The Construction of Digital Management Information System of University Documents and Archives Based on Distributed Architecture

Anying Jian[a]
Chongqing Vocational Institute of Engineering, Chongqing, China

Keywords: Hadoop, Distributed Architecture, University Archives Management, Management Information System.

Abstract: This paper analyzes the process of digital management of university documents and archives, and develops a digital management information system of university documents and archives based on distributed architecture. The system uses java programming language, SSH framework of struct2+spring+hibernate and web service technology to design the function of the system and develop its application. The data is processed by HDFS distributed storage and mapreduce computing components in hadoop ecosystem. The system can realize various management functions of summarizing, classifying and numbering files. The whole process of digital management and control of archives should be promoted by the process, so that university organizations can embed scientific archives management system into the electronic process, so that the system and digital means complement each other and make archives management more standardized.

1 INTRODUCTION

The documents and archives in colleges and universities are numerous, and the "Measures for the Management of Archives in Colleges and Universities" requires that "colleges and universities should archive paper archives and electronic archives simultaneously", and defines the scope of archiving, including more than ten kinds of archives, including the Party and the masses, administration, students, teaching, scientific research, capital construction, instruments and equipment, product production, publications, foreign affairs, accounting, etc. So in the digital age, while constructing digital archives, colleges and universities should "guarantee the construction of archives informatization and the construction of digital campus", so as to realize the integration of university management resources and promote the unified and comprehensive development of archives management in colleges and universities. With the development of the times, it is the only way to deepen the process of college archives management that how to use modern information technology to quickly complete college archives management. But at present, there are some problems in the management of documents and archives in many colleges and universities, such as huge amount of information, insufficient manual management and control ability, and insufficient automation and intelligence. Besides, relying on the limited storage capacity of local servers, many campus management systems are difficult to support the preservation of university data, which makes it difficult for campus archives management to adapt to the bottleneck of the development of electronic archives management in the times. (Zhao, 2021)

To solve the above problems, the author believes that a digital management information system of university documents and archives should be developed based on the distributed architecture of big data technology. The function involves automatic filing, quick query and safe authority, so that all kinds of digital achievements of colleges and universities can be preserved and checked for a long time. The design of this system helps colleges and universities to manage archives uniformly, collect archives of different functional departments, departments and systems uniformly, improve the

[a] https://orcid.org/0000-0002-8984-380X

Jian, A.
The Construction of Digital Management Information System of University Documents and Archives Based on Distributed Architecture.
DOI: 10.5220/0011908100003613
In *Proceedings of the 2nd International Conference on New Media Development and Modernized Education (NMDME 2022)*, pages 155-160
ISBN: 978-989-758-630-9

efficiency of archives collection and handover, and reduce the error rate. By the integration and synchronous management of landing paper archives and electronic archives, the utilization rate of university archives can be improved.

2 TECHNICAL OVERVIEW

2.1 Hadoop Ecology

The Hadoop is an open source technology to realize distributed data storage management and computing services. The Hadoop Foundation uses JAVA language to develop and is responsible for the release, maintenance and management of Hadoop ecosystem. To solve the background storage problem of massive archive data in the management of college documents and archives studied in this paper, the cloud storage platform of the system adopts hadoop to build the data distributed platform architecture at the bottom of the system. Hadoop ecosystem consists of many components, among which the core functional components are shown in Figure 1, including HDFS distributed file system component, MapReduce distributed computing component and yarn scheduling management component. In the research and development of this system, HDFS distributed file system is used to store files through the establishment of clusters. During this period, mapreduce mechanism is used to read the file business data and support offline computing operations. (Guan, 2022) Hadoop's information server nodes are mainly divided into five categories, of which master is the main server node, which is usually equipped with namenode maintenance node of HDFS and JobTracker job scheduling node of mapreduce. Mapreduce also needs to deploy tasktracker task processing nodes to process specific computing tasks. In HDFS, datanode is also needed to store data blocks in a distributed way.

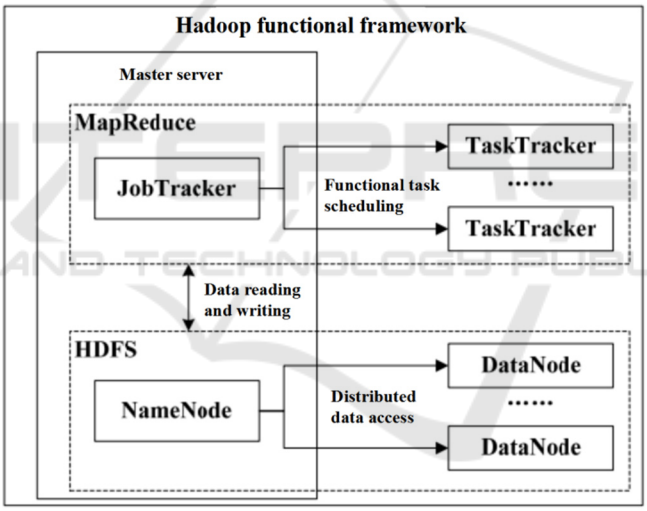

Figure 1: Hadoop core functional framework

2.2 Web Technology

The Web development technology is a core technology that combines the development of web interactive pages with the development of background servers. It can organize and publish the functions of web service application software system, and usually uses the distributed network mode of B/S structure.The Web development technology usually includes static web page technology, dynamic service script technology and web service technology. The client of Web application software system is a web browser, while the server uses the tools published by web services to schedule tasks. The client and the server need network communication, and most of the communication protocols are HTTP or HTTPS. The development technology structure of Web also facilitates the deployment and function release of the software system in the wide area network environment, and has strong support ability for multi-users. Moreover, by using the existing highly mature Web browser and Web service publishing tools, the functional reliability of the whole software system is also relatively high. (Zhao, 2022)

2.3 Development Environment

The development environment of digital management information system of university documents and archives is introduced in two parts, one is the construction of hadoop big data cluster, the other is the application environment of Javaweb technology. According to the required amount of data, this paper builds a hadoop cluster composed of four nodes. The node is named as the master node namenode, and three slave nodes datanode01, datanode02 and datanode03. These clusters are stored based on HDFS distributed storage pairs, and are built with other hadoop ecosystem components, such as flume, hbase and hive. The hadoop server cluster is developed on four clients installed with Linux system. This paper selects Centos7.8 Server release version of Linux operating system.

The Java development tool used in the JavaWeb application of this system is my eclipse, the development environment is JDK 1.8, the development language is Java, and Apache Tomcat 9.0 is selected for server construction. This system is implemented in B/S mode. The browser side uses the dynamic web page technology of HTML+CSS+JaVaScript, and uses AJAX technology to communicate with the server side. The development of the system is based on MVC pattern. The SSH framework of struct2+spring+hibernate is used as the framework, and sqlsever database is used to help manage data. Through the introduction of the above key technical theories, the overall environment, the configuration of related software and tools for the development of digital management information system of university documents and archives are determined, and the technical feasibility of the overall project is also clarified. (Liao, 2022)

3 DEMAND ANALYSIS

3.1 Functional Requirements

The user end of the digital management information system of documents and archives in colleges and universities is provided with the common user end and the administrator user end. Ordinary users include teachers and students, and ordinary user end functions mainly include borrowing, returning, booking and file retrieval. The administrator end mainly includes two main functional modules: application processing, file management and user management. (Yu, 2018) At the same time, the system has the function of electronic archives. With the increasing content and types of archives, electronic archives will become the main work of future archives management. The electronic archives in this paper need to be consulted by users in a targeted way, so as to improve the user experience and the management efficiency of archivists.

The performance requirements of the system require that when the system processes the operation request of the foreground interactive interface, it should ensure that all functions of the business logic layer are processed within 5 seconds in the normal communication network environment, and the processing results of the operation instructions are returned in time. At the same time, considering the scale of archives management business in colleges and universities, the system should support the instruction processing of concurrent operation access of at least 100 users. (Li, 2020)

3.2 Overall Design

The overall design of digital management information system of documents and archives in colleges and universities is divided into two parts: application design and data processing. This paper divides the data processing of the system into six layers. The first layer is the data source layer. The data comes from the local database server and the audit data in the information management system, as well as other data entered inside the organization, which can be divided into structured data and unstructured data. In the data transmission layer, sqoop transmits the data from the storage layer and the source layer, and flume collects the unstructured log data from the server. The data storage layer of the system consists of HDFS file storage, MySQL database and hbase database. The resources built on hadoop cluster are managed by yarn. The interaction of data query function is completed by Hive component, and the scheduling of distributed cluster needs to be completed by Oozie.

The application part of this system is divided into three layers, namely presentation layer, business layer and data layer. The whole system adopts B/S mode combined with MVC thought and uses SSH architecture of struct2+spring+hibernate for development. The web layer is the middle part between the view layer and the control layer, which is handled by struct2, in which action is used to handle all kinds of access requests and access feedback instructions, such as HttpRequest and HttpResponse from the network. In this process, the request parameters need to be repackaged and various functions of page navigation are needed. The

security permission control of this system is deployed in the presentation layer by using CAS technology, which can restrict the specific content accessed by users according to the level of users' permission, and improve the security of information and data of the system. The business layer is mainly composed of the service of spring. spring tools can effectively integrate hibernate in the data layer and struct in the presentation layer, which can efficiently solve the coupling problem of various levels in the application system. The business logic layer needs to receive the business information instruction from structaction for data processing, and at the same time, complete the related logic operations for the functional business and process business of the system. The logic layer needs to set the interface of the presentation layer, i.e., Iservice layer, which can interact the information of the action layer and the service layer, and standardize and encapsulate the interface of data interaction between the two layers. The data persistence layer is processed by hibernate, which needs to package all kinds of data interfaces JDBC lightly to simplify the persistence layer code and further improve the development efficiency of the system. While the business logic layer and data persistence layer need to set up the IDAO layer to build the corresponding data interface to the service layer and DAO layer, which is responsible for the data circulation and access. Database relational database of data layer adopts SQL Server to realize related data storage control. The overall design framework diagram of the system application is shown in Figure 2.

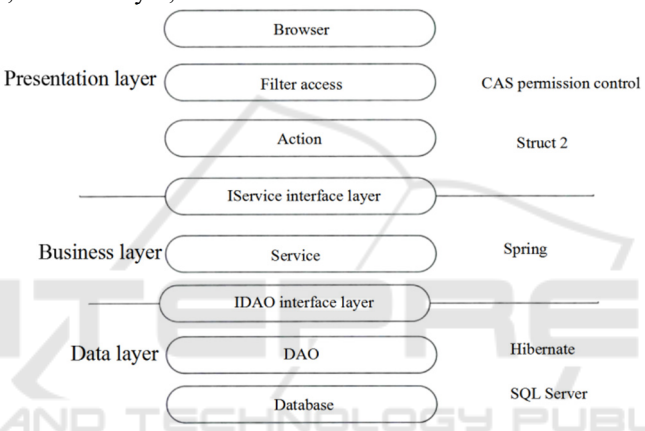

Figure 2: Overall architecture of application design

4 FUNCTION REALIZATION

4.1 Ordinary User End

The common users include teachers and students, for example, past (due) students can apply for access to their own student files through this system. After ordinary users log in to the system according to their account numbers, they can see that the functions of ordinary clients mainly include borrowing, returning, booking and file searching. You need to click the reservation function to submit the file borrowing application and reserve the borrowing time to get the borrowing information. In the file retrieval function, users can search the required files according to attributes such as time, keywords, categories, etc. When the files involved do not have access rights, the system will send out a pop-up reminder to refuse access and ensure the security of the files (Liu, 2022).

4.2 Administrator End

The administrator end mainly includes two main functional modules: application processing, file management and user management. The administrator needs to review the file borrowing applications of ordinary users and approve and reject them. The archives management includes the collection, statistics, destruction and modification of archives. In the collection of archives, administrators need to import files from various channels into the archives management information system of colleges and universities and number them. The implementation code of file import function is shown in Figure 3. This paper uses Apache's file upload component and creates DiskFileItemFactory to realize related functions, and uses file upload

parser to judge whether the data quality meets the requirements. When the system generates new files and sets aside records, the system synchronizes the records to the data interface of the university document archives information management system. If the information addition is unsuccessful, the system will try to connect to the new record database for three times and write the failed records into the abnormal information data table, and scan them every five minutes for daily processing. (Ma, 2021)

```
try{
//Use Apache file upload component to process file upload steps:
   //1. Create a DiskFileItemFactory
   DiskFileItemFactory factory = new DiskFileItemFactory();
   //2. Create a file and upload it to the parser.
   ServletFileUpload upload = new ServletFileUplad(factory); .
   //Solve the Chinese garbled file name uploaded.
upload.setHeaderEncoding("UTF-8);
   //3. Determine whether the submitted data is the data of the uploaded form.
   if( !ServletFileUpload. isMultipartContent(request){
   //Get data in the traditional way.
   retum;
//Use ServletFileUpload parser to parse the uploaded data, and the parsing result returns
What is returned is a List<Fileltem > collection, and each Fileltem corresponds to an input item of a Form form.
```

Figure 3: File import function implementation code

The user management functions include role management and permission management. The administrator needs to classify the rights of users at all levels to realize the security of data access. The administrator can also manage the account information of teachers and students on campus. The function of the administrator calling and browsing the specified information of each user is realized by the BusinessManager class, which completes the actual work by encapsulating the business logic code. The getSelfMessages class is the class used by the system to call the data of the data persistence layer. The getSelfMessages function code is shown in Figure 4. The Hibernate can use java classes such as "SelfMessages" and "user_id" in query statements to determine the mapping relationship between classes and database tables, and then obtain the table and field parameters in the corresponding database information. By this mapping relationship, Hibernate automatically synchronizes the objects with the data table records, without requiring developers to write corresponding codes.

```
try{
String strSql;
strSq1=" fr()m SelfMessages where user_ id=: user_ id and message_ text
Like: filter_ text;
Sessi()nF act()ry sf;
sf= new C()nfigurati()n(). c()nfigure(). buildSessi()nFact()ry();
Sessi()n sessi()n=sf. ()penSessi()n();
Transacti()n tx= sessi()n. beginTransacti()n0;
Query query=sessi()n. createQuery(strSq1);
query. setInteger( 'userid", user_ id);
query. setString( filtertext",    filtertxt);
```

Figure 4: getSelfMessages function code

5 CONCLUSIONS

The informatization management of university archives is very important to the improvement of the overall management level of universities, so I have built a digital management information system of university archives based on distributed architecture to support the informatization management of university industries. But my technology is limited, and I didn't combine the efficient cloud computing technology of the current era to rationalize and improve the storage capacity and response efficiency of the document archives management information system. We hope that there will be experts and scholars in the follow-up to improve this research.

REFERENCES

Guan Lige (2022). The Reform and Development of University Archives Management Mode under the Background of Big Data [J]. Technology Wind.03.

Liao Xiujuan (2022). The Information Management of University Archives under the Big Data Environment [J]. Inside and Outside Lantai.02.

Li Shuhua (2020). Research on University Archives Management Mechanism Based on Smart Campus Construction. Archives Management.02.

Liu Xia (2022). Thoughts and Strategies of Improving the Level of Archives Management in Colleges and Universities. Scientific Management.03.

Ma Wenqing, Sun Xiuyun (2021). Problems and Countermeasures of Archives Management in

Colleges and Universities in the Information Age. Science & Technology Vision.09.
Yu Chunyan (2018). The Design and Implementation of University Archives Management System Based on. NET [D]. Xi'an University of Science and Technology.09.
Zhao Donglong (2021). The Current Situation and Countermeasures of University Archives Management [J]. Archives Management.03.
Zhao Zuoyan (2022). The Research on University Archives Management under the Background of Digital Informatization [J]. Management forum.03.