Design and Implement of SDC in Digital Campus

Yunli Cheng¹, Hainie Meng²
¹Guangzhou Nanyang Polytechnic College, Guangzhou 510925, China
²Guangzhou Nanyang Polytechnic College, Guangzhou 510925, China
²Corresponding author
¹767220031@qq.com; ²412131779@qq.com

Abstract: The SDC (share data center) is the platform of collecting, processing, saving all kinds of share data, and supplying the information share services for the whole school. It carries out the data integration of each MIS, guarantees the consistency of the whole school share information, provides the support for the valid decision and for the new MIS development. After the analysis of saving and using the data at the present in Universities, this paper gives the design of the SDC’s framework, and details the function components and the consideration of design.

1. Introduction
With the development of information technology in colleges and universities, all kinds of data and information in the school have grown rapidly, which has brought many new problems to the transmission and storage of data. In particular, different transactions generate a large number of different types of data, which are respectively used in many different periods. Established and used by applications with different functions. The existing systems of various functional departments in colleges and universities are not conducive to integrating data in different systems to provide new information, and those who use data want to be able to see the comprehensive situation of all data and information [1].

The data required for these information services are stored in different departments of the application system, distributed in different storage sources, and the data is duplicated, redundant, inconsistent, etc., in order to meet information service query or decision analysis, the data needs to be carried out. Integrated processing. As one of the problems that need to be solved in the construction of digital campuses in colleges and universities, how can we solve them? This is a common concern of all. Through our years of research and practice in the digital campus, we believe that we can choose to integrate some performances at the portal level. But the key is to solve these problems at the data layer by building a university shared data center platform. After a brief system overview, this paper has carried out a series of research and discussion on the design and implementation of university shared data center platform from the aspects of university data standard construction, shared data center system architecture, system analysis and function.

2. System Overview
As a shared platform for basic data in the school, the shared data center centralizes the basic data of each business unit, realizes unified management, ensures the authority and accuracy of data, provides data support for each business application system and various service systems, and undertakes
digitization. A large amount of data, information \cite{2}, processes, and processing results on campus, system functions include:

1. Data collection and update: All data comes from various business systems, and data is collected to a shared data center through a common data extraction or synchronization method, and a data update report is also made.

2. Organization of data: According to the information standard of the school, the data updated by each department is sorted, the data is abnormal, the consistency and accuracy of the data are guaranteed, and the historical data is retained.

3. Sharing and utilization of data: By authorization, some or all resources of the data center can be shared for sharing and utilization. A shared data center makes a data usage report in real time.

4. Data mining analysis: Through the OLAP analysis of shared data and historical data, various reports are formed or analysis results are presented as needed to provide reference for leadership decision-making.

3. System structure

In order to achieve uniformity and consistency of information coding throughout the school, to ensure "who generates, who maintains", to ensure that comprehensive information reflecting the entire school is provided and to accumulate analytical data for the school decision support system developed in the future, we have designed the university as shown in Figure 1. Share the data center system design.

![Figure 1. Shared data center system design](image)

4. System analysis and function

4.1 Data source and central library

The data source is the data extraction source of the shared data center. We refer to the database of all the application systems in the school as the data source. This includes two aspects, one is the existing application system database of the school; the second is the subsequent construction. The database of the application system \cite{3}. We have designed the university as shown in Figure 2.
Figure 2. Data source and central library

The shared data center library is an integrated place for all shared data, and the shared data of all application systems is integrated here. It integrates data from the data source and keeps the update synchronized, becoming a shared data channel between the various application systems; upwards as a data source for the statistical analysis service, and provides the statistical analysis service with shared data integrated from each application system. The data of the shared data center library is derived from the data source, and the data integration tool extracts data from each application system database of the data source and stores the data according to the data type. In addition, the shared database and the data source's various application system databases are kept in synchronization with each other [4]. The data synchronization of the shared database is divided into automatic synchronization and manual synchronization. The automatic synchronization is mainly for subsequent construction of application system data, and the synchronization works through the data access layer. Implementation; manual synchronization is mainly for the original application system data, and its synchronization work is realized by the data synchronization tool.

4.2 Data integration
The data source of the shared data center library is currently mainly integrated by the data of the national standard, the basic code table, and from the existing application system. This process is the process of data extraction. The data of the existing application system does not conform to the data specification. The process of extraction is to realize the accuracy of data from the unregulated data source through integration into standardized data [5].

The prerequisite for data integration is to investigate the data source of each application system to be accessed, and the application system should ensure a certain degree of data interface. This is a process from the application system to the shared data center. First, you need to determine which data to extract from the application system. What is the meaning of this data? That is to provide the corresponding data dictionary, and determine which table corresponds to the data center, the accessible data interface mode is divided into: (1) directly open the database: only need read-only account permissions, you need to absolutely guarantee the original There are system data security and integrity, which does not affect the original system operation based on the establishment of triggers. (2) Intermediate file data source: If the application system cannot open the database to the outside world, the difference data file can be exported to the specified directory. These files can be the Access file database mode or the excel file mode. The format is agreed upon at the time of implementation.

4.3 Metadata management
The first thing to emphasize is that the metadata discussed here is not the data that constitutes the data in the general sense, but refers to a series of original basic data in the construction of digital campus.
Metadata also refers to the true meaning of each data table and field in the management system. Metadata management completes the management and maintenance of the database structure of the shared data center, that is, registers the original data of the standard table in the shared data center to facilitate future Data management.

Metadata management includes: (1) registration of the table. The Chinese name of the table name is annotated and the table is described in detail. The number of tables in a shared data center is very large and involves every aspect of the school. The registration of the table is to create a file for the shared data center for visitors to check. (2) Field registration. As with table registration, field registration also creates a file for the data structure of the shared data center for visitors to review. (3) Update the database structure. In order to adapt to the development of school informationization, it is necessary to face the times, and there are errors or errors in the face of data standards. It is necessary to update the database structure, but because of the huge implicature in the update, it must be extremely careful, not to be changed. The principle of change is to update, add, and delete field information for unused tables, and only add new operations to the already used tables.

In addition, the classification of metadata is also involved in the management of metadata. According to the classification of information subsets, the standard library can be divided into classifications familiar to business personnel to facilitate searching. There are also CheckPoint records for some data that require special attention for tracking and statistics. This is mainly for sensitive data. You need to know the ins and outs of it. Who did it at what time? Record it for later review.

4.4 Theme management
Topic management includes the creation of theme libraries and the management of theme objects. We know that for a topic, the related information is generally not obtained from a single library. To fully utilize the information, we need to build a comprehensive theme library. For example, for a teacher, a data topic can be formed, that is, information integration related to the teacher. For example, the personnel management system has basic information, file information, salary information, transaction information and other information related to the teacher, which can be classified into the teacher's subject, and so on, in the scientific research management system and the educational management system. Teacher-related information data can also be integrated to form a user-defined complete teacher theme, as shown in Figure 3. The above example is based on people. From another perspective, we can also use events as the center of the theme.

The management of topic objects includes: (1) topic object generation. According to the access permission of the application, the operation of the table that it can access is encapsulated. After the object is established, the access to the database is realized by the object, and the object is finally associated with the SQL statement. (2) Object rights management. Divide user access to topic objects. (3) View the subject object. View the corresponding application of the object, the object name, the SQL statement corresponding to the object, the xml file format corresponding to the object, the WSDL address of the accessed WebService, and so on. (4) Subject object display. Graphically display the contents of the object and object view based on the user's access rights. (5) My database. According to the user's access rights, the metadata table, fields, and data in the table, checkpoint records, and data in the database can be displayed.
4.5 Data access interface
The data access interface is for application-to-center database access and generally provides a standard set of access center database tables and view access interfaces. Let's take the xml-based WEB service interface as an example. The XML (eXtensible Markup Language) language has good scalability, follows strict grammar requirements, facilitates the transfer of information between different systems, and makes file exchange between users and applications easier. Separate from the display content, it has better preservation.

Web services, commonly referred to as WebService technologies, can be applied to the development and deployment of multi-architecture, distributed, object-oriented Java application systems cross-platform component architecture. The advantage of WEB service system is platform independence; strong scalability; strong distribution; high reliability; dynamic load balancing. The table and view for the shared data center can provide a standard WebService access interface, provide services for the user's Select, Update, Insert, and Delete operations, and provide validation for the legality of the submitted data.

4.6 Data management tool
How to quickly and efficiently process all kinds of data on a shared data center platform requires some tools to support it.

(1) Data Integration Tools
The data integration tool is a tool for data extraction from a data source (each business system database) by a shared database. It allows the user to conveniently set the system to automatically extract data from the various business system databases into the shared database. The main workflow includes: selecting the data source, ie the source database; selecting the tables in the database; setting the tables The relationship between the two; set the target database; set the target database rows and columns; set the field mapping relationship; set whether to pass the audit; set the data integration scheduling, that is, the cycle and operation of data integration.

(2) Data synchronization tool
The data synchronization tool is aimed at synchronizing between the existing business system
database and the shared database. Because the database of the original business system is built earlier, the overall design of the design and shared database must be different. From a technical point of view, it is difficult to make the subsequent construction of the business system database and the shared database like the data access layer. Maintain automatic real-time synchronization. Therefore, taking into account the database of the existing business system, the data synchronization tool is used to maintain the data synchronization between the existing business system database and the shared database.

(3) Metadata management tools
Metadata management tools are an important management tool for sharing data centers in databases. It mainly manages the description, classification and validity of the database. For the description in the database, including the database name, field name, row and column name, etc., all of them are in Chinese, so that all the characters presented to the user are Chinese expressions, and the database background definition remains unchanged. In addition, the data from different sources in the database are classified and managed, and the data extraction target is convenient downward, and different topics of data integration are formed upward to facilitate upper layer data extraction.

(4) Data consistency check tool
In the shared data center, the data consistency check tool checks or compares the data source and data target with logical errors and data rules. Once an associated error or violation of the rule is found, the system immediately disconnects the relevant database administrator. Alerts until the error is corrected. For large shared data centers, from the perspective of database logical security, once an error occurs, many related influences are extensive, so the data consistency check tool uses real-time monitoring to scan each database continuously to ensure the normal operation of the database.

(5) Data import and export tool
The data import and export tool provides an interface for interacting with external data for the shared data center. It can convert and export data in the shared data center according to external data rules. It can also convert external data according to the rules of the shared data center. And import. For example, a school can extract data from a shared data center according to the Ministry of Education's annual report and convert it to a specified format for export file submission.

5. Conclusion
In the construction of the university's shared data center platform at this stage, there have been some successful cases for the establishment of specific university data standards and the establishment of a central database. The former is mainly based on the information standard of the Ministry of Education, while the latter focuses on the platform. The management side develops a large number of tools. In addition to some of the tools mentioned in this article, the shared data center should also provide some corresponding management mechanisms to ensure the smooth operation of the platform system, such as system management mechanism, system management mainly includes application user management. And log management, in which application user management involves application management (adding and deleting operations to the application system), application rights management (dividing the application access rights to the metadata table, and dividing the access rights into read-only access and exclusive access) In addition to the functions of viewing, deleting, backing up, and exporting log records, the log management function also includes the log audit function. That is, the administrator can select a log information item that is more concerned by setting, and focus on managing a part of the log.

The data center field that can be further researched and constructed in the next stage includes how to form a historical database based on the existing shared center library, and gradually realize the transition from the data center (database) to the data warehouse, and the amount of data in the university reaches a certain level. In the future, data mining will be carried out, and on this basis, a decision support system that meets the needs of colleges and universities will be formed [7]. Another need to emphasize the security of shared data centers, although some experts and scholars have done some research in the field of data center security strategy, security technology and security
management, and also reflected in the practical application of digital campus in colleges and universities. [8], but it is still far from the ideal state. It is still necessary to further deepen the work to build a campus data center security system.

I believe that in the near future, in most universities, I will build my own shared data center platform to form a unique trusted data source, so that the entire new system and existing systems that exist in different periods can be organically integrated and maintained throughout. The data is unified and consistent, and provides reliable, sufficient and comprehensive data protection for the information inquiry and decision analysis of the whole school, laying a platform foundation for the further construction of the digital campus.

Acknowledgement

[Fund Project 1] This article is the Guangdong Provincial Department of Education's Guangdong Province Characteristic Innovation Project (Humanities and Social Sciences) -- "Research on the Innovative Talents Training Model under the Background of New Engineering" (2018GWTSCX054), phased research results;

[Fund Project 2] This paper is the phased research result of the project of “Big Data and Intelligent Computing Innovation Research Team (NY-2019CQTD-02)” of Guangzhou Nanyang Polytechnic College.

[Fund Project 3] This paper is the phased research result of the professional construction project of Computer Application Technology (NY-2019CQZY-04), the third batch of innovative strong school engineering brand in Guangzhou Nanyang Polytechnic College.

[Fund Project 4] This paper is a phased study of the Cloud Computing and Big Data Professional Committee of the Higher Education Institute of Guangdong Province, "The Research on the Construction of Big Data Major and the Training Mode of Innovative Talents in Higher Vocational Colleges under the Background of New Engineering (GDYJSKT19-13)" Results.

Author

Cheng Yunli (1980-), female, Shandong Heze, master, Communist Party member, associate professor, senior engineer, research direction: computer application technology, digital campus.

References

[1] Shen Xichen, Chen Huaichu. Standards and Standards for Information Construction in Colleges and Universities. Journal of Tsinghua University: Natural Science Edition. 2016,43(4)
[2] Jiang Dongxing, Xu Qinghong, Xie Wei, Xiang Chun, Li Gongqiang. Integration Thoughts and Practice of Informatization Construction in Colleges and Universities. Education Informatization. 2017(12)
[3] Wan Xiaojun, Yang Jianwu. Design and Implementation of Xinhuangnet Data Center. Computer Applications. 2017, 23(5)
[4] Zhao Zeyu, Wang Liming, Zhang Jilong, Zhang Chenghong, Yan. Design and implementation of the virtual campus platform of Fudan University. Proceedings of the 2003 Annual Conference of National University Informatization Research Society. October 2017
[5] Yan Li, Wu Haiyan, Feng Wei. Discussion on the construction of campus data center. Proceedings of the 2013 Annual Conference of National University Informatization Research Association. October 2017
[6] Xu Xin, Xu Bin, Chen Junliang. Design of shared data center in colleges and universities. Education Informationization. 2018(7)
[7] Song Shibin, Teng Xiaoling, Song Yaqin, Zhu Lianrong, Cui Jian. The guarantee of informationization - data center construction. Education information. 2018(7)
[8] Wu Haiyan, Yan Li. Research on Network Security System of Campus Data Center. 2018(9)