

异构数据库集成中间件的设计与实现

曾国林,傅秀芬,吕占德

(广东工业大学 计算机学院 计算机系,广东 广州 510006)

摘要:随着信息产业和通信技术的发展,企业在信息化建设过程中构建起多个异构的信息系统。为了解决网络环境中异构数据库的数据集成和共享问题,提出了一种基于XML和数据库连接池的异构数据库集成中间件模型。在此模型中设计了翻译器、分发器、集成器,分别用于数据的格式转换、分解和合并,并对模型实现过程中的关键技术给出了详细描述。为了进一步提高数据库访问效率,引入了数据库连接池技术,实现了分布式异构数据库的透明访问和联合查询。

关键词:异构数据库集成;XML;中间件;数据库连接池

中图分类号:TP311.13

文献标识码:A

文章编号:1673-629X(2011)03-0083-04

Design and Realization for Heterogeneous Database Integration Middleware

ZENG Guo-lin, FU Xiu-fen, LÜ Zhan-de

(Dept. of Computer, Institute of Computer, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: With the development of information industry and communications technology, enterprises have built multiple heterogeneous information systems in the process of building information. In order to resolve data integration and sharing among heterogeneous databases in the network, this paper, based on XML and database connection pool, provides a middleware model for heterogeneous database integration. In this model, designed interpreter, dispatcher, integrator used for data format change, data decomposition and data combination, and describes the important technologies of realizing process. To further enhance the efficiency of database access, this paper introduces the database connection pool technology, this model realizes the transparent access and joint inquiry for distributed heterogeneous database.

Key words: heterogeneous database integration; XML; middleware; database connection pool

0 引言

随着信息量的迅速增加,人们迫切需要一个简单、快速的方法来整合各种异构数据源之间的信息。如何集成、访问这些数据的一个关键问题就是要提供给用户一个良好的访问接口,将信息以“一种统一的数据格式呈现给用户”,从而屏蔽它们在系统平台、运行环境、内部组织结构等方面的异构性,使用户对任何数据库的操作就如同在本地执行^[1]。

目前,已经出现了一些异构数据集成的技术,但总的来说,它们不能实现数据透明性的访问、跨平台多模式的集成,不能完全满足用户的访问需求。这些技术主要有OMG的CORBA,Microsoft的COM/DCOM,SUN的Java RMI/EJB,它们都能够有效地解决内部之间数据集成的问题,但是每一种组件模型的内部架构以及

数据组织形式都不一样,导致各个组件模型之间相互集成变得困难。XML的跨平台性、简洁性、自我描述性^[2]使其在数据处理、数据集成等领域具有一定的优势。因此,文中提出了一种基于XML和数据库连接池的异构数据库集成中间件模型,用户只需将访问需求提交给中间件,由中间件处理与各异构数据源之间的数据交互,同时为了解决多用户条件下对异构数据库访问的低效率性,引入了数据库连接池技术,提高了访问效率,实现了用户的访问需求。

1 中间件体系结构设计

从异构数据库集成中间件的功能层次上来看,中间件位于用户层和数据层之间,向下与各数据库系统建立连接,将用户的访问需求分配到具体数据库中执行,向上为用户的查询请求提供一致的访问接口。用户的数据交互请求在中间件均以XML文档形式进行传递,通过中间件与数据库的相互交互来实现用户的查询请求;中间件将数据库操作的结果转化成XML形式反馈给用户,屏蔽了异构数据源之间的异构性,实现

收稿日期:2010-07-04;修回日期:2010-10-18

基金项目:广东省自然科学基金项目(07001802)

作者简介:曾国林(1985-),男,江西吉安人,硕士研究生,主要研究方向为协同软件和数据库技术;傅秀芬,教授,硕士生导师,主要研究方向为协同软件、数据库技术等。

了正确的访问^[3]。

1.1 XML 中间件模型

系统分为用户层、中间件处理层和数据源层,其结构如图 1 所示。

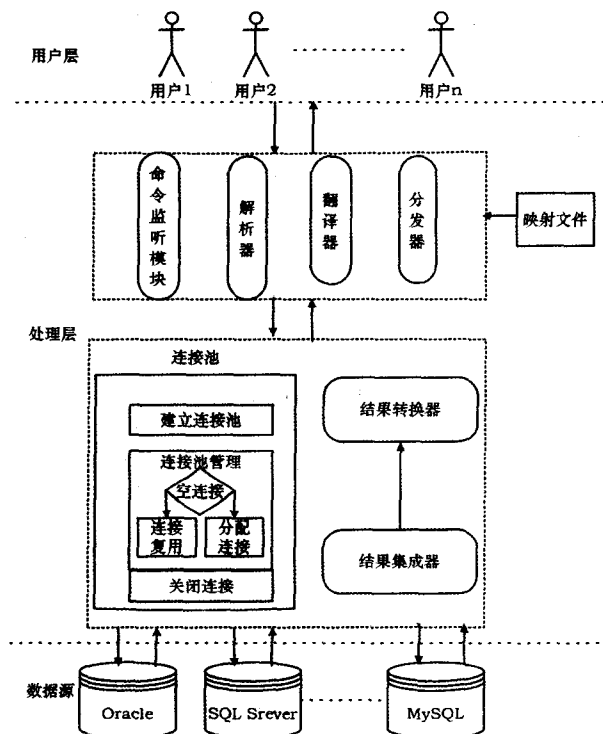


图 1 系统结构

中间件中各组件及功能介绍如下:

(1) 命令监听模块:该模块主要为用户提供访问接口,负责接收并解析用户的查询请求。

(2) 解析器:该模块实现对用户查询请求的语法解析^[4],并利用统一 DTD 描述的 XML 文件来记录各个异构数据源的描述信息,按照 XML 和数据库的映射语法^[5],将描述信息记录在映射文件中。

(3) 翻译器:该模块根据映射文件的 XML 格式,将用户的查询请求转化为 XML 查询文档,并组织成数据源能够识别的查询语句提交给分发器。

(4) 分发器:该模块参考映射文件,将翻译器提供的查询语句进行分发,使分发之后的每个子查询请求是针对一个具体数据库的查询操作,同时将分解后的各子查询之间的关联关系进行记录^[6]。

(5) 数据库连接池:数据库连接池解决了多用户情况下对数据库访问的效率下降、耗时多的缺陷。在连接池中预先建立一些连接放在内存对象中以便使用,当接收到分发器的子查询请求时,只须从连接池中取一个来与数据库建立连接,这将有效的提高数据库的访问效率。

(6) 结果集成器:该模块接收各个异构数据源执行子查询后的子结果,结合用户查询时提交的参数,对

各个子结果进行集成形成一个完整的结果集。

(7) 结果转换器:该模块根据记录的子查询之间的关联关系,对结果集进行过滤合并,将结果集转化为一个完整的 XML 文档返回给用户^[7]。

系统执行的过程为:当命令监听模块监听到用户的访问请求后,解析器对用户的查询请求进行语法解析,判断用户的请求是否符合语法规范,如果不符合则终止用户的查询;否则将正确的用户查询请求提交给翻译器,翻译器根据映射文件的 XML 格式,将用户的查询请求进一步转化为 XML 文档并组织成数据源能够识别的查询语句传送给分发器,分发器模块参考数据库映射文件,将整个查询分解为针对各具体异构数据库的子查询,具体的通信交由数据库连接池来完成。数据库服务器执行查询语句后,各异构数据库得到的子查询结果经结果转换器转化为一个结果 XML 文档并封装成一定的格式返回给用户。系统运行的流程如图 2 所示。

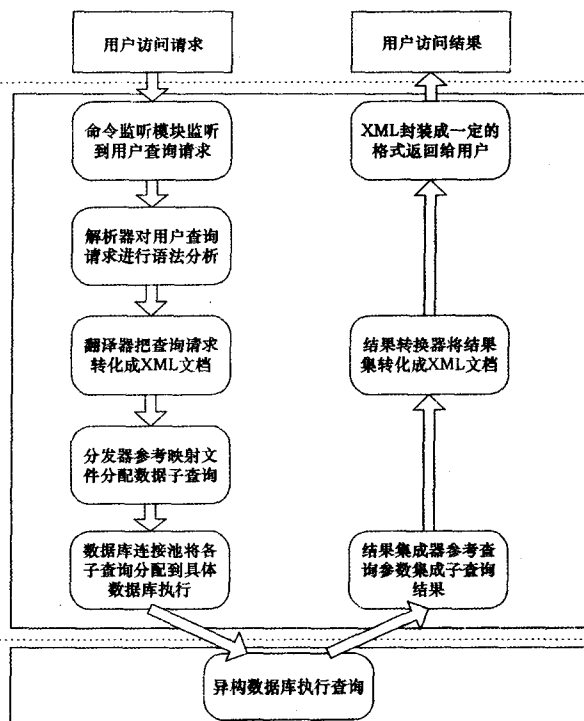


图 2 中间件运行流程

1.2 XML 中间件优势

(1) 跨平台多模式的集成。XML 具有跨平台特性,因而在异构数据库系统中应用 XML 便于在各个平台之间进行移植,这样 XML 文档就可以自由的在两个不同的系统之间交换数据,实现了异构数据库之间的直接访问。

(2) 解决了异构数据源之间的数据结构冲突。本系统采用全局访问与局部访问两层模式,把用户的全局访问分解为针对各个数据源的局部访问,屏蔽了各

异构数据源在数据组织上的冲突。

(3)实现了 XML 与数据库的互操作。XML 提供了一种通用的、自解释性好的数据格式,将用户的访问请求以 XML 的形式提交给数据库和将数据库访问结果以 XML 格式返回给用户,是将来开发应用程序的一种理想方案。

2 关键技术的实现

2.1 数据库连接池技术

连接池可看作是一些连接对象的“存储池”,连接池事先建立一些连接存放在存储池中以供用户使用。连接池通过其内部的连接机制控制连接对象的个数,对分发后的各子查询提供获取连接的接口。其工作原理如图 3 所示^[8]。

连接池的结构由线程池、连接池、数据库操作三大部分组成。线程池统一对分发后的各子查询进行线程调度,合理使用计算机内部的线程资源。连接池负责存储池中连接的建立、管理和释放。连接池在存储池中事先建立一些连接,当用户的数据访问需求被分发器分解为多个子查后,各子查询经过线程池调度只需从连接池中取出一个已建立的空闲对象和数据库建立连接。使用完毕后,用户也并非将连接关闭,而是将连接放回连接池中,以供下一个请求访问使用。数据库操作负责各子查询语句的执行,返回子查询结果,并对查询中出现的异常进行处理。通过使用连接池,将避免频繁的数据库连接时建立、关闭的开销,提高程序效率^[9]。

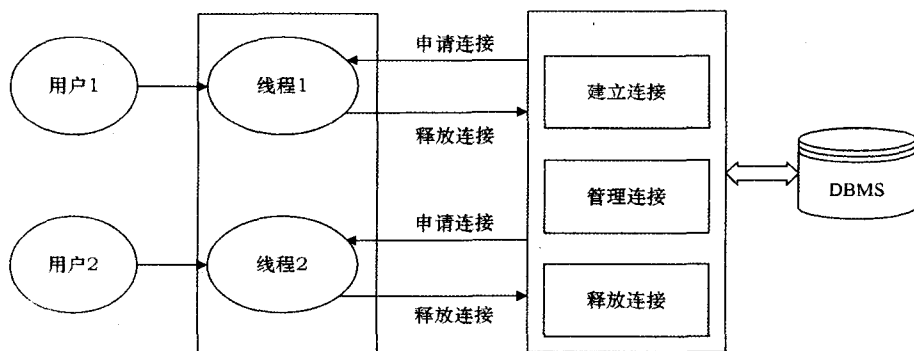


图3 连接池的工作原理

在数据库连接池中,我们是通过 JDBC 的 API 函数与异构数据库建立连接的,由于 JDBC 的 API 与其平台无关,这样不仅可以很容易把查询语句传到具体数据库中,而且可以获取异构数据库的各种表和视图信息,完善映射文件的生成。

2.2 从数据库访问结果生成 XML 文档

XML 文档是层次型结构,数据中元素之间的关系通过相互嵌套的形式体现。而关系型数据库中的数据

存储在二维表中,数据间的关系通过主、外键的方式体现,因此在由数据库数据向 XML 文档转换的过程中必须考虑数据间的对应关系。

在由数据库数据向 XML 文档转换的过程中,对于无依赖关系的单张表的映射,只需将其映射为一个单独元素,表中的每一列都映射成为该元素的子元素或属性;对于多张表之间的映射,关键在于解决表之间的依赖关系。

将关系数据库的数据转换为 XML 文档步骤如下:

(1)对于数据库中的每一张表各自创建一个与其表名对应的元素。

(2)对该表中的每一列在其对应的元素中创建一个属性。

(3)对于表中的每一主键/外键关系,在其对应的元素中创建子元素,根据表中的主键、外键关系将各元素、子元素合并成 XML 文档,然后优化 XML 文档。

2.3 XML 文档转化为数据库数据

XML 文档包含单一元素和复合元素。单一元素是指其元素不存在嵌套情况;复合元素可以包含属性、子元素,其子元素还可以是复合元素^[10]。例如下面的 XML 文档:

```
<企业>
<企业名称>华南资讯</企业名称>
<企业性质>私有企业</企业性质>
<企业领导>
<姓名>孙伟</姓名>
<性别>男</性别>
<生日>1978-08-13</
生日>
</企业领导>
<企业联系方式>
<传真>020-5651708
</传真>
<e-mail>sunwei@126.
com</e-mail>
</企业联系方式>
</企业>
```

```
<企业>
<企业名称>新泰科技</企业名称>
<企业性质>私有企业</企业性质>
<企业领导>
<姓名>张娜</姓名>
<性别>女</性别>
<生日>1975-02-21</生日>
</企业领导>
<企业联系方式>
```

<传真>020-5651802</传真>

<e-mail>zhangna@126.com</e-mail>

</企业联系方式>

</企业>

其中企业名称、企业性质、姓名、性别、生日、传真、e-mail 是单一元素;企业、企业领导、企业联系方式是复合元素。

将 XML 文档转换为关系数据库数据,采用下面的映射规则:

(1)对于 XML 中的单一元素,将其直接映射到表中对应的列。

(2)XML 中的复合元素形成表,将复合元素中的属性和子元素映射到该表中对应的列。若是复合子元素,将其和主元素间的从属关系通过 PK 码联系起来。

对于上面的 XML 文档进行映射,结果如表 1~3 所示。

表 1 企业

企业名称	企业性质	PK
华南资讯	私有企业	1
新泰科技	私有企业	2

表 2 企业领导

姓名	性别	生日	PK
孙伟	男	1978-08-13	1
张娜	女	1975-02-21	2

表 3 企业联系方式

传真	e-mail	PK
020-5651708	sunwei@126.com	1
020-5651802	zhangna@126.com	2

3 结束语

针对企业信息化过程中出现的异构数据库集成的问题,以 XML 文档作为异构数据库之间转换的桥梁,提出了一个基于 XML 中间件的异构数据源集成模型^[11]。该模型构建于用户端和数据源之间,用户和数据源之间的信息交互以 XML 格式进行,解决了异构数据源之间数据集成的异构性,给用户提供了一个良好的访问平台;为了进一步满足多用户情况下对数据库

的访问,引入了数据库连接池技术,提高了访问效率,实现了用户的正确访问。该模型具有一定的可行性和实用性,能够胜任异构数据库的集成工作,对信息系统的集成有一定的实际意义。随着网络资源的快速发展,下一步的工作是如何将本技术与 Web 相结合,实现智能性的跨站点的信息资源共享^[12]。

参考文献:

- [1] OUYANG Zhengzheng. Application of XML-based Heterogeneous Database Date Exchange Middleware in Ecommerce [J]. IEEE Software, 2009, 10(9): 132-135.
- [2] YIN Dali, JIANG Meiyu. Teaching Information Resources Integration based on Heterogeneous Data Exchange and Sharing of Technology [C]//Computer Science and Computational Technology, 2008. ISCSIT' 08. International Symposium on. [s. l.]: [s. n.], 2008: 150-153.
- [3] 于帆,王振铎,王振辉. 基于 XML 异构数据库集成中间件的设计与实现[J]. 计算机应用研究, 2007, 24(9): 185-188.
- [4] 易宝林,谭志鹏. 数据库访问柔性中间件的设计与实现[J]. 微电子学与计算机, 2006, 23(11): 166-169.
- [5] 史晔翎,黎建辉. 关系数据库模式到 XML Schema 的通用映射模型[J]. 计算机工程, 2009, 35(7): 35-39.
- [6] 甄玉钢,刘璐莹,康建初. 基于 XML 的异构数据库集成系统构架与开发[J]. 计算机工程, 2006, 32(2): 85-87.
- [7] 刘威,杨丹. 基于虚拟视图的异构数据库集成平台的研究[J]. 计算机技术与发展, 2009, 19(6): 91-94.
- [8] 孙叶枫,宋中山. JSP 中基于连接池的数据库访问技术[J]. 计算机应用, 2004, 24(6): 80-82.
- [9] 靳其兵,李晓波. 基于 JSP 的数据库连接技术的研究[J]. 计算机仿真, 2007, 24(4): 108-111.
- [10] 王蕊,陈庆奎. 异构数据库集成中间件的研究与实现[J]. 计算机工程与设计, 2008, 29(22): 5738-5741.
- [11] XIONG Fengguang, HAN Xie, KUANG Liqun. Research and implementation of heterogeneous data integration based on XML [J]. IEEE Software, 2009, 9(8): 711-715.
- [12] 李亚红,吴江,贾晖. 基于 Web Services 实现异构数据库集成技术[J]. 计算机应用研究, 2006(2): 81-84.
- [12] Martin D, Burstein M, Hobbs J, et al. OWL-S: Semantic mark-up for Web Services [EB/OL]. 2004. <http://www.w3.org/Submission/OWL-S/>.
- [13] Smith M K, Welly C, McGuinness D. Owl Web ontology language guide [EB/OL]. 2003. <http://www.w3.org/TR/owl-guide/>.
- [14] Tuscany sdo, version 2.0 [EB/OL]. 2004. <http://tuscany.apache.org/sdo-java.html>.

(上接第 82 页)

- [9] 倪志刚,洪政,刘佳. 基于服务数据对象的异构系统数据集成方案研究[J]. 计算机应用, 2007, 27: 21-23.
- [10] 吴健,吴朝晖,李莹,等. 基于本体论和词汇语义相似度的 Web 服务发现[J]. 计算机学报, 2005, 28(4): 595-602.
- [11] Paolucci M, Kawamura T, Terry R, et al. Semantic Matching of Web Services Capabilities [C]//Proceedings of the First International Semantic Web Conference (ISWC2002). Sar-dinia, Italy: [s. n.], 2002.