

基于SOA的数据交换平台设计^①

杨 丽^{1,2}, 廉东本²

¹(中国科学院研究生院, 北京 100049)

²(中国科学院沈阳计算技术研究所, 沈阳 110171)

摘要: 为解决辽河流域水环境管理中异构环境下的数据集成共享问题, 提出了按照SOA思想建立的数据交换平台的总体架构, 该平台由数据适配器、数据交换引擎和服务管理模块等组成, 并详细介绍了平台的功能模块及数据交换的流程, 最后设计实现了平台中的数据适配和数据转换两个关键模块。

关键词: SOA; 数据交换; 交换流程; 适配器; 数据映射

Design of Data Exchange Platform Based on SOA

YANG Li^{1,2}, LIAN Dong-Ben²

¹(Graduate School, Chinese Academy of Sciences, Beijing 100049, China)

²(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110171, China)

Abstract: To solve the problem of data integration and sharing under the heterogeneous environment of LiaoHe River water environmental management, we propose the overall structure of data exchange platform according to the SOA thought. The platform is composed of data adapter, data exchange engine and service administration module and so on. It introduces the functional module and the flow of data exchange in the platform. Finally, we design two essential modules: data Adaptation module and transformation module.

Keywords: SOA; data exchange; exchange process; adapter; data mapping

1 引言

近年来, 我国突发型和累积型水环境污染事件频发, 其中辽河流域水环境保护形势日趋严峻, 为保护流域水资源、监控流域污染, 辽河流域开展了水体污染控制及水体污染治理的研究, 研制了各种水环境管理信息化系统, 包括辽河流域水环境污染源管理系统、水环境质量管理体系、水环境风险评估系统、水环境预警系统和应急响应系统等。但是辽河流域大部分水环境管理模式主要以“各自为政、条块分割”方式为主, 行政辖区之间缺乏有效的沟通与环保信息交流, 往往同一跨界断面上、下游监测数据就不一致, 而且系统建设缺乏统一规划, 所采用的环境、平台和核心技术存在很大差异, 数据库管理系统、操作系统的异构性和不同的开发技术严重阻碍了水环境系统应用中

的数据交换和信息共享。如何解决数据资源的交换共享问题, 发挥数据信息应有的效能, 对辽河流域水环境综合监管系统建设影响重大。

SOA(Service-Oriented Architecture, 面向服务的体系架构)是一种组件模型, 将应用系统的不同功能单元定义为服务, 并通过服务之间定义良好的接口和契约联系起来。从功能角度分析, SOA是一种构造分布式系统的方法, 它将业务应用功能以服务的形式提供给应用程序或其他服务^[1]。SOA与大多数通用的客户端/服务器模型的不同之处, 在于它着重强调软件组件的松散耦合, 并使用独立的标准接口^[2], 不涉及底层编程接口和通信模型, 能使服务可重用, 简化开发。

XML(eXtensible Markup Language)为可扩展的标记语言, 允许用户自定义标记和属性, 并可依照所

① 基金项目:国家水体污染控制与治理科技重大专项(2009ZX07528-006-05)

收稿时间:2010-08-20;收到修改稿时间:2010-09-21

定义的标记与属性的语法来开发应用程序; XML 可以通过标记来描述数据, 或配合属性来辅助描述数据, 宜作为对象或标准的描述语言; XML 可以借助验证规则来规范一个 XML 文件的结构和内容, 具有可扩展性、结构化和平台无关性的特性, 适合网络环境下异构系统之间的数据交换^[3]。

通过对现有数据交换技术的研究, 本文提出按照 SOA 思想、以 XML 作为中间数据交换格式建立数据交换平台, 以实现对环境专题、流域水质监测、气象、水文、应急响应等水环境相关数据的交换与共享, 为流域水环境综合整治提供及时、高效的信息和技术服务支撑。

2 数据交换平台的设计

2.1 数据交换平台的设计目标

数据交换平台应能连接水环境综合监管系统中的各种异构系统, 支持不同应用系统的数据通过数据交换平台进行交换和共享, 其主要的设计目标包括:

1) 具有强大的数据适配能力, 支持多种数据类型相互转换, 包括 Oracle、DB2、SQL Server、Sybase、MySQL 等多种主流关系数据库类型和 Excel、XML、TEXT 等多种文件类型;

2) 支持即时触发和定时任务发起的数据交换请求。当需求方第一次接入系统或根据业务需要随时发出数据交换请求时, 需要支持即时触发交换方式; 而已经建立交换规则的业务系统间可以按照预先设定的周期和时间执行数据交换, 支持增量传输^[4]。

3) 支持同步和异步数据交换。当数据量不大、及时性要求较高时, 需提供同步交换, 即需求方发送交换请求后一直等待, 直到接收到数据并返回交换结果; 当数据量较大时, 传输时间较长, 此时需能进行异步交换, 采用“存储转发”或消息队列服务。

4) 保证数据在传输过程中的安全和完整性, 防止数据在传输过程中被恶意篡改或丢失。支持断点续传、分块传输, 由于提供者发送的数据打包成消息按时间顺序存放在消息队列中, 可以保证按消息产生的时间序列依次发送、接收和处理消息, 这样能保证消息传递的可靠性和时序性。

5) 提供对事物的支持。需求方只有成功的发送请求后, 才删除消息队列中的消息, 否则进行事物回滚。适配器只有成功将数据存入目标数据库后, 才给提供

方发送确认信息; 只有接收到提供方或需求方的确认信息后, 才能删除持久性存储的消息。

6) 数据交换平台应具备完善的日志管理机制。对交换任务全过程进行监控和日志管理, 具备直观、详细的运行状态日志, 并具有日志分析功能。

2.2 数据交换平台模型设计

水环境综合监管系统数据交换平台的基本目标就是为不同业务系统间提供数据交流、转换服务从而达到数据资源共享。整个平台按照 SOA 思想设计, 总体架构如图 1 所示, 主要包括适配器、交换引擎、服务管理、系统管理等模块, 分成数据层、适配层、交换层和服务管理层四个层次。其中数据层是由各应用系统的异构数据源组成; 适配层完成异构数据源与平台的数据交互, XML 与异构数据格式的双向转化, 并可按需求实现数据加密、解密、压缩和解压; 交换层提供数据抽取、数据校验、数据接收队列和发送队列的管理、数据转换、数据发送和记录数据交换的日志等, 是整个平台的核心。服务管理层完成交换标准、映射规则、路由规则的配置、交换过程的监控等。

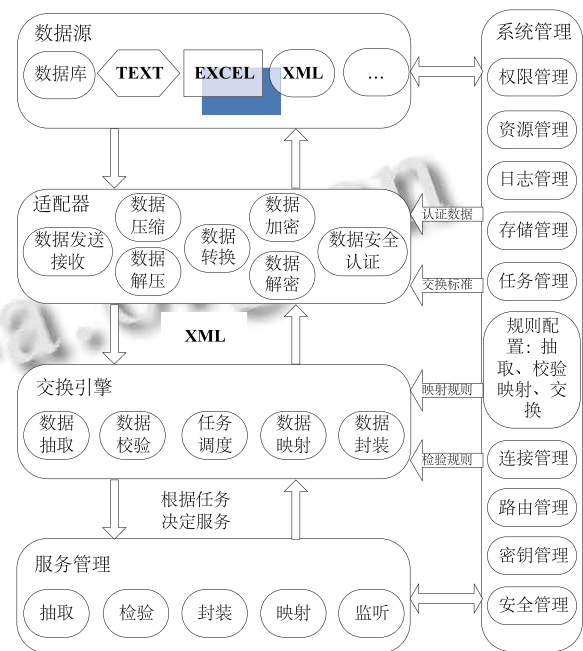


图 1 数据交换平台总体架构

2.3 数据交换流程及功能模块

该平台中数据交换的基本流程是各个业务系统通过平台的权限分配和认证后, 展示源数据模式(数据库表结构等), 数据需求方根据展示和数据提供方建立交

换契约，然后提供方通过接收适配器将本系统的数据源接入数据交换平台，适配器将数据转换成源 XML，并打包成 SOAP 消息进入接收队列后传入数据交换引擎，交换引擎对数据进行抽取、校验等后按照需求方的映射规则将源 XML 转换成目标 XML，并打包成 SOAP 消息进入发送队列后传入发送适配器，发送适配器将目标 XML 转换成数据需求方的数据格式并加载到需求方的数据源^[5]。其中 XML 是数据交换的中间数据格式，整个平台的核心数据处理流程如图 2 所示。

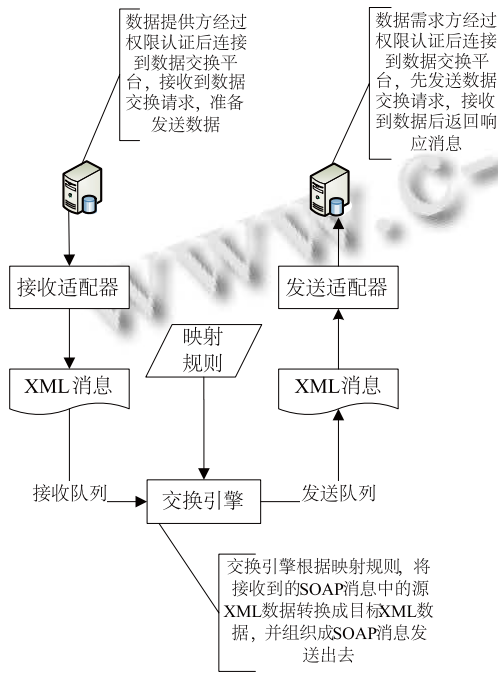


图 2 数据交换平台的核心数据处理

数据交换平台在实现过程中，其主体功能可以分为 7 个模块，这些模块根据交换平台层次的划分可以分解为多个子模块，每个模块根据其功能也可以划分到平台中的不同层次，各个模块的名称和主要功能如下：

- 1) 权限管理和访问控制模块：实现给用户分配角色和功能资源等权限，并进行权限分配的管理和维护，使用 B/S 结构实现对系统功能访问权限的控制^[6]。
- 2) 数据资源管理模块：实现对各异构系统数据资源的注册、更新和授权。数据源注册关系模式信息，并授权数据使用者，当关系模式改变时，可以进行更新操作。
- 3) 交换任务调度模块：实现“请求-应答”模式的

数据交换服务功能。在数据需求方与提供方建立交换契约后，通过对列实现对数据交换任务的调度。

- 4) 标准配置管理模块：实现对数据映射规则、交换标准、检验规则、数据安全认证、密钥等的配置管理。
- 5) 数据适配模块：实现对各种同构、异构关系数据库中数据的采集和加载，存取的数据表和字段可以灵活配置，并对数据进行安全认证；数据采集的方式支持即时触发或定时触发。
- 6) 数据转换模块：提供对数据的缓存、抽取、校验、映射和封装功能。其中映射是 XML 到 XML 的数据转换，采用 XSLT 技术，按照映射规则进行转换，其中映射规则可以手动配置，也可以根据双方的数据模型自动生成。
- 7) 日志管理和实时监控模块：记录系统在使用过程中中的重要运行信息，为管理者提供交换数据的日志和统计分析数据，以实现对数据交换进行全面监控。

3 数据交换平台关键模块的研究

3.1 数据适配器

数据适配器是各业务系统数据源与数据交换平台交互的接口，适配器从业务系统采集数据并进行预处理，将数据转换为统一的 XML 格式，并保证数据传输的效率及安全性，提交到数据交换引擎处理。当业务系统接收数据时，适配器将 XML 格式数据转换为用户需要的格式并存入相应的数据库中^[7]。适配器在

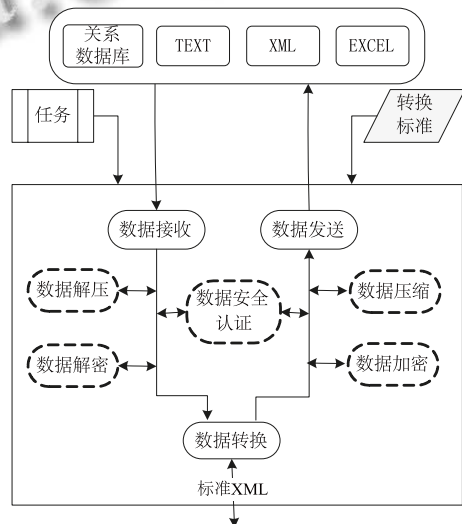


图 3 数据适配器模型图

设计时将采用组件技术可以按照平台的配置文件进行灵活的部署和扩展,数据适配器模型如图 3 所示。

数据适配器功能的具体实现包括:(1)适配器利用 JDBC、ODBC 等数据库驱动程序实现对不同数据库的访问。自动生成和创建数据库接口表、主键、索引等,支持批处理(一次可传送多条数据库纪录);支持非结构化数据和大字段数据的处理;在采集数据时,为保证数据的完整性,分为全量采集和增量采集。(2)适配器根据提供者数据源注册的关系模式产生 XML Schema,以此为交换标准将采集的数据转化为 XML 数据;接收从数据交换引擎传来的 XML 数据并转换成用户需要的数据格式。(3)按照任务可以选择对数据进行压缩、加密等以保证传输的效率和安全性。将数据按照 SOAP 协议进行打包和拆包,并把 SOAP 消息和 HTTP 协议绑定进行传输。

3.2 数据转换模块

数据转换模块是整个数据交换平台的核心,主要实现数据的缓存、抽取、校验、映射和封装功能。

适配器将含有数据的 SOAP 消息送入接收队列,等待数据转换模块的调度,这个队列能对数据起到缓存作用。转换模块从队列中得到 SOAP 消息,并消息中抽取数据,进行格式校验、数据校验,校验合格后再根据映射规则将源 XML 数据转换成目标 XML 数据,并封装成 SOAP 消息,发送到发送队列,等待适配器的处理。

其中数据映射是这个模块的主要功能,包括 XML 模式管理和映射规则管理。当双方第一次进行数据交换时,数据需求方将所需要的数据模式注册到数据提供方的模式库中,并定时进行更新,保持模式信息的同步。这样每个数据交换方的模式库中既包含本系统的数据模式,又包含曾与之发生过数据交换的系统的数据模式文件。在交换时,数据转换模块就可以根据双方的数据模式产生映射规则,并据此进行源 XML 数据与目标 XML 数据的转换。

映射规则的产生就是对交换双方的 XML 模式进行匹配,在匹配过程中,需要解决双方数据结构字段间的一些异构问题:(1)字段类型不同:将源字段的数

据按照目标字段的数据类型进行转换。(2)字段名称不同:由于交换时是以目标模式交换数据,故只需指定字段间的对应关系,直接将源数据转换为目标数据。

(3)字段拆分或组合:需要将源字段的数据进行拆分,并映射到目标实体的多个字段;或者是将源实体中多个字段进行合并或函数计算后映射到目标字段。其中映射规则可以是手动配置的,也可以根据双方的关系模式自动生成。

4 结语

数据交换平台是水环境综合监管系统建设的重要组成部分,关系着系统内部各异构子系统之间的数据交换和共享。本文从水环境综合监管系统数据交换的需求出发,研究了 SOA 架构思想和 XML 技术,并在此基础上提出了基于 SOA 的数据交换平台模型。目前该系统已经完成了方案架构设计和关键模块的研究,现在的设计主要支持关系型数据库,将进一步研究利用数据适配器的可扩展性实现 Excel、XML、TEXT 等多种文件类型的数据交换和共享。

参考文献

- 1 Rafzig D, Bamke K, Slama D. Enterprise SOA 中文版-面向服务架构的最佳实战.韩宏志译.北京:清华大学出版社,2006. 41-42.
- 2 凌晓东.SOA 综述.计算机应用与软件,2007,24(10):122-124,199.
- 3 尹芳.基于 SOA 的应急数据交换平台研究与实现[硕士学位论文].天津:天津大学,2008.
- 4 Kezunovic M. Ownership of Data and the Need for Information Exchange. Belgrade, Yugoslavia: Balkan Power Conference, 2002.
- 5 Extensible Markup Language (XML) 1.0 (Fifth Edition). [2010-3-11]. <http://www.w3.org/TR/xml/>
- 6 梅立军,付小龙,刘启新,沈锡臣.基于 SOA 的数据交换平台研究与实现.计算机工程与设计,2006,27(19):3601-3603.
- 7 郭瑜.基于 ESB 构架的交通行业数据交换平台的应用研究.计算机系统应用,2008,17(9):83-85.