Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 1, Issue 1: 114-120, DOI: 10.25236/AJCIS.010012

Research on Data Storage of Archives Information Resource Sharing Platform

XianJie Bian, Xiaomei Lu

Yancheng Teachers University, Yancheng 224007, China

ABSTRACT. *With the development of big data technology, the continuous improvement of digital archival information and the construction of archival information resource sharing platform, archival information resources will show explosive growth. How to store the ever-increasing archives information resources will become the focus of research in the process of building the sharing platform. This paper will make a comparative study of the key technologies in the construction of platform data storage system, and propose to use HDFS as the overall framework of shared platform data storage system, and then design the module of storage system. Using HDFS as the storage system architecture support of the sharing platform, not only can archives information resources be stored and processed on traditional servers, but also can realize the storage and processing of distributed archives information resources.*

KEYWORDS: *Big data, Archive information resources, Sharing platform, Data storage, HDFS*

1. Introduction

The concept of big data was presented by SGI chief scientist John Masey at the USENIX conference in 1998. He published a paper called Big Data and the Next Wave of Infrastress, which using Big Data to describe the phenomenon of data explosion. With the development of mobile Internet and big data technology, archives information resources have diversified in collection and utilization, which requires the establishment of a centralized archives information resources sharing platform for unified management. Among them, the storage of archival information resources is one of the key technologies to realize the sharing platform. Data processing based on Hadoop technology can maximize the use value of archival information contained in the sharing platform of archival information resources [1]. However, the archives information resources as an important data assets, leaving the store, is unable to the data mining research later. The focus of future research on archival information resource sharing will be to overcome the difficulties faced by archival information resources in storage space, resource sharing, and meeting the increasing demands of users. In this paper, based on the analysis of the relevant

ISSN 2616-5775 Vol. 1, Issue 1: 114-120, DOI: 10.25236/AJCIS.010012

technology of big data storage system construction, HDFS is proposed to realize the archival information resource sharing platform data storage system.

2. Key technologies for platform data storage system construction

There are two main forms of archival information resources in the era of big data. One is structured data and the other is unstructured data. For structured data, relational databases are generally used, such as Microsoft SQL Server, Oracle, etc. For unstructured data, non-relational databases are generally used, such as Redis and MongoDB. Hadoop distributed file system (HDFS) is also a good choice for archival information resource sharing platform with explosive growth of data volume. The following paragraphs will analyze Hadoop and MongoDB, two typical types of big data storage systems.

2.1 Hadoop Distributed File System

HDFS was originally developed as the infrastructure for the Apache Nutch search engine project as part of the Apache Hadoop Core project [2]. HDFS is designed to be suitable for distributed file systems running on general purpose hardware. It has much in common with the existing distributed file system. The full amount of historical archive information data can be stored in the system. HDFS has obvious advantages compared with other distributed file systems [3]. HDFS is a highly fault-tolerant system that can be deployed on relatively inexpensive servers [4] because HDFS provides applications with interfaces to move themselves around data. When the archival information resource data reaches the mass level, the external request is very close to the actual data of the operation, which can effectively reduce the impact of network congestion and improve the throughput of system data.

2.2 MongoDB

MongoDB (from the English word "Humongous", which means "huge" in Chinese) is an open source database that can be applied to information sharing platforms of various scales. As a database for agile development, MongoDB's data model can be flexibly updated as the application evolves [5]. MongoDB storage system provides the function of traditional database, namely secondary index, which is a complete data retrieval system and strict consistency. MongoDB enables applications to be more agile and scalable, and platforms of all sizes can use MongoDB to create new applications and reduce the cost of platform applications. MongoDB is a database designed for scalability, high performance and high availability. It can be extended from single-server deployment to large, complex multi-data center architectures. Taking advantage of memory computing, MongoDB can provide high-performance data reading and writing operations [6].

ISSN 2616-5775 Vol. 1, Issue 1: 114-120, DOI: 10.25236/AJCIS.010012

Through the analysis of the above popular big data platform storage systems, we find that they are more suitable for the basic database software of big data storage systems. As for the archival information resource sharing platform, the most important feature is that with the improvement of users' requirements in advance, the requirements for archival information data are becoming higher and higher, which results in the change from querying archival data requirements to viewing archival original documents. From the traditional archive file demand to multimedia file demand transformation, it is determined that the data information to be saved by the archive becomes larger and greater. The sharing platform centralizes the needs of all users for the use of archives information, and eventually develops into a massive archives information data storage system, so HDFS is more suitable for the data storage system of archives information resources sharing platform. The main advantage is to support large files, and to detect and quickly deal with hardware failures, ensuring smooth operation of the platform.

3. The overall architecture of the platform data storage system

It is necessary to strengthen the construction of three management centers for the informatization of archives management and the high efficiency of archives security. First, the construction of comprehensive management center of archival information resource sharing platform; second, the construction of archival information resource sharing platform storage service center; third, the construction of data interaction management center of archival information resource sharing platform [7]. Among them, the construction of storage service center of archives information resource sharing platform is the infrastructure. HDFS is precisely to solve the bottleneck of space, performance, availability and scalability faced by single service storage. It provides high reliable and high performance storage services for large-scale storage applications by dispersing data on multiple storage devices. The logical layer is the storage service interface of archives information resources sharing platform, which is responsible for calling archives information data.The storage system consists of two parts, one is the core part of the data storage service, which consists of the data access layer, the data layer, and the data storage processing center. The other part is the auxiliary system, which is mainly responsible for the monitoring and operation and maintenance of the file information resource sharing platform. It is mainly composed of platform operation monitoring system, data backup system and operation and maintenance management system. The overall architecture diagram is shown in Figure 1.

ISSN 2616-5775 Vol. 1, Issue 1: 114-120, DOI: 10.25236/AJCIS.010012

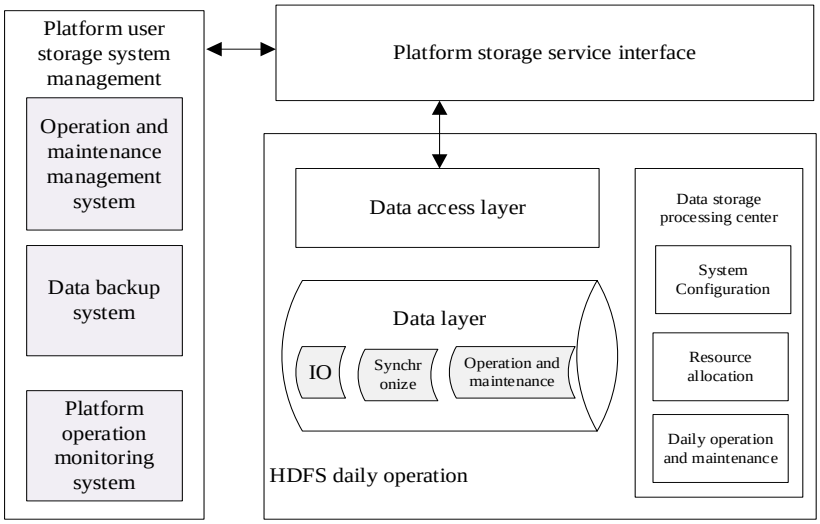

Figure 1 Overall architecture of the platform data storage system

4. Module design of platform data storage system

The module design of the platform data storage system mainly includes the design of data access and storage layer, data storage system processing center, backup monitoring system, operation and maintenance management system [8].

4.1 Data Access and Storage Layer

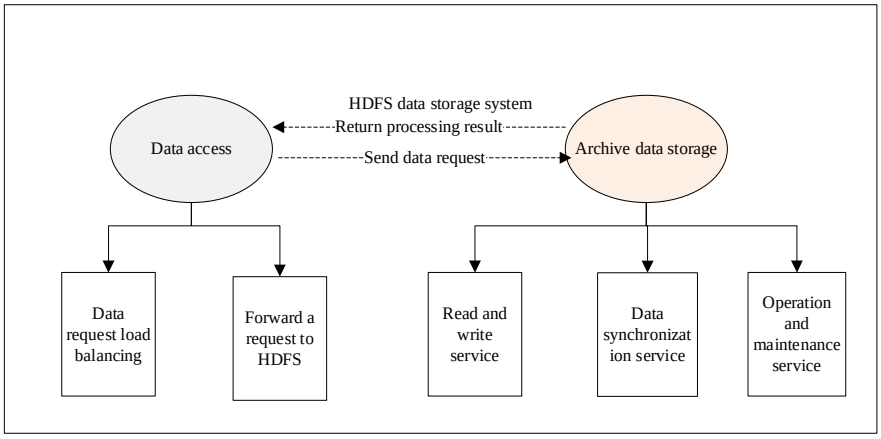

Figure 2 Data access and storage module design

ISSN 2616-5775 Vol. 1, Issue 1: 114-120, DOI: 10.25236/AJCIS.010012

The data access layer mainly provides two functions, one is load balancing the access layer of the logical layer; the other is the data layer device where the request for accessing data is forwarded. The data storage layer is the storage data, and the storage medium can support the memory or SSD. The synchronization module is used to synchronize data between different services; the read/write service is used to process user read and write requests; the operation and maintenance tool is used to switch servers, restart servers, and maintain servers. The data access and storage layer design are shown in Figure 2.

4.2 Data Storage System Processing Center

The data storage system processing center consists of three parts. The system configuration center is responsible for the configuration, maintenance and distribution of the whole data storage center; the storage resource allocation center is responsible for the quota management of the capacity, memory and other resources of different business modules of the archives information resources sharing platform; and the daily operation and maintenance center is used to issue the operation and maintenance commands of the platform data storage center.The data storage system processing center design is shown in Figure 3.

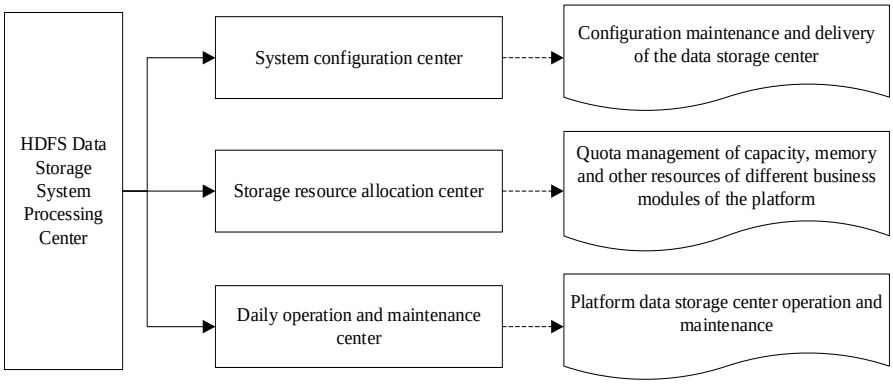

Figure 3 Data storage system processing center module design

4.3 Backup Monitoring System

The backup system is responsible for data backup and recovery of archives information resources sharing platform.The log center records the logs of all write operations; the task center manages and schedules the execution of all data backup and recovery tasks.The monitoring system reports and analyzes the key information and operating status of the platform, and monitors and alerts the abnormal situation. Daily report is to report the operation status of the platform; Supplementary

ISSN 2616-5775 Vol. 1, Issue 1: 114-120, DOI: 10.25236/AJCIS.010012

reporting is a modification of daily reporting information, which reports more dimensional information.

4.4 Operation and Maintenance Management System

The user of the platform operation and maintenance management system is the system operation and maintenance personnel. In view of the increasing number of users of archival information data and archival information, Web API technology is used to build an interactive service platform for archival information data to facilitate data business management and operation and maintenance operations, such as configuration management, fault management, business capacity expansion and other common operations. It can also view the system operation status and business operation data, and realize the data interaction of a wider range of clients (including browsers, mobile devices such as mobile phones and tablets), so as to truly meet the cross-platform needs of archival information data interaction services.

5. Conclusion

The most important foundation of archival information resource sharing platform in the era of big data is the construction of storage system, and a good storage system is the premise for the realization of subsequent data mining.This paper starts with the data storage system construction technology research of the platform, compares the commonly used big data storage system technology, and chooses HDFS as the data storage system of the file information resource sharing platform. Through the architecture design of the platform's data storage system, and the storage system of each module is designed, the platform's data storage system is finally constructed. At present, the storage system has certain expansibility and can adapt to the big data pressure of the future platform. The architecture of the storage system can serve as a reference for the storage system of archival information resource sharing platform.

Acknowledgments

This article is one of the research results of the “Research on the Construction of Archives Information Resource Sharing Platform in the Big Data Era” (Project No. 16YJA870001) of the 2016 Humanities and Social Sciences Research Fund Project of the Ministry of Education.

References

[1] X.J.Bian.Research on Data Processing Process of Archives Information Resource Sharing Platform. Archives Management, Vol. 6 (2018) No.28, p. 33-35.

ISSN 2616-5775 Vol. 1, Issue 1: 114-120, DOI: 10.25236/AJCIS.010012

[2] Y.Zhou. Study on rainfall prediction of online sequence extreme learning machine based on Storm. Xiangtan: Xiangtan University, Vol. 4 (2017) No.15, p. 108-111.
[3] X.H.Meng, H.C.Li and G.Z.Wang et al. Research on data storage service based on cloud architecture. Electronic Production, Vol. 2 (2018) No.14, p. 50-55.
[4] Z.C.Ma, P.Yang and B.H.Wang. Research and Application of Data Management Platform Architecture for Big Data. Network new media technology, Vol. 1(2015) No.17, p. 22-27.
[5] X.Hou, D.Xiao, P.D.Chen and L.Song. Visual Analysis System of Electrical Behavior Data. Computer Applications, Vol. 12 (2018) No.38, p. 77-82.
[6] C.Yang. Study on the Application of Geographic Modeling and Development of Spatio-temporal Genealogy System. Nanjing: Nanjing Normal University, Vol. 3 (2017) No.24, p. 25-31.
[7] X.J.Bian. Study on Data Interaction Service of Archives Information Resource Sharing Platform in Big Data Era. Zhejiang Archives, Vol. 11 (2018) No.17, p. 15-17.
[8] S.Q.Tang. Design and implementation of media material management system. Xidian University, Vol. 2 (2017) No.37, p. 11-13.