第32卷 第20期 计 算 机 工 程 2006年10月
Vol.32 № 20 Computer Engineering October 2006
·开发研究与设计技术· 文章编号：1000—3428(2006)20—0270—03 文献标识码：A 中图分类号：TP311

数字校园中数据同步方法的研究与实现

丁 铖[1]，张新华[2]，梁正和[2]
(1. 南京通达海软件有限公司，南京 210008；2. 河海大学信息中心，南京 210098)

摘 要：分析了目前常用的数据同步技术，提出了根据高校数字校园的特点而设计的一种新颖的数据同步方法，并在实践中证明了该方法的有效性和系统的可扩展性。还提出了该数据同步系统在实时性要求下的改进措施。
关键词：数字校园；数据同步；综合信息查询系统

Study and Implementation of Data Synchronization for Digital Campus

DING Cheng[1], ZHANG Xinhua[2], LIANG Zhenghe[2]
(1. Nanjing Tongdahai Software Ltd.,Co., Nanjing 210008; 2. Information Center, Hohai University, Nanjing 210098)

【Abstract】This article analyzes the often-used data synchronization technologies and according to specific features of digital campus, puts forward a new method of data synchronization. The method's affectivity and the system's extensibility are proved in practice. Then, the improvement measure of the system under the real time requirement is put forward.
【Key words】Digital campus; Data synchronization; Integrated information query system

随着当前教育体制改革的不断深化，招生规模的逐年扩大，各高校陆续建设了不少业务系统，如教务系统、科研系统、学工系统、人事系统、数字图书馆、校园一卡通等。这些信息系统的建设对提高高校的教学、科研和管理水平起了很大的推动和促进作用。

但随着这些“条块分割”的系统的投入使用，信息共享问题便随之而来，相同的信息(如学生的基础信息)在不同系统中需要重复录入、更新，不同系统之间信息难以充分共享，从而造成了大量的“信息孤岛”，甚至由于信息不一致导致无法确定哪里数据是正确的，而成为“信息垃圾”。因此，尽快建设“数字校园”成为摆在各高校领导面前的一个迫切需求。

数字校园建设的一大难点就是如何处理和保存数万学生和数千教职工相关的业务数据。目前一种公认做法就是实现所有数据的大集中管理，即只建立一个“中心数据库”，虽然这是一种彻底的做法，但是面对大量分离的在用业务系统，这种方法实施难度和风险都很大，目前国内外鲜有高校采用。

另一种做法就是建立“共享数据库”，实现相关基础或重要业务数据和代码的共享，此法对老业务系统的改动要求可以降到很低，从而可以快速实施和产生效益。

由于各业务系统使用的数据库通常也是各种各样的，如Oracle、Sybase、MS SQL Server等，并且版本也有差别。共享数据库中的一个关键技术就是这些异构数据库中相关数据的同步。如何实现数据的实时同步，成为此类项目建设能否成功的一个关键因素。

本文介绍目前常用的一些数据同步技术，并深入探讨作者根据河海大学数字校园项目一期的实际情况而设计的一种新颖实用的数据同步方法。

1 常用的数据同步方法

目前常用的数据同步技术主要有数据库系统本身自带的一些数据同步和复制产品、第三方的一些产品及技术。

1.1 数据复制

各大数据库软件厂商（如Oracle、Sybase）都提供了自己的数据复制产品或技术。这些方法可以满足分布式同类型数据库之间的数据同步。而对于异构数据库之间的数据同步，此法则缺乏有效的手段。

另外，它们还提供了一些技术用于数据同步，如基于Oracle本身技术的：

(1)基于Oracle行级安全策略的数据同步更新机制；

(2)基于Oracle只读实体化视图的数据更新机制；

(3)基于Oracle可更新实体化视图的数据同步更新机制。

再如Sybase为移动计算提供的基于消息的同步技术(SQL Remote)和利用基于会话的同步技术(Mobilink)。这些方法技术成熟、可靠，但是设置复杂，费用较高。

1.2 邮件

对于一些分布在不同地域的系统或在两系统之间网络不能互通时，可以通过在数据库服务器端使用数据库触发器获得更新数据后，通过Internet网络邮件服务器发送和接收包含更新数据的邮件方式来实现数据的远程同步。

这种方法可以支持异种数据库的数据更新；费用低，无需额外的网络投入；数据同步全自动化，操作简单；功能较强，所有的任务均通过描述语言书写，适合各种应用情况。但它不能支持实时性要求较高的应用；许多安全和数据同步机制都需应用程序完成，例如“两阶段提交协议”，增加了程序的复杂度和实现难度。

作者简介：丁 铖(1970－)，男，高工，主研方向：教育、房产等行业数字化解决方案；张新华，硕士、工程师；梁正和，博士、副教授
收稿日期：2005-10-24 **E-mail：**dingcheng21@yahoo.com.cn

1.3 手机短信

对于一些小数据量的传送和同步，特别是野外作业采集数据情况，可以采用手机模块来发送和接收包含数据的手机短信。此法比邮件方式实时性强，但需要手机短信通信模块和开发程序，数据处理量小。

1.4 TCP/IP Socket 通信

对于一些需要实时同步的中小型应用，可以以 TCP/IP Socket 通信方式为基础，进行类似前面邮件方式的数据同步。该方式比邮件方式实时性好，但传输受网络稳定性的影响较大，编程较复杂。

1.5 Web Service

Web Service 技术是目前较常用的解决异构数据交换问题的方法之一，它实际是一套标准，定义了应用软件如何在 Web 环境中实现互操作的一种技术规范，可以在不同的开发平台上开发符合 Web Service 标准的异构数据交换的软件。该方法编程难度和工作量大。

1.6 JMS (Java Message Service)

JMS 是 J2EE 环境下的一种分布式、异步消息处理机制，是一种低耦合通信方式。消息对象(MessageObject)的发送和接受不需要考虑对方的状态。JMS 的提供者在消息到达后立刻将它传送给客户端，而不用等待客户端的请求然后再发送；JMS 提供者可以保证所有消息被传送一次，并只传送一次。JMS 提供的这种消息机制允许系统之间或者同一系统中不同的组件之间进行可靠的异步通信。在 J2EE 构建的企业系统选择 JMS 来完成数据同步是一个自然的选择。目前 SUN、IBM 等公司均提供基于 JMS 的企业级数据交换同步的产品，但是价格昂贵。

2 数据同步的新思路

需要根据项目的要求，结合对上面各种数据同步方法研究，找到一个对现有业务系统修改少、对网络要求不高、实施费用相对低廉的方法。

一个大型的高校校园网络通常比较复杂，由若干个局域网组成，它们之间通常配有防火墙等设备进行分割，为了加强网络安全性，一般也只开放 HTTP（80）、FTP（21）等少数端口。

按教育部和校方的数据标准要求，采用成熟的 Oracle 9i 数据库系统，为项目新建了一个新的数据库（类似共享数据库），称之为“统一数据库”。

但统一数据库和各现有业务数据库中的数据项（如“专业技术职务”的代码）难以正好一一对应，所以同步时需要逐个数据项作转换。各主流数据库都是支持触发器的，并且建立触发器对原有的业务系统基本上没有影响。加上本项目的数据同步实时要求暂不是很高，于是我们研究出了一个新颖方法：数据库触发器+FTP 传输+手工编程转换数据。该方法具有以下一些主要特点：

(1)不需要修改现有业务系统程序，实施方便、快捷；

(2)不需要更改学校各级网络设备配置；

(3)不需要额外购买额外软/硬件设备；

(4)可以适应同城各校区、跨地域各校区的需要；

(5)数据同步过程全部自动化处理；

(6)实现一对多的数据同步，系统扩展性强；在数据量大的情况下，可以灵活增加配置，实现负载均衡；

(7)系统处理效率较高，可靠性较好。

3 数据同步的设计和实现

项目初期需要同步的业务系统涉及教务处、人事处、研究生院、学生处和科技处等主要综合性部门，涉及的数据库系统有 Oracle8i for Win2000、Oracle8i for Solaris 9、MS SQL Server 2000。数据同步的逻辑架构框如图 1 所示。其中，FTP 服务器为学校一级的 FTP 服务器。

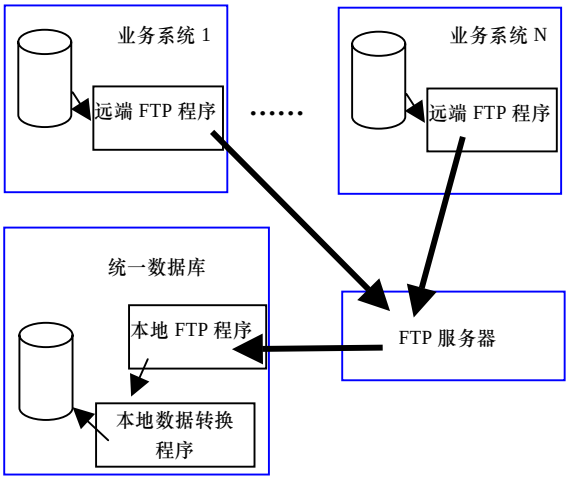


图 1 数据同步的逻辑架构框图

为了考虑通用性，各应用程序均基于 Windows 平台开发，提供工作日志的监控界面和日志文件。数据同步可以采用自动定时操作的方式，也可以用户手工即时处理的方式。

3.1 远端 FTP 程序

在业务系统的数据库上建立所需的 3 类触发器（INSERT、UPDATE、DELETE）。相关表的数据有变化时，触发器把变化了的数据放到一临时表。“远端 FTP 程序”定时查询此表，将新的记录取出后，按自定义的数据格式，整理成一数据文件，最后上载发送到 FTP 服务器中。

该临时表的定义如下：

· XH：序号，number 类型，用于唯一标识该记录的流水号，每次加 1。

· XXNR：信息内容，varchar 类型，更新信息的内容。

· SCSJ：生成时间，char 类型，该更新信息生成的时间。

· DQBZ：读取标志，char 类型，该信息是否已被数据同步程序读取(“0”—未读(初始)；“1”—已读；“2”—正被读取中，等待处理，此状态介于 0 和 1 之间，等数据文件上载到 FTP 后就改为 1)。

· DQSJ：读取时间，char 类型，该更新信息被读取的时间。

其中，XXNR 字段中的数据格式：

<操作>|<数据的类型>|<字段 1>=<值>|<字段 2>=<值>|...|<字段 n>=<值>|

其中：

<操作>：包括 INSERT、DELETE、UPDATE。

<数据的类型>：与“本地数据转换程序”中转换函数的功能对应，如 w_1。

生成的数据文件格式如下，该格式的文件结构清晰，便于程序读取和分析。

[DataNum]

DataNum=9

[Data]

Data1=UPDATE|w_1|XH=00081200|XM= 李 宏 |YWXM=LI Hong|ZYH=0801|ZYFX=1|PYFS=1|XSLB=1|XZ=4|RXRQ=20040901|BM=国贸 04_1|SFZH=510402761205221|XB=男|JG=四川省|MZ=汉族|ZZMM=共青团员|

Data2= …

程序置于各业务系统侧，操作员通过界面操作就可以支持 Oracle、Sybase、MS SQL Server 等主流数据库，及每周一次、每天一次和每小时一次等多种定时方式，基本上可以满足项目目前各种同步的需求。

3.2 本地 FTP 程序

“本地 FTP 程序”定时访问 FTP 服务器，下载新的数据文件到本机中，并把 FTP 服务器中的该文件删除。程序同样提供每周一次、每天一次和每小时一次等多种定时方式，并及时提取“远端 FTP 程序”上载的数据文件。

3.3 本地数据转换程序

“本地数据转换程序”定时逐个读取“本地 FTP 程序”新下载的数据文件，把其中新的数据信息转换并反馈到统一数据库的对应表中。

因为新老数据库数据项基本无法一一对应，这里需要详细分析现有业务系统的数据库，找到它们的对应关系（一个新数据项的数据可能来自一个老数据项，也可能来自几个老数据项），所以分别为它们分别编写转换程序。这是本同步方法的难点之一。

因为现有业务系统的建成时间不一，版本也在不断更新中，系统有学校自己找学生或教师开发的、有委托开发的，没有严格按数据标准和软件工程开发，造成系统资料不够全面或不是最新版本。所以能否正确找出新老数据库数据项的对应关系，成为项目最终能否实施成功的一个重要因素。

另外，因为项目的建设会发现现有业务系统中不规范或需要完善的地方，如组织机构名称和代码的统一，需要各业务系统做必要的完善。这个过程也比较难以协调和处理。

3.4 数据集中提取和转换

本同步还提供了“数据集中提取和转换”功能就是集中一次把各业务系统中相关业务数据全部提取出来，并传送、转换到统一数据库中。该功能可以在项目初期提取统一数据库的初始数据；也可以在日常操作时，为了确保各业务系统和统一数据库两侧的数据严格一致，强行将两侧数据重新集中同步一次。功能包括：

(1)远端集中提取数据程序：将各业务系统中所需的相关数据库表中的全部数据集中一次提取出来，组成数据文件并上载到 FTP 服务器。

(2)本地 FTP 程序（同上面的）：下载数据文件。

(3)本地集中转换程序:分别将各数据文件中的数据转换到统一数据库中。

这里的数据文件中的数据格式需和前面的有些不同，以便“本地 FTP 程序”区分处理。

3.5 系统的扩展

如图 2 所示，在项目业务量很大的情况下，可以通过增加处理机器，来达到负载均衡。更复杂的情况，则以此类推。数据同步系统扩展如图 2 所示。

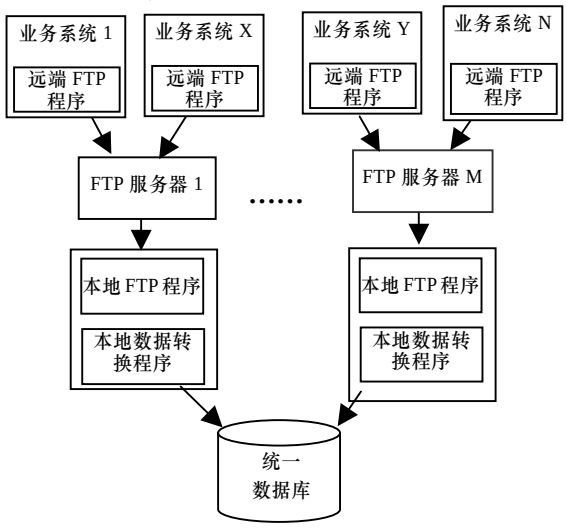


图 2 数据同步系统扩展图

3.6 进一步的改进

该方法的数据同步是异步进行的。在实时性要求很高时，可以在该方法上增加消息通知的手段，以让本地的程序及时处理。

在现有业务系统很多的情况下，用于数据同步的机器将很多，此时可以增加监控这些机器和数据同步程序运行状态的辅助管理系统，以保证同步能 24 小时不间断地正常进行。

另外，对于大文本对象数据可以编程打包处理，数据传输过程中可以将数据文件的内容进一步加密处理，以加强系统的安全性。

4 结束语

该方法的设计实现基于触发器和 FTP 服务，目前已在河海大学“数字校园”综合信息查询系统中得到成功应用，方法新颖、具有良好的可扩展性、易用性和实用性，整个同步过程可控性好，得到了用户和专家组的高度认可。

参考文献

1 丁　鲲, 严　浩, 刁兴春. 分布式数据库数据同步技术研究[J]. 海军工程大学学报, 2004, 16(5): 100-104.

2 黄显信. 利用 Socket 通信进程实现不同数据库间数据的同步[J]. 计算机与数字工程, 2003, 31(5): 28-30, 57.

3 包海峰. 基于 INTERNET 的企业数据库同步设计方案[J]. 计算机工程与应用, 2002, 38(4): 189-192.

(上接第 269 页)

参考文献

1 王文俊. 应急联动系统建设中的若干问题探讨[Z]. http://iems911.nease.net/.

2 程　鸿. 中国城市应急联动进行时[Z]. http://www.ccw.com.cn/.

3 陈秀万. 基于空间信息技术的城市应急救援联动系统研究[J]. 地理与地理信息科学, 2003, 19(4): 49-52.

4 连一峰, 魏　军. 安全事件应急处理系统分析[J]. 计算机工程, 2004, 30(7): 31-33.

5 王建鹏, 马吉苹. 基于 GIS 的城市应急系统研究[J]. 地理空间信息, 2004, 2(3): 25-27.

6 程效军, 徐文昌. 公安应急指挥系统中关键技术的探讨[J]. 同济大学学报, 2003, 31(1): 52-55.