International Journal of Multimedia and Ubiquitous Engineering
Vol. 12, No. 5 (2017), pp. 91-100
http://dx.doi.org/10.14257/ijmue.2017.12.5.08

Heterogeneous Cloud Storage Based Platform of Chinese Traditional Folk Art Resources

Xiaobo Li[1], Zhongjuan Tian[2], Ye Wang[1], Guowen Ye[1] and Zhen Ye[1, *]

[1]College of Engineering, Lishui University
Lishui 323000, China
[2]College for Nationalities (Minzu), Lishui University
Lishui 323000, China

Abstract

In an attempt to preserve and protect Chinese traditional folk art resources, it is an important and necessary task to digitize and store such information. However, because there are many types of folk art data information and the way to use these data is completely different, it is not suitable to use traditional relational database as a single storage solution. In addition, as the amount of data grows exponentially, such storage solution will also lead to performance and availability issues. In this paper, we propose a novel folk art resource data storage solution which consists of two parts, i.e. the underlying storage structure and the front-end centralized control component. The underlying data storage is composed of traditional relational database, NoSQL database and distributed file system, and different data can be stored in different storage structures according to their types and characteristics. The front-end centralized control component is responsible for handling user request and retrieving relevant data from underlying data source, and users can access this platform transparently and uniformly. We demonstrate that the construction of the heterogeneous cloud storage based platform is feasible and effective, which can play an important role in the field of the inheritance and protection of folk art resources.

Keywords*: Chinese traditional folk arts; cloud storage; heterogeneous platform; NoSQL*

1. Introduction

Each country has its own traditional folk arts, which not only reflect its unique culture tradition, but also is an important way to display its national spirit. As a unique art form, folk art reflects social life and folk culture from different aspects. Compared with other art forms, folk art is closer to people's daily life.

China, a country with a long history of civilization, has various types of folk art resources, which fully demonstrate its 5,000 years cultural deposits. However, with the transition from traditional agricultural society into modern industrial society, great changes have taken place in the people's production mode and lifestyle in the country, which leads to a profound change in the survival environment of folk art. Some types of folk arts are on the wane or even going to die out, and some folk artists is gradually aging, in the meanwhile, some exclusive techniques in folk art will be lost due to the death of those old folk artists. Therefore, how to preserve and protect the precious folk art resources becomes an urgent problem. Among many protection techniques, it is particularly important to use high-technology to digitize and store the existing folk art resources properly.

*Corresponding author: Zhen Ye, yezhen@lsu.edu.cn, +86-578-2299058 (Tel), No.1, Xueyuan Rd , Lishui 323000, P.R. China.

ISSN: 1975-0080 IJMUE

When storing folk art resources information, the involved data type includes folk art project information, folk artists information, folk art literatures library, folk art audio/video library and related news, *etc*. Currently people use traditional relational database to store above folk art resources data [1, 2]. However, such storage solution has many disadvantages, which include: Firstly, storage capacity is limited. There are many kinds of folk art resources, including a lot of multimedia data such as audio and video, and the total amount of data is very large which needs mass storage solution, but the capacity that the traditional relational database can provide is limited. Secondly, the data types are diverse, which means it is hard to store them in a unified way. Thirdly, folk art resource has various types of data, each of which has their own characteristics and the way to access them is also different. Accordingly, if we use single traditional relational database to store all those types of data, we cannot fully utilize different data's characteristic and thus cannot enhance the storage and access efficiency.

To address these difficulties, in this paper we present a novel heterogeneous cloud storage platform, where the underlying storage structure is composed of three parts. The first part is a traditional relational database, which is used to store structured data; followed by a NoSQL database, which is used to store semi-structured data; and the third part is a distributed file system, which is used to store unstructured data. Different types of folk art resource data will be stored in different storage structure according to their characteristics. In addition, the platform provides a transparent, unified interface module for data integration and data access, in order to facilitate the customers to use it.

The rest of this paper is organized as follows. Related works are discussed in Chapter 2. In Chapter 3, a heterogeneous cloud storage platform and its component are proposed. In Chapter 4, how to access such storage platform is introduced. Finally, we summarize the paper in Chapter 5.

2. Related Work

With the rapid growth of data quantity and the number of users, the traditional relational database is not able to meet the current demand for storage. Cloud storage [3], which is a basic infrastructure and key technology of cloud computing [4, 5], not only plays an important role in cloud computing, but also has been widely studied and applied in both industry and academia as an independent storage solution. Cloud storage connects to plenty of storage nodes by using series of complex technologies and forms a virtual and unified storage component in order to provide service. For the users, the underlying storage implementation is transparent, and the storage capacity of a cloud storage system can scale dynamically. Currently, the most widely used cloud storage platform is based on NoSQL [6], *e.g.* Google's Bigtable/GFS, Amazon's Dynamo, Apache's Hbase/HDFS.

Google's Bigtable [7] is a distributed data storage system which is used to store unstructured data. It is a sparse, distributed, persistent, and multidimensional sorting mapping, which uses *(row:string, column:string, time:int64)→string* to represent a key-value record. In the underlying layer, Bigtable uses Google File System [8], GFS in short, to store data and related log information. GFS divides one file into many chunks, each of which has fixed size, and distributes those chunks into many nodes evenly to achieve load balance. If we take the whole database as a big table, then we can divide this big table into many small basic tables, which are called as tablets. Tablet is the smallest unit that can be processed by Bigtable. In Bigtable, master server is responsible for distributing tablet into different tablet servers, checking newly added and expired tablet servers, balancing the loads among those tablet servers, collecting garbage files in GFS, *etc*. Tablet servers focus on serving read/write requests and divide one tablet into two when it becomes too large.

Dynamo [9], created by Amazon, is a highly available, and eventually consistency based key-value storage system. It uses many technologies to solve issues brought up by distributed environment and improve the whole system's efficiency. For example, it uses consistent hash to distribute data into different nodes evenly, which has high scalability;

and it uses vector clock to do data reconciliation and hinted handoff to handle temporary failure. In Dynamo, each computer node uses the same hash function to get a unique integer value, which is then mapped into a position, according to its integer value, within a virtual integer ring. Also, each data is mapped into a position in the ring by using the same hash function. Then, the data will be run in a clockwise direction along the ring, until the first computer node is found, and the data will be stored. To improve availability, the data will also be replicated into the following two nodes along the ring. One computer node will be mapped into many virtual nodes in order to distribute data into nodes more evenly.

Hadoop distributed file system (HDFS) [10], an open source version of GFS, is a distributed file system that can be run in common computers. Different from the current distributed file system, HDFS is more inclined to fault tolerance and device compatibility of cheap hardware. As a result, it can store and process massive amount of data in a relatively small budget. HBase [11] is a column based distributed storage database which is built on top of HDFS. It is deployed in a master/slave way. The master server manages the load balance and resource allocation. Zookeeper is responsible for maintaining the metadata and monitoring whole cluster's status to avoid single point failure. Each region server handles read/write requests to the data chunk.

NoSQL based cloud storage is suitable to store semi-structured or unstructured data. However, there exist some applications where there are many types of data, including the structured data that need to maintain ACID property and the other unstructured data. In such scenario, a heterogeneous cloud storage platform is needed to store various types of data [12]. For example, a heterogeneous cloud storage platform is used to store massive vector based geo-spatial data (*e.g*. GPS navigation and positioning points). It uses distributed file system of HDFS to store raster data, uses column based database of HBase to construct its distributed spatial index, and uses ACID compliance distributed graphic database of Neo4J to store vector data.

3. A Heterogeneous Cloud Storage Platform

Folk art resource has rich and varied forms, which means the data needs to be stored is also with large amount and various types. The folk art resource data includes folk art project, folk art talents, research literature library, audio/video library and related news information, *etc*. Folk art project data includes the origin of the folk art, its content, forms, *etc*. Folk art talents data includes both the folk artist and people who make some achievements during collecting, researching, creating and performing the folk art. Research literature library data includes published academic papers and documents on folk art. Audio/video data includes all folk art related sound music and visual information, such as mp3, MTV and many other formats. Folk art news information includes the news about government planning, policies and measures, and the news about related performance, training, conference and festivals. The news can be appeared in television, radio broadcast, newspaper, and network media. All this data contains not only structured text information, but also multimedia information, semi-structured and unstructured information, thus it is insufficient and unsuitable to store all of them in one single storage device.

In this paper, we present a heterogeneous cloud store platform used to store folk art resource (Figure 1). The heterogeneous cloud store platform consists of two parts, *i.e*. the underlying storage structure and the front-end centralized control component. The implementation of the underlying storage platform is composed of traditional relational database such as MySQL cluster, NoSQL based HBase and distributed file system HDFS. According to their characteristics, various types of data are stored within one of such data storage. The frontend of the platform is a centralized control component, which is responsible for handling user request, parsing it and forwarding it to the backend. In addition, the component is also in charge of doing user authentication, session

management, access control and concurrency control, *etc*. In the sections to follow, we will provide more technical details on the various components of the platform.

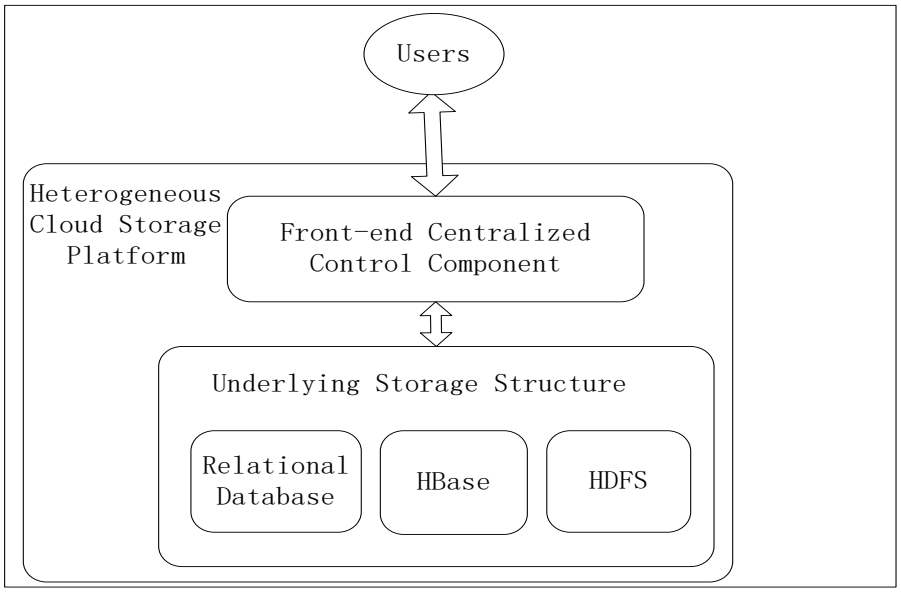

Figure 1. The Structure of Heterogeneous Cloud Storage Platform

3.1. Frontend Centralized Control Component

All requests to the cloud storage platform will pass through the frontend centralized control component. For users, the implementation of the underlying storage is transparent. Users send data query and store request to the control component by using a SQL like querying language. After receiving this request, the component can locate where the related data is stored, then parse, convert, and optimize the request, later retrieve the result from corresponding storage component and finally send this result back to the user. The control component is composed of the following parts: client interface, logic conversion, underlying data source access module, and other modules (Figure 2).

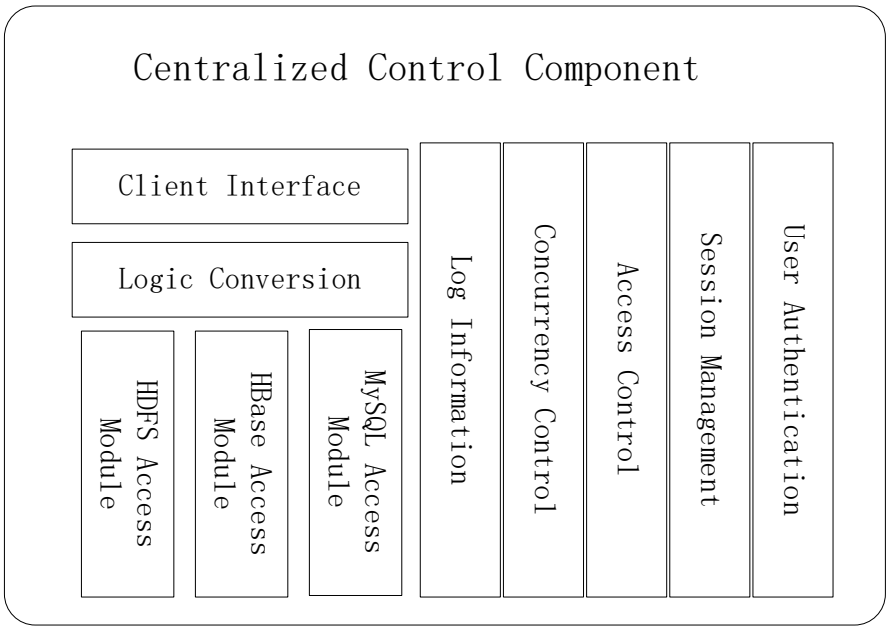

Figure 2. The Modules of Frontend Centralized Control Component

3.1.1. Client Interface

The client interface is responsible for handling user's request. The interface supports a customized query language called acSQL. For ease of use, the syntax of acSQL is similar with those of widely used SQL language. User can use acSQL to send the following command to cloud storage platform: 1. *createData*: cerate a new database Table; 2. *queryData*: data query; 3. *updateData*: data update; 4. *deleteData*: delete data from underlying storage. For each query operation, a parameter is required to identify the type of data that the query is corresponding to, since different type of data is stored in different data source.

3.1.2. Logic Conversion

After receiving user request, the logic conversion module takes over and parses it. According to its underlying data storage source, the parsed request will be sent to corresponding data access module. The request with different data type will be parsed and converted into different query statements. For the request whose data is stored in relational database, it will be converted into a series of SQL statements; for those whose data is stored in HBase or HDFS, it will also be converted into corresponding type of query statement. Whether it is relational database, or other types of storage data source, the same result can be achieved through different query statements, among which the performance is quite different. In order to make system's performance as high as possible, the module contains a query optimization sub-module that can create a most efficient query statement.

3.1.3. Underlying Data Source Access Module

This module is responsible for forwarding the request to the underlying data storage component, and retrieving the result from it. According to its target data source, the component has three independent parts, including relational database access module; distributed file system access module and semi-structured data storage access module. Relational database access module is used to forward the request to MySQL cluster and retrieve the result by using JDBC. Distributed file system access module is used to forward the request to the underlying HDFS node and retrieve the result. Semi-structured data storage access module is responsible for forwarding the request to HBase and gets back the result.

3.1.4. Other Modules

The frontend centralized control component is also responsible for managing the whole cloud storage platform, including:

1. User authentication: Only authenticated user can connect and visit the cloud storage platform.
2. Session management: An authenticated user will get a unique session ID, which means that the user can interact with the platform in the next period of time.
3. Access control: Different users have different privileges. In this platform, we divide users into administrators and normal users. Administrators can do many management tasks while the normal users only have limited authority.
4. Concurrency control: Since the cloud storage platform is composed by many components, each encountered problem will affect the performance of the whole system. Therefore, it is necessary to control the concurrent request, for the purpose of supporting as many concurrent requests as possible with good performance.
5. Record log information: All the operations will be logged, which will be used to audit and locate it once the issue happens.

3.2. Traditional Relational Database

Within our heterogeneous cloud storage platform, the traditional relational database is used to store some structured, fixed format data without having to frequently add dynamic column. For folk art resource, the data that is suitable to be stored in such traditional relational database includes folk art project related data, research literatures library for the folk art and related news information, *etc*.

Folk art project related data includes folk art projects' name, type, the origin, content and forms, *etc*.

Research literature library data includes all published papers and digitized documents that is related with folk art.

News information library includes folk art related news that is appeared in television, radio, newspaper and internet. In addition, it also includes related planning, policies, measures and meetings, performances, training, festivals and many other types of information.

3.3. HDFS Storage Structure

HDFS is a widely used distributed file system. From user's point of view, it can create, read, update, and delete a file by using directory. HDFS is a cluster system composed of many computer nodes. In this cluster, there is one name node and many data nodes. Name node is responsible for storing metadata, including directory name, the chunk number of this file, the position of each chunk. Each file is divided into many fixed size chunks, each of which has several replicas that are stored in different data nodes. HDFS's such structure is especially suitable for storing large volume unstructured data files.

In our heterogeneous cloud storage platform, HDFS is used to store audio, video and other multimedia information in folk art resources. Multimedia data plays an important role in the whole protection of folk art resources. For example, audio resources in folk music, the play videos of folk dance and Chinese traditional opera are all needed to be well stored.

The use of HDFS to store multimedia data in folk art resources can bring many benefits:

1. Each file has several replicas which are stored in different computer nodes. Thus the data is still available even when the physical failure of the computer happens. This is very important for the precious folk art resources.

2. Due to its architecture, HDFS has strong support for large volume data. When visiting multimedia information, especially large video files, the user access speed is very fast. In addition, HDFS can support high concurrent user access, which can improve the whole system's performance.

3. Capacity expansion is easy. When there is a new multimedia file coming, HDFS can divide the file into several chunks and distribute them into different nodes dynamically according to each node's load and other conditions.

3.4. Hbase Storage Structure

HBase [3] is a column based distributed storage which is suitable to store semi-structured data. The use of HBase to store data has following advantages:

1. Can store semi-structured data, with the ability to add attribute filed dynamically without stopping the computers.

2. Can be applicable to store very sparse data set. For each record, if its corresponding filed value is null, HBase will not preserve space for it, which can improve space storage efficiency.

3. Can save multi-version data. It is very convenient to query historical data records and support massive data storage.

In folk art resource, many art projects are inherited by different artistic talents. So we also need to store such artistic talents library related data. However, for different type of folk art, the inheritor's information needs to be stored is different. For example, for inheritor in folk music, the information needs to be stored includes music category; for inheritor in folk opera, the information should include the play he is skilled at and the role he plays; for inheritor in folk handicraft, it should store mentoring relationship, the art features and representative work. If all such data are stored in traditional relational database, each type of art resource should use a separate database table. When a new type of art inheritor is added, a new table needs to be created, which will cause data redundancy and increase maintenance cost. Once we use HBase to store folk art inheritor's information, each inheritor only needs to store the required fields of the type of art. By using this way, we can not only save storage space but also be convenient for maintenance management.

4. The Process of User Request in Heterogeneous Cloud Storage Platform

In our heterogeneous cloud storage platform, all user requests are processed as follows:

1. User sends the request to frontend centralized control component. The request must be with the forms of acSQL.
2. The client interface module takes over the request, checks if the syntax is correct or not, and forwards the request to logic conversion module.
3. The logic conversion module parses the request, converts it into corresponding request type, according to where the request data is stored. Then it sends converted and optimized request to the data source access module.
4. The data access module connects to underlying data source, forwards the request to it and waits for the result.
5. The corresponding storage retrieves the result and sends it back to centralized control component.
6. The client interface sends the result back to the user.

5. Conclusion

The resources of traditional folk art are highly valuable and needs to be protected carefully. Among all the protection techniques, it is especially important to digitize and store them effectively. Since traditional folk art has the characteristics of different types and huge mount, it is not suitable to store all of them within a single traditional relational database. In this paper, we propose a novel heterogeneous cloud storage platform to store Chinese folk art resources. The platform is composed by the backend storage structure and the front-end centralized control component. According to its characteristics and the way to be accessed, different type of folk art related data will be stored in different backend storage structures, namely traditional database, NoSQL based storage and distributed file system. Furthermore, the frontend of the platform is a centralized control component, which is responsible for handling user request, parsing it and forwarding it to the backend. The construction of the heterogeneous cloud storage based platform is feasible and effective, which can flexibly respond to the current needs and challenges in the field of the inheritance and protection of folk art resources.

Acknowledgments

This study was supported by grants from Humanities and Social Science Research Planning Fund provided by the Ministry of Education of China (Grant No. 13YJA760028 and Grant No. 11YJC760073), a grant from the National Natural Science Foundation of China (Grant No. 61373057), and a grant from the Public Welfare Technology

Application Project of Science Technology Department of Zhejiang Province (Grant No. 2016C31089).

References

[1] J. Jie, "A report on the development of the multi-media database for world ethno-music", Chinese Music, vol. 1, **(2006)**.
[2] A. Zhang. Discussion on the Construction of Lingnan Music Resource Database", Library Research, vol. 1, **(2013)**.
[3] D. Meyer, M. Shamma, J. Wires, Q. Zhang, N. C. Hutchinson and A. Warfield, "Fast and cautious evolution of cloud storage", Proceedings of the 2nd USENIX conference on Hot topics in storage and file systems, Boston, USA, **(2010)**.
[4] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica and M. Zaharia, "Above the clouds: A berkeley view of cloud computing", EECS Department, University of California, Berke- ley, Tech. Rep. UCB/EECS-2009-28, **(2009)**.
[5] D. Agrawal, A. Abbadi and S. Das, "Big Data and Cloud Computing: New Wine or just New Bottles?", International Conference on Very Large Data Bases, (**2010**).
[6] M. Stonebraker, "Stonebraker on NoSQL and Enterprises", Communications of the ACM, vol. 8, no. 54, **(2011)**.
[7] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes and R. E. Gruber, "Bigtable: A Distributed Storage System for Structured Data", ACM Trans. Comput. Syst., vol. 2, no. 26, **(2008)**.
[8] S. Ghemawat, H. Gobioff and S. Leung, "The Google File System. ACM SIGOPS symposium on Operating systems principles", New York, USA, **(2003)**.
[9] G. Decandia, D. Hastrun, M. Jampani, G. Kakulapati and A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall and W. Vogels, "Dynamo: Amazon's Highly Available Key-value Store", Proceedings of twenty-first ACM SIGOPS symposium on Operating systems principles, (**2007**) October 14-17; New York, USA
[10] K. Shvachko, H. Kuang, S. Radia and R. Chansler, "The Hadoop Distributed File System", IEEE 26th Symposium on Mass Storage Systems and Technologies, Incline Village, NV, USA, **(2010)**.
[11] "HBase: Bigtable-like structured storage for Hadoop HDFS", http://hadoop.apache.org/hbase/, **(2010)**.
[12] C. Chen, J. Lin, X. Wu, J. Wu and H. Lian, "Massive Geo-spatial Data Cloud Storage and Services Based on NoSQL Database Technique", Journal Of Geo-Information Science, vol. 2, no. 15, **(2013)**.

Authors

Xiaobo Li, he received his B.Sc. in Microelectronics (1990) from Nankai University (China), Master of Engineering (Research) (2004) from The University of Sydney (Australia) and Ph.D. in Pathology and Pathophysiology (2012) from Zhejiang University (China). Now he is a full-time Professor at Department of Computer Science and Technology, College of Engineering and Design, Lishui University, China. His current research interests include different aspects of bioinformatics, machine learning and data mining.

Zhongjuan Tian, she received B.A. in Music Education (1999) from Zhejiang Normal University (China), Master of Arts in Musicology (2006) from Shanghai Conservatory of Music (China). Now she is a full-time Professor at Department of Music, College for Nationalities (Minzu), Lishui University, China. Her current research interests include Chinese folk arts, folklore and anthropology of art.

Ye Wang, he received the B.S. in Electrical Engineering and Automation from Beihua University, China in 2003. He joined Zhuhai Wanlida Electric Co., LTD. in 2003 China, and also received his M.S., Ph.D in Information and Communication Engineering from Kongju National University in Korea in 2009, 2013, respectively. Since then, He has been a Lecturer in Lishui University, Zhejiang Province in China. His main research interests include Internet of Things, Mobile Internet architecture and NGN mobility management.

Guowen Ye, he received his B.E. in Physicalectronics (1989) from University of Electronic Science and Technology (China),Master of Electrical/Communications (2007) from Zhejiang University (China). Now he is a full-time associate Professor at Department of Electronic and electrical, College of Engineering, Lishui University, China. His current research interests include different aspects of :Control technology, Embedded system and Sensing technology.

Zhen Ye, he receives his Ph.D degree in Computer Science and Technology in 2013 from Zhejiang University, China. Currently he is a lecturer in Lishui University, China. His areas of interest are Distributed System, Computer Vision and Machine Learning.

