

用 XQuery 模拟数据库存储过程的多应用系统整合

李锦棠¹, 赵 勇¹, 杨文伟²

(1. 广东工业大学 计算机学院, 广东 广州 510090; 2. 广东省教育厅 考试中心, 广东 广州 510090)

摘要: 提出了一种用 XML、XQuery 模拟数据库存储过程的方法, 并在此基础上实现多应用系统的整合. 分析了大量使用存储过程的缺点, 给出了本模拟方法的模型结构和工作流程, 并给出了一些相关的实现工作.

关键词: XQuery; XML; 存储过程; 系统整合; 关系数据库

中图分类号: TP311. 1

文献标识码: A

文章编号: 1007-7162(2006)02-0098-05

在许多的企业和事业单位中, 很多业务逻辑的处理都使用到数据库的存储过程. 基于计算密集型, 处理过程要与数据库进行频繁交互的业务逻辑一般都被封装在数据库中. 数据库的存储过程有它独特的优点, 但大量使用时会发现它存在着缺点. XML 和 XQuery 的出现为解决数据库存储过程的缺点提供了改进方向. 用 XML 作为数据源, 用 XQuery 编写业务逻辑, 可使存储过程与特定的数据库分离, 实现数据与逻辑都独立于平台, 从而使有不同数据源但业务逻辑几乎相同的应用系统实现整合. 本文将详细阐明用 XML、XQuery 模拟存储过程思想和用跨平台语言 Java 实现的相关工作.

1 数据库存储过程的优缺点分析

在处理要与数据库频繁交互的复杂的业务逻辑时, 一般情况下都会使用数据库的存储过程. 存储过程从概念上讲类似于程序中的函数, 能输入参数并输出相应的结果. 数据库的存储过程与函数的最大不同点是它必须与特定的数据库绑定. 也就是说, 数据库的存储过程必须嵌入到数据库内部并利用数据库内部引擎来运行. 使用数据库存储过程的最大好处是可以把应用系统中要与数据库频繁交互的复杂的业务逻辑封装在存储过程中, 这样可以大大减少应用系统与数据库交互时, 网络通信的时间开销. 存储过程可以使数据库操作人员与应用系统开发人员的分工更明确. 但大量使用存储过程会存在以下隐患.

1) 数据库存储过程缺乏可移植性

存储过程与特定数据库的绑定使得存储过程几乎不存在可移植性. 不同厂商的数据库要用不同的开发语言编写其存储过程如 SQL Server 用 T-SQL, Oracle 用 PL/SQL. 因此在需要更换数据库时就会出现很大的问题, 存储过程必须重写.

2) 无法在多个数据库之间共享存储过程

收稿日期: 2004-11-26

作者简介: 李锦棠(1980), 男, 硕士研究生, 主要研究方向为计算机网络与分布式系统.

由于存储过程是在数据库中运行, 而且运行时很难再与外界数据库联系. 因此当需要用到多个数据库时就无法使用存储过程.

3) 增加了数据库服务器的压力

由于大量的、复杂的业务处理逻辑都在数据库服务器上运行, 必然会给数据库服务器带来巨大的压力. 而且, 在通常情况下, 为了增强系统的稳定性, 都会购买比较昂贵的设备作为数据库服务器, 想升级也不是一件容易的事情. 基于上述的分析, 提出了一种利用 XML、XQuery 模拟数据库存储过程方案, 一种新型的数据库服务器系统, 即保持数据库存储过程的优点也克服了存储过程存在的缺点.

2 体系结构

从图 1 可以看出本体系结构图从传统的两级结构(应用系统—数据库)扩展为三级(应用系统—业务逻辑处理器—数据库). 业务逻辑处理器和后台数据库构成了新型的分离存储过程的数据库服务器系统.

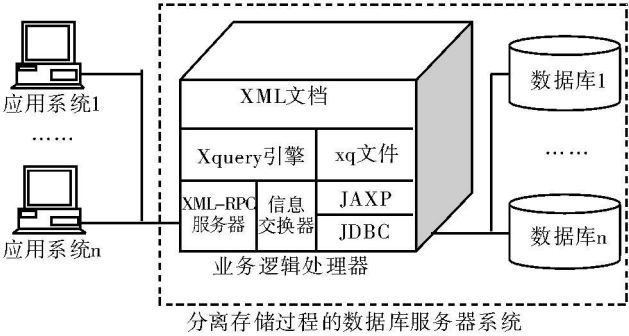


图 1 系统体系结构图

1) 数据库

本体系结构中的数据库并不是传统意义上的数据库, 它可以是多个, 甚至是异构的数据库群.

2) 业务逻辑处理器

这是本文研究的重点, 从总的结构上可以分为 3 层, 分别是顶层的 XML 数据层, 中间层是封装业务逻辑的 XQuery 语句和 XQuery 引擎这一层可称为业务处理层, 底层是运行在 JVM (Java 虚拟机) 上的一系列 Java 程序, 包括用于与应用系统通讯的 XML-RPC 服务器、用于处理中间信息的信息交换器、用于更新数据的数据更新器它使用了 JAXP 和 JDBC, 由于这一层主要负责对外通讯, 可称为对外通讯层. 数据层中的 XML 数据是后台数据库通过 XML 视图转换后得到的数据文档, 它与后台数据库中的数据相一致. 业务处理层里包含 XQuery 引擎用于解析和运行 XQuery 语句. 业务逻辑处理器通过 JDBC 与数据库相连, 目的是为了在修改 XML 文档中的数据后能马上通过 JDBC 修改数据库中相应的数据, 以保证数据的一致性.

3) 应用系统

这里的应用系统可以是一个 Web 服务器或者是一般的应用程序. 应用系统通过 XML-RPC (XML-based Remote Procedure Call, 基于 XML 的远程过程调用) 技术调用位于业务逻辑处理器上的 Java 程序, 返回的可以是一个数据流或一个特定的对象, 一般情况下返回的是 org.w3c.dom.Document 对象即 XML 视图文档. XML-RPC 可以用于在异构环境下的通讯, 它用 XML 作为信息传送的编码格式, 因此增加了应用系统实现的灵活性.

3 工作流程

系统工作流程见图 2, 具体步骤如下:

1) 应用系统通过 XML-RPC 调用业务逻辑处理器上的方法. 其方法是 client.execute(“server.callStoredProcedure”, params) 其中 params 是一个 Vector 对象, 它至少包含某存储过程的名称和具

有权限识别意义的用户名或密码等^[1]。

2) 当业务逻辑处理器的对外接口层的 XML-RPC 服务器收到来自应用系统的请求时, 会把这个信息和 params 参数表传送给信息交换器。

3) 信息交换器根据 params 中的参数识别要执行的存储过程的名称, 然后利用 XQuery 引擎激活存储过程。

4) 存储过程的运行过程需要与数据层频繁的交互, 在 XQuery 中通过 doc("data.xml") 语句来读取 XML 文档^[2]。

5) 存储过程结束后, 信息处理器会得到运行的结果, 它会根据业务的需要把结果返回应用系统。

6) 当需要更新数据时, 信息交换器会发消息给数据更新器。

7) 为了保持数据的一致性, 数据更新器会利用 JAXP 和 JDBC 同时更新数据层的 XML 数据和后台数据库中的数据。更新后台数据库时, 会根据存储过程的执行结果选择预先准备的 SQL 语句, 利用 JDBC 更新后台数据库中的数据。

4 相关实现工作

图 3 中的 studentInfo 表和 studentWill 表分别是广东省考试中心的中专录取招生系统的学生信息表和学生志愿表的一部分。XML 具有很强的数据表达能力, 特别是对于树型的数据结构。studentWill 是多主键的表, 用 XML 也能清晰简单地表达出来。把所有数据库中的所有表转换成 XML 视图文档提供了独立于平台的数据源, 就可以使用 XQuery 对 XML 中的数据进行查询和处理。

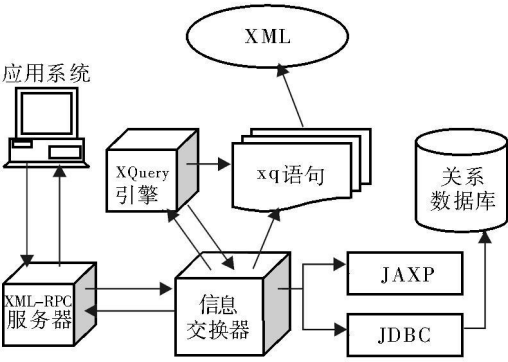


图 2 系统工作流程图

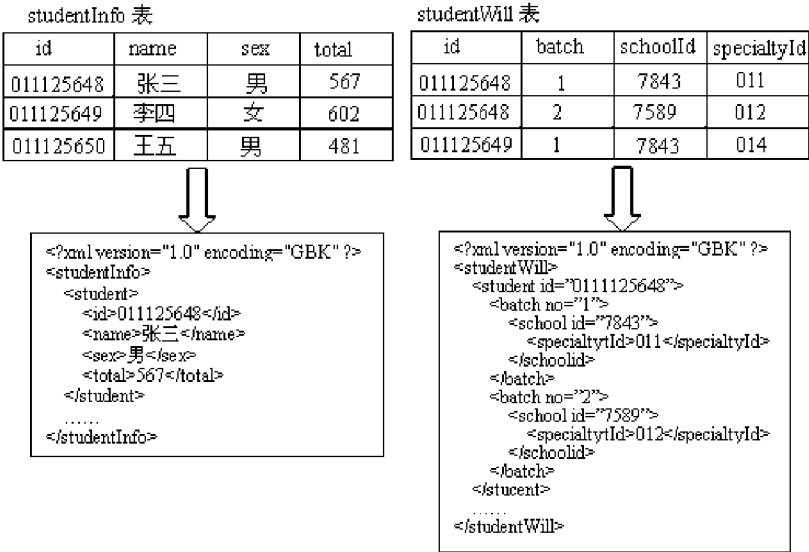


图 3 将数据表转换成 XML 视图文档

1) 将数据转换成 XML 视图(XML Views)

由于后台数据库群中可能存在异构的数据库,因此要将它们的数据转换成统一的数据结构是一个必要环节,借用文献[3]和文献[4]的思想可以将数据转换成统一的 XML 格式.一些先进的数据库管理系统拥有直接导出 XML 文档的功能.

2) 用 XQuery 封装业务逻辑

XQuery 是 XML 的数据查询语言,在撰写本文时 W3C 还没有推出其标准版, XQuery1.0 最新工作草案在 2004 年 7 月 23 日制定, XQuery 对 XML 数据有强大的查询能力,类似于关系数据库中的 SQL 语言.每一个 XQuery 查询语句都包含一个或多个表达式,其中 FLWOR 是 XQuery 最重要的表达式,它包括 for、where、let、order by 和 return 表达式^[5].执行 XQuery 语句需要一个 XQuery 引擎,现在已有多款的 XQuery 引擎可供选择例如: Yukon、SAXON、Qexo、X-Hive/DB 等.

3) 实现对外接口层

这一层包括 3 个部分 XML-RPC 服务器、信息交换器和数据更新器. XML-RPC 服务器可以使用一些轻型的处理类来实现,也可以使用 Java Servlet 实现一个 Web 服务器.信息交换主要负责中间信息的交换,可以使用简单的 JavaBean 实现^[1].用于一个 XML 文件作为系统配置文件将存储过程名称与实际的 xq 文件联系起来.这样信息交换就能更容易选择出要执行的 XQuery 文件.数据更新器又可分为两部分,用于更新 XML 数据的 JAXP 程序和更新后台数据库的 JDBC 程序.这里要预先编写好用于更新 XML 和关系数据库的语句.

5 实践结果评价

本文构想已在广东省考试中心初步实现,整合的应用系统包括:普通高考录取系统、中专录取系统和成人高考录取系统.这些系统原来都使用了数据库的存储过程,其中普通高考录取系统和中专录取系统用 Sybase 数据库,成人高考录取系统用 Oracle 数据库.这些系统有不同的数据源但它们的投档算法和机制几乎相同.虽然现在还处于试验阶段,但可以看出其能克服本文开头提到的大量使用数据库存储过程时存在的缺点,增强了系统的可扩展性.在系统开发时,前台开发人员、后台业务逻辑开发人员和后台数据管理人员可以更明确地分工.后期系统的维护工作也显得更容易,工作更明确.但与此同时也带来了新的问题.

1) 由于 XQuery 并没有提供修改 XML 的方法,当需要更新数据时就通过另外的途径,可能会影响系统的性能.

2) 由于要更新后台数据库时,需要通过网络,这也有可能降低系统的效率.改良方法是业务处理中间过程的修改不提交到数据库,等业务处理完毕一次性发送 SQL 语句.

3) 增加业务逻辑处理器是可以减低数据库服务器压力,但压力是被转移到业务逻辑处理器上.解决的办法可以同时使用多台业务逻辑处理器,将不同的业务逻辑处理放在不同的机器上执行.但要保证数据的一致性也是一个难题.

6 结束语

由于 XML 有极强的数据表达能力和平台无关性,使得它成为现在数据交换和传送的核心技术. XML 与数据库能完美地结合,使不同数据源的数据可以用统一的格式表达,从而使数据独立于平台. XQuery 的出现使本来与数据库绑定存储过程得以从数据库中脱离,使存储过程与平台无关.

虽然本思想还存在缺点,但随着 XML 技术的不断发展,特别当像 dbXML 这类专用于存储 XML 数据的数据库的出现后,以上提到的问题将能得到解决. dbXML 是专用于保存 XML 文档的数据库,这样就可以不再使用传统的关系数据库.

直接用 dbXML 存储 XML 数据,这样可以免去用 JDBC 修改后台数据库的工作,从而进一步提高系统的性能.这将是进一步研究所要做的工作.

参考文献:

- [1] Brett McLaughlin. Java & XML(第 2 版)[M]. 刘基诚译. 北京: 中国电力出版社, 2004: 285-320.
- [2] W3C. XML Query Use Cases [EB/OL]. [2003-11-12]. <http://www.w3.org/TR/2003/WD-xquery-use-cases-20031112/>.
- [3] J Shanmugasundaram, Jerry Kiernan, Eugene Shekita. Querying XML Views of Relational Data[EB/OL]. [2004-10-15]. <http://www.cs.cornell.edu/people/jai/papers/QueryingXMLViews.pdf>.
- [4] 熊家治,王飞乐,丁祥武,等. 基于 XML 的异构数据源—XQuery 查询[J]. 计算机应用与软件. 2004,21(2): 18-19, 104.
- [5] W3C. XQuery 1.0: An XML Query Language[EB/OL]. [2004-07-23]. <http://www.w3.org/TR/2004/WD-xquery>.

Study of Multi-system Integration Based on Using XQuery to Simulate Database's Stored Procedure

LI Jin-tang¹, ZHAO Yong¹, YANG Wen-wei²

(1. Faculty of Computer, Guangdong University of Technology, Guangzhou 510090, China;

2. Examination Center, Department of Education of Guangdong Province, Guangzhou 510090, China)

Abstract: This paper addresses a way of multi-system integration based on using XML and XQuery to simulate database's stored procedure. At first, it analyzes the disadvantages of stored procedure, then, addresses the model of this simulation and work flow. and some related implementation work.

Key words: XQuery; XML; stored procedure; system integration; relational database