

# 多源异构数据整合在信用系统中的应用研究

李治强 苗 放

(成都理工大学 信息工程学院 四川 成都 610059)

**摘 要** 随着诚信建设进一步推广,信用系统建设越显重要。从分布异构数据库中提取有用数据以及数据整合,是信用系统的核心技术。文中介绍了分布异构数据库的特征和连接技术,以及数据交换技术。采用了 TUXEDO/Q 消息传递机制和 SAX 文档解析标准,使信用数据可靠传输和异构数据库得以无缝连接,实现了分布异构数据源系统的信用数据整合共享。得出了一种基于 XML 和 TUXEDO 的异构信用数据有效整合的方法。

**关键词** 分布异构数据库;信用信息;数据提取;数据整合

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2007)02-0172-03

## Application and Research of Credit System Based on Distributed Heterogeneous Database

LI Zhi-qiang, MIAO Fang

(College of Information Engineering, Chengdu University of Technology, Chengdu 610059, China)

**Abstract** With the increasing development of credit construction, the credit system construction becomes more important. Extracting and integrating available data form distributed heterogeneous database, which is the pivotal technology. It introduces the characteristic connection and transfer technology about the distributed heterogeneous database in this article. Employing TUXEDO/Q message transmission mechanism and SAX documents analysis standard, to transfer the credit data reliably and connect seamlessly, and realize the system credit data integration and sharing. Then put forward the integrated method of credit system based on distributed heterogeneous database.

**Key words** distributed heterogeneous database; credit information; data extraction; data integration

## 0 引言

推广和完善以政府信用为基石的政府企业个人的综合信用系统是与时俱进的需要,更是规范道德、约束行为的需要。通过分析来自各个领域的信用信息,可便捷全面了解和考量个人和企业的社会信用水平。传统的社会信用信息极端分散异构,以及不对称等,导致难以统一管理和充分利用,已不能适应新形势下的信息化管理。怎样从软件的角度提高企业或政府部门的工作效率,应对新形势下的要求,对软件开发人员来说,也是一种需要深入探讨的问题。结合现在社会信用信息量大、实时性高等特点,对采用 XML 技术和 TUXEDO 消息中间件技术实现异构数据源系统之间的无缝连接进行了探讨和研究。

## 1 多源异构数据库系统信息整合技术的介绍

随着信息化时代的到来,越来越多的信息需要在

Internet 上整合共享。信息来源于极端复杂的多源数据库系统<sup>[1]</sup>,给信息整合提出更高的要求。而 XML 能很好地描述和传输,使信息化从软件技术上得以实现。

### 1.1 分布式异构数据库特征

分布式异构数据库<sup>[2]</sup>(Distributed Heterogeneous Database, DHD)是指分布在不同地理位置的异构数据库资源,异构的是操作系统、数据库管理系统或数据库内部表结构等。针对某一特定类型和组织内部的 DHD 系统已经比较成熟和完善。但不同组织、类型和平台的 DHD 系统之间进行同步工作,成为新的研究课题。目前,异构数据库分布计算主要有以下两种实现方式:

(1)对分布在各个结点上的各种异构数据库都建立用户交互接口,不进行任何模式的集成。各个成员数据库对全局数据进行分区域分段管理,对本区域以外的数据操作则通过远程登陆进行。此方案简单易行,但用户无法透明地访问数据,并且增加一种新的异构数据库时,必须增加用户接口,实现方法比较烦琐。

(2)采用分布式数据库系统结构,将物理上分布而

逻辑上相关的数据库通过分布式数据库系统进行管理。系统具有独立性和透明性的特点。用户对任何数据库的操作如同在本地进行,不必关心其数据模型、物理位置等细节。

1.2 XML 实现异构数据整合

XML(eXtensible Markup Language, 扩展标记语言)是互联网下的一个关键技术,它能很好地实现来源极端异构的数据描述和传输。其信息格式允许是半结构化或非结构化的<sup>[3]</sup>。为了使基于 XML 的数据交换成为可能,必须实现数据库的 XML 数据存取,并且将 XML 数据同应用程序集成,进而使之同现有的规则和技术相结合。开发基于 XML 动态应用(如动态信息发布、动态数据交换等)的前提是所支持的数据库必须能支持 XML。XML 作为一种数据库模型并从中提取信息在数据库领域逐渐受到注意。而且,XML 技术和关系数据库技术的融合能使数据更通用,能被异构系统接受,因此成为很自然的趋势。大量的 XML 文档由于不同应用之间交换数据而临时产生,而且也存在像数据库一样需要系统可靠地管理永久性 XML 数据的需要。为此,已经有许多的研究工作致力于融合 XML 应用和关系数据库。同样,参照关系数据库设计与建模的成功经验,也需要一种统一的建模方法学来针对不同的应用为 XML 数据进行建模。

XML 文档本身只是一个文本文件<sup>[4]</sup>,它需要能识别 XML 格式化信息的解析器来解析并提取其中的内容,目前有许多应用变成接口 API 可以使用,如文档对象模型 DOM,XML 简单应用程序接口 SAX, JDOM 以及 JAXP 等。

SAX 是一个公共的基于事件的 XML 文档解析标准,它提供了一种对 XML 文档进行顺序访问模式,这是一种快速读写 XML 数据的方式。当激活相应的事件处理函数,从而完成对 XML 文档的访问。

2 信用系统开发与实现

信用系统透明地实现分布式、异构数据库访问,客户端以统一的接口访问分布在各个地域的异构数据库,实现数据有效整合。以 TUXEDO 消息中间件技术实现可靠的数据传输<sup>[5]</sup>,采用 XML 技术实现异构数据库之间的无缝连接。

2.1 系统网络结构

信用系统网络结构(见图 1)由 5 个部分组成,分别是分布式数据采集终端系统(被政府信用机构授权的前置机系统,包括个人信用征信系统和企业征信系统信用数据采集终端)、资源管理中心系统、互联网网站系统、政务网网站系统和消息传输交换平台。每个

子系统联合运行,实现从分布异构数据库中抽取个人或企业信用数据,传输到资源管理中心,并在网站发布的主要功能。本系统提供了强大的企业数据查询功能、丰富的门户功能、全面的统计分析功能,以及灵活的数据传输交换流程和高效的数据资源整合共享、高效灵活标准化的数据采集功能。

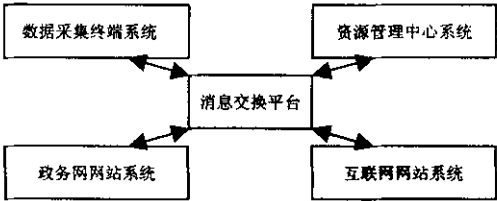


图 1 系统网络结构

2.2 系统业务流程概述

应用系统业务信用数据流(见图 2),从数据采集终端业务系统出发,直到信息发布库的过程。其中信用数据采集终端与其业务系统连接,通过信用数据采集终端完成数据采集、比对、审核,变成标准的 XML 数据格式,信用数据采集终端将 XML 数据,通过数据交换系统交换到信用数据资源管理中心系统,数据交换系统将采集的数据交换到信用数据资源管理中心系统后,经过整合、汇总生成企业的信用记录数据,被提炼、挖掘出来的企业信用数据,通过信息发布管理平台发布到信用门户中(包括:政务网和互联网网站),供各采集终端和公众访问。

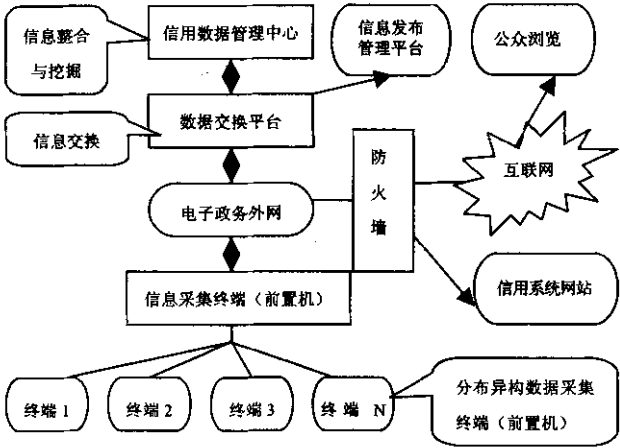


图 2 系统业务信用数据流

2.3 数据采集与传输

2.3.1 信用数据采集系统

信用数据采集系统是采用用户前置机方式,部署在各个数据采集终端(由政府授权的单位)内部,实现数据的自动提取与转换,同时支持手工录入与审核数据,通过数据交换系统安全同步地传输到信用数据资源管理中心系统。信用数据采集终端同时提供本部门信用数据维护、信用数据查询、内部交流等应用服务功

能。由于信用数据来源于各个渠道,而企业信用的可靠性也取决于这些数据,在开发中应注意开发可复用的功能模块,来有效地建立采集系统,不能是单一不变的系统。

数据采集系统采集的主要信息包括:企业基本信息、企业信用信息(企业是否欠银行贷款、欠税款、欠社保金、欠水电气费、欠其他企业货款等)、个人基本信息(包括姓名、住址、学历、社保号码、工作单位和婚姻状况等)、个人信用信息(贷款记录、偿还情况、透支、担保记录、纳税情况、诉讼和仲裁事项等)等。

除了这些企业信息外,数据采集系统还需要自己产生部分数据,作为采集过程中的操作控制。这些数据包括序号分配、操作痕迹,如查询记录、操作员编号、报告生成时间等,还有其它如:信息获取日期等。

### 2.3.2 数据抽取

数据抽取(见图 3)是系统定时自动运行,从临时库中抽取准备上传的数据形成 XML 文件准备发送到资源管理中心。数据抽取时会对成功和失败的数据进行分类日志记录。前置机通过配置文件,可设置数据抽取的频率。成功的数据会形成 XML 文件准备上传到资源管理中心。数据抽取失败的数据会进入失败数据库备份,便于后期查看。数据抽取和封装是对已通过审核、比对和校验的信用数据,从数据库中移出,形成标准格式 XML 文件,并准备上报。系统在后台运行数据抽取程序前,调用数据比对、校验程序且通过,然后将数据从临时表中移出,按照标准格式形成 XML 文件准备上报到资源管理中心。

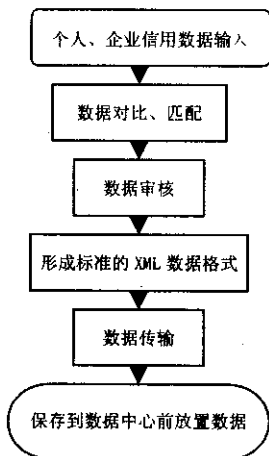


图 3 数据抽取

### 2.3.3 TUXEDO 异步交换 XML 数据

信用征信系统涉及数据采集系统众多,其数据源系统差别大,异构情况严重,需要由 XML 格式进行核心数据交换,为系统提供沟通异构系统的能力,满足各多源异构系统的各种数据信息在各个业务系统中交换

的需求。

数据交换的传输,采用消息中间件等可靠的网络传输技术;异构数据交换的关键是数据格式。采用 XML 通用数据表示格式,每个级别分级打包,彼此互不影响。XML 格式的数据是目前应用最为广泛的数据格式。在各种应用平台,各种编程语言中都有其解析器。各个部分的交换采用 XML 格式的数据大大地降低了系统各个部分的耦合程度,方便系统变更,最大可能地实现松散结合的系统。

数据传输是各终端采集信用数据整理完毕,通过数据交换系统传送到数据管理中心。消息传输交换系统是利用消息传递机制,以各成员终端为基本建设单位,依托电子政务外网实现各个信用采集点的互联互通完成数据交换。数据交换作为基础服务功能,采用 XML 格式进行核心数据交换,为系统提供沟通异构系统的能力,满足各采集终端各种数据信息在各个业务系统中交换的需求,并提供传输交换相关的日志查看、实时监控和控制。

本系统采用 BEA 公司的 TUXEDO 消息传输中间件来完成数据的异步传输,将要发送的消息或文件传送到 TUXEDO 的发送队列中。数据传输主要是提供将前置机上的文件或数据传输到资源管理中心。

### 2.4 数据整合

数据整合见图 4:在资源管理中心,XML 文件将会由 TUXEDO 消息接收队列自动发送到指定目的地,采用 XML 简单应用程序接口 SAX 快速读写 XML 数据方式。在使用 SAX 解析器解析 XML 文档时,触发

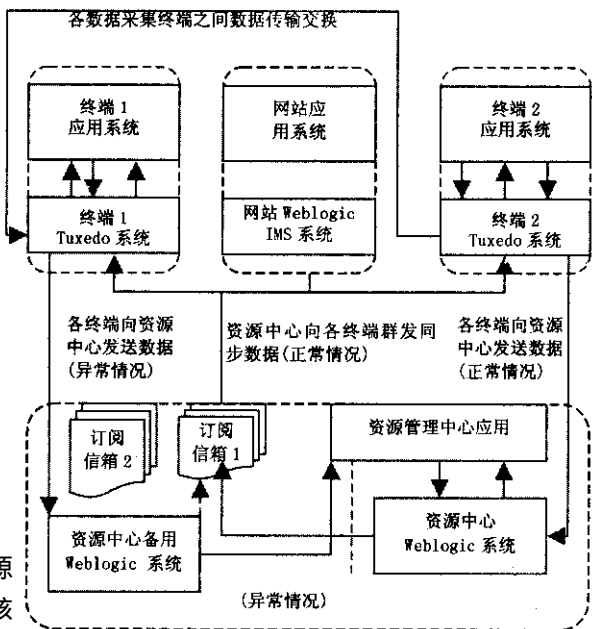


图 4 数据整合

```
cnn.Open "学生信息"; sa";"  
webserver="202.194.*.*"  
%>
```

### 4 基于角色访问的控制策略

基于角色访问控制的基本思想是将整个访问控制过程分成 3 个部分、2 个关联。3 个部分为用户、角色、权限 2 个关联为访问权限与角色相关联、角色与用户相关联 ,从而实现了用户与访问权限的逻辑分离 ,极大地方便了权限管理。

本系统中采用了根据不同用户的权限将有不同的界面 ,例如 :管理员有最高权限 ,而对一般用户而言 ,将不具有如删除不属于自己的信息、删除别人的帐号等等权限 ,通过这种角色访问控制策略也使系统的维护性大大提高。这种控制策略的优点在于<sup>[1]</sup> :

- ①使系统的整体逻辑清楚 ,各业务模块之间分工明确。
- ②使各操作员权责明确 ,防止业务的越权操作。
- ③简化了权限分配工作 ,只需根据每个操作员的职责 ,向其所代表的角色进行简单赋权操作即可。如果某个操作员的职责改变了 ,那么只需在系统中把该操作员所扮演的角色赋值给接替他的人员 ,然后把适合该操作员的角色赋值给他。
- ④在角色的基础上增加操作日志 ,增强了系统的安全功能。

### 5 实验结果与分析

实验里基本上达到了要求 ,即可实现实验客户端和服务端之间的通信 ,但是这里并没有考虑到数据库的安全问题 ,并且 ASP 本身就存在多种安全隐患 ,如 :

.....

(上接第 174 页)

并激活了相应的事件处理函数 ,完成了对各数据采集终端上传的 XML 文件的解析 ;资源管理中心系统对各终端上传的数据进行审校、汇总和整理后形成信用各种信息 ,信用信息存入数据资源管理中心 ,构成信用数据库 ;资源管理中心系统还提供统计、查询、数据管理维护、中心监控、中心对成员单位的配置管理、日志记录与查询等功能 ,完成了数据的整合共享。

### 3 总 结

采用 TUXEDO 异步传输机制和 XML 数据读取等技术 ,从分布异构数据库源系统中提取信用数据 ,转化为标准的 XML 数据传输 ,使“孤岛”信息得以整合

ASP 程序代码的安全隐患 ,ASP 的 FileSystem Object 组件对服务器数据的安全威胁等等。

- 可以采取以下几种方式加以解决 :
- (1) 对数据库文件进行加密 ,这样可对数据库里的数据起到保护的作用 ,但同时也增加了系统开销。
  - (2) 对输入内容进行有效性验证。
  - (3) 关闭没用的服务和协议。
  - (4) 维护 Global.asa 文件的安全。

### 6 小 结

Delphi 是功能强大的应用程序开发工具 ,它提供了对现在流行的 Web 技术及 Internet Express 的支持。本系统采用 B/S 模式结构 ,便于实现系统的维护 ,并且把页面中需要用到的相同的属性都集中到一个 ASP 页面里 ,方便修改 ,这样系统移植时只需修改最少的地方即可。系统中采用了基于角色的访问控制策略 ,便于业务的分割和权限的分配 ,方便了权限管理 ,也使各模块业务和权责分工明确。

参考文献 :

[1] 黄 凯 ,陈 云 ,阎如忠 ,等 .基于角色的 B/S 系统访问控制的研究与应用 [J]. 计算机工程与应用 2003 ,39(20) :227 - 229.  
[2] 侯太平 ,董爱红 .Delphi 数据库编程 [M]. 北京 :清华大学出版社 2004.  
[3] 段卫华 ,陈春玲 .ASP 技术在 Web 数据库中的应用 [J]. 电脑知识与技术 2005(2) :2 - 4.  
[4] 科南资讯 .Active Server Pages 程序设计与入门 [M]. 北京 :中国铁道出版社 2001  
[5] 网冠科技 .ASP310 时尚编程百例 [M]. 北京 :机械工业出版社 2002.

共享。该系统的建立与实施具有很大的实用价值和社会价值 ,为社会信息化进一步实施提供了技术保证。

参考文献 :

[1] 齐德昱 ,胡镜林 ,张 鹏 .多数据库中间件的模型研究 [J]. 计算机工程与设计 2005(10) :2602 - 2604.  
[2] 邵佩英 .分布式数据库系统及其应用 [M]. 北京 :科学出版社 2000 :9 - 14.  
[3] 何 慧 ,陈 博 .基于 XML 和 JMS 的异构数据交换集成的研究 [J]. 计算机技术与发展 2006 ,16(2) :70 - 72.  
[4] 田 斌 ,孙 霞 .基于 SAX 的 XML 文档解析和存储技术应用 [J]. 微机发展 2005 ,15(3) :90 - 92.  
[5] 郭 磊 ,吴春旭 .BEA TUXEDO 在移动计费系统中的应用 [J]. 计算机系统应用 2006(4) :91 - 93.