

基于 Web 服务的异构数据库 集成技术研究

徐 斌¹ 于微微² 于志涛³

(1. 滨州医学院图书馆, 山东烟台 264003; 2. 滨州医学院卫生管理学院, 山东烟台 264003;
3. 东方口岸科技有限公司, 北京 100738)

摘 要: 目前, 大多数企业在信息化过程中遗留下来大量基于各种业务流程和异构数据库的应用系统, 这些系统满足了企业某一特定的业务需求, 但由于数据具有自身的特点, 其他子系统很难使用别的子系统的数据库, 在企业内部产生了信息“孤岛”, 阻碍了企业信息化进程。对企业异构数据库进行集成成为消除信息“孤岛”的有效途径。本文提出一种基于 Web 服务的异构数据库集成方法, 以解决数据集成中的难题。

关键词: Web 服务; 异构数据源集成; J2EE; Struts 框架

中图分类号: TP311 **文献标识码:** A **DOI:** 10.3772/j.issn.1674-1544.2008.04.004

Research on the Integration of Heterogeneous Database Based on Web Service

Xu Bin¹, Yu Weiwei², Yu Zhitao³

(1. Library of Binzhou Medical College, Yantai 264003;
2. Health Management College of Binzhou Medical College, Yantai 264003;
3. East Port Technology Co. Ltd, Beijing 100738.)

Abstract: At present, a lot of heterogeneous applications and business processes remain in many companies during informatization. These applications satisfy their business requirements to some degree. As data have their own characteristics, their subsystems are hard to reuse other subsystems' data. The so-called information "Isolated Island" in the companies hinders the informatization process. To integrate the heterogeneous data source is an effective way to deal with the issue. In this paper, an integration of heterogeneous database based on Web Service is put forward to solve the problem of the integration of data.

Keywords: web service, integration of heterogeneous data sources, J2EE, Struts framework

第一作者简介: 徐斌(1982-), 男, 助教, 主要研究方向是计算机项目开发和应用。

收稿日期: 2008年3月17日。

1 引言

目前,大多数企业在信息化过程中遗留下来大量基于各种业务流程和异构数据库的应用系统,这些系统满足了企业某一特定的业务需求,但由于数据具有自身特点,其他子系统很难使用别的子系统的数据库,在企业内部产生了信息“孤岛”,阻碍了企业信息化进程。加强对企业异构数据库源的集成成为消除信息“孤岛”的有效途径。本文提出一种基于 Web 服务的异构数据库集成技术,以解决在集成过程中遇到的一系列难题,如:

(1) 异构性:多数据源系统最大的障碍是如何解决各个数据源的异构性。该异构性又分为平台异构性、数据库系统异构性、语义异构性。

(2) 透明性:数据源异构程度越大,透明性的要求就越高。透明性主要包括平台透明性、数据库源透明性和语义透明性。

(3) 自制性:每个局部数据库具有对自身完全的控制能力,同时能够决定是否为其他成员系统提供和提供多少它的功能和资源。

本文对使用 Web 服务技术实现异构数据库集成进行了简单的模拟和试验,说明这种技术路线的实际可行性。

2 数据集成技术简介

自 20 世纪 90 年代以来,数据的集成技术和异构数据的互操作就一直是研究的热点,相继出现了多种数据集成技术,如 OMG 组织的 CORBA 规范,Microsoft 的 COM/DCOM 规范以及 Sun 公司的 EJB/J2EE 规范。2000 年,微软和 IBM 联合发布了 Web 服务相应的开发规范。数据集成技术主要有以下几种。

2.1 传统的数据集成技术

2.1.1 基于 CORBA 的数据集成技术

CORBA,即公共对象请求代理体系结构,是在分布式的面向对象环境中定义的一个互操作规范。在 CORBA 环境下,它采用了一种与平台、

语言无关的接口定义语言 (IDL),规定一个分布式对象的边界以及它与潜在客户的接口,写入 IDL 中的分布式对象,可以跨语言、工具、操作系统和网络进行访问,IDL 语言是一种接口的描述语言,具有跨语言、跨系统的特点。CORBA 规范的核心是 ORB (对象请求代理)。ORB 可以对用户与分布式对象的访问细节加以屏蔽。

2.1.2 COM/DCOM 技术

随着分布式计算的流行,微软在 COM 的基础上,加上 RPC (远程方法调用)的支持,形成 DCOM 规范,它用于在分布式环境下实现异构数据和异构应用的集成。

RPC 使 COM 成为 DCOM,DCOM 依靠 RPC (远程过程调用)提供分布功能。当客户进程调用接口上的方法时,DCOM 将方法调用变成 RPC 调用。RPC 机制自动完成数据的包装和解包、数据格式的转换,并建立网络会话和处理网络应用。可以说,DCOM 提供了一种面向对象的 RPC 机制。

2.1.3 EJB/RMI 技术

EJB/RMI 是 Sun 公司提出的分布式的软件组件规范,定义了服务器端组件的编写规则、组件之间的交互规则和应用服务器对组件的管理规则,EJB/RMI 非常类似于微软的 COM/DCOM 技术规范。可以说,具有跨平台能力的 EJB 是分布式环境下的 COM。

EJB 规范包括会话 EJB、实体 EJB 以及消息驱动 EJB,它们共同完成分布式的数据、事务和业务逻辑的处理。

2.2 基于 Web 服务的异构数据集成技术

随着 Internet/Intranet 和 Web 技术的迅猛发展,传统分布式计算技术已不能很好地适用 Web 环境,主要体现在数据传输和协议方面。Web 服务为满足程序到程序的交互,正做着积极的准备,并成为 IT 产业近几年来探索的热点课题之一。

2.2.1 XML 与 Web 服务技术概述

XML(可扩展标记语言)是 Internet 上新一代

的数据表示和数据交换标准。XML 关注信息内容本身，而不关心它的实际页面展现，是 Web 上表示结构化信息的一种标准文本格式。

基于 XML 发展起来的 Web 服务技术是一种革命性的分布式计算技术。它使用基于 XML 的消息处理作为基本的数据通讯方式，使异类系统能够作为计算网络的一部分协同运行。

2.2.2 SOAP 协议

SOAP(简单对象访问协议)是 Web 服务技术数据传输的基础。虽然 SOAP 是 HTTP 中 GET 和 POST 方式的扩展，但 SOAP 协议建立在 XML 的基础上，是一个在分散和分布式环境下用于网络信息交换的标准协议。SOAP 定义了 XML 消息如何在异构环境之间传送，通过何种协议传输，并通过若干代理和路由器不加修改地传递的过程，这正是 Web 服务的核心。SOAP 协议用来传送 SOAP 消息。SOAP 消息是一个 XML 文档，包括一个必需的 SOAP 封装，一个可选的 SOAP Head 和一个必需的 SOAP Body，有比较严格的定义。SOAP 消息通过 HTTP 协议提出请求和进行传送。

3 基于 Web 服务的异构数据库集成设计与实现

结合相关技术，Web 服务在现有的实验条件下，在几台机器上利用 J2EE 平台实现异构数据源的集成，说明企业异构数据库集成的流程。

这里，我们给出基于 J2EE 和 Web 服务技术在企业系统中异构数据库的集成框架，如图 1 所示。下面介绍框架中几个主要模块的功能。

3.1 全局的数据描述

在基于 Web 服务的集成框架中，难点和重点是全球数据和局部数据的影射。全局的数据描述包括：局部数据库的信息，数据库类型（Oracle 还是 SQL Server），以及它所在位置即 IP 地址、用户名和密码等；全局表和局部表的影射；全局表的全局字段和局部表的局部字段的影射。

在总体设计框架中，统一使用数据库表作为全局数据字典进行以上各种关系的描述。通过全

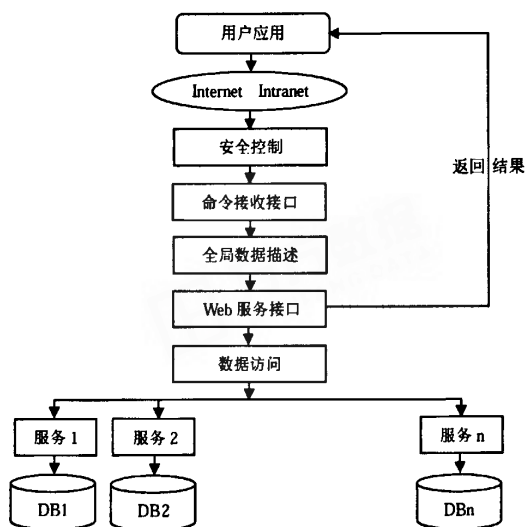


图 1 总体设计框架

局数据字典进行全局的数据描述，建立全局表 - 局部表的对应关系表和全局字段 - 局部字段的对应关系表。全局数据字典的数据结构包括如下 6 张表格。

表 1 TBL_DB_TYPE

字段	类型	约束
DBTYPE_ID	NUMBER	PK
DBTYPE_NAME	VARCHAR2(40)	-

表 1 用于对数据库产品进行描述，区分是 Oracle 数据库还是 SQL Server 数据库等。其中 DBTYPE_ID 表示数据库类型的 ID，作为这个表的主键，DBTYPE_NAME 表示数据库的名称，例如 Oracle10g, SQL Server2005 等。

表 2 TBL_LOCALDB_INFO

字段	类型	约束
LOCALDB_ID	NUMBER	PK
LOCALDB_NAME	VARCHAR2(40)	-
LOCALDB_USERNAME	VARCHAR2(40)	-
LOCALDB_PASSWORD	VARCHAR2(40)	-
LOCALDB_LAN_IP	VARCHAR2(40)	-
LOCALDB_WAN_IP	VARCHAR2(40)	-
DBTYPE_ID	NUMBER	FK

表2用于对局部数据库的描述。LOCALDB_ID表示这个表的主键ID,包括数据库的用户名、密码和数据库所在机器的广域网IP地址以及局域网IP信息,DBTYPE_ID外键依赖于TBL_DB_TYPE的DBTYPE_ID字段。

表3 TBL_GLOBALTABLE_INFO

字段	类型	约束
GLOBALTABLE_ID	NUMBER	PK
GLOBALTABLE_NAME	VARCHAR2(40)	-

表3用于对全局表的标识,描述了全局数据表的信息。GLOBALTABLE_ID表示主键ID,GLOBALTABLE_NAME表示全局表的名称。局部表的信息需要映射到该表中。

表4 TBL_LOCALTABLE_INFO

字段	类型	约束
LOCALTABLE_ID	NUMBER	PK
LOCALTABLE_NAME	VARCHAR2(40)	-
BELONGTO_LOCALDB_ID	NUMBER	FK

表4用于对局部表的信息标识,描述了局部数据表的信息。LOCALTABLE_ID表示主键ID,LOCALTABLE_NAME表示局部表的名称。局部表的信息需要映射到LOCALDB_INFO中。BELONGTO_LOCALDB_ID表示该表属于哪个局部数据表中。

表5 TBL_GLOBALLOCAL_TABLEMAPPING

字段	类型	约束
TABLEMAPPING_ID	NUMBER	PK
GLOBALTABLE_ID	NUMBER	FK
LOCALTABLE_ID	NUMBER	FK

表5用于说明全局表和局部表之间的映射,描述了一个局部数据表应该映射到哪个全局表中。TABLEMAPPING_ID表示映射表的主键ID,GLOBALTABLE_ID字段外键依赖于局部表TBL_GLOBALTABLE_INFO的GLOBALTABLE_ID,LOCALTABLE_ID字段外键依赖于局部表TBL_LOCALTABLE_INFO的LOCALTABLE_ID。

表6 TBL_GLOBALLOCAL_FIELDMAPPING

字段	类型	约束
FIELDMAPPING_ID	NUMBER	PK
GLOBALFIELD_NAME	VARCHAR2(40)	-
LOCALFIELD_NAME	VARCHAR2(40)	-
GLOBALTABLE_ID	NUMBER	FK
LOCALTABLE_ID	NUMBER	FK

表6用于全局数据表和局部数据表字段之间的映射,包括全局表中中和局部表中字段的映射。FIELDMAPPING_ID表示字段映射表的主键ID,GLOBALFIELD_NAME表示要映射到全局表中的字段名称,LOCALFIELD_NAME表示要映射的局部表的字段的名称,GLOBALTABLE_ID表示全局字段所依赖的全局表的ID,LOCALTABLE_ID表示局部字段所依赖的局部表的ID。

全局数据描述是实现异构应用集成的关键,它的主要功能是采用动态的方法建立全局表-局部表的对应关系表和全局字段-局部字段的对应关系表。

3.2 命令解析

命令解析是指系统解析用户在界面上输入的全局SQL语句,算法流程如图2所示。

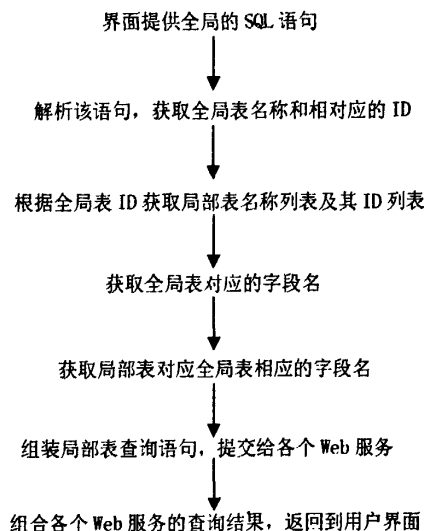


图2 命令解析流程

命令解析完成全局SQL到局部SQL的映射;访问控制是对局部SQL命令进行控制,产生访问序列并进行异常处理;数据访问的功能是执行局部SQL命令,从相应的局部数据服务中提取数据,并进行异常处理。数据访问分为两种:Web服务的直接数据访问和FTP文件传输。分散的数据以服务的形式注册在UDDI中,以供用户从接口中查询。安全机制建立在系统框架之上,主要解决访问控制的安全性和通信的安全性。系统执行

的过程为：当用户提出需要时，通过安全认证产生全局的查询语句，即通过全局数据描述中体现的全局概念、属性来产生全局的 SQL 语句。此时，将全局 SQL 语句通过 Web 服务接口提交于命令接收接口，并传送到命令解析部分，命令解析通过分析分解全局 SQL 语句，产生全局表名、字段名称，根据命令解析中存在的全局表 - 局部表的对应关系表和全局字段 - 局部字段的对应关系表，结合 UDDI 注册的服务点，产生一系列局部 SQL 语句。访问控制异常处理后，数据访问根据每个局部节点包装的数据执行局部 SQL 语句，将所获结果通过命令返回接口，提交给 Web 服务接口，并提交给用户。

3.3 Web 服务接口

Web 服务通过把本地的方法以服务的形式呈现出来供客户端以 Web 的形式进行访问。在利用 Web 服务实现异构数据源的集成过程中，按照总体设计框架，需要把实现数据库访问的 Java 类包装成 Web 服务，并把 Java 类中实现数据库访问的代码提供给用户访问。

把局部 SQL 语句发送给 Web 服务接口，然后 Web 服务在本地通过各种数据库访问技术完成数据的访问（在本示例中使用 JDBC 访问本地数据库）。

3.4 Web 服务数据访问技术

Web 服务的数据访问，即数据库的直接访问技术，是通过 SOAP、WSDL、UDDI 完成的。

(1) SOAP 协议使用 XML 作为封装和交换信息的标准，可以跨平台使用。

(2) WSDL 是一个由 XML 组成的文件，这个文件内容描述了实现程序代码对外提供的函数原型，也就是各种可供调用的函数名称以及参数等信息。

(3) UDDI 是一个公开的标准，它以结构化的方式来注册、管理企业信息以及相关的服务信息。

在异构的环境下，由于各个系统使用的数据库类型不一致，各数据库表结构可能不一样，字段名称及类型都可能不相同，因此对不同数据库

系统访问的查询命令也各不相同。在 Web 服务中，使用 SOAP 协议封装不同的数据类型，进行透明的数据传输，可实现一种统一的数据访问方式，而不需知道数据是从哪些局部系统来的和如何访问的。现在大多数 IDE（电子集成驱动器）都实现了 SOAP 和 WSDL 的 XML 文件的自动生成。例如在 JBuilder2005 中，不仅自动生成 WSDL 文件和 SOAP 文件，而且 Weblogic 服务器提供了 Web 服务代理，以一种统一的方式调用创建的 Web 服务。

4 小 结

本文结合异构企业信息系统异构数据集成这个研究课题，针对该系统数据多样性的特点，提出了基于 J2EE 平台的 Web 服务技术实现异构的数据源集成，并通过一个相对简单但是能够说明主要问题的演示程序，说明 Web 服务技术在企业信息系统集成项目上的正确性和可行性。

由于某些原因，仅给出一个样例程序来分析问题，但是该程序已经包含了全局和局部的转换和对 Web 服务的调用等，在该项目的实际开发中，需要解决和完善的问题如下。

(1) 在现有的实验条件下，由于自己的编写 Web 服务由自己调用，因此并没有将 Web 服务进行发布，然后通过 UDDI 进行定位和查找。在实际的系统中，需要将 Web 服务发布到某个注册中心上，这个中心往往是商业的，能够满足稳定和访问速度快等要求。在演示程序中，由于 Web 服务已知，因此省略了发布和查找定位这一步。

(2) 在对局部数据进行描述的过程中，如果局部数据发生变化，如增删或修改某个字段，都会对全局数据描述造成比较大的维护工作量。因此，如果要实现一个有实用价值的异构数据源的集成系统，全局表和局部表、全局字段和局部字段之间的映射必须要实现某种程度的自动化。

5 展 望

(1) 异构的数据源集成和异构的系统互操作一直是研究的热点。伴随着互联网和 Web 技术的

迅猛发展,基于 B/S 的开发模式逐步取代传统的 C/S 开发模式,“软件就是服务”逐渐成为现实。

(2) Web 服务作为资源整合技术,将不断得到发展。当前各大软件技术公司,如 Microsoft 和 IBM 都投入巨资进行 Web 服务技术规范的制定、技术的研究和产品的开发。有理由相信,Web 服务会逐步融合其他的集成技术(如 CORBA、COM 和 EJB)成为数据和应用集成的标准,是当前和未来需要关注的技术之一。

参考文献

- [1] 蔡月茹,柳西玲. Web Service 基础教程[M]. 北京:清华大学出版社,2005:234-351.
- [2] 李鹏,赵成勇. 利用 Java 技术实现基于 XML 的 Web 服务 [A]. Java 技术及应用的进展——第七届中国 Java 技术应用交流大会文集[C],2004:36-41.
- [3] 马淑娇,李晔,周俊林. 异构数据库集中的 XML 技术探讨[J]. 计算机应用研究,2004(1):94-96.
- [4] 范通让. 面向“信息孤岛”数据互操作模型的研究[J]. 无线电工程,2004(1):25-26,53.
- [5] 段永威,秦峰. 异构数据源数据转换工具的设计与实现[J]. 现代图书情报技术,2004(4):59-62.
- [6] 李冠宇,刘军,张俊. 分布式异构数据集成系统的研究与实现[J]. 计算机应用研究,2004(3):96-98.
- [7] 毛伊敏,赵秀鸟. 基于 XML 技术的电子商务异构数据集成研究与实现[J]. 商场现代化,2007(29):160-161.