

基于 Web Service 的数据集成研究及应用

苏 琪¹, 刘西林¹, 王 军²

(1. 西北工业大学 管理学院, 陕西 西安 710129;

2. 西安财经学院 信息学院, 陕西 西安 710100)

摘 要:销售计划的制定是销售企业重要工作之一,制定煤炭销售计划需集成多方面的信息。在对比分析现有数据集成技术的基础上,提出一种集成企业内、外部异构数据源的模型,以解决数据的结构异构和数据格式统一问题。模型通过调用 Web 服务获取异构数据源数据,解决了数据的实时性问题,对请求结果的处理运用了将 XML 转化成实体类对象进行存储的技术,使模型融入了数据分析功能。在保证企业外部数据源独立性的前提下实现了异构数据集成,并经过实际应用验证了其有效性。

关键词:数据集成;异构;Web 服务;实时性;反序列化

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2014)08-0051-04

doi:10.3969/j.issn.1673-629X.2014.08.012

Research and Application of Data Integration Based on Web Service

SU Qi¹, LIU Xi-lin¹, WANG Jun²

(1. School of Management, Northwestern Polytechnical University, Xi'an 710129, China;

2. School of Information, Xi'an Financial and Economic University, Xi'an 710100, China)

Abstract: Drafting sales plans is one of the important tasks for marketing companies. Many aspects of information need integrating in the process of establishing plans of coal sales. In order to solve the problem of heterogeneous data structure and normalize data format, based on the contrastive analysis of existing data integration technologies, put forward a model to integrate heterogeneous data source inside and outside the enterprises. This model calls Web service to obtain heterogeneous data to solve the real-time problem, and introduces a technology, by which XML is transformed to entity objects and then is stored, to deal with the results of requests. In this way, a real-time query and the function of data analysis are combined in the model, achieving the integration of heterogeneous data on the premise of guaranteeing the dependence of external data source. Besides, its validity has eventually been checked in practical use.

Key words: data integration; heterogeneous; Web service; real-time; deserialization

0 引 言

煤炭销售企业销售计划的制定需综合多部门信息系统的信息,不同时期构建的信息系统的架构以及所采用关键技术、开发环境等不一样,导致各信息系统之间较为独立,无法进行信息交换和数据共享,因此产生了数据集成的需求。

数据集成是将各个分散、异构的数据源,按照一定的方式组织起来,为用户提供统一的访问接口,使用户能够方便地对从各个渠道获得的数据进行集中管理。

通过数据集成,能够对用户屏蔽数据源的异构性,解决各分散的数据源的数据格式不一致、数据的结构异构、语法语义异构等问题,让用户能够透明地访问这些异构数据源。目前,已经较为成熟的数据集成解决方案主要有联邦数据库技术、中间件技术、数据仓库技术,许多学者在以上解决方案的基础上提出了一系列数据集成的体系结构和框架。国内的学者,赵毅、王浩然等,提出了一种基于 XML 的数据集成框架 XDIFrame,通过设计通用转换工具,在各异构数据源建立针对本

收稿日期:2013-09-24

修回日期:2013-12-26

网络出版时间:2014-04-24

基金项目:国家自然科学基金-煤炭联合基金(U1261111)

作者简介:苏 琪(1989-),男,河南商丘人,硕士研究生,研究方向为系统集成、管理信息系统;刘西林,硕士,教授,研究方向为运筹学、管理信息系统;王 军,博士,副教授,研究方向为信息系统、煤矿环境等。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20140424.0824.067.html>

地数据模型的 XML 包装器,使用以 XML 定义的标记语言指导数据的转换和校验,为数据访问提供了一致的 XML 数据环境^[1]。钟将,宋娟根据实际的数据需求模型,在传统的数据集成框架的数据集成中间件模块中加入本体语义描述模块,提出了基于本体的数据集成框架,有效解决了数据集成的语义异构问题^[2]。重庆大学的杨长辉,结合 Wrapper/Mediator 体系结构和数据仓库技术,提出了一种扩展的异构数据集成与分析体系结构,HDQAS(Heterogeneous Data Query and Analyzing System),在该体系结构中,采用 XML 解决语法异构问题,使用本体描述语言 OWL 描述全局本体和局部本体,通过全局本体到局部本体的映射,解决数据源的语义异构问题^[3]。国外方面,研究数据集成比较著名的项目有 Stanford 的 TSIMMIS,IBM 的 The Garlic Project,曼彻斯特大学的 Mark Turner 等人,针对传统的数据集成方案不能满足多变的集成需求,提出新的面向服务的动态集成方案^[4-7]。

当前学者对数据集成的研究大都集中在解决数据的语义异构和结构异构问题,而且要集成的异构数据源在逻辑或地理上处在同一区域^[8-9],煤炭销售企业销售计划的制定,不仅需要来自企业内部各部门的相关数据,有时还需要企业外部信息系统的数据。数据集成不仅要能缩小“信息孤岛”,而且要保证原数据源的独立性和自治性。文中从以上角度出发,结合煤炭销售企业的需求,提出一种针对逻辑上或地理上处在不同区域的异构数据源的数据集成模型,用以解决数据的结构异构和数据格式统一问题。

1 基于 Web Service 的数据集成模型

综合比较了联邦数据库技术、中间件技术、数据仓库技术。联邦数据库中每个独立的数据源都要与其他数据源建立彼此访问接口和映射关系,随着数据源的增多,数据查询效率会受到限制,主要适用于异构数据源较少的数据集成^[10]。数据仓库技术是将异构数据源的副本存到数据仓库中供用户操作,用户获取的数据对象是经过转换后的数据,对数据操作的效率较高,但数据仓库中数据实时性不强,该技术适用于对数据联机分析和决策支持系统的集成。中间件技术中,用户能直接对数据源操作,由于该模式的数据源具有自治性好、查询实时、数据源配置灵活等特点,被广泛应用于应用系统的数据集成。煤炭销售计划的制定要综合运输车辆、煤炭的最新状态信息,销售计划更多的是关于运输车辆的调度,对数据的实时性要求较高。为了完成既定的销售计划,每一个日销售计划的制定都要向月计划逼近,日计划的制定和修改又是建立在对历史销售数据分析的基础上。基于以上分析,文中提

出了以中间件技术为基础,并具有数据分析和决策支持功能的数据集成模型,如图 1 所示。

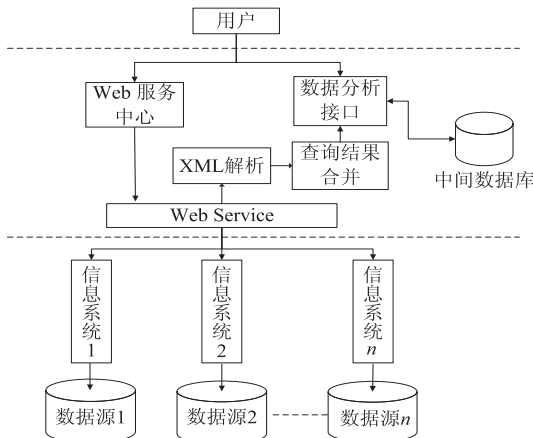


图 1 基于 Web Service 的数据集成模型

该集成框架从逻辑上可以分为三层,分别是用户接入层、中间层、数据访问层。用户接入层主要负责提供给用户一系列数据访问的接口及把数据集成的结果反馈给用户。中间层,也是集成框架的核心层,主要任务包括对用户请求的处理和用于数据分析的集成数据存储。用户通过 Web 服务中心查看可以调用的 Web Service,根据 Web Service 的 WSDL 描述,提交请求的 SOAP 消息,返回查询结果。查询结果以可扩展标记语言(eXtensible Markup Language, XML)格式的字符串呈现,中间层还需对 XML 进行解析,合并查询结果,将查询结果以新的数据模式存储到中间数据库。数据访问层直接与数据源打交道,把对数据源的操作封装成 Web Service 发布出去,并执行由中间层传过来的用户对数据源的查询操作。

与一般数据集成模型不同的是,该集成模型具备以中间件为中心的集成模型查询实时性的优点,用户通过调用 Web Service 能够直接从异构数据源获得数据,从而保证了每次查询得到的数据都是最新的数据。通过 Web Service 获得数据,不仅屏蔽了底层数据库管理系统的差异和数据源的逻辑、地理位置的差异,还降低了系统之间的耦合性和底层数据获取的复杂程度,保证了异构数据源的独立性和自治性。同时,该模型还兼备了数据仓库技术高效查询效率和数据分析功能的优越性。用户做数据分析时不必每次都访问底层的异构数据源,只需从中间数据库查询格式转换后的数据便可实现历史数据的分析。

2 模型关键技术

2.1 Web Service

模型在数据访问层应用到 Web Service 技术,Web Service 是一种服务导向架构的技术,用以支持网络间不同机器之间的数据交换和互操作。它使用 XML 描

述数据结构与数据类型,使用 SOAP(Simple Object Access Protocol)表示信息传输协议,使用 WSDL(Web Services Description Language)进行本身内容描述,使用 UDDI(Universal Description, Discovery and Integration)来发现、描述与集成 Web Service^[11]。其中简单对象访问协议(SOAP)是轻量级的协议,它是基于 XML 的、用于在应用程序之间以对象的形式交换数据的、不依赖于传输协议的通信协议,是 Web Service 的主体。SOAP 规范了消息传输的格式,提供了数据编码的标准并提供了远程过程调用的一系列规则。WSDL 用于对 Web Service 的描述,用来告知服务使用者 Web Service 的方法的命名、参数格式及返回值等信息。UDDI 是一套基于 Web 为 Web Service 提供的信息注册中心的实现标准规范,它包含这样一组规范,该规范不仅使得企业能够注册自身提供的 Web 服务,还能使得别的企业能够发现自己的 Web 服务。Web Service 为数据集成提供了灵活的访问方式,用 XML 统一了数据格式,为快速新增和部署新数据源提供了方便。

每个具体的 Web Service 都可看作一个应用程序,该应用程序向外界提供能通过 Web 调用的 API 接口,服务使用者可通过编程的方法来实现 Web Service 的功能。Web Service 的调用原理如图 2 所示。



图 2 Web Service 调用原理

2.2 请求结果处理

该集成模型使用 XML 作为数据传输的媒介,解决数据源数据格式不统一的问题。XML 是一种 Web 上标准的、结构化的数据描述语言,常被设计用于在 Web 上传输和存储数据,它的可扩展性、自描述性、平台无关性等特性造就了 XML 的优点:适用于应用程序之间的数据交换和共享。在 Web Service 调用的过程中,服务使用者通过 SOAP 消息提交服务请求,服务提供者根据请求做出响应,以 XML 格式数据返回查询结果^[12],过程如图 3 所示。

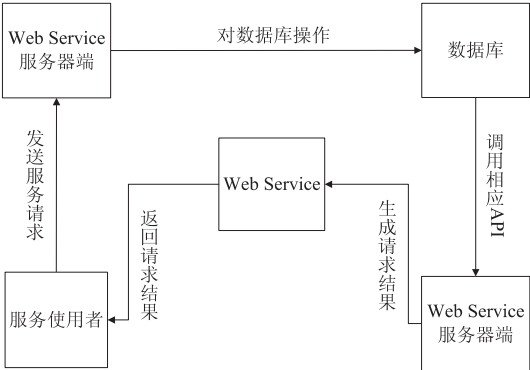


图 3 Web Service 请求过程

查询结果形成其实是关系数据库数据到 XML 的转换过程,该过程的目的是将数据对象的状态信息转换为可以存储或者传输的形式,将对象的公共字段或属性、方法的参数及返回值序列化为符合特定的 XML Schema 定义语言文档的 XML 流,从而便于异构信息系统之间的数据交换和共享^[13]。在模型的中间层,通过调用 Web Service 获得的数据并不能直接被用户识别,用户甚至不清楚合成后的查询结果的具体含义。为解决请求结果的可读性和方便存储性问题,文中采用了对其进行反序列化处理的方法。用户得到请求结果的 SOAP 消息之后,对合并后的请求结果进行反序列化处理,转变成实体类对象,还原对象的成员变量在序列化之前所表示的数据含义。例如,煤炭销售企业每一个调度计划的制定需要煤炭品种及数量、运输车辆的装载能力、客户需求品种及数量等信息,以销售计划为根基,以车辆、煤炭的状态信息为约束,制定调度计划,计划的形成如图 4 所示。

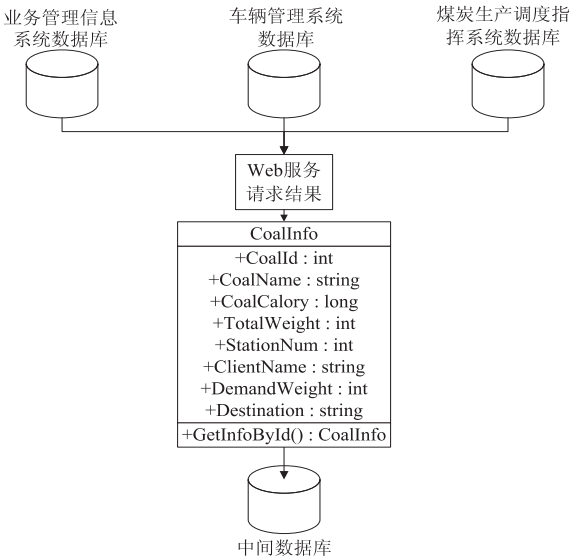


图 4 XML 反序列化及存储过程

3 应用实例

SH 集团是以煤炭销售为主要业务的销售集团,且负责煤炭的发运任务。为了降低 SH 集团煤炭运输成本,保证发运作业高效有序,需要制定有效合理的煤炭运输调度计划。调度室作业是销售集团业务的核心,该部门的主要职责是在严格执行销售任务的基础上,制定列车运输调度计划,加快列车周转,降低煤炭库存,确保销售任务按时高效地完成。调度计划的制定要综合多方面的信息,如现有空闲运输车辆及装载能力,各站台煤炭的种类及数量,销售部下发的销售指令等。由于该集团先后建立的信息系统较为独立,各系统间没有建立业务关联,因此无法实现数据交换与信息共享,运输调度计划的制定靠调度员手工统计的方

式来完成,这其中的信息有的来自企业内部,有些信息如可用运输车辆的装载能力及数量,则依靠调度员通过电话或传真等比较传统的信息传输方式从企业外部获得。SH 集团信息化水平的现状,给调度员的工作带来了极大的挑战,当运输调度工作量比较大时,调度员的任务繁重,导致其运输调度编制过程简单,调度秩序混乱,运输成本增加,不能按时完成既定的销售任务。

结合 SH 集团现有异构信息系统及需求现状,对其业务管理信息系统、生产调度指挥系统、日班生产运行系统和第三方运输单位的车辆管理系统实施了以上集成方案,如图 5 所示。

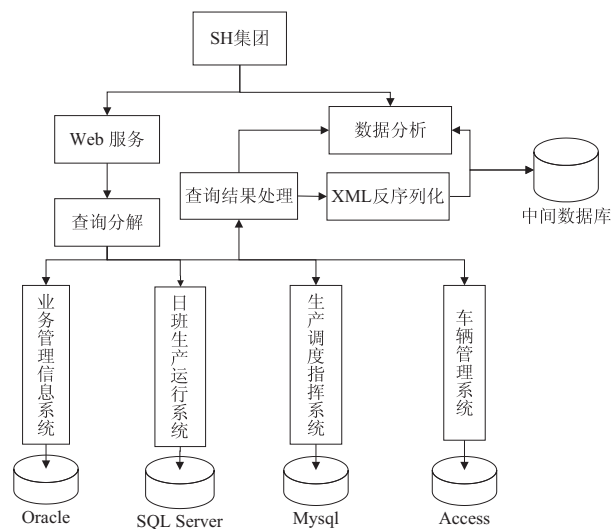


图 5 基于 Web Service 的数据集成模型在 SH 集团的应用

运用基于 Web Service 的集成方案,在原有信息系统中添加 Web 服务,调度员通过调度指挥系统直接调用该服务则可获取所需信息,制定调度计划。集成模型中的中间数据库可存放调度计划的历史信息,调度员可通过中间数据库进行数据分析,设定一系列指标,如计划完成率,根据计划完成的情况调整日调度计划,逐渐向月计划逼近。

通过该方案的实施,在保证原信息系统的独立性和异构数据源的自治性的前提下,提高了 SH 集团的信息化水平,减轻了调度人员的工作负担,为实现销售任务按时、高效的完成提供了良好的技术支持。另外,该方案还为进一步提高 SH 的管理水平奠定了基础。调度员在制定调度计划时还可把任务的轻重缓急、运输费用、库存成本等因素考虑在内,运用相关优化算法,使调度计划的制定更趋于合理、科学。

4 结束语

数据的集成是为了数据的使用,服务于企业的决策。文中通过对比分析传统数据集成方法的优缺点,

并结合 SH 集团的实际需求,提出一种针对不同区域的异构数据源的数据集成模型,用以解决数据的结构异构和数据格式统一问题。利用 XML 的自我描述性、平台无关性、可扩展性等特点,集成模型以 XML 作为消息传输媒介。为实现 XML 到关系数据库的转换,采用了 XML 转为实体类对象的存储方法。该方案已成功用于 SH 集团的数据集成,在不改变原有数据源的存在方式和存储模型的前提下,实现了分布异构数据的集成。


参考文献:

- [1] 赵毅,王浩然,庄冠华,等.一种基于 XML 的数据集成系统框架及其应用[J].计算机工程与应用,2005,41(26):181-183.
- [2] 钟将,宋娟.基于本体的异构数据集成框架[J].计算机工程,2011,37(14):44-46.
- [3] 杨长辉.基于 XML WebService 的异构数据集成系统的研究和应用[D].重庆:重庆大学,2006.
- [4] Zhu Fujun,Turner M,Kotsiopoulos I,et al. Dynamic data integration using Web services[C]//Proceedings of the IEEE international conference on Web Services. [s. l.]:IEEE,2004:262-269.
- [5] Fan Yating,Yang Jianyu,Zhu Dehai,et al. Research on the efficiency of querying historical data with the spatio-time data integration method[J]. Mathematical and Computer Modelling,2011,54(3):912-918.
- [6] Geng Rongna. Research and implementation of isomorous data integration technology based on the adapter[C]//Proceedings of 2010 international conference on computer, mechatronics, control and electronic engineering. Changchun: IEEE, 2010:354-357.
- [7] Torlone R. Two approaches to the integration of heterogeneous data warehouses[J]. Distributed and Parallel Databases, 2008,23(1):69-97.
- [8] 刺奋华,陈红星,李德玉,等. Web 环境下基于 SOAOffice 中间件的合同生成与数据抽取方案[J]. 计算机应用,2012,32(S1):73-76.
- [9] 郭文越,陈虹,刘万军. 基于 SOA 的数据共享与交换平台[J]. 计算机工程,2010,36(19):280-282.
- [10] 邓劲生,郑倩冰. 信息系统集成技术[M]. 北京:清华大学出版社,2012.
- [11] 黄序鑫. 基于 XML 和 Web Service 的数据集成研究与应用[J]. 现代计算机:专业版,2012(6):39-41.
- [12] 袁景凌,徐丽丽,苗连超. 基于 XML 的虚拟法异构数据集成方法研究[J]. 计算机应用研究,2009,26(1):172-174.
- [13] 杨小玲,甘文勇,杨隆浩,等. 基于 Web Services 的旅游信息集成技术[J]. 福州大学学报:自然科学版,2013,41(2):178-181.

基于Web Service的数据集成研究及应用

作者：[苏琪](#)，[刘西林](#)，[王军](#)，[SU Qi](#)，[LIU Xi-lin](#)，[WANG Jun](#)

作者单位：[苏琪, 刘西林, SU Qi, LIU Xi-lin\(西北工业大学 管理学院, 陕西 西安, 710129\)](#)，[王军, WANG Jun\(西安财经学院 信息学院, 陕西 西安, 710100\)](#)

刊名：[计算机技术与发展](#)

英文刊名：[Computer Technology and Development](#)

年，卷(期)：2014(8)

引用本文格式：[苏琪](#).[刘西林](#).[王军](#).[SU Qi](#).[LIU Xi-lin](#).[WANG Jun](#) [基于Web Service的数据集成研究及应用](#)[期刊论文]-[计算机技术与发展](#) 2014(8)