

物流信息系统中异构数据源的集成研究^{*})

聂瑞华^{1,2} 严粤锋¹

(华南师范大学计算机科学系 广州510631)¹ (华南师范大学网络教育学院 广州510631)²

摘要 本文通过分析当前物流信息系统缺乏有效信息协作的问题,明确异构数据源的集成是解决问题的关键。针对物流系统中异构数据源的特点并结合 XML.NET 技术,对如何实现异构数据源的集成进行了研究,提出了一个集成异构数据源的物流信息平台解决方案并加以实现。

关键词 物流,异构数据,XML,数据集成,MIS

Research on the Integrations of Heterogeneous Data Source in Logistic Information System

NIE Rui-Hua^{1,2} YAN Yue-Feng¹

(Department of Computer, South China Normal University, Guangzhou 510631)¹

(Institute of E-education, South China Normal University, Guangzhou 510631)²

Abstract The paper analyzes the cooperation problem in logistic information systems, points out that the integrations of heterogeneous data sources are the key way. Based on the XML.NET technology, presents a project of logistic information frame that can integrate the heterogeneous data source in logistic information system and realize it finally.

Keywords Logistic, Heterogeneous data, XML, Data integrations, MIS

1 引言

现代物流最大的特点是信息化,信息技术是现代物流发展进步的基础。目前大部分物流企业都建立了企业的信息系统来处理业务数据,并希望通过自身的物流信息系统与企业客户进行有效的信息交流和数据交换。但企业信息化建设过程中各自为政、各建系统缺乏统一的规划和标准。每个物流企业都只根据自身的业务需求、经济和技术等因素来决定所使用的计算机系统以及在哪些计算机系统上运行的操作系统、局域网拓扑结构、通信协议、数据库管理系统和应用程序等。这种在建立物流信息系统时的多样性,导致了物流行业中存在多种异构数据源,而如何实现异构数据源的集成是企业间实现信息交流与协作所面临的最大问题。

针对异构数据源所带来的问题,我们提出建立物流信息平台的方案,并结合 XML.NET 技术的特点着重讨论物流信息平台内部异构数据源集成的实现。

2 物流信息系统的异构问题

2.1 现有物流信息系统的状况

物流行业的信息化建设经过近年快速不断的发展,许多企业都建立了各种的信息系统,例如 ERP 在大型企业里的实施和进销存软件在中小型企业里的应用等。部分物流企业还实现了基于自身信息系统的网上发布,使得合作企业间可以共享部分物流信息。

但这些物流信息系统主要是以自我为中心的信息发布或信息收集,很少顾及合作企业的信息系统,缺乏以信息系统对接为前提的信息协作。这种状况造就了大批分散、异构和封闭

的物流信息系统。它们相互独立使得数据不能共享,信息无法有效地协作,成为制约物流行业信息化建设进一步发展的瓶颈。

2.2 异构数据源的产生

物流企业在信息化建设的初期还没有形成物流供应链的观念,没有意识到企业间信息协作的重要性,这使得当前的物流信息系统中存在大量的异构数据源,严重影响了信息的协作。

异构数据源的产生主要来自以下两个方面:

(1)体系的异构 物流企业在建立信息系统时只是考虑自身的条件,所使用的操作系统、通信协议和分布式程序都不相同,这些导致信息在传输模式上的异构称为体系异构。

物流企业所使用的操作系统大不相同,从简单的微机操作系统,如 Linux、Windows 98/Me,到复杂的网络操作系统,如 Windows 2000 Server 和 Unix 等。这些操作系统在通信协议和传输机制上都各不相同,构成了信息系统间体系的异构。幸运的是 TCP/IP 协议为这些异构的体系提供了信息交流的途径,解决了这方面的体系异构问题。

但目前物流企业的信息系统广泛采用客户机/服务器模型的分布式系统,由于分布式系统的耦合性强且没有一个统一的协作标准,不同机制的分布式进程无法相互调用与协作。这一问题已成为解决体系异构的最主要问题,也是异构数据源产生的主要原因。

(2)数据的异构 企业使用不同的数据库管理系统或其它各种文件存储格式,导致信息在存储模式上的不同即是数据异构。例如使用不同的关系数据库(如 SQL Server、Oracle 等)。

^{*})本文研究得到广州市科技项目(2002J1-C0281)资助。聂瑞华 副教授,研究方向为计算机网络及应用。严粤锋 硕士研究生,研究方向为计算机网络及应用。

很多物流企业都会根据自身企业的规模采用相应的数据库管理系统。中小型企业采用的多是 SQL Server 或 Access 等数据库系统,而大型企业则多采用 Oracle 或 Sybase 等数据库系统。从数据的存储模式上看,在本质上表示相同信息的数据,在各种不同的数据库系统中被定义为不同的存储格式;如 Oracle 8i 数据库系统与 SQL Server 2000 数据库系统在字段名称定义、字段类型定义和模式映射上就存在很大的不同导致异构数据的出现。

数据的异构中除有结构化数据源,如各种数据库管理系统外;还有半结构化数据源,如 HTML 文件和非结构化数据源,如普通的文本文件等。这些异构源在存储模式上也各不相同。

物流数据格式缺乏统一的标准,严重加大了数据交换的难度,成为当前物流数据顺利交换的重大障碍。这些异构的数据库和文件系统难以实现数据交换也是异构数据源产生的主要原因之一。所以解决物流企业间信息协作的问题,关键是集成现有物流信息系统中的异构数据源。

3 异构数据集成的解决方案

我们提出建立一个物流信息平台的方案,来解决物流信息系统中异构数据源集成的问题。物流企业可在保持原有信息系统和数据库系统的前提下,以实现企业系统间信息协作的目的。它的建立可以让信息资源在最大范围内共享和相互协作,为企业的信息系统提供基础信息,以避免低级重复。同时,通过物流信息中转、共享、协作,可以整合物流信息系统的信息资源,加强物流企业与上下游企业的合作,形成资源优化,降低成本。

3.1 物流信息平台的模型

建立物流信息交换平台把物流产业链上众多企业的信息交流有机联系起来。运用系统学原理对该平台进行整体设计、实施和管理,以最佳的结构、最好的配合,充分发挥系统平台的功效,实现物流信息交换的合理化。物流信息交换平台的模型如图1所示。

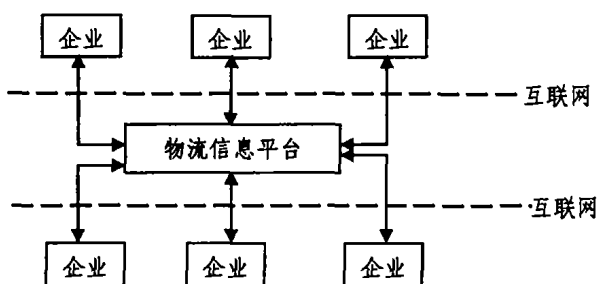


图1 物流信息平台的模型

3.2 异构数据源的集成

连通各个隔绝的物流信息系统,把分散和异构的信息资源整合起来,是异构数据源集成的最终目的。要实现异构数据源的集成,问题在于解决因为体系异构和数据异构所带来的信息传输和存储格式不同的问题。

在异构数据源的集成方案中,物流信息平台为众多的物流企业提供信息发布与信息提取功能,使企业无需改变自身使用的分布式系统和数据库管理系统,只需通过互联网接入到信息平台上,就可与供应链上其他企业进行信息协作与数据交换。

(1) 信息的发布 企业通过互联网接入到信息平台后,

可把信息系统中的数据和分布式程序都发布到信息平台上与其它物流企业共享。信息平台把企业要发布的数据转换为统一格式的 XML 文档,且当分布式程序被调用时对该进程进行 SOAP 封装发送给提取方。

(2) 信息的提取 当企业提取其它物流系统的数据或调用分布式程序时,信息平台根据提取方信息系统的需求,把 XML 文档转换为适当格式传送给提取方,并把提取分布式程序的进程进行 SOAP 封装与发送方进程进行对接,实现异构分布式进程的信息协作。

信息交换平台通过建立信息转换中心的结构来集成整个物流供应链上的企业的异构数据源,满足行业内部信息系统之间信息协作和交换数据的需要,从而为各信息系统之间实现“数据库无关”、“开发系统无关”、“网络环境无关”的信息协作提供平台基础。异构数据源的集成正是基于这一平台基础将物流行业中所有企业的信息资源整合到一个无缝、并列、易于访问的系统中,并实现信息格式的统一,这样可使多个分散且异构的物流信息系统运作起来就像一个单一的系统。

4 物流信息平台的实现技术

4.1 XML.NET 技术

建立一个好的物流信息交换平台既要考虑兼容性的问题,也要考虑平台开放性的问题,因此实现的技术是关键。只有平台同时具备开放性和兼容性,才能真正实现物流行业信息系统整体效能最大化的目的,实现物流企业间的信息协作,并有效降低信息共享成本,提高传输效率。

本文使用 XML.NET 技术来构建物流信息平台解决异构数据源的问题,它具有开放性和兼容性的特点能满足物流平台的需求。建立在 XML 和 Internet 协议上的.NET 框架正是微软为实现 XML.NET 技术与开发新型高级软件而提供的一个革命性平台。微软这一平台的设计意图十分明确,就是要将网上所有可用资源整合为一个解决方案。本文正是探讨在.NET 框架下,使用 XML.NET 技术,解决物流信息系统中存在的体系异构和数据异构的问题,实现信息协作。

4.2 体系异构的解决方式

目前分布式系统在物流行业的信息系统中被广泛地使用,而绝大多数的分布式结构是基于客户机/服务器模型的。这种基于客户机/服务器模型的分布式机制的应用程序之间如何实现进程协作成为解决体系异构的关键问题。

现有分布式计算 CORBA、DCOM、RPC 和 RMI 等多种机制,这些分布式机制的耦合性很强。它们需要进程的双方都采用相同机制才能实现进程的协作。而在现有的物流信息系统中,由于开发环境和工具的多样性,使多种分布式机制并存。要在同一物流信息平台中让多种分布式进程相互协作和通信需要引入 XML.NET 技术中的 SOAP(简单对象访问协议)来统一这些异构的分布式体系结构,模型如图2所示,它是以一个标准的 SOAP 封装方式桥接异构分布式的多个系统。SOAP 是基于 XML 的用在分散或分布环境中的交换信息的简单协议。它最大的特点是可以设计与各种协议封装结合使用。

在实现的物流信息平台中,以 SOAP 封装结合多种异构分布式体系,使各种异构的应用进程符合统一的 XML 格式,并把封装后的 SOAP 信息嵌入到 HTTP 信息中进行传输,从而使分布式应用程序之间实现进程协作与相互通信。通过 SOAP 的封装大大加强了物流平台的开发性,使不同的机制

的进程实现协作。SOAP 信息还可通过嵌入到 HTTP 信息

中,在异构的分布式体系中传输。

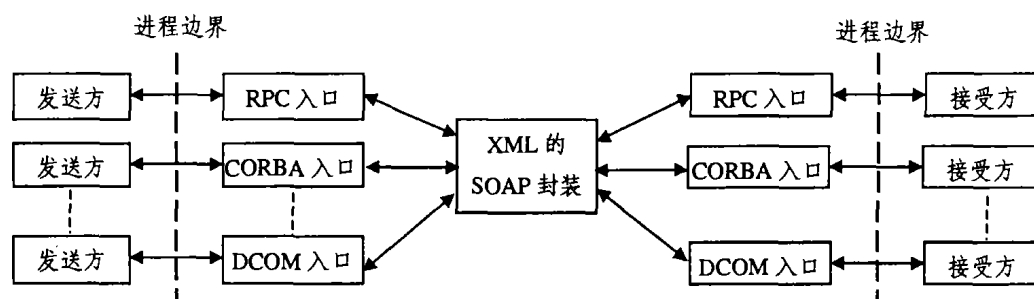


图2 统一异构分布式体系结构模型

4.3 数据异构的解决方式

物流企业的信息系统包括了各种规模不同、存储模式不同的数据库管理系统。要让这些异构数据库系统之间实现数据交换,解决异构数据的问题,物流信息平台把各种异构的数据库映射为统一格式的 XML 文档,并通过 XML 文档格式把异构数据库联系起来,以此实现异构数据库之间的数据交换。

XML.NET 技术中的 ADO.NET 作为一种中间件的形式可以实现数据库与 XML 文档之间的双向转换,具有很强的兼容性。

XML 支持把一个记录集合对象保存为一个 XML 文档,它也能够把一个 XML 文档作为一个记录集合进行处理。这样就在 XML 文档和数据库之间提供了一个桥梁。这种映射实际上是模型驱动,如果记录集合数据有变化的话,可以反映到相应的 XML 文档中,而 XML 文档中内容的变化也可以导致数据库内容的变化。ADO.NET 的体系结构如图3所示。

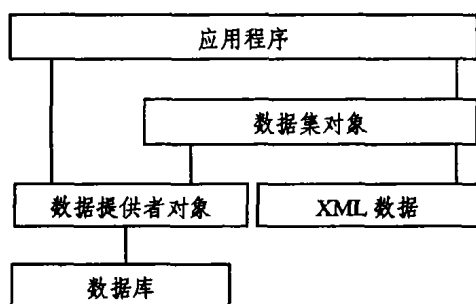


图3 ADO.NET 体系结构

ADO.NET 提供了操作 XML 文档和数据库的一组完整的类,所以通过 ADO.NET 我们不仅能访问关系型数据库中的数据,而且还能访问 XML 数据,从而实现两者间的转换。

ADO.NET 为我们提供了两种数据访问的模式:连接模式和非连接模式,后者是 ADO.NET 所特有的。相比于传统的数据库访问模式,非连接的模式为我们提供了更大的可升级性和灵活性。在该模式下,一旦应用程序从数据源中获得所需的数据,它就断开与原数据源的连接,并将获得的数据以 XML 的形式存放在主存中。在应用程序处理完数据后,它再取得与原数据源的连接并完成数据的更新工作。

ADO.NET 中的数据提供者对象包括了数据库连接接口(IDbConnection)、数据库命令接口(IDbCommand)、数据读取器接口(IDataReader)和数据适配器接口(IDbDataAdapter)等不同种类的接口。通过这些接口,应用程序就可以访问数据库、执行相关的命令操作并获取结果,获取的结果可以是以 XML 数据的形式存放在数据对象中,也可以是直接被应用程序所使用。ADO.Net 中的数据提供者对象为我们提供了以

上一些接口,通过这些接口每个应用程序都可以有不同的实现方法以访问特定类型的数据库。ADO.NET 还提供了 XmlReader 和 XmlWriter 对象实现读取 XML 文档和构造 XML 文档的能力。

ADO.NET 实现数据库与 XML 文档之间双向转换的方法如下:

(1) 从各种关系数据库转换为 XML 文档

A. 运用一个数据适配器对象(IDbDataAdapter)从一个数据库中读取模式和数据并产生数据集对象(DataSet)。

B. 根据读取的模式和数据产生数据集对象(DataSet)。

C. 根据数据集对象(DataSet)以 XmlWriter 对象生成 XML 文档。

(2) 从 XML 文档转换为各种关系数据库

A. 以 XmlReader 对象读取 XML 文档。

B. 根据 XML 文档产生数据集对象(DataSet)。

C. 根据数据集对象(DataSet)以数据适配器对象(IDbDataAdapter)生成关系数据库。

结束语 物流信息系统之间的信息协作在物流运作层上可以大幅度降低运作成本和所需时间;在物流计划层上也可以有效地降低库存成本。但现有的信息系统存在大量的异构数据源,阻碍了系统间信息的协作。我们提出一个异构数据源的集成方案,通过建立物流信息平台为物流企业解决多平台、多结构的信息协作与数据交换问题提供了一条解决的途径。通过异构数据源的集成,可以在 INTERNET 上实现物流企业间各种信息资源的交流与合作,搭建起整个物流行业的信息综合平台。

参考文献

- 1 Riordan R M. ADO.NET 程序设计. 北京:清华大学出版社,2002
- 2 张维明,肖卫东,等编. 语义信息模型及应用. 北京:电子工业出版社,2002
- 3 Sills A, Ahmed M. XML.NET 编程指南. 北京:电子工业出版社,2003
- 4 秦明森,编著. 物流作业优化方法. 北京:中国物资出版社,2003
- 5 靳强勇,李冠宇. 异构数据集成技术的发展和现状. 大连:大连海事大学,2002
- 6 王忠群,谢晓东. 基于 XML 的异构数据源的集成研究. 安徽:安徽工程科技学院,2003
- 7 桂思强. C#与数据库程序设计. 北京:中国铁道出版社,2003
- 8 蒋付军. XML 与数据库间数据传递的实现. 江苏:江苏省宜兴职业教育中心,2003
- 9 王正文. 物流信息平台建设. 江苏:南京王家湾物流中心,2002
- 10 黄晨. XML 数据库模式研究. 杭州:浙江大学图书馆,2003