

基于 ADO.NET 与 XML 的异构数据库数据交互

张倩, 王晓东

(昆明理工大学 信息工程与自动化学院, 云南 昆明 650051)

摘要: 面对信息化进程的发展, 为满足越来越多的应用中对异构数据源的访问需求, 探讨了 ADO.NET 对 XML 的集成与支持, 并结合 XML 的数据传输特点以及 ADO.NET 对数据源的访问处理优势, 描述了以 XML 为中间数据形式来实现异构数据库数据交互的方法。

关键词: XML; 异构数据库; ADO.NET; 数据交换

中图分类号: TP311.13

文献标识码: A

文章编号: 1673-629X(2007)08-0106-04

Data Exchange of Heterogeneous Databases Based on ADO.NET and XML

ZHANG Qian, WANG Xiao-dong

(Info. Eng. and Automation Coll., Kunming Univ. of Sci. and Tech., Kunming 650051, China)

Abstract: With the development of informational process, it is necessary for more and more applications to visit heterogeneous data sources. Analyses the support and integration of XML within ADO.NET. Then according to the characteristics of XML in terms of data transmission and the advantages of ADO.NET in respect of data access to different sources, describe a method to achieve the exchange of heterogeneous data sources, in which the XML is used to be the data medium, so as to meet the increasing demand.

Key words: XML; heterogeneous databases; ADO.NET; data exchange

0 引言

随着信息化和经济全球化的飞速发展, 企业及各部門间的合作日益加强, 从而使得异构数据库之间的数据交互显得越来越重要, 如何实现高效率的异构数据源中数据的交互已成为亟待解决的问题。

要对异构数据源的数据实现共享必须要达到两点: 一是实现数据转换, 二是实现数据的透明访问。为实现异构环境中的数据交流和共享, 采用了 XML 与 .NET 技术相结合的方法: 使用 ADO.NET 建立对数据库的访问; 采用 XML 作为数据传输载体; 通过套接字, 遵循 TCP/IP 协议实现数据在网络间的传输。

1 基于 XML 的数据处理特点

XML(可扩展标记语言)是一种提供数据描述格式的标记语言。该语言使跨越多个平台进行更准确的内容声明和获得更有意义的搜索结果变得更加容易。此外, XML 实现了表示与数据的分离。XML 使数据与表示及处理分离开来, 通过应用不同的样式表和应

用程序, 能够根据需要显示和处理数据。XML 一开始就建构在 Unicode(统一码)之上, 提供了对多语种的支持, 因此具有世界通用性。

从技术角度来说, XML 可以看作一种数据管理手段:

- ① 以文件系统为手段提供数据存储: XML 文件为纯文本文件, 不受操纵系统、软件平台的限制。
- ② 提供对数据的直接存取访问, 屏蔽了数据的物理结构。
- ③ 提供数据模式描述: 采用文档类型定义 DTD (Document Type Definition) 或 XMLSchema (XML 规范) 描述数据的逻辑结构。
- ④ 提供应用逻辑接口: XML 采用 SAX (Simple API for XML) 和 DOM (Document Object Model) 定义应用编程接口, 使应用程序能够访问和更新 XML 文档的样式、结构和内容。

此外, XML 在数据描述方面具有通用、开放、可扩展、自描述、跨平台和直接动态支持 Web 操作的优点, 具有适合作为数据库中间层、适于异构应用程序间的数据共享、可以进行数据检索和提供多语种支持等优点, 在异构数据转换方面具有很大优势, 但与传统数据库技术相比, XML 在数据存取的高效性、索引功能、并

收稿日期: 2006-11-03

作者简介: 张倩(1981-), 女, 重庆人, 硕士研究生, 研究方向为综合自动化; 王晓东, 教授, 研究方向为综合自动化。

发访问机制、安全机制等方面却显得不足。通过 XML 技术可降低数据库的使用条件,使其中丰富的数据资源得到利用,所以“以数据库为存储手段,以 XML 为交换载体”的数据管理模式已成为一种趋势^[1]。

2 ADO.NET 架构

Microsoft 所设计的 ADO.NET 支持断开式数据结构、能够与 XML 紧密集成、具有能够组合来自多个不同数据源的数据的通用数据表示形式;以及具有为与数据库交互而优化的功能。ADO.NET 包括两大组件:.NET Framework 数据提供程序和 DataSet,如图 1 所示。

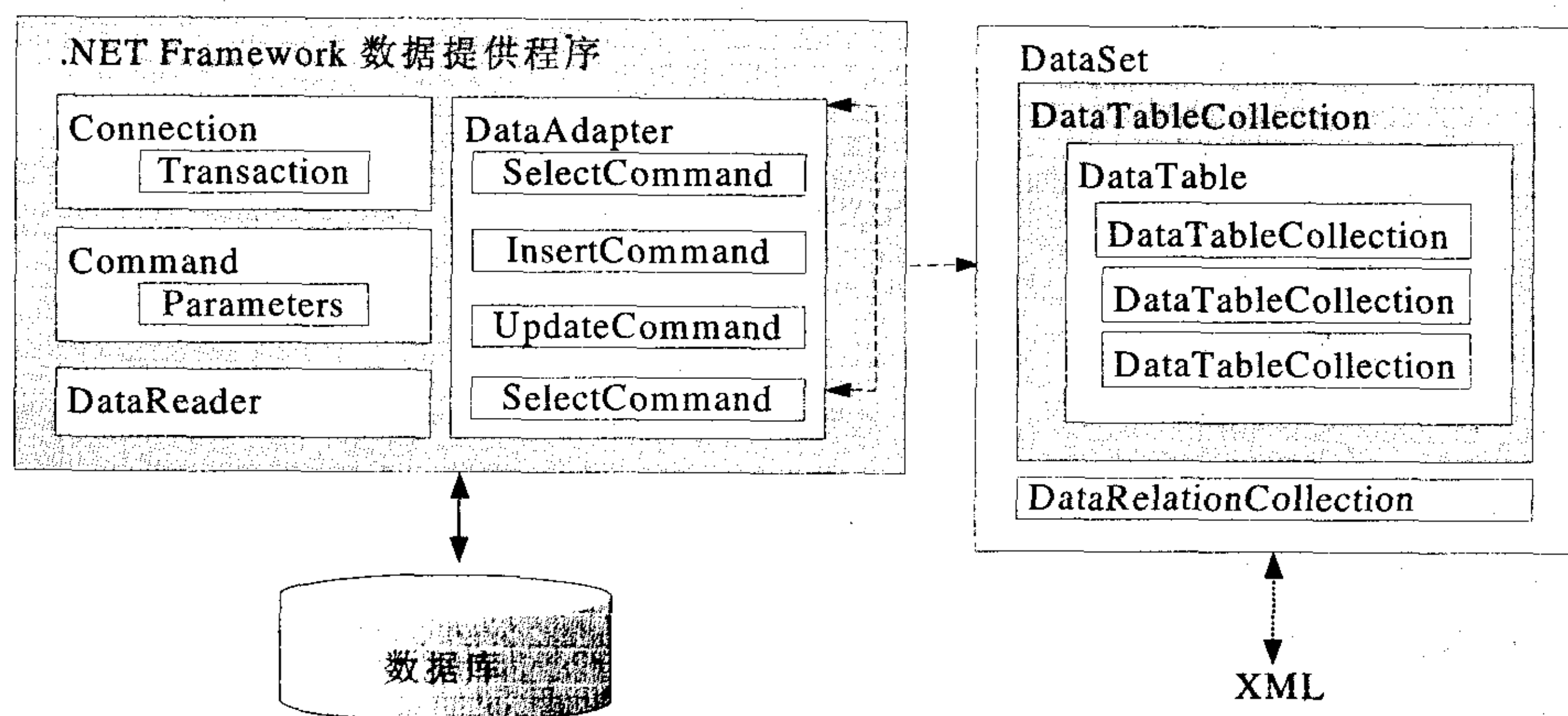


图 1 ADO.NET 结构图

2.1 .NET Framework 数据提供程序

.NET Framework 数据提供程序用于连接到数据库、执行命令和检索结果,并且使对该数据的更改与数据源保持一致,它是专门为数据处理以及快速地只进、只读访问数据而设计,是数据源和 ADO.NET DataSet 之间的桥梁。Connection, Command, DataReader 和 DataAdapter 对象是 .NET Framework 数据提供程序模型的核心元素。其中 Connection 对象提供数据源连接。Command 对象使您能够访问用于返回数据、修改数据、运行存储过程以及发送或检索参数信息的数据库命令。DataReader 从数据源中提供高性能的数据流。最后,DataAdapter 提供连接 DataSet 对象和数据源的桥梁:它使用 Command 对象在数据源中执行 SQL 命令,以便将数据加载到 DataSet 中,并使对 DataSet 中数据的更改与数据源保持一致^[2,3]。

2.2 DataSet

DataSet 是支持 ADO.NET 的断开式、分布式数据方案的核心对象,它是从数据源中检索到的数据在内存中的缓存,提供了独立于数据源的数据访问。无论数据源是什么,它都会提供一致的关系编程模型。它可以用于多种不同的数据源,用于 XML 数据,或用于

管理应用程序本地的数据。

.NET Framework 数据提供程序为 DataSet 提供与数据库的联系,即通过数据库联系、命令、参数来实现对数据源的查询,并将结果填充到 DataSet 中,从而在断开与数据库的联系后可通过 DataSet 实现对数据源的独立访问。同时还可以通过将 DataSet 中的数据通过数据联络、命令和参数映射到数据库中,实现对数据库的其他操作。

3 ADO.NET 与 XML 的集成与交互

在 ADO.NET 中,XML 是一种数据格式,提供了操作、组织、共享和传递数据的手段。而 ADO.NET 中的 XML 类集中于 DataSet 对象^[4]。一方面,可以使用 XML 流或文档向 DataSet 提供数据和/或架构信息。另一方面,对于来自不同数据源的数据,DataSet 都可以作为符合万维网联合会(W3C)标准的 XML 进行编写,并且将其架构包含为 XML 架构定义语言(XSD)架构,从而实现层次结构的 XML 数据与关系

型结构的数据模式间的转换。

型结构的数据模式间的转换。

* 从 XML 加载 DataSet:DataSet 类提供了 ReadXml 方法,可以从文件、流或 XmlReader 中读取内容,并将数据填充到 DataSet。

* ADO.NET 中可以编写 DataSet 的 XML 表示形式:DataSet 的 XML 表示形式可以写入文件、流、XmlWriter 或字符串,从而为 DataSet 的 XML 表示形式的选择提供了很大的灵活性。调用 GetXml 方法可以字符串形式返回 DataSet 的不包含架构信息的 XML 表示形式;GetXmlSchema 方法可将 DataSet 中架构信息写入字符串。调用 WriteXml 则可将 DataSet 写入文件、流或 XmlWriter。

配合 .NET Framework 提供的数据库联接,DataSet 为 XML 文档类型和后端 DBMS 管理的数据结构之间提供了一种映射机制,将不同数据库中的字段映射到各个元素和属性类型,或者将 XML 文档中的各个元素及属性映射到指定的数据库字段,实现数据的转换。其映射关系如图 2 所示。而作为 DataSet 固有的序列化格式,XML 和 XML 构架则为远程客户端之间来回传输 DataSet 内容提供了一种方便的格式,使 DataSet 成为层间数据移动的出色媒介,方便了异构系统间的数据传递^[5]。

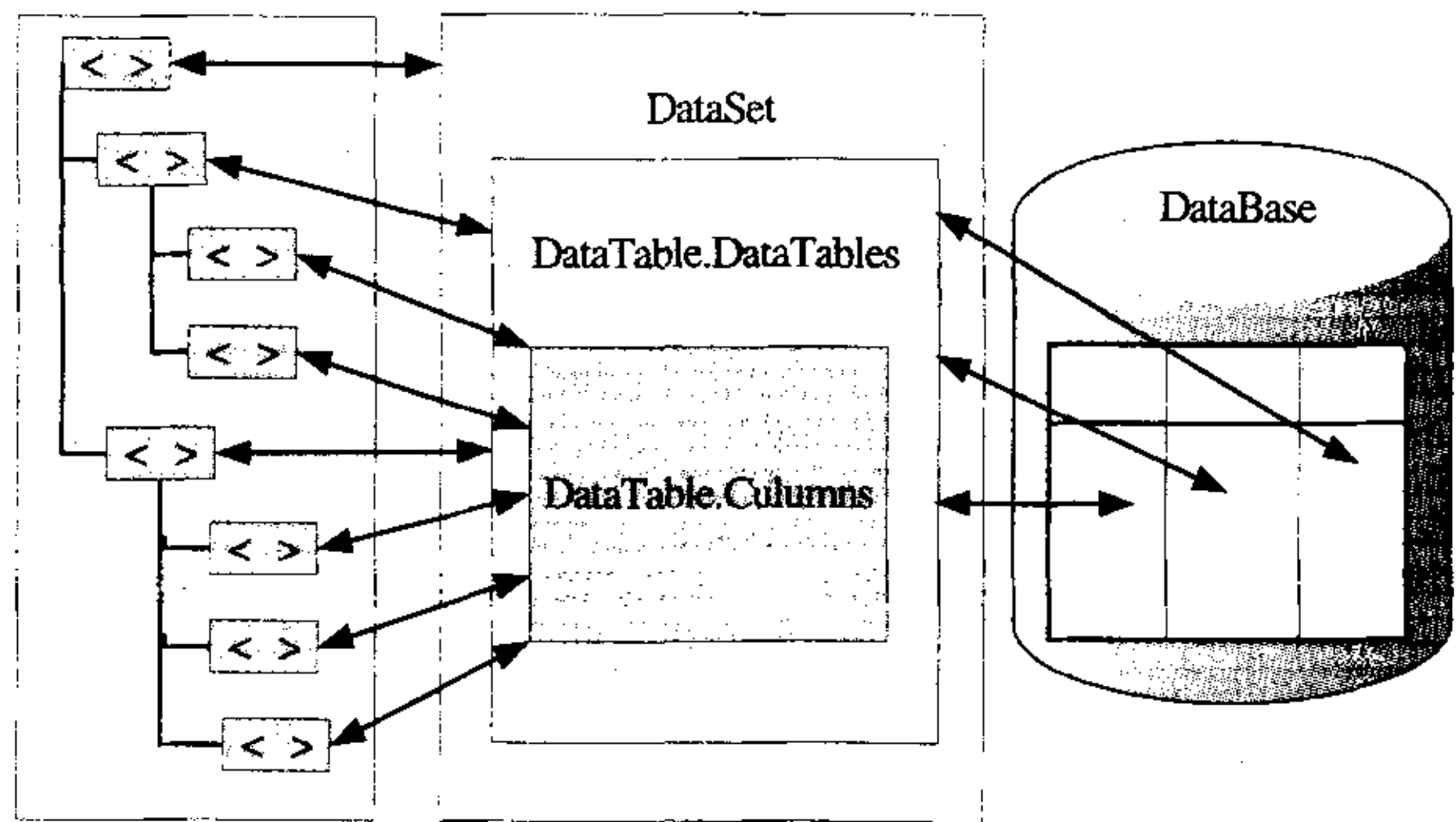


图 2 XML,DataSet 及 DataBase 间的映射关系

4 实现方法

文中提出一种多级站点的系统模式来实现异构数据库的交互(如图 3 所示)。其中,中心站点负责对各二级站点进行监控和数据转发;各二级站点间不直接通讯,而是通过中心站点实现数据交换和共享。为了实现数据的传输,采用 TCP/IP 作为底层数据传输协议,以 XML 作为数据形式,通过 Socket(套接字)方式来实现数据在网络间的移动。要在应用程序间移动 XML 数据,通信双方必须首先明确发送方需要使用什么样的信息对 XML 数据进行封装以便接收程序可以对数据进行正确处理,接收方需要哪些附加的上下文信息来完成消息处理^[6]。确定好对请求和相应的封装格式后,应用程序就可以按照约定格式对数据进行编码。在整个系统中,采用 XML 作为一种中间数据表示形式,各个站点均以其应用程序对应的数据库管理系统作为数据的永久存储机制。

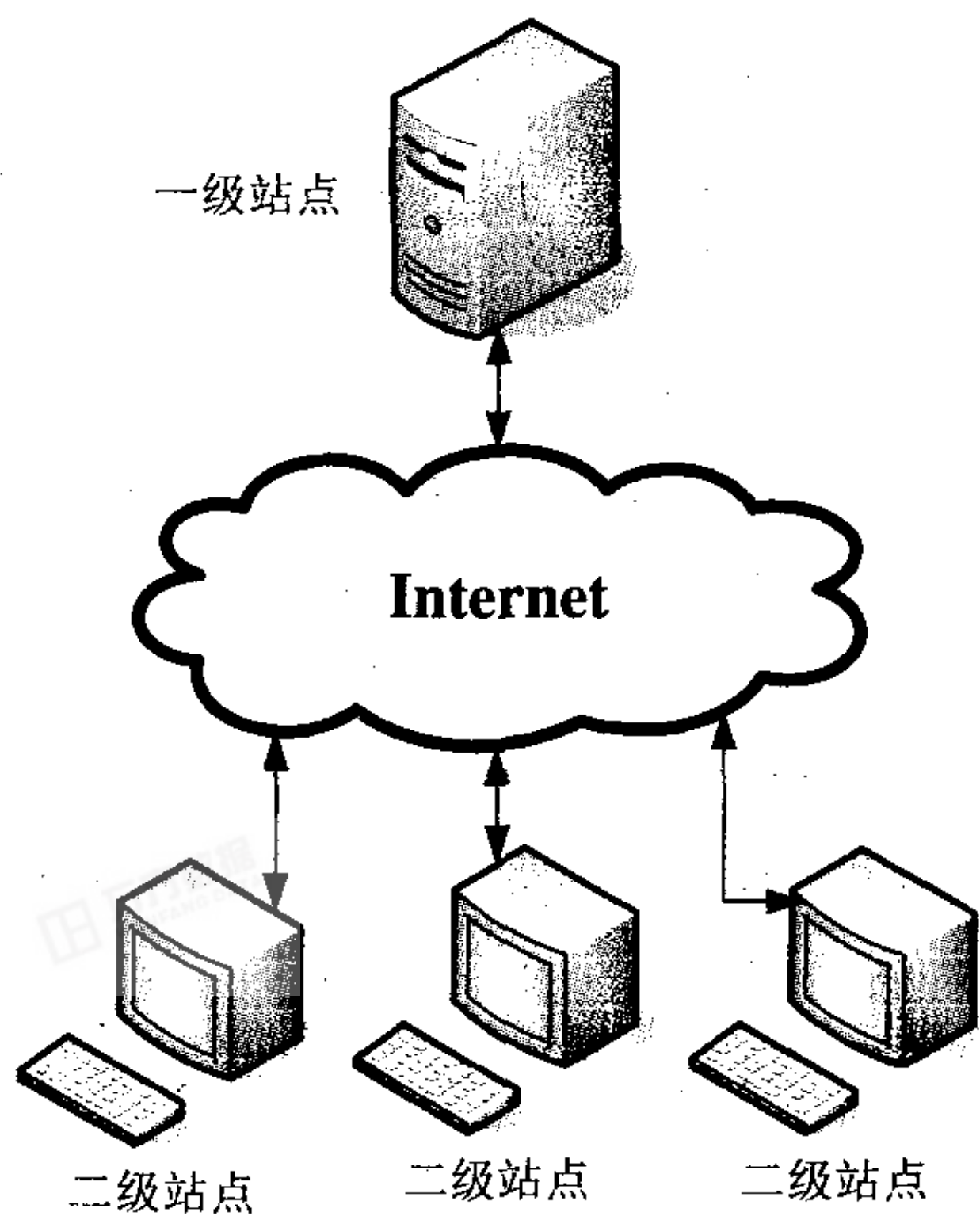


图 3 系统网络架构图

各站点功能模块如图 4 所示。

1) 接收模块:负责接收其他站点以 XML 文档发来的消息(包括数据访问请求、数据处理结果、安全认证信息等)。采用套接字对指定端口进行侦听,一旦接收到连接即开始接收消息,并通过起始标志符判断消

息的完整性。接收完毕后将消息传递给任务模块,并继续侦听。

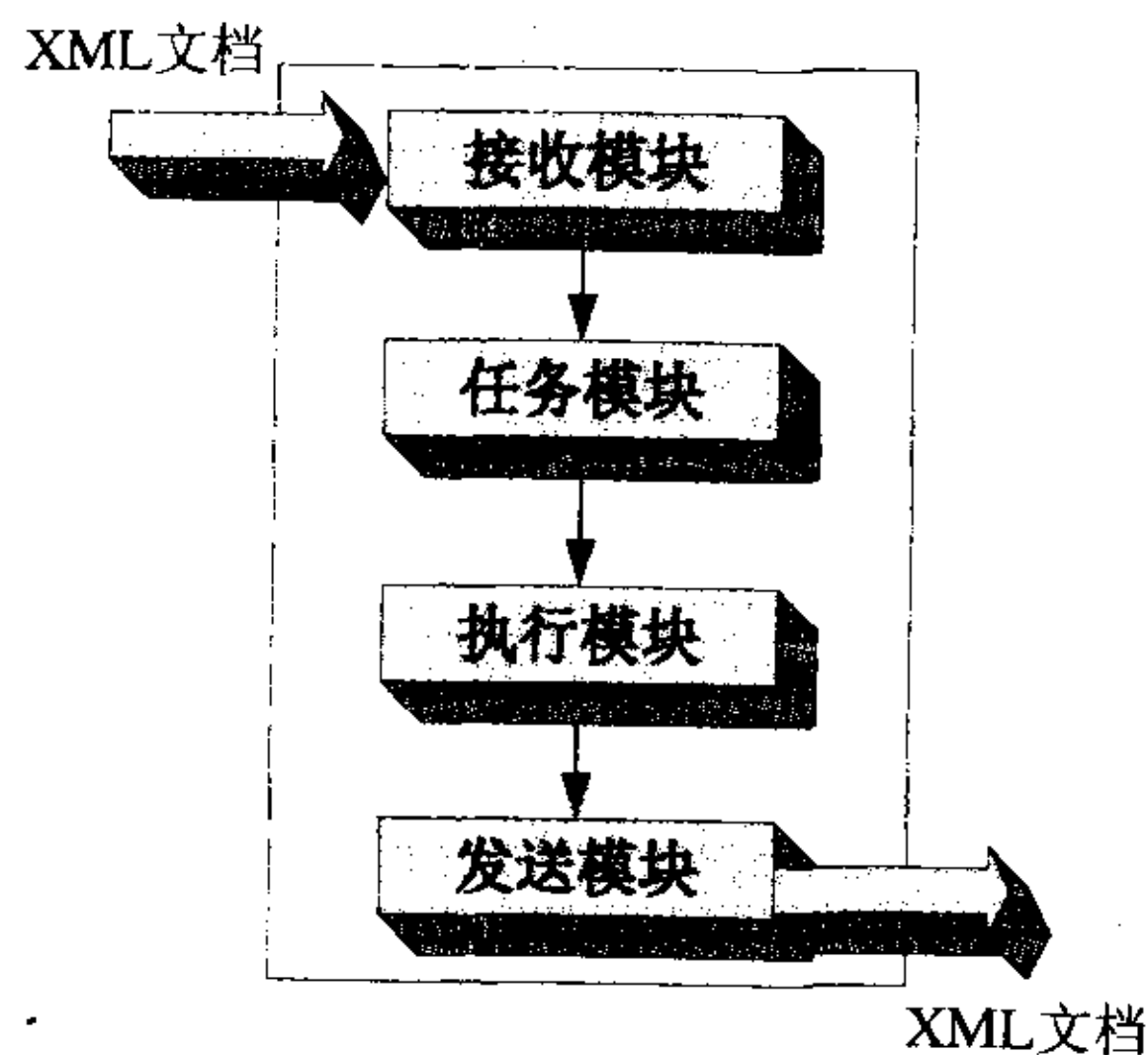


图 4 主要功能模块

2) 任务模块:负责执行任务判定和数据解析。

任务判定即根据通信双方确定的数传输规约对接收模块传来的消息文件进行解析,通过文件中指定的标识判断需要执行的任务内容(如数据请求、提交、转发等)。

数据解析是任务模块的核心。异构数据源交换的目的是要将外部数据源的数据与内部数据库管理系统中的其它数据相结合,要点是要实现获取的外部数据到内部数据库管理系统的映射。而 XML 在整个数据交换的过程中只是作为一个中间的数据表示形式,因此必须将交换文档中的 XML 数据转换为数据库管理系统支持的自然格式。数据解析即是根据 ADO.NET DataSet 与 XML 的关系(见图 2),将接收到的 XML 数据载入 DataSet,然后根据约定的数据传输格式对 DataSet 中的数据进行解析,并将数据映射到后台管理关系数据库中,从而将层次型结构的 XML 数据转换为关系型数据,实现数据的转换。

3) 执行模块:根据解析出来的数据,从本地数据库管理系统中查找出任务参数,执行任务模块指定的任务内容。

首先对数据中包含的安全信息进行认证。若安全信息认证失败则通过发送模块将错误提示信息返回请求方,通知对方任务请求失败。

通过安全信息认证后即可执行具体任务内容,如:如要求信息转发,则同样按照指定的数据发送协议将解析出的数据通过 DataSet 重新编写为 XML,通过数据发送模块转发到指定站点。

如要求对站点内部数据进行访问,则通过 ADO.NET 实现对数据库的操作,并返回执行结果,步骤如下^[7]:

(1) 连接数据库:在程序中建立到数据源的链接。在同一个程序中可以使用多个数据源,或者多个连接对应同一个数据源。

(2)数据库操作:将 DataSet 中的解析结果作为参数,构建 SQL 命令,通过 Command 对象对数据源执行命令。

(3)填充 DataSet:将查询结果生成数据集。

(4)关闭数据连接。

(5)在 DataSet 上进行所需要的操作:DataSet 是任意数目的表在内存中的表示形式,以及这些表 and 任何约束之间的关系,因此在断开与数据库的连接之后仍可独立地对数据结果进行访问。

(6)根据需要可将 DataSet 的变化更新到数据库,或自定义 SQL 语句,直接通过 Command 更新数据库。

若需要向请求方返回操作结果,则将结果信息按指定格式封装为 XML,由数据发送模块发送到相应站点。

4) 发送模块:它与接收模块一起实现与外部数据的联系,不同的是发送模块负责将信息封装成 XML 数据,通过网络套接字发送到指定站点。

5 结 语

该系统中各站点每个站点都可以拥有相互独立的 DBMS,故各站点既可以自主管理自己的内部数据,又可以通过网络传输实现对外部异构数据的有条件共享。通信双方通过 ADO.NET 对 XML 数据进行组织和解析,并根据解析情况执行相应的业务处理,可方便地实现 XML 数据与后台管理数据库的关系型结构数据间的相互转换,从而使系统具有较大的灵活性。系统中还可以 XML 格式传递辅助系统任务的配置信

息,以保证相应站点数据的同步更新,便于数据任务的统一管理。该系统在结构上便于向上扩展,实现更大范围的数据集成与共享。目前,该方法已在云南某钢厂得到成功应用。

笔者创新点:利用 ADO.NET 作为数据访问和数据转换的渠道,将传统数据库技术与 XML 相结合,有效地实现异构数据库的交互。同时提出将数据访问、传输与转换的方法应用于系统辅助信息,在企业应用中有助于数据任务的统一管理,更加灵活地实现企业间异构数据库数据的共享。

参考文献:

- [1] 黄乐丹. XML 技术与数据库技术的比较及应用[J]. 温州职业技术学院学报, 2005, 5(1): 46 - 48.
- [2] Wildermuth S. ADO.NET 实用指南——面向 Internet 世界的数据库访问技术[M]. 周靖译. 北京:清华大学出版社, 2003.
- [3] Beauchemin B. ADO.NET 本质论[M]. 周靖译. 北京:清华大学出版社, 2003.
- [4] Esposito D, Wintellect. ADO.NET 中的 XML 特性[EB/OL]. 2004 - 01 - 01. <http://www.microsoft.com/china/MSDN/library/data/dataAccess/XMLFeaturesinADO.NET.aspx?mfr=true>.
- [5] 朱敏,王开建,苏博. 基于 XML 的企业网络数据集成模型研究[J]. 微计算机信息, 2005(8): 21 - 23.
- [6] Dick K. XML: 管理者指南——构筑 XML 的企业级应用[M]. 邓尚贤译. 北京:清华大学出版社, 2003.
- [7] 尹立宏. Visual C#.NET 应用编程 150 例[M]. 北京:电子工业出版社, 2003.

(上接第 105 页)

4 结束语

由于互联网的动态性、不确定性和一些其它原因,导致 Web 服务的服务质量不断变化,这给基于 QoS 的服务发现提出了挑战。文中结合 BP 神经网络提出了一种基于服务质量预测的 Web 服务选择方法,并详细介绍了基于 QoS 预测的服务选择框架和服务质量预测模型。下一步工作将继续完善服务质量预测模型,优化 BP 网络的训练算法,研究更合理有效的训练样本集更新策略。

参考文献:

- [1] Vu L H, Hauswirth M, Aberer K. QoS - based Service Selection and Ranking with Trust and Reputation Management [C]//In: Proceedings of the International Conference on Cooperative Information Systems (CoopIS 2005). Agia Napa,

Cyprus: [s. n.], 2005: 466 - 483.

- [2] Ran S. A model for web services discovery with QoS[J]. ACM SIGecom Exchanges, 2003, 4(1): 1 - 10.
- [3] 杨胜文,史美林. 一种支持 QoS 约束的 Web 服务发现模型[J]. 计算机学报, 2005, 28(4): 589 - 594.
- [4] Day J, Deters R. Selecting the best web service[C]//In: Proceedings of the 14th Annual IBM Centers for Advanced Studies Conference (CASCON). Canada: [s. n.], 2004: 293 - 307.
- [5] 胡建强,邹鹏,王怀民,等. Web 服务描述语言 QWSDL 和服务匹配模型研究[J]. 计算机学报, 2005, 28(4): 505 - 513.
- [6] Liu Y, Ngu A H, Zeng L Z. QoS computation and policy in dynamic web service selection[C]//In: Proceedings of the 13th international World Wide Web conference. New York, USA: [s. n.], 2004: 66 - 73.
- [7] 袁曾任. 人工神经网络及其应用[M]. 北京:清华大学出版社, 1999: 66 - 68.