

信息化校园建设中数据库同步问题的研究

郭大春, 田农乐

(浙江万里学院 信息化办公室, 浙江 宁波 315100)

摘要: 针对信息化校园建设中的数据库同步问题, 在分类总结当前数据库同步技术及同步方案各自的特点的基础上, 提出了一种数据库同步方案, 并对其中的方案进行了设计。

关键词: 信息化校园; 数据库同步; 同步方案

中图分类号: TP311.131

文献标识码: A

文章编号: 1674-7720(2011)06-0073-03

Research of problem about database synchronization in the construction of campus information

Guo Dachun, Tian Nongle

(Information Office, Zhejiang Wanli University, Ningbo 315100, China)

Abstract: Aimed at the problem of database synchronization in the construction of campus information, the paper classifies current database synchronization technology and methods and sums up the characteristics each other, finally presents a kind of database synchronization method and designs it.

Key words: campus information; database synchronization; synchronization scheme

目前, 由于大多数高校的网络应用系统存在各自为政, 没有统一规范的信息、接口标准, 信息难以交换, 形成信息孤岛, 难以在更高的层次上进行信息的分析处理等操作, 因此各高校都开始进行信息化校园建设, 而信息化校园建设的关键是建立公共数据中心, 其核心技术是数据库同步技术。如何高效、安全地保持数据库之间数据同步, 是信息化校园建设中需要考虑的重要问题之一。针对这种现象, 本文对各种同步技术与策略进行分析研究。

1 数据库同步技术

数据库要实现同步, 首先需实现数据库某几个表或者整个库的复制, 其中最重要的是复制过程中采用的变化捕获方法。不同数据库产品支持不同的变化捕获方法, 如快照法、触发器法、日志法、API法、影子表法、控制表法。其中, 触发器法和日志法是最基本的变化捕获的方法, 目前流行的商用数据库都提供触发器机制或日志分析工具。

1.1 基于快照法

快照是数据库中存储对象在某一时刻的即时映像。通过为同步对象定义一个快照方法, 可以将它的当前映像作为更新副本。基于快照法不需依赖特别的机制, 不

占用额外的系统资源, 管理和操作也非常容易, 而且在同步初始化和崩溃恢复时是必须的。但由于它无法区分同步对象中哪些具体项发生了改变, 因此效率很低。Oracle 和 SQL Server 明确提供了基于快照的方法。事实上, 所有的同步方案都隐含地要用到快照法。

1.2 基于触发器法

在源数据库为同步对象创建相应的触发器, 当对同步对象进行修改、插入或删除等 DML 命令时, 触发器被唤醒, 将变化传播到目标数据库。基于触发器法克服了基于快照法的主要缺点, 提高了效率。但是触发器法占用的系统资源比较多, 对较复杂的同步任务需要非常复杂的配置和实施, 管理极不方便。Oracle 的对称复制综合应用了基于快照和基于触发器两种方法。

1.3 基于日志法

数据库日志作为维护数据完整性和数据库恢复的重要信息, 其中包含了全部成功提交的操作记录信息。基于日志法就是通过分析数据库日志的信息来捕获同步对象的变化序列。传输的上下级之间仅在进行同步时, 仅对相对于上次同步传输后修改的数据进行传输。传输上下两级之间日志发生变化的记录从而提高传输速度, 减少了传输信息量以及带来的一系列问题。大多

技术与方法 Technique and Method

数数据库都有日志,利用日志不仅方便,也不会占用太多额外的系统资源。但是基于日志法也存在一些缺点:首先,一些数据库系统不公开其日志的格式,除非厂家提供相应的日志分析工具或接口,否则要开发一个基于日志的变化捕获程序非常困难;其次,尽管都是利用数据库日志获取变化,但不同数据库系统在具体细节上还是存在很大差异,这会异构数据库带来新的问题;最后,很多情况下 DBA 对数据库日志的管理已经很繁重、很复杂了,而基于日志法无疑会加重这种负担。DB2、SyBase、SQL Server 等主要商用数据库都采用了该方法。

2 数据库同步方案

随着数据库同步需求的出现和增长,人们提出了各种同步解决方案,像常用的基于 JMS 和基于 Web Service 的数据库同步方案和技术主要由数据库端触发器和客户端应用程序共同完成。该类型的同步技术具有以下优点:在业务过程中获得同步数据,实时性高,对同步服务器的性能要求高。

2.1 基于 JMS 的数据同步方案

采用 JMS 实现应用系统数据库和公共数据库之间的数据同步,在这种运行模式下,应用系统数据库和公共数据库都可以在自己需要数据时或资源空闲时将变化量提交或从相应的主题取得其他数据的变化量进行更新。应用系统将自己的变化量提交到服务器后,就完成了自己的任务。客户端可以进行其他的操作,不需要等待服务器端的消息处理结果,即使是网络故障或者服务器崩溃也不会造成数据的丢失或不一致性,消息会保存在消息队列或主题中直到被最终接收。比早期采用的套接字和 RMI 方式更加可靠和高效。

当应用系统数据库需要向公共数据库请求更新时,在客户端将更新信息发送到 JMS 服务器,服务器端的消息驱动 Bean(消息驱动 Bean 是能够接收 JMS 消息的特殊的 EJB 组件)会监听并接收这条信息,它调用实体 Bean 或会话 Bean 来完成具体的数据库更新操作。实际上消息驱动 Bean 是一个在服务器上实现某些业务逻辑的 JMS 消息使用者,当客户端发送一个 JMS 消息到达时由容器来调用它,客户端由 JMS 通过发送消息给 JMS 目的地(队列或主题)来访问消息驱动 Bean,如图 1 所示。

在发送消息之前,按照需要同步的数据表来定义应

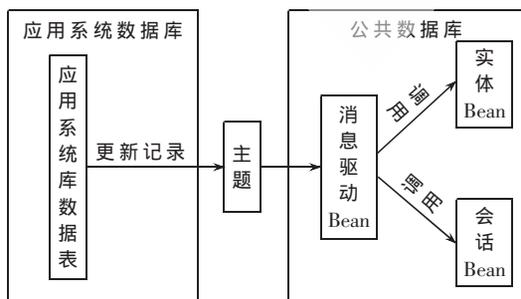


图 1 基于 JMS 的数据同步的模型

用系统数据库与公共数据库之间传送消息的格式,可以采用 XML 消息格式。当有消息到达公共系统时,服务器端的消息驱动 Bean 会监听到消息,它对接收的消息进行解析,取出相应的数据库字段,并调用相应的 EJB 组件来完成业务的逻辑处理,更新公共数据库。

采用 JMS 方案的数据同步技术能大大缩短应用系统与公共系统之间的响应时间,保证数据准确、高效地传输,更易于在各种平台之间移植。

2.2 基于 Web Service 的数据同步方案

现有主要平台供应商都支持 Web Service 的标准协议,可以使用任何语言编写 Web 服务。Web Service 完全屏蔽了不同软件平台的差异,不用考虑应用和具体 Web Service 所在的环境,并且使用绑定到 HTTP 的 SOAP 消息能不受阻拦地穿越现有的防火墙软件。借助于 WSDL 和 UDDI,同步功能能够以服务的形式被描述、注册并被不同的应用发现和使用。

方案主要由应用系统数据库、数据同步客户端程序、Web 服务器、Web Service 程序和公共数据库五部分组成。

数据同步程序负责从应用系统的数据库中提取同步数据以及负责与 Web 服务器中的 Web Service 程序进行通信,调用其发布的服务。当进行数据同步时,根据服务器端的 Web Service 程序的 WSDL 服务描述文档的调用信息,把客户端要调用的数据同步函数名和输入参数等信息加入到一个 SOAP 请求消息中,当向外发送时,该 SOAP 请求消息又被嵌入到一个 Http post 请求中,通过 HTTP 发往 Web 服务器,等待服务器端的处理和回应,如图 2 所示。

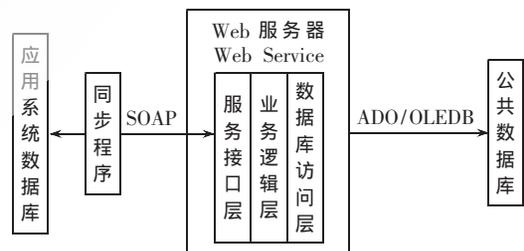


图 2 基于 Web Service 的数据同步的模型

基于 Web Service 的数据库数据同步系统方案能够比较好地解决数据库同步问题,但是由于局限于某些特定的技术,对各种技术实现及其策略缺乏宏观的比较、评价和优化,影响了实际使用效果。

2.3 浙江万里学院的数据库同步方案

下面以公共数据库从应用系统数据库中同步数据为例来说明浙江万里学院的数据库同步方案,为了实现数据的准确性同步,首先需要在应用系统库中建立一张中间表,该中间表是在数据同步过程中同步程序的操作表,公共数据库需对中间表有一定的操作权限。

公共库也会在其数据库中建立相应的中间表,其表

技术与方法 Technique and Method

结构与应用系统数据库提供的中间表表结构相同,并周期性地将应用系统数据库中间表的数据同步到该中间表中。通过数据清洗将数据及时地更新到公共数据库中的目标表中,进而实现了应用系统中的数据可持续地与公共库保持同步,如图3所示。

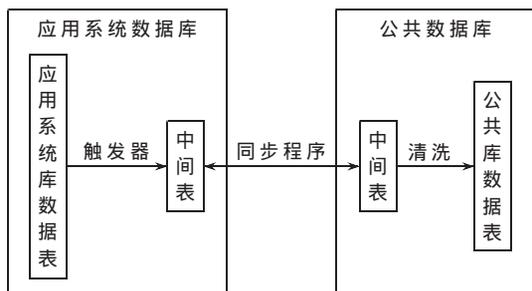


图3 浙江万里学院的数据同步的模型

该中间表的表结构除了具有应用库数据表中需同步的字段(字段名、类型和长度必须相同)外,还须具有标识字段。该字段是用来存储每条记录的增加、删除与修改等操作方式信息。为了能准确获取应用库中变动的数据(需要同步的数据),可利用数据库系统提供的触发器机制,为需要同步的数据表建立 INSERT、UPDATE 和 DELETE 三类触发器。每当用户增加、修改或删除数据表中的数据时,将自动触发对应的触发器,在触发器中编写代码把变动的数据保存到中间表中,并把操作方式记录到相应的标识字段中,以供公共库读取或操作。

根据如下的字段内容来设计应用系统的同步策略,任何一张提供数据变动信息的中间表都包含以下两个字段,具体如表1所示。

中间表查询应用库数据表变动情况,按照 FLAG(变动状态)、TIME(变动时间)为数据变动标准,进行数据增量同步。

更新和同步策略如表2所示。

当应用数据库提供的数据库信息有变化时,根据表结构把更新数据和对数据的操作都保存在相对应中间表

表1 举例字段列表

字段中文名	字段名	类型长度	备注
时间	Time	DATE	变动时间,DATE 类型
有效标记	Flag	VARCHAR2(1)	变动状态: 'n' 为新增记录; 'm' 为修改记录; 'd' 为删除记录

表2 更新和同步策略表

	数据操作	对应的标识状态变化
公共库更新	新增	time=sysdate and flag='n'
	修改	time=sysdate and flag='m'
	删除	time=sysdate and flag='d'

中,以供应用系统的使用,中间表内只需保持业务数据的最新状态,公共库可根据表结构及其数据自行决定采用何种策略来保持同步。

针对目前各种数据库同步技术和数据库同步方案,浙江万里学院信息化校园建设采用的数据库同步方案,有效地解决了数据库同步中存在的问题。文中对其中关键部分进行研究和设计,下一步研究重点是数据库同步具体设计的完善。

参考文献

- [1] 姜浩,王强.基于 Web Service 的数据库同步技术的研究与应用[J].计算机工程与应用,2007,43(专刊):162.
- [2] 章章荣,张军洲,诸葛隽.基于 Web Service 的异构数据库同步系统设计与实现[J].计算机技术与发展,2009,19(12):222.
- [3] 姚刚,邓江沙.JMS 在数据同步问题中的应用[J].长沙电力学院学报,2006,21(2):70-71.

(收稿日期:2010-10-29)

作者简介:

郭大春,男,1977年生,硕士,工程师,主要研究方向:网络与信息系统。

田农乐,男,1967年生,硕士,高级实验师,主要研究方向:网络与信息系统。