

Research on Data Integration System Based on Web Services

Shufu Cao^{1, a *}, Linru Li^{1, b} and Quange Yi^{2, c}

¹College of Information & Technology, Hebei University of Economics & Business
Shijiazhuang, China.

²Shijiazhuang University of Economics Shijiazhuang, China.

^acaoshufu@heuet.edu.cn, ^blilinru@heuet.edu.cn, ^czhouyqg@163.com

* College of Information & Technology, Hebei University of Economics & Business Hebei
Shijiazhuang, China.

Keywords: Data Integration; Web Services; Star-Topology

Abstract. At present aiming at the problem of heterogeneous database within the application system in universities, put forward a set of star data integration scheme based on Web Services. In the scheme, without changing the original application system structure and the database in universities, on the basis of this add a service agent, complete the data transfer among different systems and the registration work to integration center. Data integration center in the central location, provide registration and release service for each application node through Web Services, as the same as provide access services for other application nodes.

Introduction

With the development of "Digital Campus", the present majority of domestic universities have completed the earlier stage construction of information. Since lack of unified management and planning^[1] at the beginning of construction, each business segment in school is usually establish an independent application system based on their business needs, so there are many different functions application system in the same universities. These platforms and database technology used in these application systems are diverse from each other^[2], database is short of uniform data formats and standards, as the same time the interface and the standard of each system are not the same, resulting in data information is independent among the various of business segments, can not be shared^[3], and also caused a waste of resources.

In order to realize the sharing of the data and information within the various of business segments on campus, to build a complete and unified, efficient and stable campus, as the same time avoiding duplication investment in the development of software for various applications, universities urgent need to integrate the various of relevant data resources. At present aiming at the actual domestic situation of the development of information technology in universities, put forward a set of data integration solutions having star structure. It is based on Web Services technology, can excellent integrate heterogeneous data among the business segments within universities^[4], so that it can be shared among different systems, enabling smooth workflow among departments.

Analysis of Heterogeneous Database

According to the institutional settings the universities is basically divided into teaching faculties, scientific research supplementary agencies, regulatory agencies and logistics agencies etc. Each agency is also divided into a number of departments in charge of specific business management. These business departments is now basically established its own application system for the management of the department's operations. Such as the personnel department in regulatory agencies establishes a personnel management system, academic departments establishes educational management system, scientific research department establishes management systems. As these implemented functions are different in these application systems and the development time is different, many different development

languages and database technologies are used . These cause the lack of unity interface among these application systems, data transmission and sharing can not be achieved, so effective integration is need.

By the status quo analysis to university informatization system, summed up the data needed to integrate are divided into the following categories:

- The sharing public data provide for most applications system, such as students basic information, staff basic information etc. As soon as generation these data have many features such as large quantity, small changes.
- The cross-system visits data. For example personnel system access some of the data in educational administration systems and scientific research systems in order to understand a teacher's teaching and research conditions. These data have many characteristics such as smaller amount of data and must abide by the same data pattern when visits.
- The cross-system batch transfer data, for example educational management system batch transmits student course-choosing and get credit data information to student management system. This kind data has larger amount of transmission.
- The unstructured data generated by the daily office system. This data is mainly electronic document to be exchanged among the various departments of the school.

To public data shared by multisystems and the cross-system batch transfer data, since a larger amount of transmission, database replication method can be used, to copy data from one database to the other databases in need, update and modify regularly data.

The cross-system visits data is usually shared data pattern that application system directly accesses other applications , so integrated approach can be used to integration.

Integration Solution of Universities Data

As a distributed computing technology, Web Services use XML, SOAP, UDDI and WSDL technologies to construct a loosely coupled computing environment, and then integrate the different development platform and different database technology application systems, provide data transmission sharing each other.

System Architecture Design:

A set of star integration solution with multinodes and the integration center is designed for universities. The system architecture is shown in Fig.1.

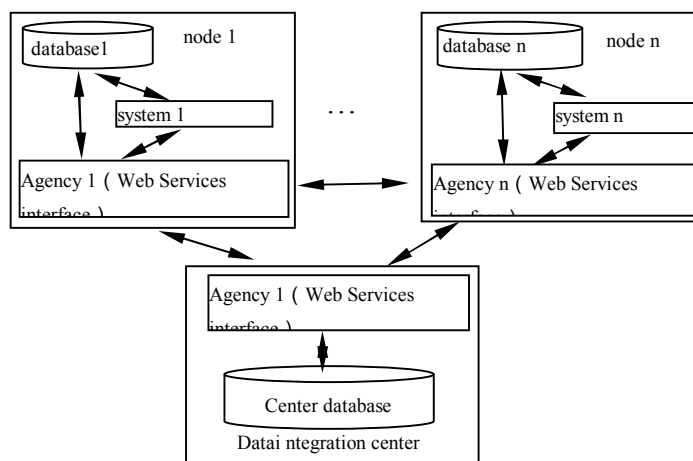


Figure 1. The universities data integration system architecture

In this scheme, in addition to the application node contains the application system and the corresponding database, also provide released data mode in local, so that other applications nodes complete data access services to the the node. At the same time access interface of application nodes even register the local access service and release to the data integration center.

Data integration center in a central location, in addition to package standard Web Services access interface, but also establish a central database used for storing public data and documents. Data integration center implements the registration and release of application nodes through Web Services, At the same time it shields a part of the heterogeneity of each application node systems to provide access services for other application nodes.

Web Services shield effectively the heterogeneity of application system and database among the application nodes. The data transfered among applications Nodes convert to standard XML document and transmit through simple object access protocol(SOAP).Data not only can be transmitted directly among application nodes,but also be forwarded through the data integration center.This improves the flexibility and efficiency of the entire system.

Data Replication Interface Between Application Nodes:

The realization of data replication among application nodes uses data pull ways, that is the objective database issue initiatively to the source data. The data is obtained from the source to the local. In this way the source and destination nodes must be able to record what data needs to be replicated, what data has been copied but changed and needs to be copied again, in order to avoid repeat replication and missing data.

Data replication queue adopts a trigger mode recording. A trigger is a set of SQL statements which were pre-defined and compiled, stores on a server, generally is used for achieve a particular function. The trigger is close relationship to data tables and can protect the data in the table. When the operation affectes data which is protected by a trigger, trigger executs automatically.

As increase triggers in the source database need not modify existing applications system, data replication, data replication queue can adopt to set a trigger means to achieve. By setting a trigger, it can trigger the corresponding action when the source data change and record the change of the data in the data replication queue.

Data Integration Center:

Data integration center mainly performs some functions such as the registration and management of data model and access for each application node, store and forward the data of each application node, user privilege management of the entire system and transaction coordination of data transfer among application nodes etc. user permissions to manage and and other. In addition, central database is used for storing public data and documents.

(1) Web Services Description

Web service description is to describe data operation provided by database as service using Web service description language(WSDL)and generate WSDL documentd. Other applications can use directly all services provided by it accordance to all services information described by WSDL document.

WSDL document abstract description the request / response messages when services are accessing, then bind these messages to specific access protocol and message format, and at last define the service access point of specific deployment. WSDL document does not relate specific operation of the service in the database, achieve the separation of service and specific operations, making the process change in the internal database is transparent to the outside world. That is, as long as the invocation interface provided by services does not change, it will not affect the calling of service.

(2) Web Service Registration

Web Service registration is to record the related information provided by database in the service registry, in order to provide technical detail for the query and calling service of the other applications.

In universities data integration, system use private UDDI registry to implement web service registration. API is divided into two logical parts: query API and release API. Query API is that application can quickly locate the candidate business entities , Web Services calling specification, then based on the information obtained from initial calling get more detailed information. Release API create various types of application tools, on this account business technical personnel interact with a UDDI registry directly, manage release information from business entity and technology model

structure. When release Web Services through release API, access must be authorization, but not for anonymous access. Each UDDI choose and use a license agreement API compatible with the release API, at the same time it also provides registration mechanism. Before using release API, users must be registered on the UDDI. Query API is the same as the release API and it realizes via SOAP message based on HTTP protocol to achieve, but the query API does not need any authorization.

(3) Data Model Management

Data mode used to describe the sharing data information such as the data models, data types, providers and Web Services interface addresses. Data model and access service must register in data integration center in order to unified management, registration, release, query, modify, disable, and delete etc.

Data integration center storage the registered data according to provider and type. Not only does the data pattern store in the XML database directly, but also generate the corresponding table structure in a relational database of the central database through the mapping XML to RDBMS. Later the data transferred to the data consolidation center can be stored in relational tables of the central database automatically according to the mapping.

Data integration center managers configure access permission for the data mode registered in data integration center, and then publish the data model to other nodes agent through Web Services interfaces of data integration center. Users can search for the required data mode through these parameters such as the data mode providers, type, model name, model access permission and so on. After retrieving the pattern, users can obtain the data pattern WSDL (service description) of data access Web Services interface registered in the data integration center if permissions allow.

System Management Services:

System management services is used for realize the system configuration management and status monitoring. Configuration operating parameters of the various parts of the data integration center through system management service, service start ,stop control and monitoring the whole system running.

Conclusion

In the paper, at present aiming at the problem of heterogeneous database within the application system in universities, put forward a set of star data integration scheme based on Web Services. The data can be transmitted among application nodes directly and also be forwarded via data integration center. This scheme solves the problem of sharing and exchange in campus heterogeneous database effectively, improves the flexibility and efficiency of the entire system.

Acknowledgements

Fund Project: Science and Technology Research Projects of Hebei Colleges and Universities (QN20131136)

References

- [1] Li Chunyu. Research and Implementation Data Integration Layer of Educational Management System Based on SOA [D]. Changsha: Central South University, 2009 In Chinese
- [2] Jiang Wenguang. Discussion the Problems and Countermeasure of University Information Construction [J]. Computer CD Software and Applications, 2014 (24): 251-252 In Chinese
- [3] ZhangLei. Design and Implementation of Digitalization Campus Data Integration System [D]. Wuhan: Huazhong University of Science and Technology, 2012 In Chinese
- [4] Wu Yucheng, Xu Jinqun. Integration Research of University Digital Library Data Based on SOA and Web Services Architecture [J] Information Science, 2012 (7): 1030-1033. In Chinese

