

Research and Implementation of Data Integration of University Application Systems

Shijie Wu

Information Engineering Department
Tianjin Maritime College,
Tianjin, China
397656413@qq.com

Abstract—With boosting development demands for university information construction, the number of university application systems had been dramatically increased. However, many universities were lacking of pre-planning in the information construction, which had resulted in mutual independence of application systems, dispersion of information resources and low utilization rate. In order to improve the usage efficiency of data resources in current application systems, to eliminate the phenomenon of "Information Island", the data in current application systems needs to be integrated. This article analyzed the necessity of data integration of university application systems, deeply analyzed multiple data integration modes, proposed the implementation policy of data integration and provided references for the data integration of university application systems.

Keywords—application systems; data integration; data resources; implementation policy; information construction

I. INTRODUCTION

For years of development and accumulation, the elementary hardware environments in universities had been substantially improved, different kinds of application systems had been put into operation and remarkable achievements had been made in the information construction. However, certain critical problems still need to be studied and solved. For instance, certain universities were lacking of pre-planning during the information construction; information resources of universities were relatively dispersed; the utilization rate was relatively low; there were a lack of shared data center platforms etc. Additionally, the disunity of information criterions had led to current disordered data, which significantly influenced the utilization efficiency of data resources in universities. To fully utilize current data resources and improve the utilization efficiency, the data integration has become the key issue in the university information construction [1].

II. NECESSITY OF DATA INTEGRATION IN UNIVERSITIES

A. Distribution between Application Systems and Data

In order to fulfill development demands for university information construction, universities had successively completed the design and development of many application systems in the field of educational administration, finance, human resource, scientific research etc. A large amount of elementary data had been applied and accumulated in above systems. Due to the

differences in development times, developers and storage devices of above data, it was extremely difficult to access to the data management, which resulted in the repetition and inconformity of data. These phenomena disabled the direct comparison and analysis of data in each system, and severely impeded school leaderships from inquiring and using critical data [2].

B. Low Utilization Rate of Information Resource

So far, after the information construction and accumulation in universities, architectures of Megabit and Gigabit network had been created and certain redundant mini servers had been applied as application servers. However, due to relatively dispersed system environments in early development, data exchanges were lacked among systems and sharing of information resources could not be realized, which resulted in the lack of completed and real-time data in information unified platforms and various information management systems in universities.

C. Low Decision Making and Supporting Ability

Current data sharing in universities could not meet the requirement of integrated development and utilization of information resources. Though there was a large amount of simple and repetitive application data, application data supporting management and decision was in severe shortage. Although enormous information resources were reserved in the data, no effective tool could be applied for full adoption and utilization, nor had the appreciation function of information resources been fully developed during the management and decision processes.

With boosting accesses to computers in universities, system administrators had to attend to more complicated operations, while massive manual operations had to be applied in many complicated operations. The data processing and analysis therefore became more complicated; and it became more difficult to collect required data from each system for application. Thus, the collection and feedback of external information could not be realized immediately or accurately; massive valuable data could not be generated by application systems nor delivered to decision-making sections. Current application system platforms could not be compatible with development kits, which resulted in the small range of use of each application system [3].

Due to the disunity of standards and criterions for data definition, no unified data interface could be utilized for sharing common data sources, which resulted in the problem of "Information Island" that reflected isolations

among different application systems. How to solve "Information Island" and how to achieve the efficient and fast integration of data resources in application systems became critical issues for the university information construction.

III. DATA INTEGRATION MODES

A. Data Concentration Mode

To fulfill requests from each department and to make data management and maintenance more convenient, a shared data center in the institution that is applied to save the data of each department shall be constructed in strict compliance with information construction criterions set by each university. A unified design for the application platform shall be applied; meanwhile application subsystems shall be constructed under a unified framework. Advantages of above method contain high coupling degree among each subsystem and easily-implemented seamless connection between the data center and subsystems. However, the development circle has to be extended as data exchanges involve operations between the data center and each department; user demands cannot be perfectly controlled and user flexibility will be reduced.

B. Shared Data Center

The integration of information resources is based on invariable application systems. Data with shared value shall be extracted from application systems, and shall be applied in the shared data center construction. R/W of data is still operated under original application system of each department, only upon access request from other departments, the application system would then release the shared data to realize the interconnection at data level through early deployment [4].

C. Distributed Resource Integration

Distributed resource integration is similar to the shared data center: For local operation, only local data access is involved; when the interconnection with the data center is required, data access to other departments shall be applied. The realization of the distributed resource integration is to construct the resource index database that is similar to the search engine, which is to collect and record data of each department and determine the practical location of data in the network simultaneously. However, each department still keeps the original data. Above is the main difference between the resource index database and the shared data center.

IV. IMPLEMENTATION POLICY OF DATA INTEGRATION

The highly-independent data in university application systems is generally kept in separate application system database. In order to realize the effective data sharing, data in each system shall be cleaned, converted, exchanged and stored in the shared data center. The shared data center exchanges data with each system by data exchange mode and performs as the important channel for data exchanges of each system [5].

A. Data Exchange Modes

There are two options for data exchanges between application systems and the data center:

1) *Directly connect with ESB (Enterprise Service Bus) to realize the access to the application system database.*

2) *Create data exchange intermediate tables in the application system database, realize the sharing with objects connected with the database of shared data center and activate the data exchange via ESB client.*

B. Direct Access to Application System Database

The access to the application system database is realized by utilizing ESB data exchange tool, which can be applied in application system databases that have the same structure as the shared data center and small data exchange capacity. Operation authorization of the application system database is required for the data synchronization.

C. Data Exchange Intermediate Table Creation

The data exchange intermediate table that is created in compliance with the data exchange requirements of application systems and the shared data center is a kind of data field table. All data corresponding to fields, behaving as the link between application systems and the data center during the synchronization operation is originated from the shared data center and application systems with intermediate tables.

The data center adopts data intermediate tables to realize the data synchronization and push in application systems, not directly read or write data. Thus, following cases can be referred to when adopting data intermediate tables to perform the data integration.

Case I: If the application system vendor didn't approve the direct access to the database during the data exchange, user could adopt the data intermediate table to perform the data exchange operation.

Case II: Due to the data security importance, users were not recommended to read or write the data of system database directly when integrating certain critical application system data. It is recommended to upload or download intermediate files or to create data exchange intermediate tables to realize the data exchange between application systems and the shared data center.

The data exchange is practically realized by implementing data intermediate tables. Special attentions shall be paid to concerning the data security and integrity during the data integration of university application system. Certain exchange method shall be determined according to practical environment to ensure the data exchange security [6].

D. Integration of Isomorphic Databases

Isomorphic database refers to the concept that the software and version of the data center database are exactly the same as that of application system databases. For instance, both the shared data center and the educational administration system adopt Oracle, therefore, direct data exchange can be realized between these two systems. In consideration of the application data security and independence of shared system data, it is recommended to realize the data integration by creating data intermediate tables in application systems.

E. Integration of Heterogeneous Database

Heterogeneous database refers to the concept that the database vendor of the data center is different from that of

application systems. For instance, the data center adopts SQL server or MYSQL, while the application system adopts Oracle. There are generally two methods for realizing the data integration of heterogeneous databases: One is to create data intermediate tables to perform the data exchange between the system with heterogeneous database and the data center database. The other is to utilize the data exchange tool to configure the data exchange, to extract the shared data for operations including cleaning, converting, storage sharing etc. and to realize the data synchronization and exchange.

F. Data Exchange Policy

The data exchange plays an important role in ensuring the data synchronization between the shared data center and application systems. It can ensure the accuracy and real-time capacity of the data exchange, and also guarantee the correctness and authority of the data in the shared data center. The data exchange can further improve the office efficiency of each department in the university, promote the entire management level of the university and provide real-time and accurate data support for the leadership decision-making.

It is better for users to select data exchange modes by considering practical cases. It is recommended to adopt data intermediate tables and create triggers in case that application systems require to realize the extraction and push of data exchange with the shared data center. In case that application systems are only requested to provide data but not acquire data of other application systems from the data center, the data shall be provided through view creation by application systems, and later the data exchange tool shall be applied for synchronization to the data center [7].

The increment synchronization approach is generally applied in data intermediate tables for improving the data exchange performance. Meanwhile, in order to save the hardware resource cost during the data exchange and to ensure the system performance will not be weakened by the creation of data intermediate tables, application systems shall implement the periodical data reading policy to complete the data synchronization. Considering the importance of data exchange, elaborative analysis shall be implemented on the process and approach of data exchange, and performance of data exchange tool shall be taken good care of.

V. PRACTICAL APPLICATION OF DATA INTEGRATION

With the continuous development of the information construction, universities had successively developed a variety of application systems. Due to their independence, system data was relatively dispersed. The data integration that covers multiple aspects shall be applied to realize the data sharing. Regarding to the practical condition, the application of data integration is analyzed as below.

A. Connection with Heterogeneous Database

The connection with integrated application system database shall be solved before the creation of a unified identity vault. JDBC drive and eDirectory drive shall be applied to realize the data exchange on the directory server of the unified identity vault when Identity Manager is accessing to Oracle database.

Metadirectory engine module plays an important role in realizing the data sharing among systems under the architecture of Identity Manager. The drive of Identity Manager will perform the information synchronization with the directory tree of central identity vault via the interface provided by Metadirectory engine module, which will finally realize the data sharing.

Connections between the unified identity vault and each application system contain two data channels belong to the publisher and the subscriber. The publisher channel bears the data sent by application systems to the unified identity vault, while the subscriber channel bears the data sent by the unified identity vault to application systems. Each channel contains separate policies for defining processes of data exchange and load.

B. Configurations of Policy and Filter

The definition of sending and receiving update modes of IDM is realized by Policy, which will realize the data interconnection and synchronization in the management of application systems. The definition of each policy exists in the publisher and subscriber channels, while procedures of each policy may vary in the data exchange application. Certain policies will only be applied under specific operations. The execution sequence of policies is as below:

Event Conversion Policy – Major common tasks are accomplished by customized filtering.

Matching Policy – Locate objects from targeted memory blocks, but formers correspond to non-associated objects in original memory blocks.

New Policy – Conditions that must be fulfilled by the defined location during the new object creation.

Planning Policy – Determine the title and location of new objects which must exist in the unified identity vault and connected applications.

Command Conversion Policy – Convert commands sent by Identity Manager to targeted memory blocks by Add or Replace command.

Outline Mapping Policy – Unification of the outline mapping definition between the unified identity vault and connected systems.

Output Conversion Policy – Data format exchange in the subscriber channel.

Input Conversion Policy – Data format exchange in the publisher channel.

IDM filter is responsible for the filtering of information in the configuration filter, which only permits required information. For the filter design, firstly define the filter, only synchronize required objects, and then define relevant operations executed by IDM on permitted filtered objects [8].

C. Synchronization of Discrepant Data

How to reduce the amount of synchronized data at every turn is the dominating procedure regarding to the massive amount of data synchronization, which will reduce the bandwidth loss and improve the data synchronization efficiency. The way to solve this problem is to adopt Identity Manager. Identity Manager Driver will perform the periodical database scanning to acquire the discrepant data. If the scanning result is null, then no synchronization operation will be executed.

D. Single Login System

The single login system is created on basis of the unified identity vault. It is applied to unify user identities in the unified identity vault and to realize the access to application systems with mutual trust in the platform by single login. The campus network can also be accessed to by logging from external networks. Thus, the system must be created on basis of reverse proxy service.

Form filling is adopted in realizing the identity authentication, which will simplify the access to Web applications. When the identity authentication is inputted in the user login page, information will be stored in the reverse proxy server automatically. Later, when the user logs in the same page and performs the identity authentication by adopting the same Web application form, the application system will quickly extract the user identity authentication information and fill the form automatically. Meanwhile, only users authenticated by the reverse proxy server can access to the service information in each campus network [9].

E. Data Sharing Platform

A data sharing platform shall be created after the creation of a unified authentication platform. It is a substantial part of the application system and lays the foundation of data exploration and big data warehouse in certain degree. The construction of the data sharing platform mainly adopts the data extraction technology, which is to extract the required data from each application system, convert the extracted data by pre-specified rules and import the converted data to the shared database or data warehouse, fully or complying with the planned increment[10].

VI. CONCLUSION

The integration of application systems is definitely a complicated task and an essential procedure in the university information construction. Only the interconnection between each system at the data level is realized, can digital campus and unification of educational administration be accomplished. Establish premium

criteria by considering four aspects of human resource, finance resource, property resource and management resource, support the unified information standard, fully realize the campus information collection and analysis from multi-dimensions and views, and provide comprehensive information resource for the further development of universities.

In practical cases, the integration of application systems shall be gradually completed in compliance with the entire planning and construction scheme, which will finally realize the efficient integration and optimized polices of data resources, realize the efficient sharing and sufficient utilization of data resources, improve the efficiency and benefit of administrative and service work of the university, and provide strong support for the leadership decision-making.

REFERENCES

- [1] Chen Min, Zhou Jiangbo, and Gao Hui, "Research on Integration, Sharing and Utilization of Large-Scale Data Resource", E-Government Affairs, pp. 44-50, March 2013.
- [2] Guo Dan, "Research on Information Integration of University Educational Administration", Northeast Petroleum University, 2012.
- [3] Feng Xiaojuan, "Data Integration in University Information Construction", China Education Innovation Herald, 2012, pp. 6-11.
- [4] Pang Jiu, "Brief of Data Integration in System Integration", Electronic Technology & Software Engineering, 2013.(11): pp. 150-156.
- [5] Chen Gang, "Summary of Intercepting Filter Realization Policy", Modern Computer (Professional Edition), 2012.(24): pp. 120-126.
- [6] Deng Jiechao, "Research on Resource Integration in Digital Campus Construction", Guangdong Technology, 2007.(2): pp. 34-35.
- [7] Yang Mu, Hu Yanmei, and Luo Ting, "Design and Realization of Data Exchange Center in Digital Campus", Modern Educational Technology, 2006: pp. 56-59.
- [8] "Concept of Digital Campus", CERNET, <http://www.edu.cn>.
- [9] Luo Xinmao, "Study of Construction of University Teaching Resource Database", Academic Journal of Longyan University, 25(3): pp. 127-129.
- [10] Xu Xin, "Research on Application Integration of Digital Campus", New Technology of Library And Information Service, 2005: pp. 61-66.