the security performance of a specific cloud digital archive from a macro perspective. Because the operation platform of this topic is based on the abstract cloud digital archives security evaluation model, not a specific archive, the index system established has certain subjectivity, and its hierarchical relationship is fuzzy. Therefore, this paper uses the two-level fuzzy comprehensive evaluation method to quantify the fuzzy relationship evaluated (i.e. to determine the membership degree). The security score value of cloud digital archives is obtained by statistical operation, and its security performance level is determined. The whole example and calculation process focus on the operation method of cloud digital archives security assessment.
4.1 Establish Hierarchical Structure According to Index System

Before using fuzzy comprehensive evaluation method to conduct comprehensive evaluation, it is necessary to show the known evaluation model index system in the form of hierarchical structure diagram, and clarify the hierarchical structure system between the general objective and the sub objectives of each level. It is the hierarchical structure relationship of cloud digital archives security evaluation system in Figure 1.

![Hierarchical structure of cloud digital archives security evaluation system](image)

**Figure 1** Hierarchical structure of cloud digital archives security evaluation system

4.2 Determine the Set of Evaluation Factors

It is the first step to determine the evaluation factors, that is, the specific objects to be evaluated. As far as this paper is concerned, there are two levels of indicators that need to be evaluated and multi-level fuzzy comprehensive evaluation should be adopted, that is, from the low level to the high level, the single factor evaluation of each level is the multi factor comprehensive evaluation of the lower level. The index set and corresponding weight set of main factor layer and sub factor layer are defined as follows:

In the cloud digital archives security evaluation system, the main factor layer \( U = U_1, U_2, U_3, U_4, U_5, U_6, U_7 \). Among them, \( U_1, U_2, U_3, U_4, U_5, U_6, U_7 \) respectively represent seven indicators in the indicator system, including information system security, cloud digital technology security, digital archives security, laws and regulations, organization internal evaluation, management and control security, objectives and feedback. The sub factor layer is composed of the first level indicators and the second level indicators.

After determining the factor set, it is necessary to draw up the corresponding evaluation set to determine the subordinate relationship between the final index and the comments. Evaluation set is
grade set, which is usually represented by $V = (v_1, v_2, \ldots, v_m)$, $v_j$ is the evaluation grade of each index, and each grade will correspond to a fuzzy subset. Generally, the grade $m$ of comments is ($m \in \mathbb{N}^+$), considering the difficulty and feasibility of calculation, this paper defines the evaluation set as five levels, namely $V = (v_1, v_2, v_3, v_4, v_5)$, among which the corresponding five comments are {very safe, safe, relatively safe, general, unsafe).

4.3 Experts Score and Establish Evaluation Matrix

First, determine the score of different grades. Five levels of the comment set have been identified above: $V_1$ (very secure), $V_2$ (relatively secure), $V_3$ (secure), $V_4$ (general), and $V_5$ (insecure). According to the needs of calculation and statistics, the above five comments are assigned 5 points, 4 points, 3 points, 2 points and 1 point.

Then, the single factor fuzzy evaluation matrix is established. In multi-level fuzzy comprehensive evaluation, it is necessary to start from low-level single factor evaluation to high-level. The proportion of safety comments corresponding to the scores of five experts for each index in the five experts' opinions is calculated, and the judgment matrix of secondary index safety is obtained as follows:

$$R_1 = \begin{bmatrix} 0.4 & 0.4 & 0.2 & 0 & 0 \\ 0.6 & 0.4 & 0 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 & 0 \\ 0.2 & 0.4 & 0.4 & 0 & 0 \\ 0.8 & 0.2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} 0.2 & 0.6 & 0.2 & 0 & 0 \\ 0 & 0.2 & 0.8 & 0 & 0 \\ 0.6 & 0.4 & 0 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0.2 & 0 \\ 0 & 0 & 0.2 & 0.4 & 0 \\ 0.2 & 0.8 & 0 & 0 & 0 \end{bmatrix}$$

$$R_3 = \begin{bmatrix} 0.4 & 0.4 & 0.2 & 0 & 0 \\ 0.6 & 0.4 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0.8 & 0.2 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 & 0 \\ 0.6 & 0.4 & 0 & 0 & 0 \end{bmatrix}$$

$$R_4 = \begin{bmatrix} 0 & 0 & 0.6 & 0.4 & 0 \\ 0.6 & 0.4 & 0 & 0 & 0 \\ 0 & 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0.6 & 0.2 & 0 \end{bmatrix}$$
The comprehensive evaluation vector is calculated, and the comprehensive evaluation conclusion is obtained. Through the above calculation, the evaluation matrix of the first level index as the evaluation factor is obtained. At this time, it is necessary to transform the known weight vector \( w \) obtained by analytic hierarchy process (AHP) into fuzzy vector \( B \) on the evaluation set through synthetic calculation, and \( B \) is called comprehensive evaluation vector, and its calculation method is \( B = W \cdot R \). \( R \) is called synthetic operator of comprehensive evaluation. Generally, \( M(\cdot,+) \) type is selected for calculation, i.e. weighted average type. The comprehensive evaluation vector is obtained by multiplying the known index weights of the same level with the judgment matrix

\[
B_1 = W_1 R_1 = (0.2099, 0.1212, 0.1962, 0.0801, 0.3177, 0.0749) = (0.5802, 0.3457, 0.0740, 0, 0)
\]

In the same way, it can be concluded that:

\[
B_2 = W_2 R_2 = (0.2091, 0.5538, 0.1864, 0.0507, 0)
\]

\[
B_3 = W_3 R_3 = (0.5161, 0.4102, 0.0723, 0, 0)
\]

\[
B_4 = W_4 R_4 = (0.1497, 0.4538, 0.3387, 0.0578, 0)
\]

\[
B_5 = W_5 R_5 = (0.3184, 0.3860, 0.1946, 0, 0)
\]

\[
B_6 = W_6 R_6 = (0.3960, 0.4436, 0.1604, 0, 0)
\]

\[
B_7 = W_7 R_7 = (0.0629, 0.3483, 0.5888, 0, 0)
\]
In order to keep consistent with the synthesis operator, this paper still uses the weighted average principle when calculating the final evaluation results, optimizes the obtained comprehensive evaluation vector, and then obtains the score value of the security performance of the cloud digital archives. The calculation formula of the weighted average principle is as follows:

$$c = \frac{\sum_{j=1}^{n} b_j^k \cdot j}{\sum_{j=1}^{n} b_j}$$

In which $m$ represents each comment grade, and $k$ represents the coefficient to be determined ($k=1$ or $k=2$), in order to control the influence of higher value of $b_j$ on the results.

In this paper, $k=1$ is taken. According to the rating values of each grade drawn up above, the total rating value of the security of cloud digital archives is:

$$c = 5 \times 0.3403 + 4 \times 0.4029 + 3 \times 0.2355 + 2 \times 0.0056 = 4.0348$$

The evaluation levels set in this paper are shown in Table 1:

| Total score | Level         |
|-------------|---------------|
| 4.0-5.0     | Very safe     |
| 3.0-4.0     | Security      |
| 2.0-3.0     | Safe          |
| 1.0-2.0     | Commonly      |
| 0-1.0       | Unsafe        |

According to the comparison, the security evaluation level of the cloud digital archives is "very safe".

5. Conclusion
Since the research platform of this paper is a virtual cloud digital archives security assessment model, the expert scoring data in the fuzzy comprehensive evaluation stage has certain virtuality and randomness, so this paper does not summarize and discuss the security issues that need to be explored and analyzed after the evaluation results. In practical application, the archives staff should have the ability to effectively use the security assessment system, and report the results to the leadership of the archives and the leaders of the information technology department in a timely manner after conducting a comprehensive security assessment for each specific index, so as to discuss the solution to the problem. At the same time, listen to the opinions of technical experts from other archives, universities and related industries, summarize the opinions, and jointly determine the solutions to the security problems of cloud digital archives. Gather the strength of the people, so that the cloud digital archives is always in a good state of operation.

References
[1] Fu Rongxiao. Thinking About Digital Archives. Archives Science Communication, 2001, 000 (005): 26-28
[2] Yu Lijuan. Overview of Digital Archives Construction Abroad. Archives of China, 2003 (03): 23-24
[3] Meng Yu, Zhang Tao, Zhao Zhengwen. Construction and Research of Digital Archives Cloud
Platform. Microprocessor, 2013, 034 (006): 49-52, 57
[4] Liu Rong. Function and Implementation Technology of Digital Archives. Archives Research, 2002 (03): 49-51
[5] Tang Yanfang. Construction of Archives Information Service Platform in Digital Archives. Archives Research, 2006
[6] Yang Zhiyong, He Hongjia. Research on the Construction of Government Digital Archives Based on the Connotation of "integration". Archives Science Communication, 2012 (06): 70-73
[7] He Jiasun. Current Document Reading Center and Digital Archives Center. Archives Research, 2003
[8] Jin Gengda, He Jiasun. Discussion on the Mode of Digital Archives -- Research on Integrated Management and Service of Electronic Records Based on Metadata (part 2). Archives Science Communication, 2005 (05): 54-58
[9] He Jiasun. Current Document Reading Center, Document Center and Digital Archives. Archives Research, 2003, 1: 32-35