

# 基于 Web 服务的数据同步机制的研究与应用

陈红亮,程文青,吴 砥

(华中科技大学 电子与信息工程系,湖北 武汉 430074)

**摘 要:**在大型分布式信息系统中,各个子系统并非同时集成,在其投入运行时,存在与信息系统数据不一致的问题;或者,分布式系统在运行一段时期以后,该系统中各个子系统间的数据就可能出现差异。文中针对分布式系统中的数据缺乏一致性和同步性的问题,提出了一种基于 Web 服务的同步策略,论述了该同步策略的实现机制,介绍了该同步机制的触发方式以及同步信息的结构和传输算法。

**关键词:**同步;Web 服务;分布式系统;SOAP

**中图分类号:**TP39

**文献标识码:**A

**文章编号:**1673-629X(2006)09-0160-03

## Implementation and Application of Data Synchronization Strategy Based on Web Services

CHEN Hong-liang, CHENG Wen-qing, WU Di

(Dept. of Electronic Info. Eng., Huazhong Univ. of Sci. & Tech., Wuhan 430074, China)

**Abstract:** Due to the time disparity when subsystems integrated into the information system, the data discrepancy is an essential problem. In some another case, data inconsistency among subsystems emerges after a period of running in a distributed system. Aiming at the deficiency of data consistency and synchronization in distributed systems, presents a synchronous mechanism based on Web service and discusses its implementation. Furthermore, introduces the trigger way of the synchronous mechanism as well as the structure and transmission algorithm of the synchronous information.

**Key words:** synchronization; Web service; distribution system; SOAP

### 0 引 言

随着计算机技术的日益发展和网络规模的不断扩大,各种分布式应用系统平台已经得到了广泛的应用。在这些分布式应用系统平台中可能要涉及多个文件的管理或者多个数据库的管理,在系统的实际运作中,可能会产生各种不完整和不同步的数据和信息,这就需要一种同步机制来保持分布式系统内数据和信息的完整性和同步性。例如:在一个分布式应用系统中,由一个中心数据库来记录分散在各个不同点的文件系统的信息,可能某个点的文件系统发生了变化,而这种变化却没有在中心数据库中反映出来,这时数据和信息的一致性和同步性就是这个系统要迫切解决的问题。

Web 服务为分布式计算机领域带来了很好的异构能力,Web 服务的目标是实现分布式环境中各个组织内部及各个组织之间任意数量的应用程序或者应用程序组件

能够以与平台和语言无关的方式无缝交互<sup>[1]</sup>。而 Web 服务协议族中的核心协议——SOAP(Simple Object Access Protocol)是一个基于 XML(eXtensible Markup Language)的、用来在分布式环境中交换信息的协议。SOAP 的一个关键特征就是可访问性。SOAP 会话是通过“绑定”另一个更低层的协议而实现的,而且大部分应该是绑定 HTTP,FTP 或者 SMTP 协议,因此 SOAP 会话可以轻松跨越防火墙等界限<sup>[2]</sup>。而且 SOAP 本身是和平台和语言无关的,故文中采用带有附件的 SOAP 实现分布式系统中同步数据的传输。

### 1 同步策略

文中将以图 1 为例讲述同步机制在分布式系统中的实现策略。假设图 1 是一个集团公司的总部和下属的  $n(n > 1)$  个子公司,总部每个子公司都配置了一台具有固定 IP 的服务器,其它电脑通过集线器或交换机与服务器相连,各个子公司的服务器通过通信网络联接起来。

#### 1.1 同步机制的调度方式

分布式系统提供同步服务时,可以采用两种本质上不同的调度方式:事件触发和时间触发。事件触发方式是系统同步服务的执行,是由某个事件行为决定的,其发生时

收稿日期:2005-12-20

基金项目:教育部科技司项目(教技司[2004]104 号)

作者简介:陈红亮(1980-),男,新疆昌吉人,硕士研究生,研究方向为远程教育、电子商务;程文青,副教授,研究方向为远程教育、电子商务、宽带网络通信技术等。

间具有随机性。也就是说,同步在时间上表现为不可预测性,而时间触发方式则不同,它是通过一个 Timer 控件进行驱动的,系统的行为不仅在功能上得到确定,而且在时间上也是确定的。

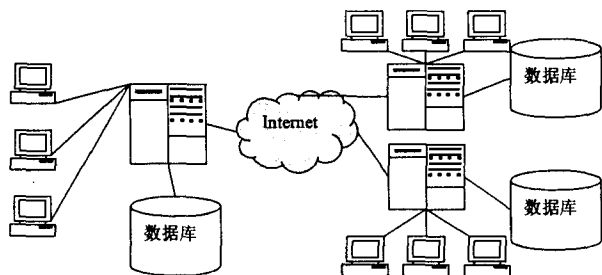


图1 同步机制用例

若图1中的分布式系统采用基于时间的触发方式,则系统根据自身反馈的统计信息来启动 Timer 控件。这些统计信息包括网络利用率、数据流量、空闲时间等;若图1中的系统采用基于事件的触发方式,那么系统为要同步的数据信息设定级别,当关键信息发生变化时,系统就捕获了这个事件并触发了信息同步的执行。当其它非关键数据发生变化并累积到一定数量时系统才触发同步服务。

从应用的特点上看,事件触发系统使得系统获得较大的吞吐量和反映的灵活性。而时间触发系统则可以获得更好的可预测性和可靠性<sup>[3]</sup>。分布式应用系统应根据自身系统的特点在提供同步服务时采用这两种调度方式中的一种,或同时采用上述两种调度方式。

## 1.2 同步信息的结构和调度算法

在分布式系统中,如果没有好的协议机制和调度算法,同步服务的可靠性是没法保证的。采用 SOAP 实现分布式系统中同步数据的传输,在 SOAP 消息的发送端和接收端可采用 DTD(Document Type Definition)来规范和识别同步信息。有两种类型的 SOAP 消息,带附件的消息和不带附件的消息<sup>[1,4]</sup>。当同步信息量相对较少时,可直接把同步信息置于 SOAP 体中,同步信息的结构如下所示:

```
<soap-env:Envelope xmlns:soap-env="http://schemas.xml-soap.org/soap/envelope/">
  <soap-env:Header/>
  <soap-env:Body>
    <books:GetAllBooks xmlns:books="http://eie.hust.edu.cn">
      <books:book id="2-1234-4455-4">
        <books:name>J2EE 企业应用开发</books:name>
        <books:publisher>电子工业出版社</books:publisher>
      >
      <books:price>60</books:price>
      <books:category>计算机类</books:category>
      <books:description>J2EE 企业应用开发用书</books:description>
      <books:author>陈亚强</books:author>
      <books:author>刘晓华</books:author>
    </books:book>
  </soap-env:Body>
</soap-env:Envelope>
```

```
</books:GetAllBooks>
</soap-env:Body>
</soap-env:Envelope>
```

当同步信息量相对较大时,按上述同步信息的结构把数据信息写入一个或多个文件,并由这些文件生成一个加密的压缩包。在传送 SOAP 消息时,把这个经过加密的压缩包放入 SOAP 消息的附件中,当 SOAP 消息到达接收端时就可以通过一系列操作获得这些同步信息。高级语言已经提供了处理带附件 SOAP 消息的 API,如 Java 中的 SAAJ(SOAP with Attachments API for Java)<sup>[5]</sup>。通过这些 API 可以方便地实现包含同步信息的 SOAP 消息。SOAP 消息在两个节点间的传递过程如图2所示。客户端可以是一般的 Java GUI 程序(也可以是 JSP,Servlet 等)。客户端通过 SOAP 消息和 Servlet 容器里运行的 JAXM(Java API for XML Messaging) Servlet 进行交互,JAXM Servlet 是服务提供者,EJB 容器里运行的是业务组件,它们为 JAXM Servlet 提供服务。

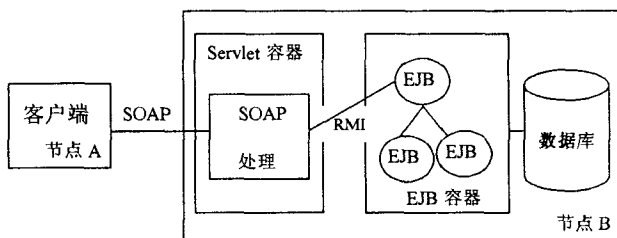


图2 SOAP 请求传递过程

分布式系统的各个子系统分布在通信网络的随机位置上,它们的物理位置是没有规律的。为了使同步信息能有效可靠地更新分布式系统的各个子系统,首先,要为信息更新选定一个起始点,这个起始点的数据信息是最新的并且是完备的。在图1中,可选择集团公司的总部为数据更新的主节点,主节点发送 SOAP 消息至系统中的其它节点,这些节点把未同步的信息传输至主节点,对主节点进行插入或修改操作时,主节点只记录有插入值或修改值的列名和内容;删除某条记录时,首先要将原来存放在主节点中对该条记录进行插入和修改的相关信息删除,然后再单独增加一条该记录的删除信息<sup>[6]</sup>。这样就选定并更新了主节点。

其次,要为分布式系统的各个子系统设计逻辑结构。可选择总线型结构、树型结构和环型结构来逻辑区分各个子系统。在总线型结构中,主节点采用轮循的方式依次更新该结构中的其它节点;在树型结构中,树的根节点为更新主节点,可采用深度优先搜索算法或广度优先算法来更新节点,已经更新的节点可按相同的算法搜索未更新的节点并更新之;在环型结构中,主节点先更新逻辑上较近的节点,然后在选择另一个逻辑上较近的节点进行更新。同时,已经更新的节点以同样的方式更新其它节点。

## 2 基于 Web 服务的同步机制的应用

上述基于 Web 服务的同步机制已经在《面向 21 世纪

教育振兴行动计划》“现代远程教育工程”中央财政专项项目“现代远程教育关键技术与支撑服务系统天地网结合专项”课题“网络教育资源描述、注册与管理”系统，国家教育部行动计划[教技司(2001)193号]基金资助项目中得到了应用。该系统是实现教育信息资源建设规范化和教育信息资源有效共享的基础平台，它能实现标准化网络资源的建设、注册、管理、发布与检索，使网络教育资源能够在 Internet 上实现共享，提高资源的有效利用率，从而提高网络教育资源的服务质量。该系统总体框架如图 3 所示。

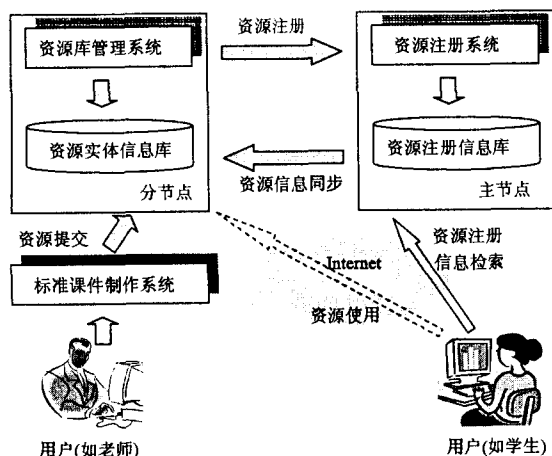


图 3 网络教育资源描述、注册与管理系统的总体框架

### 2.1 同步机制的应用课题介绍

该系统包含 3 个子系统：标准网络资源制作系统、资源库管理系统和资源注册系统。这 3 个子系统分别实现教育资源的制作、资源实体的保存及发布和资源信息的注册及检索。资源建设者使用标准网络资源制作系统制作自己的资源，并将制作好的资源保存到资源实体库，实体库管理员对资源进行审核，审核通过则在网上发布，并提取资源的注册信息在资源注册系统进行注册；资源使用者则通过注册系统查询自己需要的教育资源，注册系统将实体位置返回给用户，供用户浏览或者下载。该系统采用了 J2EE 标准进行开发，是一个具有多层结构的分布式应用系统。大量采用了组件技术、Web 服务技术、中间件技术

等软件技术，使整个系统具有极佳的可移植性和可维护性。

### 2.2 同步机制在课题中的实践应用

在该分布式系统中，存在多个资源注册系统，而每个资源注册系统同时又对应着多个资源库管理系统。首先，资源库管理系统以基于事件的触发方式把资源制作者或者管理员对资源的任何修改通过 SOAP 消息服务反馈到资源注册系统，实现资源注册系统和资源库管理系统之间的信息同步。再者，多个资源注册系统在逻辑上排列成树型结构，采用基于时间的触发方式依据广度优先算法来执行资源注册系统之间的信息同步。从而实现了分布式系统中各个子系统之间的信息同步，保持了信息的一致性。

## 3 总结

文中针对分布式系统中各个子系统之间可能出现的数据缺乏同步性和一致性的问题，提出了一种基于 Web 服务的同步策略。在实际的课题应用中表明，采用文中基于 Web 服务的数据同步机制能较有效地实现分布式系统中数据的同步性和一致性，消除了分布式系统运行的隐患，使应用系统更加健壮、可靠。

### 参考文献：

- [1] Chappell D A, Jewell T. Java Web 服务[M]. 北京：中国电力出版社，2003.9-10.
- [2] 柴晓路，梁宇奇. Web Services 技术、构架和应用[M]. 北京：电子工业出版社，2003.
- [3] SUN. J2EE 1.4 Platform Specification Proposed Final Draft 2 [M]. 北京：机械工业出版社，2003.
- [4] Horstmann C S. Gray Cornell, Core Java 2, Volume I: Advanced Features[M]. 北京：机械工业出版社，2000.
- [5] 陈亚强. 用 JAXM 开发 Web 服务[EB/OL]. <http://www-900.ibm.com/developerWorks/cn/webservices/ws-jaxm/part1/index.shtml>, 2003.
- [6] 赵 炜，张 浩，陆建峰. 基于 SOAP 协议的远程工程数据传输[J]. 计算机应用与软件，2002, 19(9): 6-8.

(上接第 159 页)

务请求的情况的处理，尽管采用“池化资源”技术（线程池和数据库连接池），但还是有可能达到服务器服务极限的可能，此种情形下，简单地抛弃请求不可取，那么最好采取的处理策略是：指示服务忙暂缓响应来让客户端稍后再发请求；如果不想抛弃或拒绝请求，那么最好的办法就是升级服务器硬件，这样就可以通过动态调整线程池大小来满足大批量的并发服务请求；另外的方法是对客户端数量的限制。总之，本方案的优点：其一，和其它方案相比，所需软硬件的投入很少，而使企业的业务技术水平发生质的提高；其二，收费信息实时性的提高可使企业得到及时的收支统计信息，为企业决策提供了及时的信息支持。

### 参考文献：

- [1] 黄嘉辉. Java 网络程序设计[M]. 北京：清华大学出版社，2004.
- [2] 武严娟，张素伟. OA 系统间数据互通网关设计与实现[J]. 计算机应用，2004, 24(1): 52-54.
- [3] 李 争，陆正中. Jbuilder 精髓[M]. 北京：电子工业出版社，2005.
- [4] 邵 波，王其和. 计算机网络安全技术及应用[M]. 北京：电子工业出版社，2005.
- [5] Delaney K. Microsoft SQL Server 2000 技术内幕[M]. 北京：北京大学出版社，2002.