

# Data Storage and Research on Distance Education Platform

Huang Qinghua  
Qingyuan Polytechnic  
Qingyuan City, Guangdong  
Province, 511510, China

---

**Abstract:** Distance education resources have the characteristics of large amount of data and rapid growth. Low-cost storage and content-based efficient retrieval of these massive data resources are a major problem in the construction of distance education cloud platforms. Based on Hadoop, this paper designs a set of storage and retrieval methods for the massive resources of distance education, which solves this problem. Distribute the load of video and audio streams to the network. For this reason, this paper proceeds from the reality of the application of the distance teaching platform, fully considers the human factors of the user's order and the technical environment of the video service system and conducts in-depth research on key technologies such as data storage, database update, and resource synchronization in the resource service system. And put forward the following improvements and innovations: use of cloud storage technology; automatic update of educational information in the database; automatic download of resource updates; synchronous update technology, which greatly reduces the workload of managers.

**Keywords:** Data Storage, Distance Education Platform

---

## 1. INTRODUCTION

The advent of the information society has made it more urgent for people to update knowledge, and people are increasingly aware of the necessity of lifelong learning. The rapid development of the information society and the rapid expansion of the Internet have made distance education a trend. The system is a new generation of teaching technology that combines computer network and multimedia technology in different places or at different times. Cloud computing is a computing model for sharing resources. It aggregates computing, storage, network, software, and other resources that are physically scattered across the Internet through virtualization, distributed computing, and other technical means to realize the logical concentration and integration of resources. Use dynamically and flexibly and provide these resources to Internet users in the form of services. Compared with the traditional model, cloud computing has powerful information storage and processing capabilities, and can provide convenient, flexible, on-demand rental, and cost-effective information services.

Significant improvements have been made to security and management. It is manifested in two aspects of technology and demand processing changes. In addition, authentication and authorization are enhanced in terms of security. It provides enhanced management capabilities, improved XML database management and new command-line tools. The component model is an architecture and API set established for developers to define software components, so that developers can build application systems through dynamic combination of software components. The component model consists of two main components, components, and containers. Components are basic software parts with reusable characteristics. Containers are used to store and arrange components to realize the interaction between components. A container can also be used as a component of another container.

To meet the strategic development of the Open University of China, some people proposed to build a distance education cloud based on cloud computing. By building a high-performance computing environment, it can quickly store, distribute, and push massive digital resources, realize high-

quality distance teaching transmission, and provide users with a personalized, one-stop, integrated learning environment and working environment, support personalized learning and individualized teaching, promote the development of learners' advanced thinking ability and group wisdom, and improve the quality of education. Use GUID and XML configuration files to improve resources Packing and transport functions. Establish a GUID identifier and XML configuration description file for each resource bundle, which can record and track the resource content and version of the resource bundle content in detail.

Improve the rights management function, which can flexibly set different rights for teachers, teaching managers, etc.: filter word management function, shielding words that should not appear on the teaching platform. It provides a multi-layer distributed application model, component reuse, and uniformity the unique security model and flexible transaction control, as well as the support for many middleware technologies, not only reduce the development work to a considerable extent, but also enable developers to launch creative customer solutions to the market faster, and the solutions are independent of platforms. Will not be bound by any one vendor's products and APIs. The emergence of the J2EE system not only facilitates the development of distributed applications, but also has incomparable advantages compared with the traditional Internet application model. Use the server's local disk to store data.

## 2. THE PROPOSED METHODOLOGY

### 2.1 Fast Storage Technology for Data Platform

The data is closely integrated with the application system, and the data capacity is relatively limited (about tens of TB). It can be expanded through DAS (Direct Attached Storage) technology, but the installation and debugging of system software is complicated. It is mostly used for personal computers and servers carrying small businesses. In the system, the provincial school teaching platform web service provides a public access page, and all teaching points provide a unified access page, and the background points to the server; the provincial school resource library server stores non-video

resources in the platform, and the server function of the teaching point can be controlled by the platform management server replaces; the video resource server stores all video resources; the resource management server manages the directories and storage locations of all resources, and records the basic information of each teaching point server; the directory center database stores the data information of each teaching point, which is synchronized with the program server coordinates the work and completes the automatic updating of resources.

The J2EE architecture is a multi-layer distributed system. In this architecture, the teaching resource library and teaching management library are all stored and managed in the form of database. Since teachers, students, teaching resources and teaching management are scattered in different geographical locations, the distance education platform is essentially an integrated platform of distributed database resources. Therefore, if the J2EE multi-layer structure is adopted when constructing the distance education platform, the user interface, business logic and data can be well separated. MapReduce is an easy-to-use software framework. Applications written based on it can run on large clusters and process PB-level data sets in parallel in a reliable and fault-tolerant manner.

A MapReduce job usually divides the input data set into several independent data blocks, and the map tasks process them in a completely parallel manner. The framework will sort the output of the map first, and then input the result to the reduce task. Usually, the input and output of the job will be stored in the file system. The entire framework is responsible for scheduling and monitoring tasks and re-executing failed tasks. Composed of multiple storage devices, different storage devices need to use technologies such as cluster technology, distributed file system, and grid computing to realize the collaborative work between multiple storage devices, so that multiple storage devices can provide external the same service, and provide larger, stronger, and better data access performance.

Without the existence of these technologies, cloud storage cannot be truly realized. The so-called cloud storage can only be an independent system one by one and cannot form a cloud-like structure. The central objects in this business logic module are students and teachers. It mainly describes that after students choose courses, teachers decide which students to choose according to the students who choose courses, and finally teachers give credits to students. The applicable objects of this functional module include students, teachers, and administrators, including 3 basic processes. Other column clusters are used to store various information of remote resources. The meta column cluster is used to store the basic information of the resource. Since the basic information may include resource title, introduction, and author, etc., three columns (meta: title, meta: info, meta: author) are designed to represent these three types of information respectively. The text format content of educational resources is saved to the c-text column cluster. Since it may contain text attachments such as teaching plans, slides, and test questions, 3 columns (c-text: plan, c-text: slide, c-text: test) are designed to express.

## 2.2 Data Sharing and Storage Strategy of Distance Education Platform

Apply the snapshot difference algorithm to compare the generated new snapshot with the last snapshot and output the result to the incremental file. This step is completed by the snapshot difference module, which calls a certain algorithm in

the snapshot difference algorithm library to complete this process, and its input is a snapshot file, and its output is an incremental file. The middle school student curriculum and grade management function module of the teaching platform is developed using ShouIts and Hibernate technology and hibernate is used to operate data objects. Here, a HibernateUtil class is defined to be responsible for initializing Hibernate. It creates a global SessionFactory instance, and provides utility methods for creating Session instances, closing session instances, opening/closing transactions, and recreating SessionFactory instances. And all methods are static methods. In the second step, due to the large number of rows in the ResourceTable table and the large number of keywords in each row, many keywords need to be compared for each retrieval, resulting in a performance bottleneck. It can be processed by MapReduce cluster on HBase. The specific method is as follows: store the keyword column clusters of the ResourceTable in multiple HBaseRegions in a dispersed manner, process multiple HBaseRegions in parallel through the map method (that is, compare the keywords), and finally summarize the keyword comparison results through the reduce method.

The method of configuring HBase to use MapReduce. To prevent the loss of system data in unexpected situations such as power failure, and ensure the efficiency and security of data transmission, this function is introduced. Different from other resumable uploads, this system adopts the upload method based on web Service. Because this method uses XML to transmit data, it is easy to expand and migrate, and because it uses port 80 of the WEB service, it can freely penetrate the firewall without hindrance when transmitting data. Before Ajax, web-based applications had to submit entire pages to validate data or rely on complex JavaScript to check forms. While some checks are simple enough to be written in JavaScript, others are not and cannot be written entirely in JavaScript.

Also, every validation routine written on the client side must be somehow rewritten on the server, since it is possible for the user to disable JavaScript. After completing the above configuration, when querying the keyword column cluster, HBase will use the MapReduce method to perform the query in parallel, thereby improving the efficiency of the query. Considering two factors, the system adopts WS. Security to ensure the data security of Web services. It defines SOAP extensions that allow the passing of security tokens. The framework built with WS-Security can exchange security messages in a heterogeneous Web service environment, so it is very suitable for heterogeneous distributed resource library systems.

## 3. CONCLUSION

Aiming at the large amount of data of distance education resources and the characteristics of rapid growth, this paper designs a set of storage and retrieval methods for massive distance education resources based on Handoop, using the idea of distributed storage and parallel computing. Compared with the traditional shared storage method, this method not only has low cost, but also supports efficient content-based retrieval and improves the recall rate. The database is updated automatically. Adding this function makes the management and operation of the platform more convenient and simpler for grassroots managers. The paper proposes a solution and completes the functional program design.

#### 4. ACKNOWLEDGEMENT

**Subject: 2018 Guangdong Open Distance Education  
Research Fund Project: Internet plus background of  
distance education and occupation education integration  
path YJ1817**

#### REFERENCES

- [1] Li Yongqiang. Research and development of heat metering system based on wireless network and remote monitoring [D]. Jilin University.
- [2] Gu Xu. Mobile phone information remote storage and interaction platform based on 3G network [D]. Hunan University.
- [3] Zhu Xiaoli. Research on storage method of massive distance education resources based on Hadoop [J]. Journal of Shandong Radio and Television University, 2012(3):4.
- [4] Li Hanwen. Research on Remote Monitoring System of Truck Crane Based on Internet of Vehicles Technology [D]. China University of Mining and Technology, 2022.
- [5] Jiang Jiandong, Zhou Qian, Pan Bosong, et al. Remote Cloud Monitoring and Traceability Research and System Design of Tea Processing Process [J]. 2022.
- [6] Bei Yayao, Wang Xiaogang, Song Tao, et al. Remote Analysis and Management Platform for Eddy Current Data of Steam Generator Heat Transfer Tubes in Nuclear Power Plant: CN, CN102541009 B[P]. 2012.
- [7] Xiong Yan. Research on storage of ECG database based on Hadoop[J]. Biomedical Engineering Research, 2016(35):177.
- [8] Chen Shuo; Shen Li; Li Jinhui; Li Fengqiang; Yuan Chen; Research on the selection of power grid unstructured data and real-time data storage technology [C]// 2017 "Electronic Technology Application" Smart Grid Conference. 0.
- [9] Wang Ping. Research on the application and optimization management of cloud storage technology in network data security [J]. Newspapers and Periodicals: Part 2, 2018.
- [10] Du Dongxu. Research on distance teaching platform suitable for massive data storage in the western region [D]. Lanzhou University.
- [11] Zeng Ke. Research and Design of B-S Architecture Sewage Remote Monitoring System on Virtual Platform[J]. Electronic World, 2017(20):2.
- [12] Hu Haiguang. Application research on data storage and security of drilling engineering projects [D]. China University of Geosciences (Beijing), 2012.
- [13] Gu Xu. Mobile phone information remote storage and interaction platform based on 3G network [D]. Hunan University.
- [14] Wen Liquan. Research and Application of Remote Monitoring System Based on Embedded Web [D]. Dalian University of Technology.
- [15] Dai Cheng, Ye Yan, Liu Taijun, et al. Research and Implementation of Remote Automatic Control System Based on WEB [J]. Radio Communication Technology, 2014, 40(3):4.
- [16] Wang Xiuqing, Liu Qing, Zhao Jimin, et al. Research on remote monitoring system for greenhouses [J]. Journal of Tianjin University of Science and Technology, 2018, 33(1):4.