

基于 JMS 的分布式数据同步^①

吴高峰^{1,2}, 丁君辉¹, 徐远兵¹

¹(湖南中烟工业有限责任公司, 长沙 410014)

²(中南大学 信息科学与工程学院, 长沙 410014)

摘要: 针对分布式网络应用环境下数据同步问题进行了研究, 提出一种基于 JMS 与 XML 技术的数据同步模型, 对模型采用的主要技术、工作流程、实现方案做出了详细分析. 通过实验验证了所提出的同步模型在数据传输过程中的实时性、一致性和完整性, 最后对模型进行了总结并对后续工作进行了展望.

关键词: 分布式; 数据同步; JMS; XML

Data Synchronization Mechanism Based on JMS

WU Gao-Feng^{1,2}, DING Jun-Hui¹, XU Yuan-Bing¹

¹(China Tobacco Hunan Industrial Co., Ltd., Changsha 410014, China)

²(School of Information Science & Engineering, University of Central South of China, Changsha 410014, China)

Abstract: In order to solve the problem of data synchronization between different distributed network environment, a method of data synchronization based on java message service and xml technology is presented. This paper introduced the major technology, work process in detail. Experimental results show that the proposed method is real time efficient and reliable. Finally, we analyzed the system's characteristic and given some prospects to this mode.

Key words: distribution; data synchronization; JMS; XML

随着信息化和网络技术的迅速发展, 以数据分布存储和分布处理为主要特征的分布式数据库系统得到了广泛应用. 但分布式系统在提高系统灵活性、可靠性和扩展性的同时也带来了一些新问题, 其中如何使处于不同地理位置的多个数据库能够进行数据同步, 维护数据一致性是需要解决的重要问题.

目前解决分布式数据库的同步的方法主要有以下几种:

(1) 使用数据库厂商提供的数据库同步功能. 当前, 数据库厂商通常在数据库管理软件中提供成熟的数据同步工具. 此种方案对于同种类型的数据库实现简单, 可靠性高. 缺点是对于不同类型的数据库实现较难, 而且往往需要数据库管理人员人工操作, 不能完全实现分布式数据库间的自动同步;

(2) 基于 Web 服务的分布式数据同步方法. 此种方案由于采用了 SOAP 作为传输协议, 可以很好的穿

透防火墙, 可以在不同语言和平台之间相互通信. 缺点是需要额外设计数据一致性机制; 对一个数据库中的数据同步到多个数据库这种应用模式支持差, 实现困难, 逻辑复杂.

本文提出了一种基于 JMS 的分布式数据同步解决方案和实现方法. 该方案不仅可以方便地实现数据在应用系统和数据中心之间的多节点双向数据同步, 而且 JMS 自身机制在一定程度上保证了数据同步过程中的数据一致性和可靠性.

1 相关技术

1.1 JMS

JMS 是 SUN 公司提出的 java 消息规范, 它提供给应用程序创建、发送、接收和读取消息的接口, 可以通过 JMS 访问独立于厂商的消息服务系统.

JMS 是异步处理的, 消息发送者可以发送一个消

① 基金项目: 湖南中烟科技支撑计划(KY2011XX0002)

收稿时间: 2014-05-12; 收到修改稿时间: 2014-06-06

息而无需等待响应. 消息发送者将消息发送到一条虚拟通道(队列或主题上), 消息接收者监听或订阅该通道. 该消息可能最终转发给一个或多个消息接收者, 这些消息接收者无需对消息发送者做出响应.

JMS 支持两种基本的消息传递模型, 第一种是点到点的消息传递(P2P), 消息生产者向消息队列发送消息, 消息消费者从消息队列中提取消息, 每一个消息只有一个消费者, 但可以有很多的消息生产者. 第二种是发布/订阅(Pub/Sub),消息被传送给多个接收者, 消息发送者向主题队列发布消息, 消息接收者订阅主题, 每个消息可以有多个订阅者.

使用 JMS 有以下几个优点: 减少应用程序的耦合; 支持集群和失效转移, 具有良好的扩展性和兼容性; 提供分布式、事务性信息处理能力; 异构性.

1.2 XML

在数据集成领域, 由于 XML 作为多种应用程序之间的桥梁, 有效解决了异构数据源之间数据交换的难题, XML 做为异构数据交换具有自描述性、可扩展性和跨平台性等特点. XML 已经成为系统之间交互的标准格式, 不论数据库采用何种类型, 只要封装的数据满足 XML 格式, 就可以在系统间进行传递.

本系统用 Dom4j 来对 XML 读写, Dom4j 是一个非常优秀的 Java XML API, 具有性能优异、功能强大和非常容易使用的特点, 同时它也是一个开发源代码的软件.

2 系统设计

本系统将各个分散的、独立的数据库通过 JMS 连接起来, 在它们之间通过 XML 实现数据交换. 主要由以下几部分组成:

(1) 数据变化捕获组件: 用触发器实时捕获源数据表中发生变化的数据;

(2) 数据同步组件: 在数据同步发送端, 从数据源获取

变化数据, 并将这些数据进行处理和转换为 XML 文档, 然后将其发送到 JMS 服务器中的队列或者主题中; 在数据同步接收端, 接收 JMS 服务器中队列或者主题中的消息, 解析 XML 文档, 根据规则更新目的端数据库;

数据同步模型如图 1, 图 2 所示. 图 1 采用点对点

(P2P)消息传递方式,图 2 采用的是订阅/发布(Pub/Sub)消息传递方式.

对于 P2P 数据同步模型, 适用于将多个应用系统中的数据, 由下向上向数据中心汇聚.

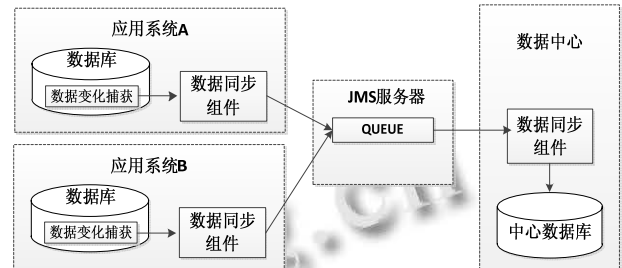


图 1 基于 JMS 的 P2P 模式数据同步

对于 Pub/Sub 数据同步模型, 适用于将数据中心的的数据, 由上向下分发到多个应用系统.

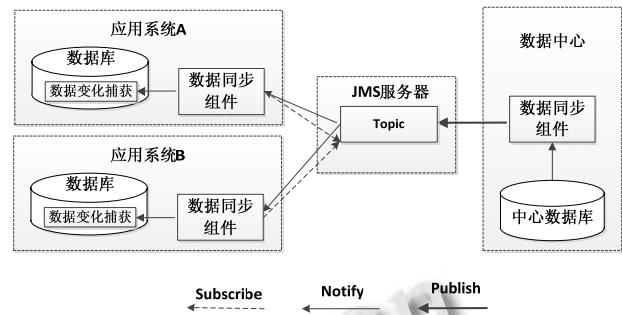


图 2 基于 JMS 的 pub/sub 模式数据同步

3 关键模块设计与实现

3.1 数据变化捕获组件

为了准确获取各应用系统的数据库中变动的数据(需要同步的数据),为需要同步的数据表建立 INSERT,UPDATA 和 DELETE 三类触发器. 每当用户增加、修改或删除数据表时, 将自动触发对应的触发器, 在触发器中编写代码将变动的数据保存到一个专门记录触发事件的差异数据缓冲表中, 该表包含原表所有的字段, 且字段类型完全一致, 并且增加一些同步描述字段, 新增字段和格式说明如表 1 所示.

3.2 数据同步组件

数据同步组件由数据同步发送端和数据同步接收端组成.

3.2.1 数据同步发送端

数据同步发送端程序检查差异数据缓冲表中有无更新

的数据。当同步程序发现差异数据缓冲表中有新数据时，根据数据更新时间的先后顺序构造一个包含表元数据和内容的 XML 文档。流程如图 3 所示。

表 1 差异数据缓冲表新增字段描述

字段名称	字段类型	字段说明
SEQUENCE	INT	主键，操作序号， 每增一条记录， 自动加 1
SOURCE_TABLE	VARCHAR	用户操作的数 据表名称
PRIMARY_KEY	VARCHAR	用于更或者删除目标表中 与该主键值相同的内容
OPERATION	CHAR	操作类型，分别 使用 0,1,2 表示 INSERT、UPDATA、 DELETE 操作
TIMESTAMP	TIME	时间戳

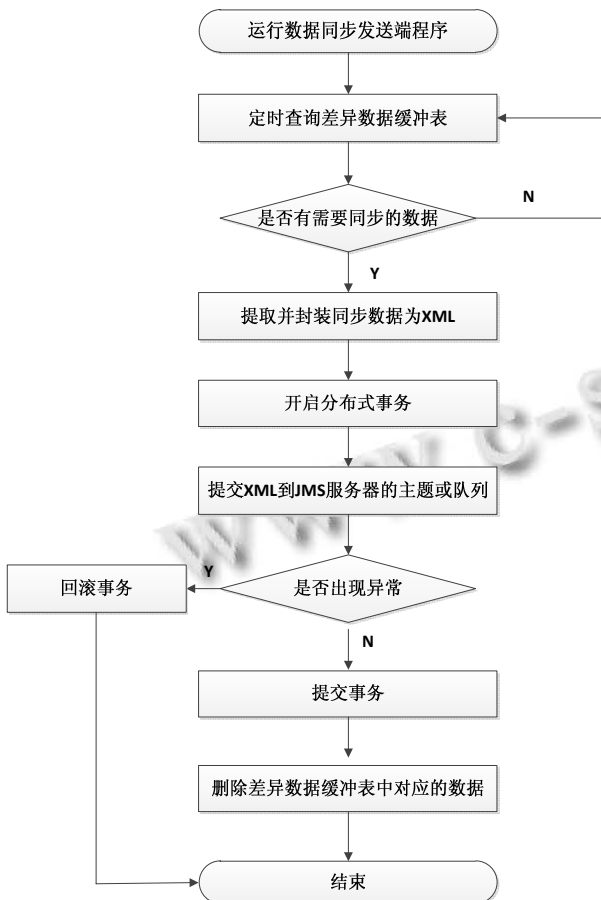


图 3 数据同步发送端流程

(1)构造待传输的 XML

Syn_content=DbAccess.selectObjects(S_cache.class,“select * from S_cache”,BaseConfig.maxline);
参数 S_cache.class 映射数据缓冲表中的各列，BaseConfig.maxline 是配置文件中可配置的最大行数，通过修改该配置，可以将数据表中的字段批量构造为 XML。当数据表中的数据小于该值时，将所有的数据构造为 XML 文档 XML 格式如下。

```
<?xml version='1.0' encoding='UTF-8'?>
<root>
<Data_xml>
<info>
<SEQUENCE>值</SEQUENCE>
<SOURCE_TABLE>值</SOURCE_TABLE>
<PRIMARY_KEY>值</PRIMARY_KEY>
<OPERATION>值</OPERATION>
<TIMESTAMP >值</GENETATE_TIME>
</info>
<data>
.....
</data>
</Data_xml>
<Data_xml>
.....
</Data_xml>
.....
</root>
```

(2) 将生成的 XML 文档发送到 JMS 消息队列或主题，当 XML 正确进入消息队列或主题后，删除数据缓冲表中的已构造完成的数据记录；

(3)循环执行上述步骤。

3.2.2 数据同步接收端

数据同步接收端接收到 XML 消息后，解析 XML，根据 OPERATION 值执行对应操作，当 OPERATION 值为 2 时，直接删除目标数据库中与 PRIMARY_KEY 值相同的数据记录。

为了保证数据同步的可靠性，设计接收差异数据缓冲表，依次读取 XML 文件的若干块数据，当 OPERATION 为其它值时，执行数据库操作，先写入接收差异数据缓冲表中。最后一起将接收差异数据缓冲表中的数据写入目标数据库。

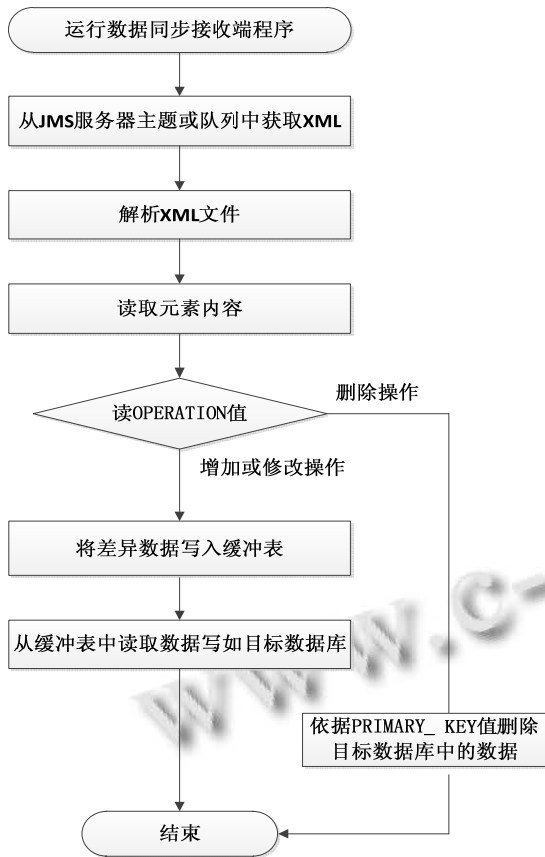


图 4 数据同步接收端流程

流程如图 4 所示:

- (1)从 JMS 消息队列接收消息或者订阅主题;
- (2) 解析 XML;
- ①取 Data_xml 节点为根节点;

②取<info>子节点,并取得<info>各子节点的,即主表附加字段的值. 如果 OPERATIO 值为 2, 则根据 PRIMARY_KEY 的值, 删除目标数据库中对应的数据; 如果 OPERATIO 值为 0 或 1, 则取<data>子节点, 并依次取的<data>各子节点的值, 即依次取的主表各字段的值;

(3)根据 OPERATION 值构造数据库操作语句, 将数据写入目标缓冲表中;

(4)将目标缓冲表中的数据拷贝到目标表中, 确认无误后删除目标缓冲表中的数据.

4 安全机制

由于在整个数据一致性维护过程中, 可能会因为网络、数据等诸多不可预料的因素而造成维护失败, 因此为保证整个数据的有效性和正确性, 系统采用以

下机制:

(1)事务

将整个数据一致性维护封装在一个事务中, 当对数据库的所有操作都成功时才提交事务, 否则回滚事务.

(2)持久化

消息采用 JMS 的持久化机制来实现持久化, 将所有的消息写入数据库中, 即使 JMS 服务器因故障崩溃或者网络中断, 恢复后, 原有的消息不会丢失, 而且在发送 JMS 消息时, 选择 PERSISTENT 模式, 该模式下, 能够让没发送完的消息继续发送出去.

本模型是将变化的数据封装为 XML 消息发送到 JMS 中间件的主题或队列中, 然后消息接收者再从主题或队列中取出消息进行操作, 在此过程中, 由于受到网络不确定因素的影响, 不能保证接收到的消息顺序与发送时一致, 从而造成数据的不一致性. 特别是在执行数据更新操作时, 这种不一致性造成的影响是显而易见的. 为了避免这种现象, 采用以下解决方案: 数据同步接收端解析收的 XML 消息, 当 OPERATION 值为 1, 即为修改类型时, 查找目标数据库中 PRIMARY_KEY 与 XML 中 PRIMARY_KEY 值相同的数据, 比较时间戳 TIMESTAMP, 如果 $TIMESTAMP > TIMESTAMP$ 则修改表中数据并更改时间戳为 $TIMESTAMP$, 如果 $TIMESTAMP < TIMESTAMP$, 则表示这是一条延迟的修改信息, 记录已经更新, 不需要处理这条消息.

5 实验

5.1 功能验证

为了验证本文提出的方法的正确性和合理性, 对系统进行了正确性验证, 在 Pentium Dual-Core 2.8G/4GRAM 配置下, 应用系统向数据中心采用 P2P 模式进行数据更新操作, 结果显示数据正确同步. 由数据中心向应用系统采用 pub/sub 模式进行数据更新, 结果显示数据被正确同步到各应用系统. 实现了数据的双向同步.

5.2 性能分析

设 N 为分布式应用系统个数, C(S)表示在分布式应用系统 i 进行数据同步处理的代价, C(T)为在分布式应用系统 i 的数据传输量, C(R)表示数据中心进行数据同步处理的代价, α 表示在网络中单位数据的传输代价,

则进行数据同步的总代价 C 为:

$$C = \sum_{i=1}^N C(S_i) + \sum_{i=1}^N \alpha C(T_i) + C(R)$$

为了验证本文提出方法的有效性,我们设计了一组实验,在模拟环境中验证系统的性能.以 P2P 模式数据同步为例给出测试结果.从图 5、图 6、图 7 可以看出,数据量小于 1600 时,同步时间稳定增长,整个同步时间在 10s 内完成.数据量大于 1600 时,呈线性增长.

本文提出的方法为了保证可靠性和通用性,同步过程中有缓存、事务、协议转换、文件转换等步骤,对同步时间会产生一定影响,但适用于小规模、数据类型多、可靠性要求高的同步环境.

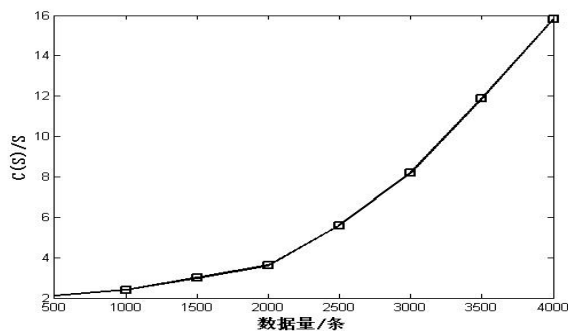


图 5 随同步数据量增加, $C(S)$ 变化情况

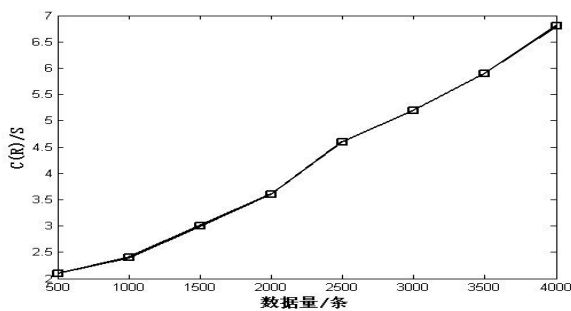


图 6 随同步数据量增加, $C(R)$ 变化情况

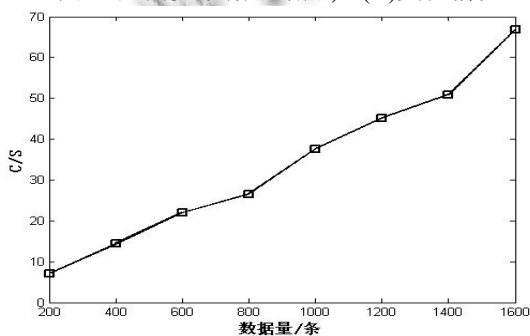


图 7 数据同步总代价

6 结语

本文首先介绍了 JMS 和 XML 相关基础概念,然后提出了基于 JMS 的分布式数据同步方法,详细阐述了对其中关键模块的设计实现方法,最后通过实验验证了方法的正确性并分析了性能.实验结果表明,本文中所提方法有效可行,具有较好的性能和可靠性.下一步将进一步考虑数据传输过程的安全性,及大数据的分割及压缩方法,以更好的适应大量数据的同步.

参考文献

- 1 夏克俭,张琰,巢群,等.XML 在数字化校园数据同步平台中的应用研究.计算机工程与设计,2002,24:483-486.
- 2 Jaechun JS. Synchronization mechanism for providing data consistency. 2010 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT). 2010. 506-510.
- 3 <http://activemq.apache.org>.
- 4 <http://java.sun.com/products/jms>.
- 5 ActiveMq in action, Bruce Snyder Dejan Bosanac Rob Davies.
- 6 Sun Microsystems. JMS Specification.
- 7 Zhang HW, Xiong Z, Ouyang YX. Data consistency maintenance of multi-database in distributed system. Computer Engineering and Applications, 2004, (23).
- 8 Tanenbaum AS, Van Steen M. Distributed Systems Principles and Paradigms. Beijing: Tsinghua University Press, 2005: 26-35.
- 9 王玉标,饶锡如,何盼.异构环境下数据库增量同步更新机制.计算机工程与设计,2011,32(3):948-951.
- 10 Lin LY, Huan CH. An integrated information system for real estate agency based on service-oriented architecture. Service-Oriented System Engineering, 2008, 52: 184-189.
- 11 贾艳燕,娄燕飞,杨树强,等.分布异构多数据库中多副本一致性维护研究与实现.计算机科学,2006,33(11):184-186.
- 12 贾焰,王志英,韩伟红,等.分布式数据库技术.北京:国防工业出版社,2000.