Research on the Construction of Library Data Integration System in Big Data Era

Yongwang Xi

Zhejiang University of Media & Communications, China.

Abstract. With the rapid development of network technology and information technology, the era of big data has come. Every day is accompanied by a huge amount of data. The library is also undergoing different changes every day in the processing of information resources and data resources. Knowledge and information are of little use for library data. Only by integrating library resources and big data can a system with greater value be formed. This article also hopes to effectively integrate library management with big data to ensure the further improvement of the level of library management, so as to promote the construction of library data integration system in the era of big data.

Keywords: Big data, Library data, System construction.

1. Introduction

With the increasing popularity of the concept of a smart library, the traditional information service of the library has gradually shifted to a user-centered personalized service model. For this reason, the data center of the library is showing a trend of massive growth. Massive data includes user privacy behavior information, library management system logs, and operational monitoring data. With the rapid development and wide application of network technology and information technology, the data information has also shown an explosive growth trend, marking the beginning of the era of entering the big data era. The library has always been greatly affected by the environment and technology. At the dawn of the era of big data, not only has the development of the library ushered in more opportunities, it has also brought challenges to the library [1].

2. Big Data Integration Based on University Library Resource Services

In the era of big data, the information resources of university libraries exist in the form of big data. Based on the consideration of personalized service needs of college users, it is urgent to apply related technologies such as cloud computing to handle big data to information services. In order to help university libraries, achieve "one-stop" information resources services. The so-called "one-stop resource service" refers to the deep integration of data from various databases or knowledge bases, restructuring or heterogeneous data, making it easy for university users to access various physical and virtual collections. Resources. The entity's collection data resources generally refer to self-built information libraries of university libraries, including bibliographic databases, characteristic databases, etc. The virtual collection resources include all available network data resources. The information service of university libraries based on big data is the use of big data such as library collection information, network information resources and user information resources. It uses MapReduce, Hadoop, NoSQL, cloud computing and other technologies to acquire, extract and integrate in a timely and rapid manner. , Combine these massive and complex digital information resources together to achieve search from different types of information resources, and use a unified data standard to express the results of the query, and truly integrate the different structural data, and through the integration of complex data Achieve user management and use of data[2].

3. The Analysis of the Pattern of Library Digital Resources Fusion under Big Data Environment

3.1 Content of Library Digital Resources Integration in Big Data Environment

At the dawn of the era of big data, small data possesses a much smaller value. Moreover, with the continuous development of cloud computing technologies and big data technologies, it is possible to process and analyze massive amounts of data information, and to mine the potential of data information. value. The library's effective fusion of digital resources is the fusion of large-scale data resources. It integrates all the data in the two different dimensions of time and space to ensure that all data information can form a very complete system. Then we will fully develop the digital resources and mine the potential value of digital resources. With regard to the time dimension of library digital resources, the integration of digital resources by libraries is to integrate all the data accumulated in the corresponding time period; in terms of the spatial dimensions of library digital resources, the library is to integrate digital resources. The data in the libraries in each area is effectively integrated. The digital information integrated in the library contains various digital resources such as business data, user information, industry information, and external data. In the context of big data, data integrity is also reflected in the "Multi-purpose" characteristics of the data. Therefore, in the process of integrating digital resources in libraries, we must also integrate data resources that differ in the same subject but from different sources, so as to ensure the integrity of digital resources [3].

3.2 The Idea of Library Digital Resources Integration in Big Data Environment

Big data is widely used in various fields, and many big data processing technologies and processing methods have also been formed. Although big data processing methods are different, they basically include data collection processes, data organization, and processing processes. Data statistics and analysis processes, data mining processes, etc. In the process of large data processing in the library, the same process should be followed. Based on this process, the effective integration of digital resources is completed. There is a close relationship between library digital resources and the services provided by libraries. It is impossible to integrate library digital resources only to meet the requirements of users. It is necessary to complete the integration of library services to ensure that user needs are met [4]. Therefore, the idea of integrating digital resources in libraries under the big data environment is:

(1) Library big data collection process. Library big data collection collects data on different dimensions of library time and space. When collecting data, ensure the integrity and reliability of the collected data. In this way, libraries are required to have more advanced storage technologies and complete infrastructure.

(2) The organization and processing of library big data. In the data information collected in the previous period, there will be many worthless and duplicate data information. Therefore, such information should be distinguished from useful information. Before the data information is imported, some valuable information and duplicate information should be added. Clean it up. In addition, according to the requirements of the fusion system, the collected data information is classified to ensure that the data information has a uniform format, so that the data information has usability and value.

(3) Analysis and application of big data. According to the library's needs for data information, statistics and analysis of the collected data are used to deal with the practical problems encountered by the library. It can also fully carry out data information mining work, obtain the potential value of data information, and play a predictive role of data information, and apply the corresponding results to the specific work of the library.

3.3 The Construction of Mapping Relationship of Scientific Data Integration Mode in the Era of Big Data

Regardless of whether XML-based middleware construction or semantic (model)-based middleware construction is used, both methods involve the mapping of original data and user query intent. Currently, there are two basic methods for constructing mapping relationships between

patterns: The Global-as-View (GAV) method and the Local-as-View (LAV) method. The GAV method maps local views of local data sources to the global view. The global mode is described as a set of views on the source schema. User queries act directly on the global view. The advantage of the GAV method is that the query efficiency is relatively high. The disadvantage is that the mapping relationship constructed by using this method has poor stability and is not suitable for the dynamic change of the data source. Because any local data source changes, the global view must be modified, difficult to maintain, and costly. The LAV method maps a global view to a local partial view on each data source, each data source schema is described as a view on the global schema. When a user submits a query, the mediation system decides how to answer the query by integrating different data source views. This approach can be seen as using views to answer queries. The advantage of this method is that the mapping relationship has good stability, which is suitable for the situation of relatively large changes in the information source. The disadvantage is that the information may be lost and the information query efficiency may be low [5].
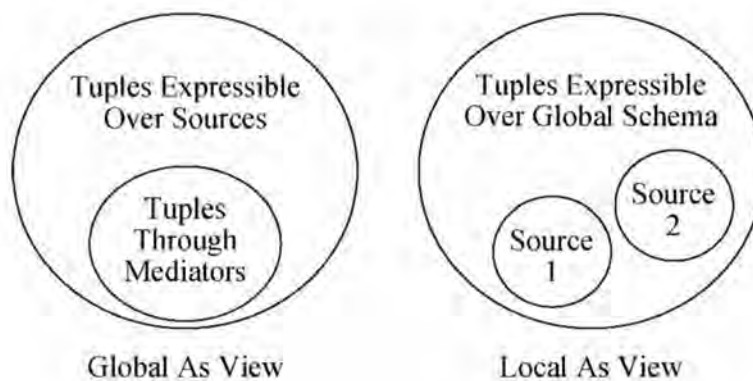

Fig. 1 Comparison of Library's GAV Method and LAV Method under Big Data

4. Design and Integration Strategy of Library Big Data Resource Integration Platform

4.1 Design of Library Big Data Resource Integration Platform

Big data resource integration refers to the process of sharing or merging data from two or more applications to create an enterprise application with more capabilities. Therefore, the design of library big data resources integration platform should adopt multilevel system structure design to ensure that the system platform has strong expansion ability and operates in loose coupling degree. At the same time, the addition, modification, and deletion of any functional model cannot reduce the overall functionality, ease of use, and controlled of the Big Data resource integration platform.

The system architecture of the library's big data resource integration platform consists of four parts: management operation layer, data per-cleaning and filtering layer, data integration layer, and big data resource layer. The management operation layer is the user interface to the management and application of the platform system. The administrator completes the control, management, maintenance, and application of the platform system for the integration of big data resources through the operation of the layer. The data per-cleaning and filtering layer is based on the defined data cleaning and filtering rules, and performs errors and availability checks, data quality analysis, data filtering and cleaning on the collected large data resources to ensure high data value for data integration. Density and personality. The data integration layer is the data that has been pr-cleaned and filtered in the temporary database. By reading the data source, parsing the data conversion rules, and loading the system, the converted data is written to the main database, and finally the library big data resources are completed. Integration. The big data resource layer is mainly composed of big data resources, temporary databases, main databases, and application system databases collected by different terminals and monitoring devices. The big data resources collected by the library are temporarily stored in a temporary database, and the data is cleaned and filtered according to

predefined rules, and imported into the main database after the data integration operation. The data synchronization mechanism ensures consistency of data between the application system data and the main database, providing data security, high-value-density, controllable and easy-to-use data support for library big data applications.

4.2 Utilize Cloud Computing Technology to Ensure the Efficiency and Economy of Data Integration

With the transformation of service models and the continuous development of user needs, the total volume of library data will show a trend of massive growth and progression. Reliance on increasing the number of IT infrastructure equipment, improving the performance of data processing and storage, enhancing the scientific of data storage and search, and reducing the user service QOS standards can no longer effectively solve the problem of the integration of big data resources in the future complex big data environment. problem. The library must rely on cloud computing technology support to provide secure, efficient, economical, and reliable cloud service protection for big data integration. According to the importance of data and the impact on library users' service quality, big data resources can be divided into service system operation and security monitoring data, user individual characteristics and social relationship data, user reading needs and behavioral privacy data, and user location information. Data with high security level such as behavior prediction and data with low security level such as system operation log, reading service data, user group demand data, service environment analysis data, and customer relationship guarantee. These big data resources have the characteristics of rapid increase in total data volume, timeliness, and difficulty in storage and search. Therefore, libraries must integrate cloud computing technology with the needs of big data integration to ensure the efficiency and economy of data integration.
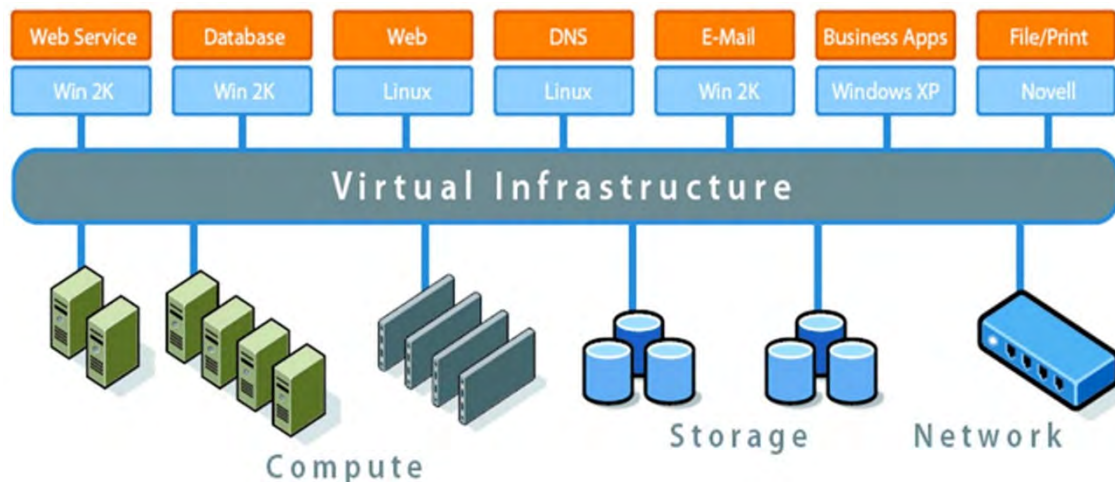

Fig. 2 The composition of the framework of the Big Data Library

4.3 Big Data Storage Architecture

Storage of big data the method adopted in the storage of big data is the distributed storage method. The Hadoop architecture is used to store the collected data information in the Hadoop platform. After the data information is processed, the data is stored in a distribution. Database. The Hadoop platform includes a big data distribution storage framework and a big data processing framework. The Hadoop cluster can be expanded to hundreds of servers. It can ensure that the capital investment in the data storage process is lower and the data query process can be ensured. To achieve "low delay" and "high concurrency" effect. When storing big data, you should follow the principle of tiered storage, that is, you should store data information hierarchically according to different physical attributes, life cycle, and business use maps of data information.

4.4 Big Data Storage Information Processing

The processing of big data in big data processing, it is mainly in accordance with the actual needs of the library to complete multidimensional analysis of data, real-time analysis, and information

mining. In the processing of big data, the Map Reduce model is used to accomplish big data processing. In Map Reduce model for big data processing, the main ideas are "Map" and "Reduce". Tasks, and assign each Map task to each node to complete the execution. After each Map task is completed, a certain intermediate file can be obtained, and the obtained intermediate file is regarded as the Reduce task input data information. Reduce is the merging step, which actually aggregates the results of previous different Map tasks and forms the final output. The application of big data Big data applications refer to the results of data analysis and processing, which are applied in all the work carried out by libraries to ensure the improvement of library service standards and promote the innovative development of library services. The library can apply fusion related data, carry out various forecasting services and information consulting services. In addition, the library can also visualize the processing results of the data information in front of the user through the application of visualization tools to provide users with visualization solutions.

5. Conclusion

Under the big data environment, the sharing of digital resources is gradually improved, and the library digital resource integration platform integrates all kinds of data information to ensure that the library data information can be combined into a complete organism, which can further promote the digital library. The co-construction of resources and the development of sharing work, the era of big data, the scientific, sh-arability, value, and ease of operation of data structures are the operational security and efficiency of relational library systems, user service models and content changes, and improved service capabilities. Important factors for the effectiveness of customer relationship management. How to build a scientific big data architecture and use scientific and rational strategies to collect, clean, convert, and integrate low-value-density, massive, ordered, and decentralized big data resources, so that the data resource structure is reasonable, high-value-density, and Management and ease of use are the key to the library's ability to gain new market insights and predictive analytic, scientifically match, optimize service resources, and satisfy readers' individual needs. The library must aim at improving the total value of readers' individual service support and the value of big data resources, and build a scientific big data integration, management platform, and service resource integration system. The library should combine the needs and characteristics of the user service system to carefully analyze the structural characteristics, value attributes, controlled, and availability of the library's big data resources, and efficiently perform data for different data sources, data formats, properties, and application objects. Integration can enhance the correlation between data, eliminate islands of information and increase the breadth and depth of knowledge discovery, and provide reliable and large-scale library service decisions, system management and operations, overall optimization of service resources, and customer relationship management. data support.

References

[1]. Cheng Xueqi, et al. Application of Big Data and Discussion on Scientific Issues. Mathematical and Big Data Science Forum. Beijing: Chinese Academy of Sciences, Vol. 5 (2013) No. 23, p.135-139.

[2]. Jiang Shan, et al. Wang Gang. Enlightenment of Big Data to the Library. Library Work and Research, Vol. 7 (2013) No. 36, p.152-159.

[3]. Qin Hong, Qian Guofu, Zhong Yuanyuan, et al. Comparative Study of Three Discovery Service Systems. Journal of Academic Libraries, Vol. 9 (2012) No. 15, p.267-274.

[4]. Li Guojie, et al. the Scientific Value of Big Data Research. Chinese Computer Society, Vol. 10 (2012,) No. 32, p.221-226.

[5]. Wang Yuanzhuo, Ruan Xiaolong, Cheng Xueqi, et al. Network Big Data: Status and Prospects. Chinese Journal of Computers, Vol. 12 (2013) No. 22, p.1125-1138.