

Research on Data Heterogeneity in Enterprise Information Integration

Ruizhen Duan, Zhongwen Wang, Mingshan Chi, Xiaoqiu Xu

*Rongcheng College; Harbin University of Science and Technology, Shandong,
264300, P.R. China
wangzhongwen1979@qq.com*

Abstract

In view of the problem of data heterogeneity in enterprise application integration, we proposed a solution, which is based on XML, WSDL and Web Service technology to achieve the elimination of data heterogeneity, which can reduce the complexity of the system, and has certain significance for the information communication between enterprises and enterprises.

Keywords: XML; heterogeneity of data; Web Service

1. Introduction

With the continuous improvement of enterprise information, there are all kinds of application systems. However, due to the heterogeneous data of network and operating system, it is difficult to share information between different systems.

Enterprise data heterogeneity can be divided into three levels. The bottom is the platform layer, such as different computer, operating system or communication protocol, and the system can adopt different programming languages and compiling environment, the middle layer is the system layer, such as different database management systems, and even the file system, they are based on different data model, providing different language, and the top is data semantic layer, because of the different database is independent design, there are great differences between different data semantics, so different system uses different ways to describe their data, the same information can be used different patterns, even if the same information in the same system can also have different data models[1].

Web service provides a distributed computing, through the XML and SOAP protocol to transmit information, and it can also through the enterprise internal firewall. Web service is a cross platform, which has important significance for the transformation of heterogeneous multi system environment and legacy system in the enterprise. Through the UDDI registration center, each enterprise can share a large number of services and information, and it also shield many problems that caused by heterogeneous. Under the web service framework, every data source can create a web service, and then we can use the WSDL to register the service center. Such an integrated system can send a search request to the registry and select the appropriate data source, and then obtain the data from the data source through the SOAP protocol[2].

Enterprise information integration system can integrate different data sources and information to a user interface which can understand the document. When the system has interoperability, there is another big problem between data transfer and conversion. This paper puts forward using XML to eliminate the isomerism data from different system model, and using simple object access protocol SOAP and Web Services description language WSDL to eliminate the heterogeneity between the components. Both of the ways of interaction to thoroughly remove the island which is brought by the data heterogeneous.

2. Research of Eliminating the Heterogeneity Data

Heterogeneous data model, namely the data in the different storage mode. Storage mode mainly includes the relationship model, object model, object relation model, document nested pattern and so on. The relationship model (relational database) is the main storage mode. At the same time, even if is the same kind of storage model, their model structure may be existed differences. Type and the data model heterogeneity mainly has the following several forms: heterogeneous data sources, data naming standard is not unified, data synchronization difficult. At the same time, there are different in each enterprise application system of the programming language, operating system and database, which are differences in the definition of data types.

Below for data model heterogeneity and the heterogeneity between components we put forward the solutions.

2.1. Eliminate Data Model Heterogeneity

XML as the current internationally recognized standard of data exchange format, with simplicity, openness, extensibility, and have self description function, which can distinguish data and display. It also provides the man-machine interactive language, and it is very suitable for heterogeneous data exchange between applications. The Web Service provided "standard XML data format of remote access interface + content of the XML data remote readout interface + unified standard encoding data", It can laid a solid foundation for conversion and definition of the data format. And the foundation make the sharing of remote information come ture.XML as a new network information exchange standard is content oriented, and it has a better structure and semantics, it also has good scalability, simple and easy to master.

When network customers must pass information between different databases, need to get the most from the network server to download the data published to the web client processing, and when the network client on the same data varies for different user interfaces, XML has the incomparable superiority to HTML.

XML is designed to transmit and store data, and the focus of XML is the content of the data, which is designed to transmit information, However HTML is designed to display data, the focus of HTML is the appearance of data, which is designed to display information. Therefore, for one perspective, XML is independent of the software and hardware of the information transmission tool. This makes XML into a different data model into the middle of the bridge, and the public data network data processing and transmission mode[3].

Now the most widely used data model is relational data model. And most of the application end use the way of Web browser, most of these browsers can transport by XSL to global database management mode to the XML document format conversion and interface design. HTTP as a communication protocol, XML as the format of service requests and structured data said in a distributed system transmitted between the client and the server. Global database schema by the ORB communication mechanism between each local database management system (called IDL compiler dedicated database interface conversion between XML and IDL. Shall be the responsibility of the IDL compiler to produce special database interface, according to the format of the sender receives the data, and the data format for the said into the receiver[4].

The conversion between global data schema and local database data mode, and data processing mode is shown in Figure 2-1.

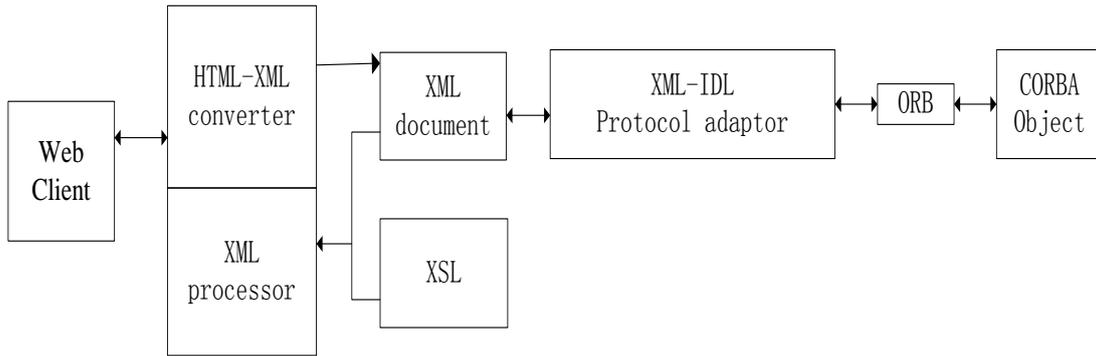


Figure 2.1. Diagram of Different Data Mode Transition

2.2. Elimination of Inter Component Heterogeneity based on SOAP and WSDL

SOAP is divided into simple type and complex type when the data is encapsulated. In order to complete the encapsulation of remote call, the SOAP packet needs to include the information as follows: the communication protocol, the address of the service program, the port number, the service activation method, the status information and so on[5]. The following figure is the request and response of SOAP.

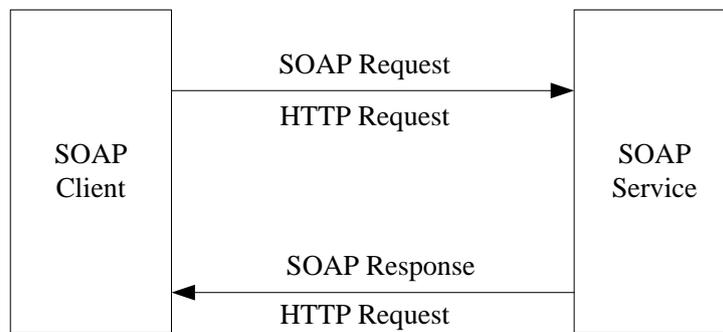


Figure 2.2. The Request and Response of SOAP

SOAP does not define any application semantics, such as programming model or specific semantic implementation, it is just the definition of a simple mechanism, through a module packaging model and the data of module in a specific format code rewrite code mechanism to represent the semantics of the application. It defines the message transmission standard between the service provider and the service requester, and uses the XML to format the message, and transmits the message with the underlying transport protocol (such as SMTP, HTTP). SOAP specification actually is defined by a message framework. A complete SOAP message mainly includes the following four elements: SOAP Envelope, SOAP Binding, SOAP Encoding Rules and SOAP RPC Representation. Although these four parts are part of the SOAP as a whole, but they are functional and they are independent of each other.

The remote invocation model of SOAP is the use of the SOAP rule for the state (stack) of client calls, and then transfer to the server through the grid protocol, and the call stack is established again at the server end. Finally, the server is implemented by a specific activation call service.

In heterogeneous distributed environment, SOAP is used to implement the integration of heterogeneous data system. The system is based on the data processing function in the existing system, and is released on the Web server, which provides the interface of the department data processing service. Users can implement data by calling services to realize the integration of systems[6].

WSDL is a logical interface for the application system, which describes the services provided by the application system, and allows all the application systems to communicate and integrate with WSDL. It is a document that is composed of XML, and describes the functions which are provided by the program. After the WSDL interface is implemented, it can be invoked by other applications. Because of the abstract definition of service access point and message is separated from the specific service deployment or data format binding, it can be used for the abstract definition again. The message is the abstract description of the exchange data, and the port type is the abstract set of operations. Specific protocol and data format specification for a particular port type can be used to bind.

Figure 2-3 is a research background based on SOAP and WSDL. It can be seen from the graph that SOAP, WSDL and distributed components work together, so that different clients can access different components, thus providing greater flexibility and software reuse. The grid service is encapsulated into a SOAP object, which is passed to the Web server, the Web server forwards the received SOAP object to the grid service request server; the grid service requests server receives the SOAP request, calls the grid service, operates the heterogeneous components and then returns the result to generate the corresponding SOAP response. After the Web server receives the SOAP response, the result is returned to the client by means of the HTTP response[7].

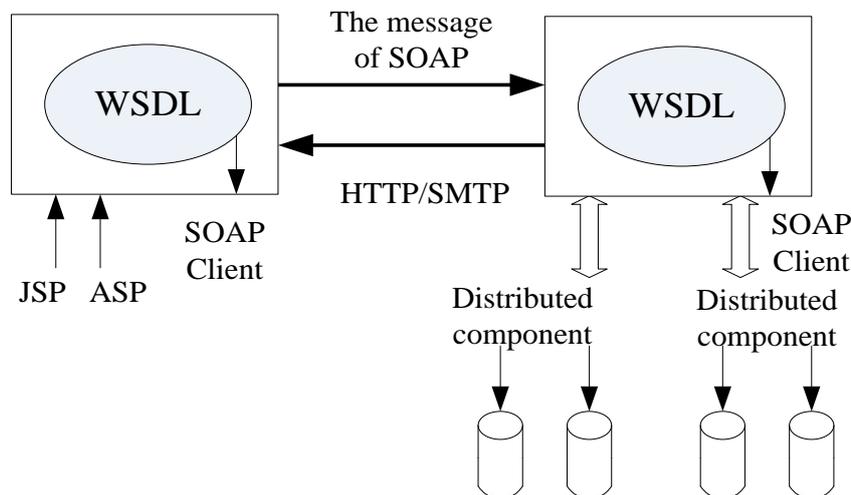


Figure 2.3 Diagram of Different Data Mode Transition

Every node information source in the integrated framework is published in the way of grid service. The Web application client or other application client calls the grid service interface from different platforms and different software development environment, thus obtains the data from different nodes. And then it is realized transparent access to services.

Users through the client's service browser can see the unified service view, access to resources and operations, call system to provide services and so on. The communication between the client and the server is used the grid service interface, and the transport protocol is SOAP protocol. Their interaction data through the XML language to describe the unified, so that it is easy to understand the system, and then the client can be used to explain the return of data and display.

Among them, the resources from every node system are published in the form of grid services. Only when the user sends out the service request, the service is transmitted directly to the user through the data integration platform. When resource format changes or new resources are added to the node system, it only needs to be published by the new grid service, which realizes the low coupling between the system.

The client of the enterprise subsystem uses JSP for develop. When it visits to the module, the SOAP client object is initialized, and the request is sent to the SOAP server through the XML server. In fact, the request is sent to the listener. The following is the SOAP request program fragment:

```
Content Type:text/xml;char set="utf-8"  
Content Length:xxx  
SOAP Action:http://www.Appserver.com/pw/GetAmmount.wsdl  
<SOAP Env:Envelope xmlns:SOAP ENV="http://schemas.xmlsoap.org/  
soap/encoding">  
<SOAP ENV:Body>  
<m:Get Ammount xmlns:m=""/>Web/Get">  
<In Parameter>3.6</In Parameter>  
<m:Get Ammountm>  
</SOAP ENV:Body>  
</SOAP ENV:Envelope>
```

In order to enable the COM component to deal with the SOAP message, the server side has installed the SOAP Toolkit4.0 Microsoft tool. We can use the Microsoft WSDL Generator in the toolkit and use the COM component to create WSDL files, WSML(Web Service Meta Language) files and a network service listener. There are two listeners as follow: ASP and ISAP IDLL. And we use ASP here. Because it can provide greater flexibility and easier to edit and debug. The following is the core code of the ASP file.

```
<%WSDL File Path=Server.MapPath ("Get Ammount.wsdl")  
WSML File Path=Server.MapPath ("Get Ammount.wsml")  
Set Soap Server=Server.Create object ("MSSOAP.Soap Server30")  
Soap Server.Init WSDL File Path, WSML File Path  
Soap Server.Soap Invoke Request, Response,""%>
```

Soap Server.Init can load WSDL files and WSML files, it can also initialize SOAP server object. So that it can provide the service information for the SOAP server object. Then the SOAP server object is called by the SOAP client, and the related data is obtained from the Request object. SOAP server communicate with COM component, and get results from the COM component, and then return the results to the SOAP client. SOAP client interact with other JSP pages, so that the JSP client and the COM module can complete the communication.

In the WSDL file, every parameter of the COM component has a pair of <message> elements, and has the corresponding data type attribute. Data types in XML and Schema SOAP are compatible with the built-in data types. And the <service> element defines the port of service. It also specifies the location of the service with the location property. In other words, it is the URL that is the end of the SOAP service listener in the ASP.

3. Conclusions

In this paper, we analyze the problems of data integration in enterprise information integration and the paper gives the corresponding solutions to the two problems of data model and component based on SOAP, XML and other standard Web Service technology, and the integration of grid technology. These measures have a big significance to solve the problem of information island in the enterprise.

References

- [1] Y Dali, J Meiyu, "Teaching Information Resources Integration Based on Heterogeneous Data Exchange and Sharing of Technology[J]", *Computer Science and Computational Technology*, vol. 12, (2008), pp. 150-153
- [2] L Wen, "The Research and Application of Information Resources Integration Model Based on Grid Service[D]", North China Electric Power University Master's Thesis, (2011)
- [3] S T, "Research and application of heterogeneous database integration based on XML[D]", Fudan University Master's Thesis, (2012)
- [4] M. Yunfeng, W. Yuan, "Research and Implementation of Heterogeneous Data Integration Based on XML and Web Service[J]", *Computer Technology and Development*, no. 11, (2010), pp. 42-50.
- [5] C Tianhuang, L Fan, "Using SOAP Extension to Implement the security of SOAP Message[J]", *Journal of Wuhan University of Technology(Transportation Science&Engineering)*, no.03, (2007), pp. 518-524.
- [6] S Yiting, "Research and Application of Web Services in Enterprise Application Integration[D]", Dalian Maritime University, (2007)
- [7] H Huijun, D Lan, X Chongyang, D Kesheng, "Research on Virtual Enterprise Information Systems Based on Web Services[J]", *Manufacture Information Engineering of China*, no. 17, (2010), pp. 13-17.