

基于 Web Services 的异构数据库数据集成的研究

蔡振闹, 张艳超, 王立

(温州大学现代教育技术中心, 浙江温州 325035)

摘要: 针对高校多业务系统异构数据库的特征, 提出基于 Web Services 的校园异构数据库数据集成的框架体系结构, 并对数据集成的关键技术进行研究和设计, 为校园数据共享和互联互通提供一种技术方法。

关键词: Web Services; XML; 异构数据库; 数据集成

中图分类号: TP311.13 **文献标识码:** A **文章编号:** 1006-0375(2008)04-0025-05

高校在信息化过程中建立了大量基于各种业务流程和异构数据源的应用系统, 这些系统满足了高校特定职能部门的业务需求, 但由于数据自身的特点, 各职能部门业务系统间数据共享相当困难, 这样在校园内部就产生了“信息孤岛”, 阻碍了数字化校园进程。

对学校异构数据源进行集成是消除“信息孤岛”的有效途径。本文提出一种基于 Web Services 的异构数据库数据集成解决方案, 主要解决数据集成中的下列难题:

(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、数据库系统异构性和语义异构性;

(2) 透明性: 数据源异构程度越大对透明性的要求就越高。透明性主要包括: 平台透明性, 数据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, 同时能够决定是否提供和提供多少它的功能和资源服务于其它成员系统。

1 Web Services 及其相关技术

1.1 XML

XML (Extensible Markup Language) 是 W3C 组织发布的一种在互连网上交换数据的标准, 是定义标记语言的一种元语言。XML 提供一种与平台无关的格式, 实现了数据格式与数据内容的分离, 具有机器可理解的强大数据描述能力, 可在任何两个遵守 XML Schema 的应用间进行数据交换和消息传递, 遵守不同 XML Schema 也可通过 XML 文档转换技术 (XSLT) 进行有效的数据交换, 它作为一种具有可扩展性、结构性和平台无关性的描述数据的方法, 已成为异构系统的

收稿日期: 2007-02-19

基金项目: 温州大学 2007 年度校级科研项目 (2007L027)

作者简介: 蔡振闹 (1981-), 男, 浙江平阳人, 助理工程师, 硕士, 研究方向: 数据库, 计算机网络

信息交换的事实标准。

1.2 Web Services 架构

Web Services 体系结构基于三种角色（服务提供者、服务注册中心和服务请求者）之间的交互，交互涉及发布、查找和绑定操作，这些角色和操作一起作用于 Web 服务构件，即 Web 服务软件模块及其描述^[1]。在典型情况下，服务提供者托管可通过网络访问的软件模块（Web 服务的一个实现）。服务提供者定义 Web 服务的描述并把它发布到服务注册中心或服务请求者。服务请求者使用查找操作从本地或服务注册中心检索服务描述，然后使用服务描述与服务提供者进行绑定并调用 Web 服务实现交互。服务提供者和服务请求者角色是逻辑结构，因而服务可以表现两种特性。Web 服务体系结构图（如图 1）显示了这些操作和提供这些操作的组件及它们之间的交互。

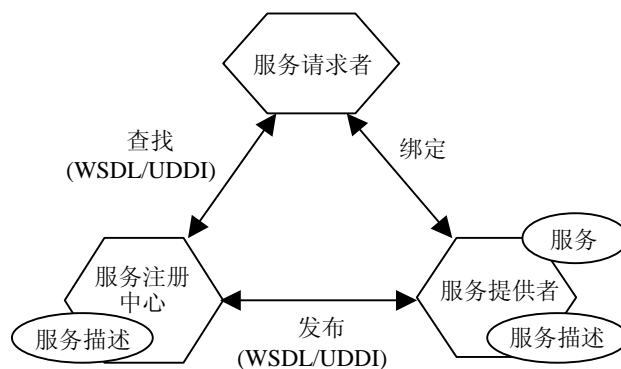


图 1 Web 服务体系结构

Fig. 1 Web Services Architecture

1.3 Web Services 相关标准协议

构建 Web Services 的三大技术基础是简单对象访问协议 SOAP，统一描述、发现和集成协议 UDDI，Web 服务描述语言 WSDL。首先，服务提供者所能提供的服务接口用 WSDL 和 WSFL 描述出来，然后使用 UDDI 在服务注册中心将这些服务接口进行注册；服务注册中心则使用 UDDI 注册这些服务接口，并接受服务请求者的查询；服务请求者通过使用 UDDI 在服务注册中心查询，得到所需的服务接口的描述文件，然后绑定到这些服务，最终完成调用。

SOAP 协议：SOAP 是一种简单、轻量级的协议，用于在 Web 上传输、交换 XML 数据。客户应用程序正是通过 SOAP 协议来访问 Internet 上的 Web 服务的。

UDDI：UDDI 是分布式 Web 服务的信息注册规范。Web Service 可以根据这一规范进行注册，以便被其它需要该服务的用户发现。UDDI 真正实现了 Web Service 信息访问的“一次注册，到处发布”。

WSDL：WSDL 是一种描述 Web 服务的 XML 语言，它定义了描述 Web 服务接口规范的标准格式。

1.4 Web Services 应用优势

利用 XML 统一数据描述格式，可使校园信息描述规范化。此外，XML 实现了表现形式和数据的分离，这允许它能够对来自校园不同应用系统的异构数据进行集成和处理。

使用 SOAP 协议统一数据交换格式，可保证校园信息在异构平台之间的无障传输。SOAP 协

议可以穿过任何防火墙，并且 SOAP 数据包中包含着以 XML 编码的数据，易于分析和使用。此外，SOAP 具有很好的伸缩性，能同时为非常多的用户服务。

WSDL 文件用于说明消息格式的表示法以 XML 架构标准为基础，这意味着它与编程语言无关，而且以标准为基础，因此适用于说明从不同平台、以不同编程语言访问的 Web Service 接口。此外，WSDL 还定义了服务的位置，以及使用什么通信协议与服务进行通信。

Web Service 是一种动态的集成方案^[2]，所有的服务都可以通过 UDDI 标准动态地被发现、绑定和使用，容易适应系统的变动，提高系统的灵活性和伸缩性，克服了过去使用的基于 RPC 和 API 的集成技术的缺陷，这也满足了松散耦合的要求。

2 异构数据库数据集成方案

2.1 异构数据库数据集成体系结构

异构数据库数据集成是共享或者合并来自两个或多个应用系统后台数据库的数据，创建一个具有更多功能的数字化校园应用的过程。异构数据库数据集成是数字化校园应用集成的基础和关键，异构数据库数据集成有两层含义：交互、共享^[3]。从交互角度来说，就是实现异构系统间的基于消息处理的数据交互，从一个数据源将数据移植到另外一个数据源来完成数据集成，是应用系统间的业务流程整合的基础。从共享角度来说，就是把不同来源、格式、特点性质的异构数据库数据在逻辑上或物理上有机地集中，为系统存储一系列面向主题的、集成的、相对稳定的、反映历史变化的数据集，从而为系统提供全面的数据共享。异构数据库数据集成体系结构如图 2 所示：

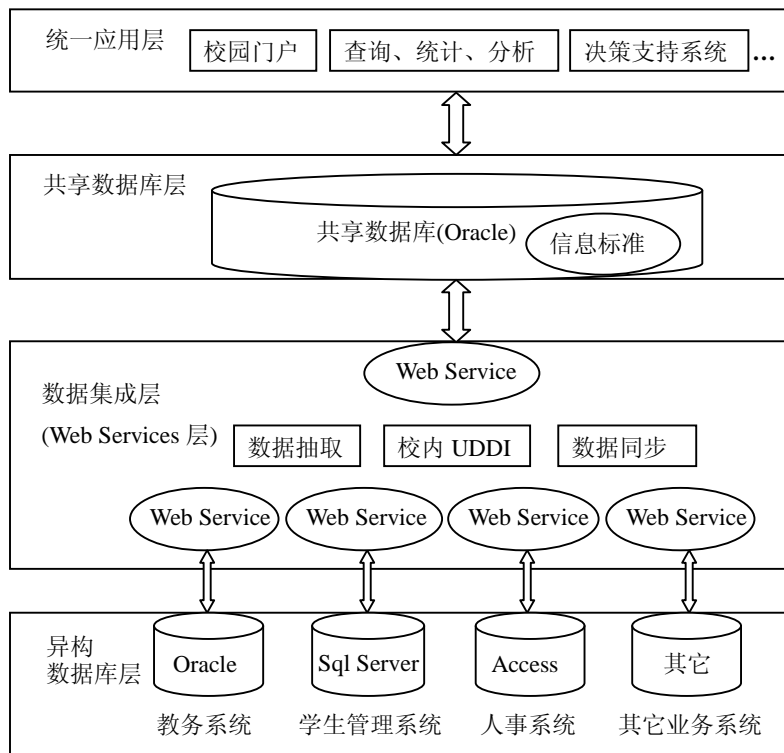


图 2 异构数据库数据集成体系结构

Fig. 2 Data Integration Architecture of Heterogeneous Database

异构数据库层：指校园内各职能部门业务系统所使用的数据库，这些数据库运行的环境以及数据格式、语义等可能存在很大的差异。如教务系统采用的数据库是 oracle 数据库，学生管理系统后台采用的是 Sql Server 数据库，人事管理系统后台采用的是 access 数据库等。

数据集成层：主要由数据抽取、数据同步、校内 UDDI 组成。它衔接着分布的异构数据库和共享数据库，对上层应用屏蔽掉底层数据库管理系统（DBMS）的异构性、分布性、复杂性以及数据结构的不一致性，同时还要解耦共享数据库与异构数据库。数据抽取实现异构数据库数据的清洗、转换、加载到共享数据库。数据同步实现异构数据库和共享数据库间数据的同步。校内 UDDI 汇聚数据集成层的所有 Web Services，这些 Web Services 完成异构数据库和共享数据库间的通信、交互、共享，采用基于 XML 和 Web Services 中间件的设计思想，以保证异构数据库集成层的通用性和高可重用性。

共享数据库层：根据已经制定的校园信息标准和校内共享需求，采用大型数据库 oracle 创建共享数据库平台。共享数据库集中存储异构数据库中需要集成和共享的数据。通过异构数据库集成层获取并存储需要的共享数据，为顶层——统一应用层（可能是门户系统、综合数据查询应用系统或是决策支持系统）实时提供共享数据，保证系统的高可用性。

统一应用层：基于以共享数据库为特征的校园异构数据库数据集成，进行校园应用集成，如创建校园信息门户、查询、统计分析系统等。满足数字化校园纵深发展的需要，开发决策支持系统以及数据仓库相关应用。

其中，异构数据库数据集成层，即 Web Services 层，是本文研究的核心和关键，下面着重讨论数据集成层的研究和设计。

2.2 基于 Web Services 异构数据库数据集成层的设计

为了实现对底层异构数据库复杂性的屏蔽，达到良好的灵活性、可配置性，解耦共享数据库层与异构数据库层，并保证异构数据库集成层的高可重用性，提出一种异构数据库集成层设计模型^[4]，如图 3 所示。

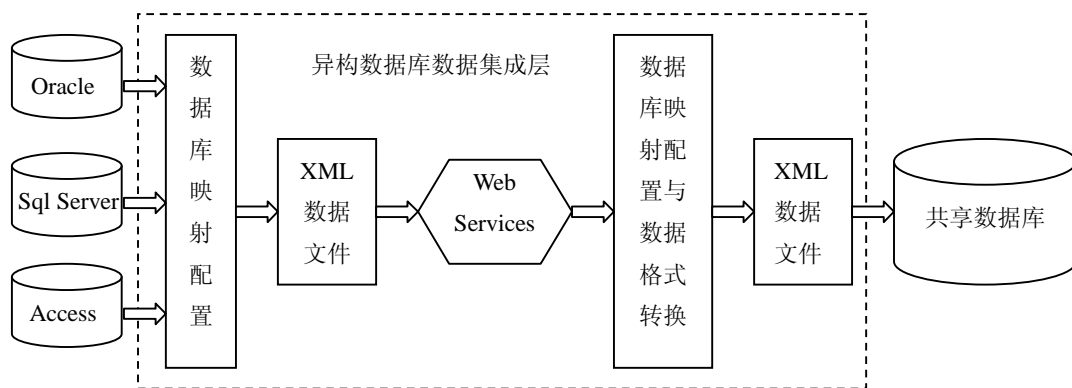


图 3 异构数据库集成层设计模型

Fig. 3 Design Model of Integration Layer for Heterogeneous Database

整体流程设计说明：

(1) 根据各分布的异构数据库结构信息和共享需求，动态的配置数据库共享设置，生成映射配置文件；

(2) 由程序根据配置文件自动从异构数据库中抽取设定共享的数据，并将其生成为 XML 文

件表示;

(3) 将 XML 文件包装成为消息, 通过调用校内 UDDI 中的特定 Web Service 将其发送到数据集成应用服务器上;

(4) 服务器端异步接收消息, 还原成 XML 文件, 并根据配置文件, 采用 XSLT 技术实现数据格式转换, 生成符合共享数据库格式要求的 XML 文件;

(5) 解析 XML 文件, 利用 XML 文件数据信息更新共享数据库, 最终实现异构数据库集成共享和同步刷新.

3 结束语

本文提出基于 XML、Web Services 技术构建跨平台低耦合的校园异构数据库数据集成的解决方案, 并对其中关键部分异构数据库数据集成层进行研究和设计, 提供一种数字化校园异构数据源数据集成的有效方法, 该方法能够有效解决校园数据共享和交换的问题, 实现校园数据的互联互通. 下一步研究重点是基于 Web Services 的异构数据库数据集成的具体实现技术的完善.

参考文献

- [1] David B, Hugo H, Francis M, et al. Web Services Architecture: W3C Working Group Note 11 February 2004. [EB/OL]. [2007-10-08]. <http://www.w3.org/TR/ws-arch/>.
- [2] 何勇, 陈世平. 基于 Web Service 的校园数据共享的设计与实现[J]. 计算机应用与软件, 2005, 22(10): 64-66.
- [3] 崔伟. 基于 XML 和 Web 服务数据集成的研究[J]. 计算机与数字工程, 2007, 35(2): 27-31.
- [4] 刘杰, 蒋朝惠. 基于 XML 的异构数据库集成[J]. 贵州工业大学学报: 自然科学版, 2007, 36(5): 49-53.

Research on Data Integration of Heterogeneous Database Based on Web Services

CAI Zhennaο, ZHANG Yanchao, WANG Li

(Center of Modern Educational Technology, Wenzhou University, Wenzhou, China 325035)

Abstract: Aiming to the characteristic of heterogeneous database of multiple application system in universities, the authors in this paper proposed the data integration architecture of heterogeneous database based on Web Services. Besides, they also researched and designed the key technology of data integration, and provided one technical method of data sharing and communication in campus.

Key words: Web services; XML; Heterogeneous database; Data integration

(编辑: 王一芳)