

The platform design of the cloud instructional resources based on the hadoop framework

Xu Dong
Beijing Jiaotong University
Beijing, China
11120358@bjtu.edu.cn

Wei Huiqin(The Tutor)
Beijing Jiaotong University
Beijing, China
hqwei@bjtu.edu.cn
Zhang Xiaoxin
Northeast Normal University
Changchun, China
zhangxx777@nenu.edu.cn

Abstract—With the continuous development of Internet technology, the national colleges are actively building own teaching platform and the instructional resource database. And the construction of the instructional resource platform has played a good role in improve the teaching quality. Under the background of rapid development of cloud computing, using cloud computing technology to build the cloud instructional resources platform is a new topic in educational fields. This article aims at using hadoop cloud framework to build a opening, flexible and feasible cloud instructional resources platform.

Keywords—component;cloud computing, instructional resources,hadoop, resources sharing

I. INTRODUCTION

With the science and technology developing continuously, the information technology has a revolutionary influence on the development of education. *The development plan of the educational informationization for ten years of China* points out that by 2020:the information-based learning environment of the high quality educational resource that everyone can enjoy will be basic completed; various kinds of educational digital resources has become increasingly rich and are widely Shared; the public service platform of the high quality education resources will be built; the resources construction mechanism of the government guidance, participating, sharing constantly improve; the digital divide will be significantly narrowed; the high quality educational resources informationization environment will be basic formatted. *The Development plan* stressed that priority of development of educational informationization is the useful method to reduce the digital divide and promote quality education resources to be shared.

The concept of digital campus is put forward in 1990 by Kenneth Greene who is the professor of the Claremont Colleges. After decades of development, the digital campus has also made great progress in China. As an aspect of the digital campus construction, the construction of instructional resources platform has got the significant development[1], but also it still has some problems which need to be solved:

- First, the management mode of the platform lack innovation and it is unitary; the main force for constructing and maintaining the resource platform is the teacher, the beneficiary of the platform are rarely involved into the platform management.

- Second, the process for producing resource is single; the resources that be developed or collected will be saved to repository, and accumulated the simple learning resources, but don't have the continuous update resources. The feedback mechanism of the resource usage is lacking.
- Third, the costs of the construction of the resources and platform is huge,and the design for the resource sharing is ugly. Every school makes effort on their own instructional resource platform. Assuming that each institution needs 200,000 RMB for the platform, in China, there are 1,200 schools and the whole cost at least 240 million RMB[2]. From the data, the size of the digital campus construction can also been seen.

With the continuous development of the Internet, cloud computing, as a new computing model[3], with its high reliability, large scale, visualization, scalability[4] and a good ability for the development of the Internet has entered a new era. The core of the cloud computing infrastructure is resources sharing.Its basic principle is to use a local or remote server cluster for Internet users to provide services. Based the shortage of teaching resources platform under traditional mode(B/S structure), using hadoop framework which is a cloud computing open source project supported by the Apache foundation builds the model of the creating resources as a starting point, resource access method of cloud computing based on web-service encapsulation instructional resources sharing platform. In this platform, the user involved into the development and maintenance of resource and it will integrate with the user's evaluation mechanism. Resources will be constantly optimized in use process.

II. CLOUD COMPUTING AND THE HADOOP FRAMEWORK

Cloud computing is a computing model through the Internet providing scalable resources, it refers to a service business that makes a large number of various types of resources work together through the cloud computing platform, and commonly provide data storage and access capabilities for the external through the application of the cluster[5], Internet technology[6], distributed file system[7], and other functions.

Hadoop's core component include the Distributed file system (Hadoop Distributed File System, HDFS) and MapReduce framework of data processing, which are used

to store and analyse a large of data sets[8]. Hadoop, as one of the open source cloud computing platforms, has made a great development both at home and abroad, people from all walks of life gave it a high evaluation for its excellent data processing and storage capacity, including Stanford university at abroad, university of California, Berkeley, the domestic first-class university, Tsinghua university, RenMin university and research institutions launched numerous academic studies on it. Figure 1 is the whole architecture design of hadoop

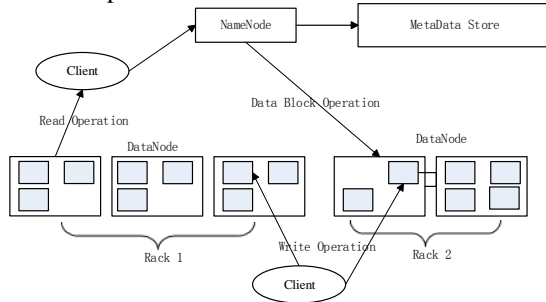


Figure 1. Hadoop's overall architecture design

The hadoop platform allow users to construct their own distributed cloud computing platform at the least cost. Users can easily build their storage of huge amounts of data on the cloud platform, but also can carry out for mass data processing applications. Integrating the characteristics of the instructional resources, we can summarize the advantages of the hadoop cloud platform for the instructional resources:

- The high stability of the resource platform. After the resource platform come into service, It will produce large amounts of data access, it needs good handling and response speed to give a good user experience, which has a high requirement for the stability of data access. The hadoop distributed computing center jump mechanism is used to judge whether data node (the data node) to work properly, namely every time takes the initiative to report to the master node current running state. If it is over a certain time threshold, it will be determined that data node has failed, then the master node (the name node) via meta data node for data block data is copied to the other standby node, to ensure the stability of data access.
- Resource platform has great data processing ability and speed of access to resources. The data of the resource platform has great data processing ability, data processing effectively and speed of access to resources. Traditional data processing tasks in the distributed system is a mobile data, hadoop break the routine method, which distributes the data processing tasks to the node, thus forms the high efficiency of the platform. Accessing the data, users only need to access the location of the resources in the primary node query, the master node do not participate in the rest of the transfer task, replaced by data node, thus providing the efficiency of the data access.

- Instructional resources platform has good expansibility. As the instructional resources have been developed, for storage capacity of the platform, it will inevitably put forward new requirements, need to constantly extend resources storage nodes, to meet the need of users to access. Hadoop distributes and finishes computing task between the available computer cluster data. Cluster nodes can be easily extended to thousands of nodes. When a new machine is connected to he cluster, you just need to copy the configuration confidence information of the hadoop, establish SSH(Secure Shell) access between nodes. The nodes will automatically do the data replication work according to the configuration information and quickly is transformed into the data nodes, which overcomes the the poor scalability problem under the traditional B/S architecture.

III. THE DESIGN OF THE INSTRUCTIONAL RESOURCES PLATFORM OF THE CLOUD COMPUTING

A. The principles for the design

In order to ensure the smooth operation of the platform design task, first statement about several design principles:

- Compatible with Sharing principle: the construction of instructional resources of the cloud platform follows the principle that *integrating the existing resources, sharing all resources to everyone*, and making full use of existing resources, constantly adding new resources.
- Safety principle: it has identity authentication mechanism, and establish user authority data table, to ensure that different users can only use the corresponding function in the system.
- Opening principle: the instructional resources of the cloud platform is an opening, standard cloud platform. Hadoop technology overcomes the problems of hardware expansion, thus resource platform needs opening, scalable data format. According to the different application requirements, it provides an open interface in light of the general data format, and accesses resources through access to the interface and secondary develop.

B. The design of the whole framework

With the rise of construction of sharing resource, many colleges and universities are actively building their own sharing repository. The non-unified standard, non-unified storage format of the construction of the resources, non-unified platform development technology and a series of factors among colleges and universities led to that sharing problems between the platforms. Good scalability of hadoop can overcome the lack of extensibility of the traditional B/S structure. The instructional resources of the cloud platform will store the data into the cloud resources database based on hadoop, using the powerful storage

capacity and flexible extensibility. Besides storage function of instructional resources, the the cloud platform of the instructional resource, the functions of the resources re-development, supporting diversified learning and boutique display on the basis of the instructional resources are necessary, which can achieve optimize resources during using of resources and form a good ecosystem for resources. Instructional resources of the cloud computing platform framework is shown in figure 2.

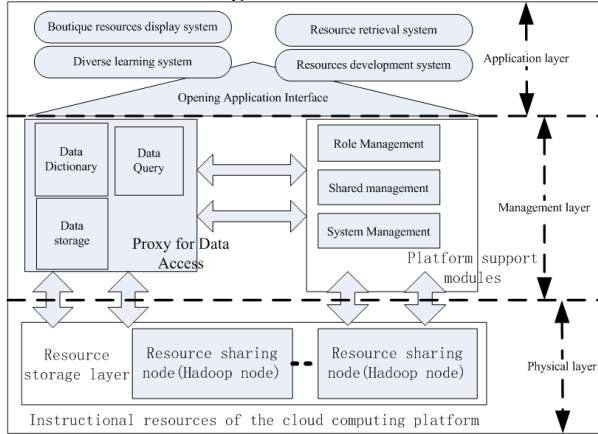


Figure 2. Instructional resources of the cloud computing platform framework

C. The design of resource storage

Instructional resources is the core of the cloud computing resources platform, is the main object of cloud platform management. The key point of the establishment of the cloud platform is: whether the optimization of storage design, utilization of the instructional resources is in line with the instructional rule. Therefore, the design of the storage of the instructional resource should not only have good instructional resource organization and management mode, but also need to support the diversification learning in a certain extent.

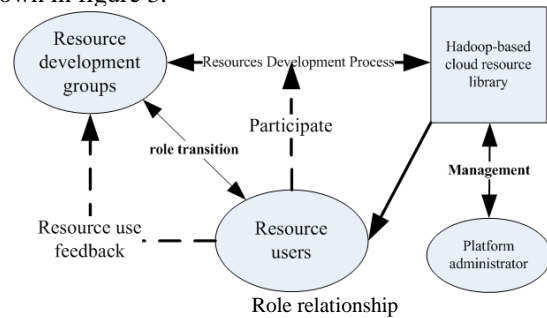
In the cloud platform of instructional resources, the resources are classified according to the professional categories, the construction and management of the resources are divided on the basis of the professional code. Using unified professional code, library code and course system can reduce the repeatability of the construction of the learning resources, so cloud platform adopt the unified professional code of the instructional resources to construct the learning resources. Using unified professional code to build the professional library and the professional curriculum system, the specific modules are refined to the various professional to be responsible for sorting and building. Thus, on the base of the optimization of the resources retrieval, the user can also learn autonomously according to the professional, library and courses.

According to the hadoop's file storage rules *write once, read many times*, when uploading process of a resource file is completed, the next time it is need to delete the current file first and then write. Using this feature, you can provide multiple revision version for resources, save changes every

time for a revised version. Resources being accessed, the user can choose different versions according to different requirements. Every resource at the same time, according to the selected professional code in the longtime is stored the parent level fold named with code number, and then according to the resource library number the resources are deposited in the corresponding folder. To optimize the search speed of the resources, it avoids through the API interface provided by the hadoop file system to find every time whether there is a corresponding resources, so using multilevel XML files to store the basic information of the resource file, when finding, it is just according to the keywords to search XML documents to determine the specific location of the document.

D. The design of the resources feedback and role relationship

During the process of the use of resources to optimize resources, the cloud platform of the instructional resources adopts the new type of resources development and management mode with the support of the team resources, to redefine the resource user roles. In the foundation of adding resources users' direct feedback to the production team of resources, it allows users to directly involved in the resources forming process. And the resource user and the development team roles are interchangeable. Developers can use the resources already available in the repository for the secondary development of the resources. Role relationship is shown in figure 3.



E. The design of resource accessing

The concept of the cloud computing is that the resource that the users need not run on a personal computer or terminal. So the instructional resources and computing resources needed by the user which all provided by hadoop large-scale cluster. The instructional resources that the users process is not stored locally, but is saved the repository of the platform. The platform system center is responsible for providing persistent data storage management and data maintenance, while maintains strong enough computing power and large enough storage space. Users can anytime, anywhere through any terminal equipment are connected to a cloud of the instructional resources platform.

In the past resource platform, most of the ways of access to resources is based on the browser. Therefore, in order to maximum in a way compatible with the previous access to resources, we can reform the WEB content of the

existing instructional resources platform of cloud platform. Main content is firstly the switch interface of the resource access, which uses the cloud platform to provide comprehensive resources access interface and upgrades the existing resource access interface according to the requirements of the cloud platform. Followed by adding new resources to access the page, the cloud platform shows its own characteristics.

IV. DEPLOYMENT OF THE PLATFORM

The storage construction of the platform resources is the core content, which provides platform for autonomous learning at the same time, opens all resource access interface for the users, realizes the role of the new relational schema, develops multi-user types of the application systems. The platform uses PaaS (the abbreviation of Platform-as-a-Service, means Platform as a Service) model of development, unified enterprise team resource construction and the management permissions of colleges and universities, realizes that management mechanism that the ordinary users change into the professional team change user.

Realizing resource access interface through an open application platform, can synchronously operate the selectivity and points of learning resources through online broker, to ensure the consistency of Shared data between platforms and platforms, between client and platform, and the application of resources construction and sharing of consistency between systems. The platform achieves the mechanism of the scores and excellent resources recommendation, to encourage builder of the resource. And it provides ranking system at the same time in the total platform, such as assessment, and integral, attention ranking mechanism.

The cloud computing platform based on hadoop will be according to the physics, the management of the instructional resources, application of three layer architecture approach for deployment. The physical machines as the carrier build the cloud storage cluster of the physical instructional resources of the cloud platform, according to the different categories of resource for planning a physical node, the node to distinguish between different types of resources, such as video storage nodes need efficient processing speed and greater bandwidth to support the watch and download of the videos, then some of the smaller documents can be stored in some nodes made up of commercial PC resource. In the cloud platform of the instructional resources, hardware resources (including network, host, storage, etc.) is maintained by the management, the way of the resource access of hardware resources is encapsulated interface, which provides a friendly, open application interface for the application layer.

Application layer on the basis of the management of the interface, builds localization cluster system in accordance with the requirements, in accordance with the school requirements for the level of resource nodes. Schools can

according to their own professional characteristics and the existing resources display platform, organize the form of organizational resources for showing, so that to avoid a lot of inconvenience that the monotony unified display system brings for teaching. In order to make full use of existing instructional resources, providing the institutions existing resource platform and cloud platform to integrate instructional resources conversion interface agent realizes the professional resources co-construction and sharing.

The sharing of the learning resources uses both online and offline sharing mechanism. The resources Shared library that the team or individual upload synchronize to the cloud storage cluster, and finally the resources are spread by the management, the management assigns to each access to the cluster.

V. SUMMARY

As the increasing in the field of country education investment, the pace of the modern technology accelerates into the classroom. Cloud computing into the field of education is the inexorable trend in the future. The cloud instructional resources platform based on hadoop will provide the open and safe resources construction platform for the national colleges and universities and the service system of instructional resources for the society, which is completely conform to the state for the purpose of the education resource construction of public service. With the using and development of the cloud computing technology in the field of education, it will bring an important part of the change for the future way of teaching.

ACKNOWLEDGMENT

We would like to thank all the friends who believed in my opinion and give us help and encouraged. Also, we would like to thank all teachers who shared their ideas and give us advices.

REFERENCES

- [1] Zhang Xiangyang. Based on cloud computing platform architecture and application of teaching resources functional studies [J]. Coal technologies. 2012.31 (1): 239-242.
- [2] Wang Lu Ming. From traditional computer-assisted instruction to the cloud to support teaching Development Studies [J]. Private Technology. 2011 (1): 62-63.
- [3] Kou Yuanyuan Network Teaching Platform Development Status and Trends [J]. Electronic design engineering. 2011 (6): 122-124.
- [4] Zhangjia Gui. Luolong Tao. Based on the concept of cloud computing platform for building digital teaching resources [J]. Modern Educational Technology. 2011 (3): 100-101
- [5] Bend steel. Deng Jianqing. Han Yun Lu. Linux cluster technology research [J]. Application Research of Computers, 2005 (5): 100-101.
- [6] Sutanto D. Special Section on Smart Grid Technologies [J]. Journal of Electronic Science and Technology. 2011(1): 1-2.
- [7] Tobbicke R. Distributed file systems: focus on Andrew File System/Distributed
- [8] Di Yongdong. Hadoop Distributed File System (HDFS) Reliability and Optimization Research [D]. Huazhong University :1-2.