

数据资源服务化技术的研究与实现

王孝满^{1,2}, 周晓明², 毛宇光¹

(1. 南京航空航天大学 信息科学与技术学院, 江苏 南京 210016;

2. 中国电子科技集团公司第二十八研究所 信息系统工程重点实验室, 江苏 南京 210007)

摘要:为解决网络环境下分布式异构数据资源集成共享面临的资源成长性、用户不确定性问题,实现企业内部、跨企业的数据集成共享。文中依据面向服务的思想,采用 SDO 技术构建通用数据访问模型,并结合 Web 服务技术实现数据资源的服务化封装,提出了数据资源服务化的概念模型,建立灵活、敏捷的数据资源集成共享机制,最后提出了数据资源服务化的总体集成架构,列出了一次执行流程,验证了其可行性和松耦合性,能够很好地满足分布式异构环境的企业数据资源集成需求。

关键词:数据资源服务化;面向服务;服务数据对象;Web 服务

中图分类号:TP301.2

文献标识码:A

文章编号:1673-629X(2011)03-0079-04

Research and Implementation of Data Resource Serving Technique

WANG Xiao-man^{1,2}, ZHOU Xiao-ming², MAO Yu-guang¹

(1. College of Information Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China;

2. Science and Technology on Information Systems Engineering Laboratory, The 28th Research Institute of China Electronics Technology Group Corporation, Nanjing 210007, China)

Abstract: To resolve the problem of integration and sharing of heterogeneous, user uncertainty, changeable data resource, implement data sharing and integration inside and outside of enterprise. According to the thought of service oriented architecture, build an unified data access model based on the technology of service data object, propose a solution of data resource serving based on the technology of Web service, propose a concept model of data resource serving, build a flexible and smart sharing and integration mechanism of data resource. At last, introduce an integrated framework of data resource serving, list an execution process, experiments show executable and loose couple, meet the requirement of heterogeneous data resource integration in distributed environment.

Key words: data resource serving; service oriented; service data object; Web service

0 引言

数据集成的主要方法有构建标准数据模型,采用 XML 技术,结合人工智能技术等方法^[1],这些方法和实现技术在解决数据共享需求的快速变化、数据资源和用户的动态性问题方面存在着不足,并且构建的系统耦合性高,可重用性低。文中提出将 SOA(面向服务架构思想)^[2]引入到数据资源的集成共享中,SOA 的相关技术与理念可以很好地解决上述问题。

随着 SOA 的发展,基于 SOA 理念的数据集成技术被提出。文献[3]基于 SOA 分析了数据映射关系

和数据访问流,开发了流程数据集成的工具箱。SOA 关于服务技术的核心—Web 服务被运用到数据集成中,通过 Web 服务实现资源的虚拟化成为一种技术途径和方法,国内如文献[4]介绍了基于 Web 服务的数据集成,文献[5]使用 Web 服务实现异构数据集成。国外文献[6,7]分别提出了基于 Web 服务的数据集成方法和架构,后者应用于英国国家健康服务工程。SOA 的业务数据编程模型技术—SDO 技术也被运用到数据集成中,文献[8,9]提出了利用 SDO 技术进行异构数据集成。

文中在前人研究成果基础上,依据 SOA 思想,利用 SDO 技术构建通用的数据资源的访问方法,结合 Web 服务技术将各类异构的数据资源虚拟地封装成能够在网络上发布、查询、绑定调用的服务,构建一个灵活、敏捷的数据集成共享机制。

收稿日期:2010-07-18;修回日期:2010-11-07

基金项目:国家重点实验室基金项目(9140C8301011001)

作者简介:王孝满(1986-),男,安徽明光人,硕士研究生,研究方向为语义 Web 服务和本体;周晓明,研究员,研究方向为软件体系架构;毛宇光,副教授,研究方向为数据库理论。

1 数据资源的服务化概念和方法

1.1 数据资源服务化概念

定义 1:数据资源服务化是指将各种形式的数据资源访问方法封装成服务,以资源服务的形式提供数据访问、查询支持,实现数据资源虚拟化。数据资源服务化本质上是将服务技术引入到数据集成共享过程中,使数据资源能够像服务一样在网络上发布、查询和访问,使数据资源的应用具有灵活性、松耦合性和跨平台能力。这些服务带有数据资源的特征,数据集成演化为数据资源服务的集成。

定义 2:数据资源服务是指将各类数据资源访问和管理方法封装为服务,提供数据资源灵活、快速的访问方式。

数据资源的服务化过程是数据资源以服务的形式虚拟化的过程,其目的是屏蔽底层差异,对上层应用呈现一致的数据视图,并通过 Web 服务动态、灵活的部署发布机制实现数据资源的“即插即用”,面向需求动态构建数据共享环境。

1.2 数据资源服务化方法

数据资源的服务化是依据 SOA 思想,将各种数据资源构建形成一个服务系统(如图 1 所示)。

其核心是通过数据资源服务化封装将数据资源虚拟化,形成一个逻辑资源服务,并将服务注册到数据资源注册中心,数据资源用户可以通过服务发现/查找获得所需要的数据资源信息,绑定相关服务,通过服务调用实现对数据资源的访问。数据资源用户可以通过对不同的数据资源服务的组合、服务流程的编排实现复杂的数据集成。其具体步骤如下:

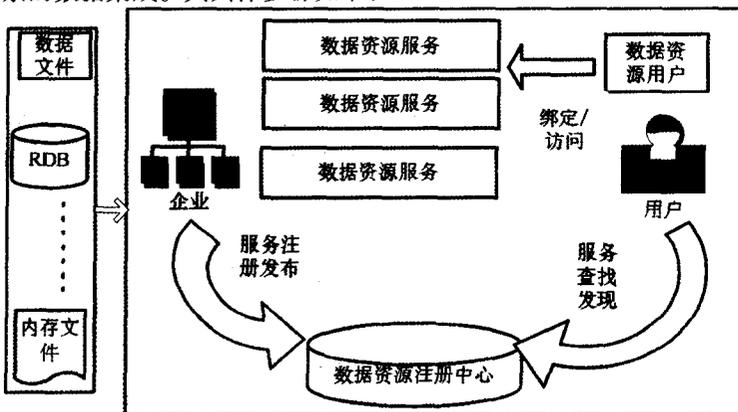


图 1 数据资源服务化方法

- 1) 依据数据共享需求,建立一个数据资源注册中心,提供数据资源管理注册服务功能;
- 2) 对各类不同形式的数据资源的访问管理方法接口实施服务化封装,开发部署对应的数据资源服务;
- 3) 数据资源提供者向数据资源注册中心注册、发布数据资源服务;

4) 用户按照需求,结合资源领域本体和服务语义信息,通过语义 Web 服务发现技术获得相关的数据资源服务,资源服务的发现方法采用文献[10,11]中提出的服务发现匹配方法;

5) 绑定数据服务调用接口,通过服务调用实现数据使用。

2 数据资源的服务化设计

2.1 数据资源服务化总体思路

总体思路:针对分布异构数据资源虚拟化需要解决的问题,采用 Web 服务技术将各个数据资源封装为数据资源服务并部署在数据资源注册中心,提出的数据资源服务概念模型见图 2。

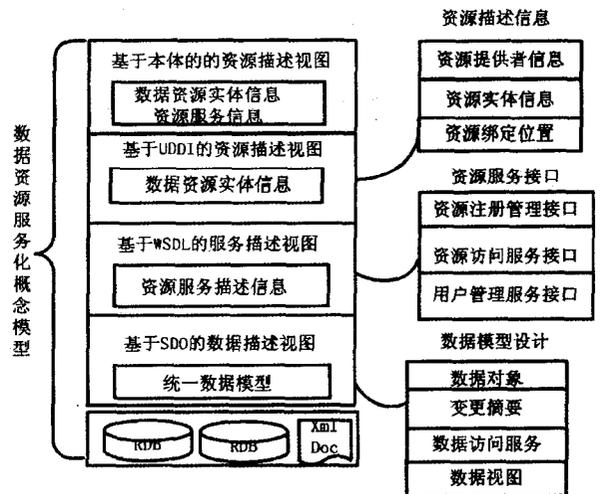


图 2 数据资源服务概念模型

针对数据资源服务化所需要解决的异构数据资源通用访问问题、资源动态发布、定位问题,数据资源服务化主要由四个部分组成:

一是基于 SDO 的数据视图,即采用 SDO 技术为用户建立统一的数据模型,提供统一的数据访问接口,这是数据资源基础访问表示模型;

二是基于 WSDL 的服务描述视图,即按照 WSDL 标准对数据资源相关服务描述,将数据资源服务分为三组服务接口,分别为数据资源注册管理、用户权限管理以及数据资源的访问接口;

三是基于 UDDI 的数据资源描述视图,即是注册、发布数据资源服务的信息模型;

四是基于本体的资源实体描述信息和资源服务信息,结合领域本体信息和数据资源服务本体可以实现基于语义的资源发现和定位,这个部分是数据资源虚拟化、逻辑化和按需发现、按需服务的基础。资源服务

本体采用 OWL-S^[12] 描述生成语义服务本体,资源实体信息本体采用 OWL^[13] 语言描述。

2.2 数据资源服务化的架构设计

数据资源服务化利用 SDO 标准针对各种异构数据资源统一数据访问接口和数据表现形式,形成一个统一的、跨平台的数据访问模式,实现了异构数据资源的统一的透明访问。而数据资源提供者可以将可共享的数据以服务接口的形式暴露出来,通过标准的服务描述语言 WSDL 加以描述,实现了数据资源集成的可扩展性以及数据资源的动态集成。数据资源服务化的总体架构如图 3 所示。

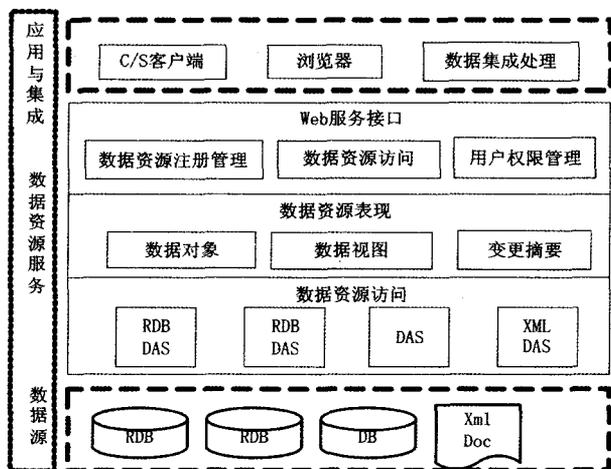


图3 数据资源服务软件体系架构

数据源层:数据源层是存在于网络上的数据资源,其形式包括关系数据库、XML 文件等;

应用层:各种应用程序可以无差别地利用 Web 服务提供的接口访问数据资源,应用层的应用程序充当数据请求者的角色,发出各种数据请求即服务调用的参数,透明地访问数据资源,接收统一格式的数据。

数据资源服务层:该层包含数据资源访问和服务封装两大部分,利用 SDO 编程模型构建统一的数据访问接口、访问方式及数据表示方式,Web 服务的服务功能接口以 SDO 提供的接口功能为基础,封装组合后形成功能丰富的服务接口,供应用层使用。

3 数据资源服务化的实现

3.1 数据资源描述及服务接口设计

资源描述:

数据资源描述模型即数据资源元数据,注册到数据资源中心,形成数据资源目录的信息,其具体的描述规范包括:

1)资源实体:资源名称,资源类型,资源描述,资源存储方式,标准规范,资源标识。

2)资源提供者:提供者姓名,提供者描述,隶属组

织。

3)资源绑定信息:URL 地址,服务器 IP,资源访问逻辑名,资源访问端口,访问用户名,访问密码。

这些描述信息,通过数据资源管理服务的注册接口,注册到数据资源服务中心,完成数据资源服务注册。

服务接口:

数据资源服务是由数据资源管理、数据访问、用户权限管理三类服务接口组成,数据资源的提供者、用户通过这些接口实现对数据资源的管理和访问。

数据资源管理服务接口是为数据资源的提供者使用的,一般在资源服务发布的服务描述文件(WSDL)中不公开发布,主要是实现数据资源描述信息的注册、删除、更新。

用户管理服务接口用来实现数据资源访问用户的管理,提供对数据资源访问用户的增加、删除、更新、验证和用户信息获取功能,验证与信息获取接口是对外发布的,供各类用户使用。其他的接口都是数据资源提供者使用的,不对外公开。用户分为三种角色:管理员、授权用户和普通用户,数据提供者可以创建管理员、授权用户和普通用户角色,它们权限依次降低,权限的差别体现在各自角色拥有的数据访问权限。

数据访问服务依据用户身份提供对数据资源的信息查询、更新、删除功能,数据访问服务依据用户身份分为三类,其中管理员接口具有对数据资源操作的所有权限,不仅能够实现数据资源的数据定义操作(DDL),而且可以实施 DML(数据管理操作),授权用户能实施数据资源的增、删、改操作,普通用户只能进行数据资源查询。

3.2 基于 SDO 的通用访问模型

SDO(Service Data Object)服务数据对象编程模型实现了 SOA 的业务数据编程模型,由数据图(Data Graph),数据对象(Data Object)和数据访问服务(Data Access Service, DAS)构成。数据资源服务所有接口的实现的基础是基于 SDO 编程模型的通用数据访问技术,SDO 中的数据对象格式的数据是数据资源数据表现形式,DAS 能够以离线的方式从数据资源抽取数据形成数据图,DAS 也负责创建用于数据对象和数据图的元数据,DAS 介于数据资源和 SDO 数据对象之间,它屏蔽了数据的异构性,实现了业务逻辑和底层数据的分离。参考文献[9]提出了基于 SDO 的数据资源通用访问中间件的实现方法,如图 4 所示。

文中采用了 Tuscany 开源社区提供的 SDO 和 RDB DAS 工具^[14],基于 java 语言,实现基本的数据访问接口,供上层的服务功能组合使用。DAS 是一个接口类,创建实例时,对于数据资源是关系数据库如 oracle 需

要抽取其元数据,即数据模式信息,如表结构、字段名、类型等;数据资源是 XML 文件需要加载文档结构信息文件。

先向提供者申请,提供者授权后将用户信息发布到数据资源注册中心,然后调用获取用户信息的服务获得带初始密码的用户信息和数据资源相关信息,用户也可以调用更新服务修改密码。

数据资源访问:用户获取权限信息后,先调用用户验证服务,验证通过后,发出调用命令,如 sql 命令查询服务,调用对应权限的数据访问服务,实现数据访问。自此完成了数据资源的注册发布,用户的授权管理,数据访问的全部过程。

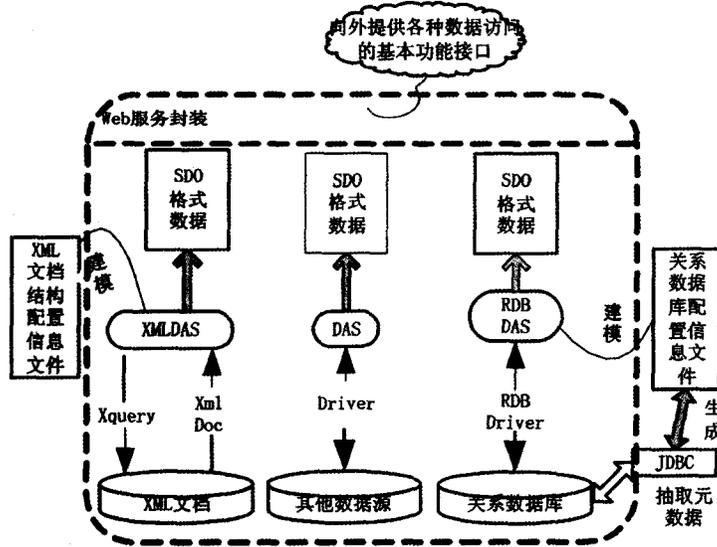


图 4 数据资源访问接口服务封装

3.3 数据资源服务化流程

数据资源服务的应用流程具体如图 5 所示。数据资源提供者、数据资源使用者及数据资源注册中心三者之间相互协作,以服务操作为单位,按时间先后顺序完成了数据资源注册、用户授权和数据访问的过程。

4 结束语

文中利用 SDO 与 Web 服务技术实现数据资源的服务化虚拟封装发布集成,并以 WSDL 文件和语义本体描述的方式发布数据资源,提出了数据资源集成共享的总体架构,具有跨平台、松耦合动态性、可重用、计算机可理解等特点。较好地实现了分布式异构环境下的企业数据资源的共享和集成。

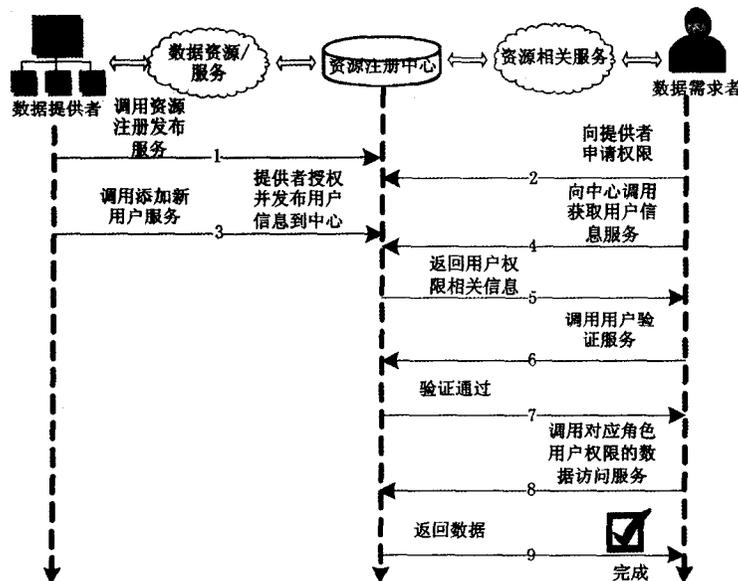


图 5 数据资源的发布、访问流程

数据资源注册:调用数据资源管理注册服务的接口方法,将数据资源实体信息、数据资源绑定信息、数据资源提供者信息一并注册到数据资源注册中心,实现数据资源的注册。此时对应生成了数据资源的名字,发布了数据资源的相关数据访问服务。

用户注册:数据资源注册成功后,数据提供者可以创建各种用户角色。数据需求者如果需要获取权限,

参考文献:

- [1] Halevy A, Rajaraman A, Ordille J. Data Integration: The teenage years[C]//In VLDB. [s.l.]:[s.n.], 2006:9-16.
- [2] 王紫瑶,南俊杰,段紫辉,等. SOA 核心技术及应用[M]. 北京:电子工业出版社,2008.
- [3] 傅向华,王志强,明仲. 基于 SOA 的流程与数据集成[J]. 小型微型计算机系统, 2007,28(10):1817-1821.
- [4] 岳昆,王晓玲,周傲英. Web 服务核心支撑技术:研究综述[J]. 软件学报,2004,15(3):428-442.
- [5] 黄浩,欧阳柳波. 基于 Web Services 的语义异构数据集成设计与实现[J]. 计算机工程与设计,2009,30(19):4406-4410.
- [6] Hansen M, Madnick S, Siegel M. Data integration using Web Services[C]//In:Proc. of the VLDB 2002 Workshop Efficiency and Effectiveness of XML Tools and Techniques and Data Integration over the Web (EEXTT). Hong Kong: Springer-Verlag, 2003:165-182.
- [7] Zhu F, Turner M, Kotsiopoulos L, et al. Dynamic data integration using web services [C]//Proc. of IEEE International Conf. on Web service. Washington, DC, USA:IEEE Computer Society,2004:262-269.
- [8] 郑垒,曹宝香. 基于 SDO 的异构数据集成研究与应用[J]. 计算机技术与发展,2009,19(11):163-166.

```
<传真>020-5651802</传真>
<e-mail>zhangna@126.com</e-mail>
</企业联系方式>
```

</企业>

其中企业名称、企业性质、姓名、性别、生日、传真、e-mail 是单一元素;企业、企业领导、企业联系方式是复合元素。

将 XML 文档转换为关系数据库数据,采用下面的映射规则:

(1)对于 XML 中的单一元素,将其直接映射到表中对应的列。

(2)XML 中的复合元素形成表,将复合元素中的属性和子元素映射到该表中对应的列。若是复合子元素,将其和主元素间的从属关系通过 PK 码联系起来。

对于上面的 XML 文档进行映射,结果如表 1~3 所示。

表 1 企业

企业名称	企业性质	PK
华南资讯	私有企业	1
新泰科技	私有企业	2

表 2 企业领导

姓名	性别	生日	PK
孙伟	男	1978-08-13	1
张娜	女	1975-02-21	2

表 3 企业联系方式

传真	e-mail	PK
020-5651708	sunwei@126.com	1
020-5651802	zhangna@126.com	2

3 结束语

针对企业信息化过程中出现的异构数据库集成的问题,以 XML 文档作为异构数据库之间转换的桥梁,提出了一个基于 XML 中间件的异构数据源集成模型^[11]。该模型构建于用户端和数据源之间,用户和数据源之间的信息交互以 XML 格式进行,解决了异构数据源之间数据集成的异构性,给用户提供了一个良好的访问平台;为了进一步满足多用户情况下对数据库

的访问,引入了数据库连接池技术,提高了访问效率,实现了用户的正确访问。该模型具有一定的可行性和实用性,能够胜任异构数据库的集成工作,对信息系统的集成有一定的实际意义。随着网络资源的快速发展,下一步的工作是如何将本技术与 Web 相结合,实现智能性的跨站点的信息资源共享^[12]。

参考文献:

[1] OUYANG Zhengzheng. Application of XML-based Heterogeneous Database Date Exchange Middleware in Ecommerce [J]. IEEE Software, 2009, 10(9): 132-135.

[2] YIN Dali, JIANG Meiyu. Teaching Information Resources Integration based on Heterogeneous Data Exchange and Sharing of Technology [C]//Computer Science and Computational Technology, 2008. ISCSIT' 08. International Symposium on. [s. l.]: [s. n.], 2008: 150-153.

[3] 于帆,王振铎,王振辉.基于 XML 异构数据库集成中间件的设计与实现[J].计算机应用研究,2007,24(9):185-188.

[4] 易宝林,谭志鹏.数据库访问柔性中间件的设计与实现[J].微电子学与计算机,2006,23(11):166-169.

[5] 史晔翎,黎建辉.关系数据库模式到 XML Schema 的通用映射模型[J].计算机工程,2009,35(7):35-39.

[6] 甄玉钢,刘璐莹,康建初.基于 XML 的异构数据库集成系统构架与开发[J].计算机工程,2006,32(2):85-87.

[7] 刘威,杨丹.基于虚拟视图的异构数据库集成平台的研究[J].计算机技术与发展,2009,19(6):91-94.

[8] 孙叶枫,宋中山. JSP 中基于连接池的数据库访问技术[J].计算机应用,2004,24(6):80-82.

[9] 靳其兵,李晓波.基于 JSP 的数据库连接技术的研究[J].计算机仿真,2007,24(4):108-111.

[10] 王蕊,陈庆奎.异构数据库集成中间件的研究与实现[J].计算机工程与设计,2008,29(22):5738-5741.

[11] XIONG Fengguang, HAN Xie, KUANG Liqun. Research and implementation of heterogeneous data integration based on XML [J]. IEEE Software, 2009, 9(8): 711-715.

[12] 李亚红,吴江,贾晖.基于 Web Services 实现异构数据库集成技术[J].计算机应用研究,2006(2):81-84.

[12] Martin D, Burstein M, Hobbs J, et al. OWL-S: Semantic markup for Web Services [EB/OL]. 2004. <http://www.w3.org/Submission/OWL-S/>.

[13] Smith M K, Wely C, McGuinness D. Owl Web ontology language guide [EB/OL]. 2003. <http://www.w3.org/TR/owl-guide/>.

[14] Tuscany sdo, version 2.0 [EB/OL]. 2004. <http://tuscany.apache.org/sdo-java.html>.

(上接第 82 页)

[9] 倪志刚,洪政,刘佳.基于服务数据对象的异构系统数据集成方案研究[J].计算机应用,2007,27:21-23.

[10] 吴健,吴朝晖,李莹,等.基于本体论和词汇语义相似度的 Web 服务发现[J].计算机学报,2005,28(4):595-602.

[11] Paolucci M, Kawamura T, Terry R, et al. Semantic Matching of Web Services Capabilities [C]//Proceedings of the First International Semantic Web Conference (ISWC2002). Sar-dinia, Italy: [s. n.], 2002.