

多数据库中间件上推规则与上推机的设计与实现

赵 洁¹, 魏 超², 齐德昱²

(1. 广东工业大学 经济管理学院, 广东 广州 510520)

2. 华南理工大学 计算机科学与工程学院, 广东 广州 510640)

摘要: 研究了局部数据库中上推数据的内容、获取和描述, 定义出描述局部数据库中上推数据的 XML 模式, 作为集成机和局部数据库上推代理之间的通信规范, 以此为基础实现中间件中的上推机。

关键词: 多数据库中间件; 局部数据库上推描述; 可扩展标记语言

中图分类号: TP311.138 文献标识码: A 文章编号: 1007-7162(2006)04-0017-06

1 研究背景和基础

多数据库系统 (Multidatabase System MDBS) 在多个自治、分布、异构的数据库系统之上构成全局系统管理层, 屏蔽了不同数据库在物理上和逻辑上的差异, 提供一致的外部接口, 使用户能实现对异种数据库的透明访问。

1.1 国内外多数据库中间件的研究进展

多数据库集成非常广泛, 现在国内外这方面的研究处于一种无规范的滞后状态, 采用的一些方法主要有以下几种: 1) 直接数据库访问接口; 2) 分布对象; 3) 数据发布; 4) 扩充 SQL 功能; 5) 数据访问代理 (如虚拟数据库 VB); 6) Web Service 方式。

这些方法均有值得借鉴之处, 但大都就事论事, 只是权宜之计, 也未提高到中间件的解决路线, 研究还有待于标准化、规范化^[1-2]。

1.2 多数据库中间件模型

为解决上述问题, “多数据库中间件”项目提出了一种多数据库中间件模型, 它能支持异源数据的“上推 受访 集成”的模型化, 同时支持数据上推的“热插拔”、事件驱动、敏感识别、安全保密控制、并发控制、访问过程控制、输出整合等功能, 从而, 使异源独立分布数据的存取/复用/集成, 能像单一关系数据库那样走向模型化, 兼容标准 SQL 与面向对象数据库访问接口, 建立“DBMS 的 DBMS”, 模型如图 1 示, 主要分为 4 个逻辑部分: 1) 上推规则; 2) 上推机; 3) 集成机; 4) 分布数据链接集。

在此模型中, 必须考虑到异构的局部数据库间存在着巨大差异, 如计算机操作系统、网络通信协议、DBMS 数据结构的差异等。其中如何屏蔽这些差异, 形成统一的数据模式是多数据库集

收稿日期: 2005-12-09

项目基金: 广东省重大科技专项资助项目 (2003A1030404)

作者简介: 赵洁 (1979-), 女, 助教, 主要研究方向为网络集成与运算。

成的关键技术之一. 本课题是上述项目的子课题, 主要研究上推规则, 并在此基础上设计、实现上推机.

2 上推规则: 国内外相关的研究分析

2.1 目前的 XML化数据库产品

目前数据库产品的“XML”化工具主要采用两种策略, 把 XML应用到关系数据库中:

1) 查询语言的扩充. 首先在关系数据库之上建立虚拟的 XML视图, 然后在其上定义查询. 这种方式是单方向的, 只能从关系数据库中转换数据输出到 XML文档, 而且现在还只能做到查询, 对插入、更新、删除语句的实现是其发展趋势.

2) 标准的映射地图, 生成数据库模式和 XML模式的匹配映射. 这种方式是双向的.

这些方法都是根据本身数据库特点而推出的 XML描述和转换, 并未规范化、标准化.

2.2 异源数据的定义描述

在对异源数据的集成与上推的研究中, 大部分工作都将重点放到 API或 GUI上, 这方面所应用的主要方法在下面列出, 但从语言方面考虑的例子不多.

1) 新型数据访问中间件: 通过让用户建立虚拟的关系数据库来解决异源数据问题. 通过管理元数据的全局目录来屏蔽局部数据库的差别.

2) 定义异类外部数据库的视图合集: 主要通过代理组成的变换模式和由虚类组成的概念模式来完成局部数据源到集成数据库的映射.

3) 此外, 在虚拟数据库 (VDB: Virtual Database) 也有了数据发布的语言化和主动性支持机制, 使用了多种不同机制的描述语言来定义不同的数据源.

3 局部数据库的上推规则的分析与设计

上述语言形式基本上都是图形化且面对用户, 并未从局部数据库的主动性方面进行考虑. 欠缺数据编辑支持 (仅仅是发布支持), 没有数据主自定义事件的机制, 也没有数据改变感知、数据模糊查询、数据智能发布、安全保密控制、并发控制、过程控制等方面的描述. 这些问题是完善的主动式数据发布所需解决的.

定义 1 上推是数据主自行按某种描述方法, 对所需公开的 (供第三方访问的) 数据及其操作进行描述. 其中数据主是指本地数据库中数据的拥有者或者被授权者, 只有数据主才能够上推本地数据库中的数据.

一般的发布是指数据主将数据从自己的数据源中提取出来, 供第三方访问. 上推则不同, 它只是对欲公开供第三方访问的数据进行描述 (非真正提取数据), 具体的数据访问, 由专门的数据访问代理根据相应的“描述”进行. 这种描述基于异源数据上推描述语言, 它是一种形式化的描述语言, 使异源数据呈现为一致的 SQL兼容视图.

定义 2 规则是定义局部数据库上推数据模式的方法.

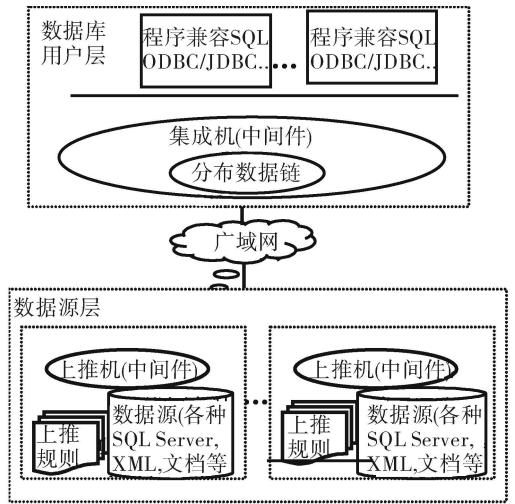


图 1 多数据库中间件架构

可以把模式的信息定义在一个配置文件(例如 XML 文件)中, 作为后继的局部数据库代理和上层集成机进行数据操纵、集成的规范。

3.1 模型的逻辑结构

通过对 XML 和数据库区别与联系的分析, 采用 XML 来描述局部数据库, 所形成的文档以数据为中心。数据库模式到 XML 模式映射使用 DTD, 建立 XML 模式和数据库模式之间的映射时综合采用两种方法的优点: 1) 基于表的映射; 2) 关系对象的映射, 完成从局部数据库到 XML 的映射^[3-5]。

局部数据库代理须与上层集成机服务器进行双向的通信。局部数据库代理向上层上推数据, 另一方面, 集成机向下方局部数据库代理发送请求命令, 要求其下拉数据。如图 2 所示。

3.2 上推规则的实现

局部数据库代理和集成机之间的数据传送和通信是双向的: 局部数据库代理向上要上推数据; 集成机向下要发送请求和命令给局部数据库代理下拉数据。通过以上研究分析, 定义下面两个 XML 模式的 DTD 文件, 作为局部数据库代理和集成机之间的通信规范: rule_response.dtd——上推数据 XML 模式, rule_request.dtd——下拉数据 XML 模式。此处略去模式的具体内容。

举例说明。根据上推数据 XML 模式, 一个局部数据库中的两个表数据封装成如下形式(已经过处理和简化)

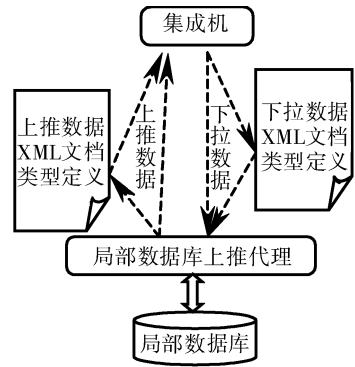


图 2 上推下拉数据示意图

```

< DataObject>
  < ObjectType> DBMetadata< /ObjectType> < ObjectBody>
    < DataBase DataBaseName= "数据库名" >
      < Tables>
        < Table TableName= "表名 1" >
          < Columns>
            < Column>
              < ColumnName> ID< /ColumnName>
              < DataType> IDENTITY < /DataType>
              < MaxLength> 10< /MaxLength>
              < Scale> 0< /Scale>
              < IsNull> false< /IsNull>
              < DefaultValue> < /DefaultValue>
            < /Column>
            < Column>
              第 n 列的数据
            < /Column>
          < /Columns>
        < /Table>
      < /Tables>
    < /DataBase>
  < /ObjectBody>
< /DataObject>
  < Row> 3 songwei 273.80< /Row >

```

```

    < Row> 第 n行的数据 < /Row>
  < /Rows>
< /Table>
< TableTableName= "表名 2">
  表 2的数据
  < /Table>
< /Tables>
< /DataBase>
< /ObjectBody>
< DataObject>

```

根据下拉数据 XML 模式,对 SQL 语句的 XML 描述举例如下

```

< SQLs>
  < SQL>
    < SQLCommand> DELETE FROM DEPT_PRO WHERE ID = ? < /SQLCommand>
    < VALUES Num= "1">
      < VALUE> 127< /VALUE>
      < VALUE> 128< /VALUE>
    < VALUES>
  < /SQL>
< SQL>
SQL 语句 2描述
< SQL>
< /SQLs>

```

4 局部数据库上推机的实现

上推机为中间件形式,负责接收“集成机”(见图 3 中的服务器端)发来的数据访问消息,根据“上推规则”访问数据源,并将结果回送到相应的“集成机”。该中间件也位于数据源所在的 LAN,是数据源信任的软件体(相当于智能代理)。

4.1 整体结构分析

如图 3 上推机包含以下模块: 1)数据库连接模块:负责调用不同的 JDBC Driver 连接不同类型数据库。2)推拉数据模块:负责上推、下拉数据,并接收请求对数据库进行操作。3)XML 转换模块:把数据转换成 XML 形式。4)数据 DOM 解析模块:对不同类型的数据按照定义的格式进行封装。5)通信模块:与多数数据库服务器端通信,包括侦听和发送器。听器接收服务器端发来的请求,发送器发送代理端处理的数据。

服务器端和代理端之间的通信是以封装的数据对象进行 socket 消息传输的,采用了 TCP/IP 的 Socket 和 java 线程技术。在服务器端的 ServerThread java 类能够侦听局部数据库代理端的消息并做出回应,一旦代理端有活动,即分配一私有端口给代理端。这样代理端和服务器端可利用唯一的私有端口生成新的 Socket 进行通信,相应的两端都有侦听器(listener)和发送器(sender)来互相发送接受消息。

4.2 基于 java 类的设计与实现

设计的 java 类分别是 package com. MdbxmlData 数据处理主要类; package com. MdbxmlClient 局部数据库代理端主要类; package com. MdbxmlServer 多数据库服务器端主要类; package com. MdbxmlUtils 公共类。

1) 数据封装类: 数据须进行统一封装处理, 再放入多数据库服务器的数据缓冲区中, 见表 1。

2) 代理端主要类, 见表 2。

3) 服务器端主要类: 服务器端对数据的操作是基本上是代理端的逆过程, 可使用上面两个包中的类。主要的类包括: ServerListener 是多数据库服务器端侦听器; DataTable 是多数据库服务器端发送器。

4) 重要的通用类: 通用类 TypesUtil 作用是对局部数据库的数据进行类型转换, 得到一致的 JDBC 数据类型。例如: CHAR、NCHAR、CHARACTER, 统一转换成 CHAR 类型; DECIMAL、DEC、NUMBER 统一转换成 DECIMAL 类型。

4.3 开发环境

系统环境 Windows2000server, 使用 java 语言, 结合 XML 技术编程。工具包括 Jbuilder8.0 java 开发包 JDK 1.4.1 使用 SQL Server 2000 DB2、Oracle 8.1.6 辅助测试; XML、DTD 和 Schema 编写工具为 XML-SPY 3.0 和 UltraEdit-32

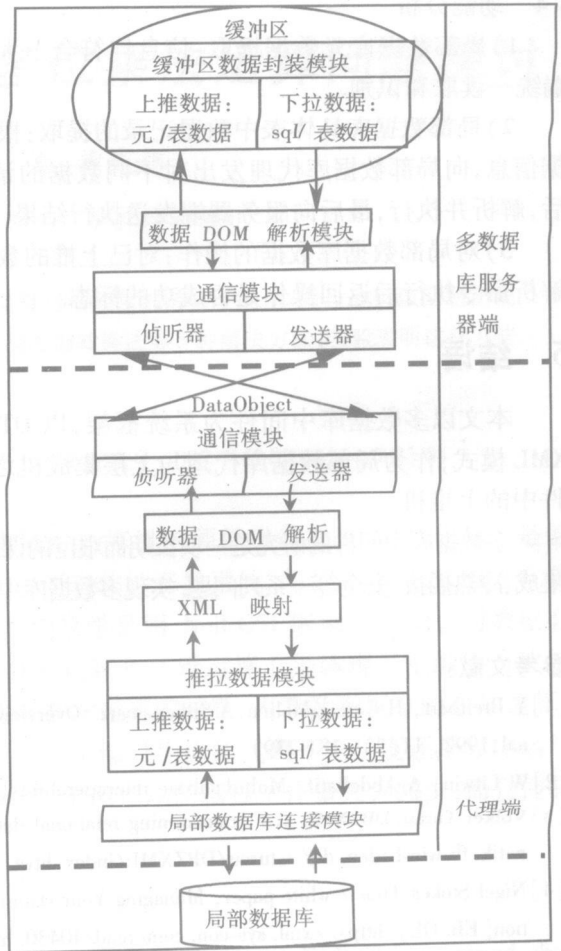


图 3 局部数据库上推代理系统中整体结构

表 1 数据封装类

类名	功能描述
DataColumn	封装一列数据
DataTable	利用 DataColumn 封装一个表的数据
DataSet	包含上推的 DataTable 对数据进行统一的描述
DataAdapter	利用上述几个类对数据进行操作
XmHelper	读取 dataset 中的数据, 形成局部数据库中数据的 XML 描述文档

表 2 代理端主要类

类名	功能描述
ClientListener	代理端侦听器
ClientSender	代理端发送器
LDBinb	处理局部数据库的信息, 包括驱动程序, 用户名, 密码等
SysDb	建立和保持与局部数据库的连接
DBHelper	读取上推的局部数据库数据, 封装到 DataSet
SqIType	对各数据类型进行分类, 形成统一数据类型
XmDoc	各类型数据的 DOM 封装、解析、处理

程序中使用 JDBC 进行数据库联接; XML 文档在生成和传输过程中以 DOM 对象存在, 用 DOM 技术对其进行操作; 一服务器可能连接多个局部数据代理, 使用多线程来进行同步通讯。

4.4 功能分析

1) 局部数据库元数据提取: 信息均符合上面定义的 DTD 格式的 XML 文档形式, 在服务器端统一读取和识别。

2) 局部数据库具体表中数据记录的提取: 根据已上推并存储在服务器端缓冲区的表的元数据信息, 向局部数据库代理发出对不同数据的请求命令。代理接收到以 XML 文档打包的命令后, 解析并执行, 最后向服务器端发送执行结果。

3) 对局部数据库数据的操作: 对已上推的数据进行操作, 包括查询和更新。局部数据库代理解析命令执行后返回操作是否成功的标志。

5 结语

本文以多数据库中间件为系统框架, 以 DTD 的方式定义了描述局部数据库上推数据的 XML 模式, 作为局部数据库代理与上层集成机之间的通信规范, 以此为基础实现多数据库中间件中的上推机。

多数据库中间件的研究是一项长期而艰巨的课题, 在此框架下的局部数据库上推描述涉及到模式集成、冷热插拔、安全等一系列问题, 实现多数据库中的局部上推中间件还需要进行深入研究和完善。

参考文献:

- [1] Y Breitbart, H Garcia-Molina, A Silberschatz. Overview of multidatabase transaction management [J]. VLDB Journal 1992, 11(2): 181-239.
- [2] W Litwin, A Abdelatif. Multidatabase interoperability [J]. IEEE Computer 1986, 19(12): 10-18.
- [3] Volker Turau. DB2XML 1.4 Transforming relational databases into XML documents [EB/OL]. <http://www.informatik.fh-wiesbaden.de/~turau/DB2XML/index.html> 2006-04-17.
- [4] Nigel Stokes. Oracle white paper: Managing Your Data the XML Way: Data Transformation, Exchange and Integration [EB/OL]. <http://xml.sys-con.com/read/40430.htm>, 2006-04-17.
- [5] 罗思群, 钟华, 阮彤. 基于元素树的 XML 文档与关系数据转换 [C] // 全国软件技术研讨会论文集. 大连: 大连出版社, 2001: 253-257.

Design and Realization of the Local Database Push-pull Mechanism and Up-pusher in Multidatabase Middleware

ZHAO Jie¹, WEI Chao², QI De-yu²

(1. School of Economics and Management, Guangdong University of Technology, Guangzhou 510520, China

2. School of Computer Science and Engineering, South China University of Technology, Guangzhou 510640, China)

Abstract This paper analyzes the content extraction and description of the up-pushed data in local database. It defines the XML schema of the up-pushed data description in local database by DTD which is used as the up-push rules and communication standard between the agent of local database and the integrator. Based on it the up-pusher is realized.

Key words multi-database middleware; local database up-push description; XML