**分布式异构数据库同步更新的研究与应用**

Xin Yingxiu

摘要：由于不同的设备，不同的系统和异构数据库之间存在着信息“孤岛”的问题，而且设备的数据可以在业务点完成实时更新.这篇文章提出了一个基于JMS和XML的数据同步平台方案来实现数据的实时更新，并对工作流系统进行了描述，还对框架结构和使用的主要技术进行了详细的分析。经实验证明，该方案具有方便和实时等优点.

关键字：数据同步 XML JMS 数据库 实时更新

0介绍

随着计算机技术的广泛应用和和网络技术的飞速发展，各种各样的网络信息系统被应用在学校，工厂和研究所中。例如，示波器内部使用，信号发生器，数据采集和研究研究所的其他可以通过无线WIFI或局域网的访问互联网的设备。这些系统可以通过技术提高研究的效率，研究人员在家里就可以实时监控仪器数据，并提高他们的研究效率.但是，这同时也变成了一个“信息孤岛”的问题。也就是说，很多设备具有互联网功能，但是他们不能够更新或及时共享彼此之间的数据，这就导致了管理的困难。对于分布式网络设备，如何保持异构数据库的同步，输入信息更新之间的高效率和安全性，是我们需要关注和考虑的主要的重要问题之一。异构网络操作系统是不同的，或者说数据库系统是不同的，在相同的内部的异构数据库之间也有不同形式。通过组件，保持数据库之间的同步，并提高系统和数据的可靠性和传输效率.同时，为维护系统带来方便。

目前，分布式异构数据库的同步更新有很多方法，并且在国内很多领域都有研究。但是各自领域的研究重点是不同，许多研究者和研究人员都进行了研究，但很少取得成果。

Gong Tao在“异构数据库实时同步的研究”这篇文章中提出了“返回SQL语句的分解”。笔者主要针对数据冲突的研究，尤其是同步技术的研究提出了新的思路.此外，笔者还解释了同步的表结构的存在的重要作用，但它却没有说明表结构如何冲突，领域如何冲突等，或者说作者无法对同步数据到XML的映射和XML映射结构表到系统的映射作出分析，因此，一般情况下，其系统在跨平台和可扩展性方面的问题具有一定的限制。

XiongXian在他的文章中给出了一个在Oracle，SQL服务器，DB2等主流数据库之间的同步方案，最终在同步系统中提供了解决方案：触发trigger和AOP.AOP技术旨在为没有trigger支持数据库提供支持.笔者提出面向消息的中间件的“发布/订阅”模式，通信的解释认为它可以满足时间，空间，过程的要求。但是文章中没有涉及到表结构的冲突、综合模式同步，映射文件同步的问题。所以这个方法在某些方面也受到一定的限制。

Yang Peng在他的文章中对引入触发器捕获和日志表同步的组合进行了详细介绍，并通过在日志表的更改与时间标记同步战略。战略结合使用相同的触发来捕捉的变化的数据，提高了同步系统的性能;笔者使用的中间件：XML和SyncML，它可以根据该协议去传输同步数据，完全解决了异构数据库或不同设备的互连问题之间的数据同步，而且系统易于扩展和同步升级，以便适应多个数据源。但是该系统也暴露出一些问题，如：文章并没有对XML映射给出具体设计和数据同步的实现代码。

上述三种方案，主要代表当前异构数据库在分布式计算的普遍实现和应用，它们的优点和不足主要表现在以下几个方面：

(1)每个数据库是建立在用户界面上的：各种分布在每个节点的异构数据库都是建立在用户界面上的，但他们没有集成，并且他们每一个都能够分段管理和远程登录。这个方案很简单，但当异构数据库的增加新的内容时，用户和访问规则必须增加的用户界面，以便维护和使用复杂的方法;

(2)在异构数据库之间建立全局的模型。这个系统致力于对物理上分散分布，逻辑上相关的集中管理的部分进行集中管理模式设计，使得在任何数据库上的操作和运行都像是在本地实现一样。他们不必关心物理数据模型和其他底层细节.由于的数据库集成，当用户面对这样的异构数据库时，他们想要建立全球架构就面临很多困难。

(3) 分散和集中管模型。不同自治数据库和数据库的构件的集中性管理，采用数据传输技术来解决分布式数据处理和多节点集群的一致性问题。

(4) 一般的采集模式忽略系统。

(5) 忽略映射同步数据到XML上的同步性能的影响。

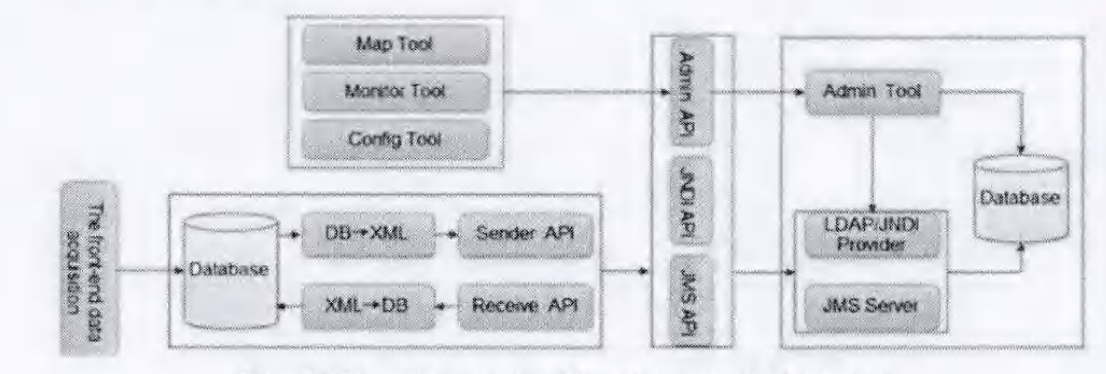
在本文中。为了有效地解决数据包流，减少流量（如无线传感器网络），数据同步，从根本上解决可以整合各种业务终端数据的分布式异构数据库的“信息孤岛”的问题，实现实时数据的同步更新，我们采用帧同步平台JMS和基于XML技术。

1系统框架和应用

图一给出了系统框架结构和主要模型，在图中我们可以看到：

1. 前端设备，包括设备的数据采集，数据存储。数据库，数据库接口API;
2. 同步数据平台，它是系统的核心部分，它主要包括消息服务器管理API，API JNDI，JMS API;
3. 管理中心。它主要包括管理工具，LDAP /ANDI服务，JMS服务器和数据库。

其中消息服务器是整个系统的核心，它遵循JMS（Java消息服务器）标准的JMS服务器服务。首先，JMS服务器上的消息队列通常使用“发布/订阅”消息队列的模型。在该模型中，发送到消息终端的目标数目可以无限制，并且每个消息可以任意次数被发送到到接收端，接收端的数目也可以不限数目。

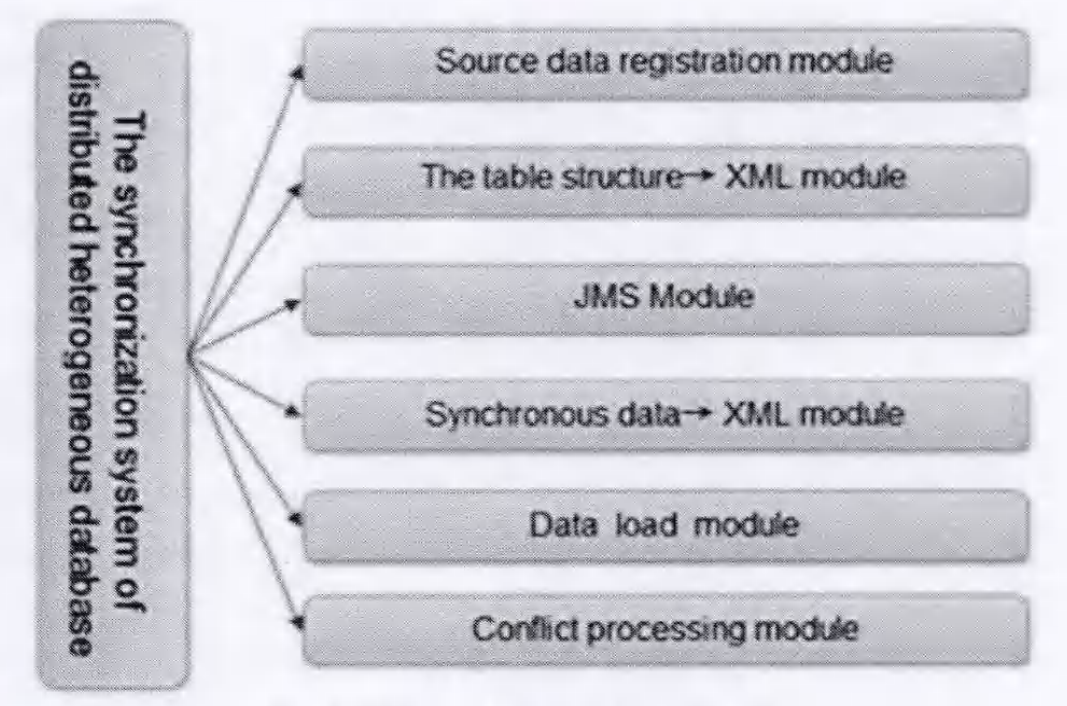


图一 系统体系结构图

从系统功能的角度看，由图1中的区块图可以看出，该系统的功能模块有数据来源寄存器模块，映射XML形成模块，消息监视模块。数据被映射到XML模块，数据读取模块和冲突处理模块。如图2中所示，各功能模块的任务是：源数据模块：这一部分主要是读取数据库的登录信息，如：ID，密码，IP地址等..该模块可以登录，删除注册用户信息。

1. 中心DB→XML适配器模块：源数据库注册后，将提取所有用户表的结构，如：登记表名，字段名，字段类型，并把它变成在数据库中的Table\_Info.xml文件，该文件被称为发件人的API，也基于映射文件。
2. 监视模块（监视工具）：根据新信息，监测模块进行实时采集，并刷新异构网络的信息，它是同步的基本要求。该模块是：当源数据库形成的数据发生变化时变化，通过一定的方法来提取同步数据。
3. 同步数据到主任务的XML模块：是对同步数据获取模块抽取设计的延续，并将其转换为Data\_xml.xml文件。Data\_xml文件的每个节点包含数据库的源类型，源数据库名，源表名。
4. 冲突处理模块：分析并记录维修故障系统的操作。
5. 数据下载模块：这个模块读取数据，并传送到目标表。除去C-TabChange表相应的控制信息后进行读取;如果读取错误，消息回滚，并将相信的控制信息保留在表C\_TabChange，进行异常处理和便于系统维护。

2472字



图二 系统功能分区

2 系统处理和工作原理

2.1系统实现流程

1）对于源和目标数据库之间的关系，首先，系统读取源数据库和目标数据库的登录信息和所有用户的表结构信息

2）在源数据库中创建（c\_change）记录列表，并创建源数据库中存在的共享所有表信息的触发器数据，当DML对源表操作时。触发器被唤醒。在源表中提取同步控制信息并写入表中记录;

3）行读取的记录表中的记录，根据来自数据库的数据中提取的记录的内容并生成Data\_xml.xml文件。根据表中UPDAITETYPE值来生成XML文件格式设计的表

4）基于源和目标表的元数据，建立映射文件（MappingFile.xmI），映射文件需要有层次结构和规律，在有冲突存在的情况下尽可能做出处理

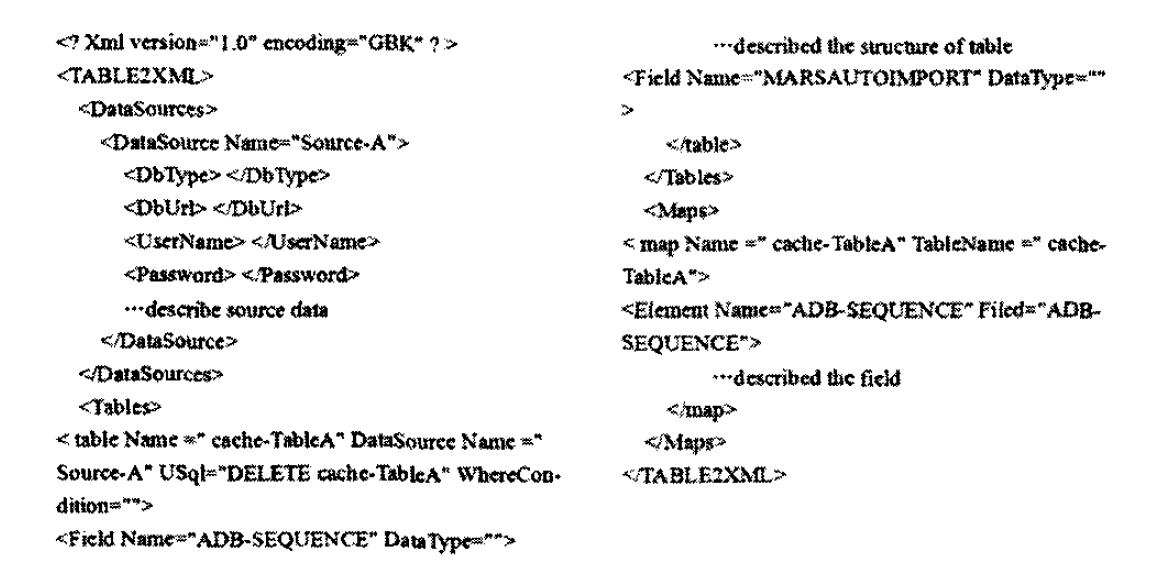
5）使用DOM解析XML文档，根据Dataxml在源表中找到该目标表信息的映射文件节点信息，并确定是否有各种复杂的映射模型.我们根据DataXml.xml和同步之前的映射文件来结合和还原SQL语句，同步数据被装载到目标表

6）数据成功同步之后.删除源记录在表中相应的信息记录.如果由于网络或其他不明原因造成破坏和中断，停止一切操作，当前变化记录事务回滚，等待同步操作的下一个周期，并保存同步操作记录。

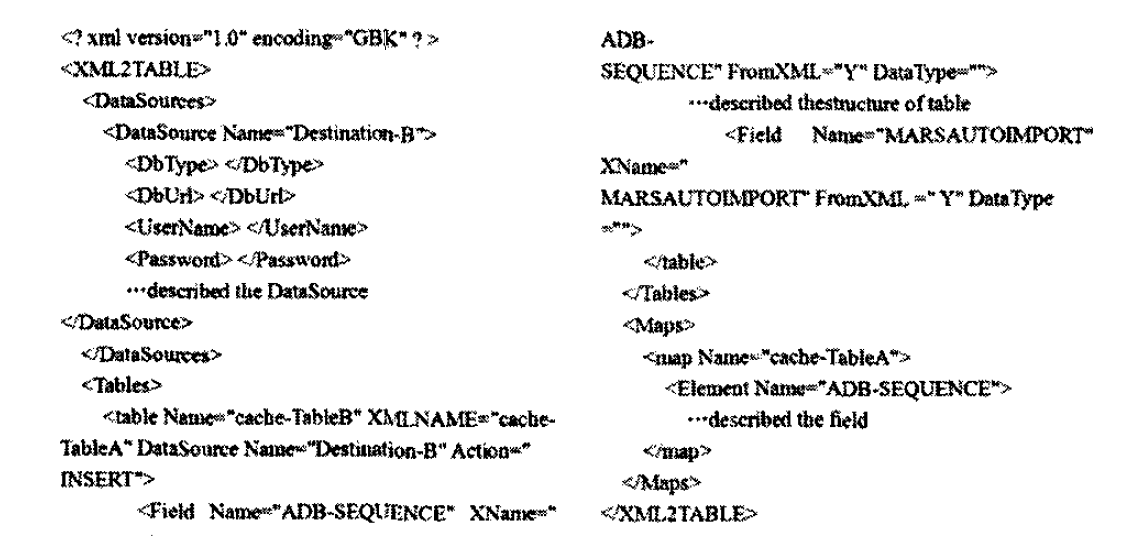
2.2 系统实现的主要代码

异构设备之间的数据同步：该缓存S-table将转换为映射，这个映射XML导出并且以缓存表D-Tlab1e的格式转换XML为映射。

导出映射的格式：



引入映射的格式：



3总结

随着信息技术发展，大网络设备由于环境因素和数据库选择类型的差异，造成了很多信息孤岛。然而，很多数据组合起来构成相对独立的信息服务，管理体系和数据库系统.这就要求要求我们必须保证每一个数据在数据库节点中的一致性.这篇文章提出建立一个安全的，通用的平台来解决分布式异构数据库同步的问题.本文中的同步方法提出“映射文件和JMS消息''捕获同步数据，同时，使用XML作为中间数据表示格式，由于XML易于使用，这就可以有效地避免异构数据库标准之间的统一问题，以确保专业的数据和数据共享，数据库和空间数据库的一致性具有重要意义。

参考资料：

[1]Gong Tao .Research on real time synchr0nization technology of heterogeneous database [D]．Huazhong University of Science and Technology．2010

[2]Xiong xian .Design and implementation ofheter0geneous database synchronization system of JAVA based on XML[D]．Shanghai JiaoTong University．2007

[3]Yang Peng．Heter0geneous database change capture and synchronization strategy[J]．Computer engineering，2008，34(1 6)：53．55

[4]Shen Min，Xu Hua—hu，et al．Implementation of data synchronization for distributed heterogeneous database[J]．Computer Engineering and Application，2005，41(5)：184—186

**分布式异构数据库的同步机制研究**

ZHANG Zhenyou

摘要 对于不同的物理位置分布的异构数据库，它是如何保证异构数据库的用户数据同步是一个非常关键的问题。本文为分布式异构数据库和提取模块，数据集成模块，数据更新模块分析数据同步系统，并且提出了解决方案的具体实现。

关键字 数据同步机制；最小变化集；最小更新集

一、前言

分布式异构数据库系统的研究是综合了不同类型的数据库系统的应用研究，研究的主要方向之一是分布式异构数据库系统之间的数据同步。设计数据同步的目的是为了使异构数据库的系统表数据与每一个站点数据对应同步。它可根据用户的要求，维持数据的一致性。这种同步没有改变原始异构数据库系统，因此让许多节点自治以实现数据库之间的数据共享是可以实现的。

二、分布式异构数据库同步机制的理论分析

异构数据库的数据同步的根本目的是在不同类型的数据库的不同节点上，保持相同的表数据相同。由于异构数据库系统的自主性，系统使得各个异构数据库中的数据只能由第三方系统维持同步数据传输的一致性。

异构数据库数据同步的实质是，当用户数据在节点数据库上发生变化时，同一时间其他异构数据库也做同样的变化。

同步数据的收集是异构数据库同步机制的基础。它在用户的需求下对于每个需要同步的数据表建立了三个触发器包括INSERT，UPDATE和DELETE触发器。每次当一个用户操作带有触发器的数据表时，触发器将被触发。因此数据库中的这些数据变化可以被发现，通过一个被称为数据库改变集的数据库。

异构数据库同步机制的关键是管理同步数据。目的是根据同步组，当触发器取得同步数据，并根据该同步组以获得变化最小集。再以每个节点的数据库完成不同的操作，这样就获得了各种异构数据库的最小更新集。

异构数据库同步机制里边的得到的更新用的数据，是每一个异构数据库的最小更新集，被用来更新各种异构数据库对应的数据表。

三、用于异构数据库的同步机制的设计

3.1设计数据提取模块

在此以测试表的SQL Server数据库为例，使用触发器来跟踪和提取测试表中的变化集。

如果测试表是遵照数据同步表，来创建test\_info\_insert数据表和test\_info\_delete表，但它有一个额外的列，这一列记录与测试表中的系统的当前时间。如果在“触发的内部”写入SQL语句来定义触发器，它就可以将数据集写到test\_info\_insert数据表和test\_info\_delete表。

将记录写入test\_info\_insert数据表的SQL语句如下所示：

1. 提取插入的数据记录：

//创建一个叫做test\_insert的触发器

CREATE test\_insert TRIGGER ON test

//入触发器的类型为添加触发

FOR INSERT AS

//在test\_insert触发器中已经包含了读取从已插入的数据表中读取数据并且把它放入到test\_info\_insert中

INSERT INTO test\_info\_insert SELECT \* FROM INSERTED

2.提取删除的数据记录

//创建一个叫做test\_delete的触发器

CREATE test\_insert TRIGGER ON test

//入触发器的类型为删除触发

FOR DELETE AS

//在test\_delete触发器中已经包含了读取从已删除的数据表中读取数据并且把它放入到test\_info\_delete中

INSERT INTO test\_info\_delete SELECT \* FROM DELETED

3.提取更新的数据记录

//创建一个叫做test\_ update的触发器

CREATE test\_ update TRIGGER ON test

//入触发器的类型为更新触发

FOR DELETE AS

//在test\_insert触发器中已经包含了读取从已插入的数据表中读取数据并且把它放入到test\_info\_insert中

//在test\_delete触发器中已经包含了读取从已删除的数据表中读取数据并且把它放入到test\_info\_delete中

INSERT INTO test\_info\_insert SELECT \* FROM INSERTED;

INSERT INTO test\_info \_delete SELECT \* FROM DELETED

3.2数据集成模块的设计

数据集成的目的是得到所有的异构数据库的最小更新集。因此，它必须首先将所有单个数据库变更集放到系统数据库。然后根据删除同步组中相同记录集，所以它得到总变化最小集，并有不同异构数据库的最小更新集。

在此模块中，它具有四个子模块，它是同步表管理，同步集集中，生成最小变化集，生成最小更新集。同步表管理的主要功能是在系统数据库中建立了同步数据表对应的的表结构，它被用于存储由数据采集模块得到的变化集。同步集集中的主要功能是根据同步组复制各种异构数据库中的数据到系统的数据库中。生成最小变化集的主要功能是处理的修改集的同步组，并删除冗余。生成最小更新集的主要功能是使用所产生的最小变化集区分不同的异构数据库更新集。插入最小变化集的过程如图1所示。插入最小的更新集的过程如图2所示。

图一 插入最小变化集

图二 插入最小的更新集

四、总结

数据同步的问题是目前信息管理的一个重要问题，传统的数据库复制技术有许多缺点。本文所提出的同步机制解决了数据的提取和数据更新和最小变化集可以通过减少同步数据来提高系统的效率。因此，我们希望在异构数据库数据同步研究方面提出更多的想法。

参考资料：

[1] WangYiHe etc. Distributed database systems [M].beijing: publishing house of electronics industry, 2008,6.

[2] YanHao distributed database data synchronization technique [J].jnaval engineering university journal, 2007,6 (3) : 11-15.

[3] DingBaoKang. Distributed database system realization technology[M]. Beijing: science press, 2006,8.

[4] huashan etc. SQL Server data replication [M].beijing: higher education press, 2006.

[5] P.S Bernstein. Hipman D and Rothnie j.carol carroll oncurrency. In a Distributed Control System for Database (SDD - 1) [J].j ACM Database System, Trans on 2007,4.

[6] Ceri, s.d. istributed bear - flying and System [J].j McGraw Hill,2006,8. -

[7] Traiger I.L.T ransactions and Consistency in Distributed bear [J].jIEEE Transactions on Software Engineering, 2006,6.