Research on the Construction of Unstructured Data Center Based on Cloud Storage in Universities——Taking Shaanxi Normal University as a Case

Qi Xu

Network Information Center, Shaanxi Normal University, Xi'an, Shaanxi 710062, China.

xuqi@snnu.edu.cn

Abstract. In view of the large number of scattered and unstructured data problems in the process of comprehensive informatization in higher education, the article proposes to build a cloud storage platform to provide users with private cloud storage with high capacity and high reliability, and to exert aggregate value of various resources, Provide a space for active sharing and collaboration, realize the precipitation and sharing of teaching resources, and then build the ideas of unstructured data centers in universities. Taking the construction of an unstructured data center based on cloud storage at Shaanxi Normal University as a case, the article describes in details on plan design, implementation, application effect, and the post-planning.

Key words: Unstructured Data, Cloud Storage, Education Resource Center.

INTRODUCTION

With the coming of the era of big data, massive digital resources in universities are increasing rapidly. In this case, there are unstructured data such as documents, forms, images and other unstructured data, which contain excellent teaching resources, management information, research results, etc. Therefore, it has become one of the important research topics in universities how to build unstructured data centers in universities, to store and manage the data in a unified way, to realize the sharing and utilization of these high-quality resources and to excavate the value behind these data.

ANALYSIS AND STUDY OF UNSTRUCTURED DATA IN UNIVERSITIES

Definition of Unstructured Data

Data in computer information system can be divided into structured data and unstructured data. Unstructured data is irregular or incomplete data structure. There is no predefined data model, which is not convenient to use database two-dimensional logic table to express data. It includes all kinds of office documents, texts, pictures, XML, HTML, reports, images and audio/video information, etc.

Unstructured data is information that either does not have a pre-defined data model or is not organized in a pre-defined manner. This result in irregularities and ambiguities that make it difficult to understand using traditional programs as compared to data stored in fielded form in databases or annotated in documents. [1] So the storage, retrieval, publication and use of more intelligent IT technology, such as massive storage, intelligent retrieval, knowledge mining, content protection, information value-added development and utilization, etc.

The Present Situation of Unstructured Data in Universities

The Challenge of Data Storage and Management

In the new wave of science and technology, universities have also ushered in the "big data age". All kinds of business systems have been launched in succession: MOOC system, educational system, OA system, mailbox system and so on. The shaft type business system broke out massive small files, video and pictures. Secondly, teachers and students have massive pictures, videos, various documents, reports and so on, universities have produced and accumulated a large number of unstructured data, but there is no unified planning and management of these data, which are basically decentralized storage, management and use. This traditional, chimney type management idea leads to disordering management of unstructured data. It is difficult for data mining, management and analysis, which makes it difficult to embody the value of this part of data.

The Challenges of Data Use and Sharing

The university teachers, students, administrative staff and various departments have produced a large number of unstructured data resources, which are hard to find, easy to lose, and cannot meet the needs of unified access and sharing of data resources and collaboration rights. In addition, university users use different terminals (mobile phones, Pad, notebook computers, desktop computers), and cannot achieve unified access to data between different terminals. Data exchange and sharing between teachers and teachers and students are hard to satisfy.

The Challenges of Data Security and Analysis

There is no unified planning and management of unstructured data in universities, which makes these data not only inconvenient for individuals to store, use and share, but also cannot obtain the value of data teaching and scientific research for universities, resulting in the waste of a large amount of value of data. Secondly, these data storage and transmission have no security mechanism, it is easy to be attacked by computer virus, Trojan horse, network attack and so on. Once security problems occur, these data will be lost and damaged.

RESEARCH ON THE CONSTRUCTION OF UNSTRUCTURED DATA CENTER IN UNIVERSITIES——TAKING SHAANXI NORMAL UNIVERSITY AS A CASE

Design of Unstructured Data Center Based on Cloud Storage

In view of the current status of unstructured data, Shaanxi Normal University began to carry out the construction project demonstration and related testing work in 2016 and plans to conduct unstructured data center construction in two phases. In the first phase, we plan to create a whole-school teaching resource center based on cloud disk, to realize the centralization and sharing of teaching resources, and to provide campus cloud disk service for teachers and students. The second phase plans to build a unified unstructured data center in a multi cloud environment and distribute unstructured data in a multi cloud environment, such as desktop cloud, communication cloud, and business cloud, to store, manage and use unstructured data in a multi cloud environment. Through collecting, analyzing and forecasting these data, we can provide data support for school management and decision making, and finally achieve the construction of unstructured data center. The results of this center are as follows:

Shared Teaching Resource Center

Gather all the teaching resources, and the courseware and materials produced by school teachers, so as to achieve one-stop teaching, lesson preparation, teaching and data searching.

Provide Private Cloud Service for Teachers and Students

The learning resources and open curriculum resources of all parties are brought together to achieve self-learning and interaction based on cloud disk. It provides multi version security, access anytime and anywhere for documents and information generated by teachers' scientific research and teaching process.

Teaching Archives Information Center

It forms the collection and management of teaching archives and documents, including the collection, sorting, querying and archiving of teaching administrative system and teaching administrative documents. Integrate with various application systems such as teaching management to form an independent unstructured document center, provide independent document sharing and knowledge publishing services, and better serve the digital campus.

Implementation of Desktop Virtualization Applications

Integrate the platform and desktop virtualization applications to achieve fast backup and recovery of desktop files. When users are out of business, the data center documents can be accessed directly by cloud disk instead of starting the virtual machine.

Unified Storage and Archiving of All Kinds of Business System Attachment Documents

At present many business systems (such as OA, e-mail, multimedia video system) the attached document directly stored in the database, or NAS / SAN data storage, these attachments increase very rapidly, leading to the high performance storage space occupied powerful, and easy to fill, it is difficult to complete the backup protection of various business systems; Attachments data are scattered and stored, and they cannot be managed uniformly, which is not conducive to compliance check, tamper proofing and retrieval.

Intelligent Large Data Analysis Service

Through the construction of unstructured data centers in schools, the education and teaching resource data, various types of document data, business system data, and archival data will be converted, cleaned, and integrated in the analysis background to establish a complete, rational, and rigorous data system. The school's unstructured in-depth data realizes data management and feedback functions and provides data support for school management decisions.

Unstructured Data Center Architecture Diagram Based on Cloud Storage

As shown in Figure 1, based on the service mode of cloud computing technology, the unstructured data center of university is composed of four layers: the hardware infrastructure layer, system layer, data application layer, and application access layer. As the top-level service application access layer of the data center, a unified portal can provide teachers and students with multiple access services such as WEB and mobile clients, as well as standardized API interfaces and open application components. The data application layer is responsible for the core tasks of the data center. The system deployed at this layer can implement core services such as document organization, user organization, log system, and object storage cluster, etc. The system layer is mainly responsible for implementing node management, access scheduling, copy control, and data storage. The lowest level of the hardware infrastructure layer is the storage server and network infrastructure. It can integrate different brands, types, configurations of servers and storage resources into a unified data storage space.
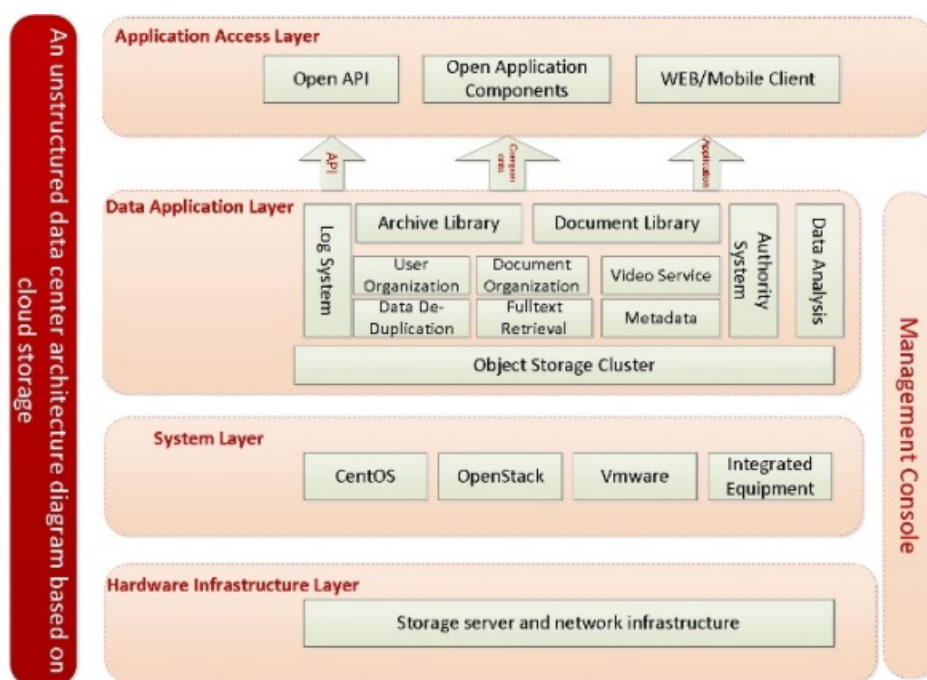
FIGURE 1. Unstructured data center architecture diagram based on cloud storage.

Application of Unstructured Data Center in Universities Based on Cloud Storage Platform

Construction and Implementation

Universities can use existing storage resources to build cloud storage clusters, or directly purchase server storage devices, and server-side storage nodes form a cloud storage cluster environment to provide cloud storage services. In the first phase of construction, Shaanxi Normal University purchased 8 integrated server storage devices, using distributed data processing mechanism to achieve parallel processing of data, support dynamic expansion of platform hardware, eliminate single point failure, and damage any hardware equipment, and do not affect business continuity. The server storage device is configured with two 2.0GHz six-core Intel Xeon E5-2620 processors, 64GB high-speed memory, a 120GB SSD cache disk, a 120GB SSD full-text search cache disk accessory, and 24 6T NLSAS hard disks for data storage. The operating system and software system use SSD hard disk storage to separate the system from user data and improve security and performance.

On the software platform, it provides cloud service for teachers and students to achieve multiple accesses to multi terminal (PC, mobile phone and PAD) of campus users, and provides personal storage space and public resource access directory to realize the sharing, access and search of diversified resources based on the application; Online editing module can realize online editing of Microsoft Office documents; security of user data in anti-virus system security system; illegal content management and control system can effectively isolate illegal data in the system and ensure the safety and compliance of system management; Provides system performance monitoring, including service status, storage space, disk status monitoring and performance monitoring, etc.; access log analysis system, mainly used for log collection, statistics and analysis, mainly including online number monitoring, file classification quantity statistics, files Operational change statistics, file download count statistics, user activity statistics, user statistics of non-school use systems, etc. Through log push and analysis, we can make statistical analysis of user behavior, data life cycle and trend of change, so as to improve system operation and maintenance efficiency.

Application Situation and Effect

Taking the cloud storage data center built by Shaanxi Normal University as a case, the school has planned multiple entrances. The platform integrates with the school's unified identity authentication, provides WEB version and multiple client access methods, and integrates with the school portal system, single sign-on access, improve user access experience. The first phase of the project was formally launched in June 2017 with an implementation period of approximately 3 months. The total capacity of the school's cloud storage system is 1152T. Due to the adoption of the three-copy technology, the actual usable capacity is 384T, and supports >=6400 concurrent users.

Provide teaching staff cloud disk service assisted teaching and research.

There are currently more than 3,000 registered teaching staff users, and each staff member is allocated 100G network disk space to provide school private cloud storage services. Up to now, the total number of file uploads has reached 1.72 million. There are more than 50 online users each day. Users can share these resources by two ways, internal sharing and link sharing. The internal sharing is used to share resources to internal users, and link sharing provides a method for external user to visit the internal resources. These two ways have many application scenarios, for example, the teacher share resources to class students for self-learning, collect student's work, add comments and etc. Absolutely all resources access is based on many controls of access privileges, such as access passwords, access deadlines, access rights, access number and etc.

Providing a department document library to achieve collaborative work.

Set up a department document library for secondary school units to achieve hierarchical management and access to documents. You can create sub-departments and designate administrators to manage sub-sector documents to ease the pressure of higher-level management personnel. Each department can classify documents according to needs and increase Document management flexibility and autonomy. In collaboration office, it provides fine-grained authority configuration management function, supports eight kinds of fine grained access rights, supports document owners to quickly and flexibly share documents to other people. The feature of group files provides a way of file collaboration. To better support this way, the platform has a series of useful functions, such as multi version of security protection, message reminding, anytime and anywhere access services. These methods actively promote the ability of teaching, scientific research and office assisting.

Provide quality resource center services.

According to the school resource bank construction plan, use the school's cloud storage system to establish related resource libraries by category, such as quality courses, teaching and research resources, subject resources, and knowledge bases, etc., for the relevant departments and professions to share and publish, teachers can research the quality resources generated during the teaching process will be shared and knowledge will be released. Students can learn and interact autonomously based on these resources. In addition, all parties gather learning resources and open curriculum resources to link the current library of schools, such as excellent courses, MOOC resource pools, etc. These resources are integrated into the school's unified resource center, not only enriching the school resource center, but also effectively save the existing storage resources of the school, improve the utilization of resources, provide a unified platform for the sharing of quality resources for teachers and students.

Provide security and stability platform services.

The system administrator can clearly understand the operation of the entire cloud storage platform through a visualized view. It includes the status of viewing server storage, the number of online users, the number of files, and operation logs. When the server storage fails, it is automatically given Administrators make alarms to ensure the security and stability of cloud storage. All user operations on the cloud storage will be recorded in the audit log, so that it is well documented, compliant, and safe and controllable. It is an unstructured data center. Provide unified storage management and efficient and convenient use of the platform.

Post-Planning of Unstructured Data Centers in Universities Based on Cloud Storage

Based on the existing construction, Shaanxi Normal University will continue to optimize and improve the unstructured data center. The main plan for the second phase of construction will be explored in the following aspects.

Deep Integration Business System

The school plans to build a unified, unstructured data center in a multi cloud environment. The school will be dispersed in non-structured data such as desktop cloud (personal data), communication cloud (mail, IM, etc.), and business cloud (OA, MOOC, etc.). Storage, management, and use provide centralized management of document storage for school applications. At present, the construction of the school cloud storage platform is an open platform, which provides a series of standardized interfaces for third parties, such as file operation API, directory operation API, user organization API, authority control API, backup and recovery API and so on. In the case of authorization, third party applications can connect with the cloud storage through interface development, integrate the unstructured data scattered in the business environment, and finally achieve unstructured data storage, management and use, helping the construction of school information.

Providing Intelligent Data Analysis Services

In addition, when we build the school's unstructured data center, we can use these valuable data to carry out large data analysis through the acquisition, analysis and prediction of a large number of unstructured data accumulated in the study, scientific research and life of teachers and students, so as to achieve intellectual decision-making for school teaching and scientific research work. Through the unstructured data analysis system, it provides school-wide data query and detection, data report and report query services, and provides powerful data support for the use of unstructured data in schools and decision-making. Through data analysis, teachers and students can know which type of teaching resources are the hottest, the most concerned by students, and the characteristics of high-quality teaching resources. This way provides data support for improving the production of teaching courseware and resources, and then improves the quality of school teaching resources.

SUMMARY

In the process of accelerating the integration of information technology and education, there are a lot of unstructured education and teaching resources, how to give play to the aggregate value of all kinds of resources, and to integrate the wisdom of teachers and students at the same time, to provide an active sharing and cooperation space, which is the goal of the construction of the unstructured data center in universities. It extends to collect and analyze these data, make them play a valuable role and provide decision support for the future development of universities. In addition, we should strengthen the supervision and system construction of unstructured data in universities, prevent and stop the occurrence of illegal content and tort, and reduce the risk of data mining. These are the contents to be considered in the construction of the data center in the next step.

REFERENCES

1. Unstructured data: en.wikipedia.org/wiki/Unstructured_data.
2. GAO Hong-qing1, Wang Hao2, et al. Research and Implementation of Educational Resources Integration based on Cloud. Modern Educational Technology. Vol. 20 (2010) No. 3.
3. Xing Cong-cong1, Xiao Tong2, et al. Framework of College Digital Resources Integration Based on Cloud Computing. Journal of Tianjin University of Science & Technology. Vol. 26(2011) No. 3.
4. Chen Qiao1, Shi Quan2, Shao Ye-qin3, et al. The Research of Personal Cloud Data Cabinet Construction under the Big Data Era. Modern Educational Technology. Vol. 23 (2013) No. 10.
5. Chen Wan-jing1, Liu Qing-qing2, et al. On the Construction of Instructing Cloud Resources in University. Modern Educational Technology. Vol. 23 (2013) No. 6.
6. Xin Hua-eheng1, Chen Xiang-dong2, et al. Cloud storage-oriented unstructured data storage. Journal of Computer Applications. Vol. 32 (2012) No. 7.
7. Bain Genqing1, GAO Song2, Shao Bilin3, et al. Security Structure of Cloud Storage Based on Dispersal. Vol. 45 (2011) No.4.