

Internet 环境中的数据库信息发布途径*

赵洪彪 张素琴 周立柱

(清华大学计算机科学与技术系 北京 100084)

摘要 当 Internet 成为信息社会最重要的传递手段以后,应用系统的范围和客户端的结构就都多样化了. 与 Internet 的集成成为数据库研究者面临的非常急迫和重要的研究问题. 因为多数已经存放在计算机系统内的数据是驻留并由数据库系统管理的,并且大量即将进入计算机系统的数据也将通过数据库系统管理,目前的数据库系统并不支持与 Internet 的集成,它们将成为未来的遗留系统(Legacy System). 该文介绍了作者利用 Internet 实现 TCPIMS 系统数据库信息发布工作的设计思想,描述了基于 WWW 视图的 Internet 信息结构和基于脚本描述文件的改造工具.

关键词 数据库, Web 客户, WWW 视图, 脚本文件, HTML 文档.

中图法分类号 TP311, TP393

当 Internet 成为信息社会最重要的传递手段以后,数据库系统和 Internet 的集成就成为摆在数据库研究者面前的非常急迫和重要的研究问题. 首先,多数已经存放在计算机系统内的数据是驻留在数据库中的,其次,大量即将进入计算机系统的数据也将通过数据库系统管理,然而,目前的商用数据库管理系统并不支持与 Internet 的集成,随着 Internet 的迅速发展,如果不设法实现目前的数据库系统与 Internet 的集成,那么,建立在传统的数据库管理系统之上的应用系统将成为未来的遗留系统(Legacy System). 本文介绍了作者所完成的,利用 Internet 进行的数据库应用系统信息发布形式的改造工作,并且探讨和分析了几种主要的集成概念和途径,提出了进一步的研究问题.

1 工作背景

我们所承担的国际原子能机构技术合作项目管理系统 TCPIMS^[1](如图 1 所示)是一个从 IBM 主机系统改造过

来的采用客户/服务器结构的信息管理系统. 系统第 1 阶段的开发工作是从主机-终端模式到 Client/Server 模式的降级改造,将大部分程序从主机系统降级到微机平台上,利用 SQL Server 管理数据. 此时,系统的结构是基于 LAN 的一般的 Client/Server 系统. 本系统第 2 阶段的开发工作是利用 Internet 实现数据库信息的发布,取代打印文本邮寄的传统方式,利用浏览器和电子邮件更加经济、快捷地完成信息发布任务.

这种改造对于新一代信息系统的构造具有一般意义. 它本质上是数据库系统与 Internet 的集成问题. 这种集成具有下面一些明显的优点: (1) 利用 Web 浏览器作为客户程序,大大减少开发工作量. (2) 模块性更强的系统结构,通过 Web 客户访问的系统结构更易于扩充和修改. (3) 利用 Web 客户可以使不同应用系统具有公共的用户界面,大大降低用户的培训费用.

为了满足用户的需求,我们需要解决下面的几个问

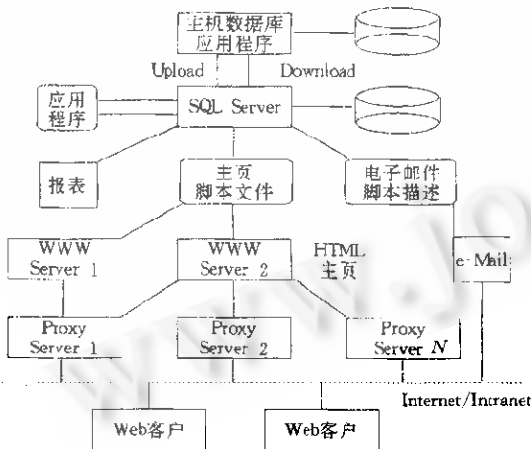


图1 TCPIMS系统结构

* 作者赵洪彪,1964年生,博士后,主要研究领域为数据库系统,软件工程. 张素琴,女,1945年生,副教授,主要研究领域为语言编译系统,面向对象技术. 周立柱,1947年生,教授,博士导师,主要研究领域为数据库,信息系统.

本文通讯联系人:赵洪彪,北京 100084,清华大学计算机科学与技术系

本文 1997-05-26 收到原稿,1997-08 19 收到修改稿

题:(1) 如何扩充 HTML,用以描述信息的结构和内容。(2) 如何从数据库生成所需要的主页和电子邮件。(3) 如何解决电子信息的更新。

2 系统的改造

2.1 主要困难

数据库系统在最初设计时没有考虑 Internet 的需求,因此,Internet 与数据库系统的集成需要采用新的技术手段,而不是简单的接口问题。在 TCPIMS 系统的数据库信息发布方式改造过程中,我们体会到在数据库系统与 Web 客户的集成过程中有这样一些困难:

(1) 有态与无态的矛盾。HTTP 协议是无态的(Stateless),而 Client/Server 系统中,客户和服务端之间的协议是有态的(Stateful)。在有态协议中,消息的含义依赖于前面消息的含义,有态性的优点在于,通过存储状态信息,减少网络上的信息传输以提高效率。而在一个无态协议中,消息的内容不依赖于前面的交互历史,它通常适合简单的查询。

(2) 能否利用 Web 浏览器实现通常的客户功能,包括 GUI、外部工具的调用、对库函数的调用和对设备操作和系统服务的调用。目前,这些难以利用 HTML 来模拟,而需要使用其他技术。

(3) 服务功能分布的问题。系统需要提供 HTML 界面和一些帮助信息,同时支持对数据库的访问,维护不同客户的数据视图,因此,数据库系统与 Internet 的集成需要模型的指导。

我们通过 WWW 视图基本上解决了上述(3)中的问题,利用改造工具解决了上述(2)中的一部分问题,利用维护 WWW 视图的元信息部分地缓解了上述(1)中的矛盾。

2.2 WWW 视图

WWW 服务器的 HTTP 守护进程所创建的子进程只能处理 1 个请求,因此,当用户数目较多时,守护进程会出现饱和,不能迅速创建子进程,提供服务。同时,当请求的数目很大时,许多进程并发执行,竞争处理器资源,服务器的 CPU 能力和传输带宽都是有限的,负荷过重就会影响服务响应时间,因此,根据应用系统中使用群体和任务的不同,将他们关心的信息形成 WWW 视图,放到不同的代理 WWW 服务器上,有助于提高 Internet 方式下信息发布的效率。系统在支持用户要求的主页访问和电子邮件功能的同时,采用代理服务器,实现代理服务器的一个关键技术在于如何实现 WWW 视图(如图 2 所示)。

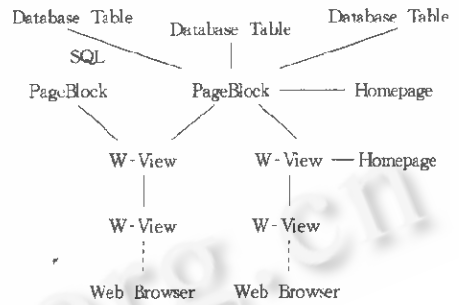


图2 WWW视图

WWW 视图 W-View,由一些数据库表通过 SQL 语句形成 PageBlock,PageBlock 是从数据库中形成 WWW 主页的基本单位,它相当于 HTML 格式的报表,从 PageBlock 通过 SQL 语句形成 WWW 服务器上的 HTML 主页和 WWW 视图 W-View。

2.3 视图更新

系统在处理脚本定义文件时,将数据库表、PageBlock 和 W-View 定义的数据依赖关系记录在系统表 DataRelation(名称、标识、源数据最近更新时间、源数据次近更新时间、服务器标识、上层结点、下层结点)中,在数据库数据更新时,利用触发程序修改源数据最近更新时间,而 PageBlock 和 W-View 的更新则通过视图更新程序,以定时刷新的方式来完成,以保证 HTML 的 Homepage 主页与数据库的数据一致。

2.4 改造工具

数据库信息发布形式的变化需要对应用系统前期的多数交互模块进行改造,为此,我们首先设计和实现了一些改造工具、表格生成工具、电子邮件生成工具、视图生成与维护工具(如图 3 所示)。工具的输入是脚本描述文件,选用不同的脚本描述语句组成文件,可以完成不同的转换功能。脚本语句的语法与文献[2]中所述系统的脚本语法类似,部分脚本语句的描述如下。

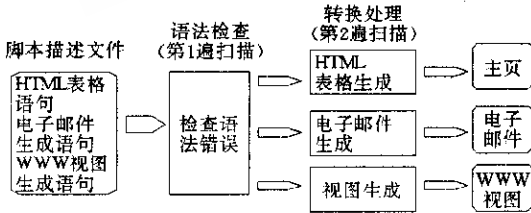


图3 改造工具

脚本描述::=(脚本语句标志)(脚本语句)
脚本语句::=表格生成语句|邮件生成语句|视图生成语句|图象描述语句

表格生成语句::=(表格标题)(数据源描述)|总计定义

表格标题::=正文信息

数据源描述::=SQL 语句

总计定义::=总计属性|...|总计属性

邮件生成语句::=(接收者地址)(信件内容)|(附加信息)

视图生成语句::=VIEW 语句

图象描述语句::=图象文件指针|图象信息说明

HTML 生成工具的功能是,将应用系统所生成的报表转换成 HTML 形式,放到 WWW 服务器上,供用户浏览。HTML 生成工具支持标题语句、格式语句、查询语句和聚合语句等主页描述语句。HTML 生成工具使用两遍扫描法,第 1 遍检查语法错误,第 2 遍处理其中的脚本语句,生成相应的 HTML 主页,并将其安装到 WWW 服务器上。

电子邮件生成工具 电子邮件生成工具根据相应脚本的描述语句确定电子信件的内容、发出者和接收者,支持通过 word 工具进行浏览和修改,电子邮件生成工具可以处理地址语句、聚合语句、查询语句和结尾语句等电子邮件描述语句,然后自动生成相应的电子邮件发出。

WWW 视图生成工具 本工具能够处理视图定义语句,描述和安装在不同的代理服务服务器上的视图。WWW 视图的源数据可以来自不同的 WWW 服务器的 PageBlock,或者直接来自数据库表,或者来自以前定义的 WWW 视图。

3 结 语

数据库与 Internet 系统的集成只有在满足这样的一些限制的前提下才是有意义的:(1) 保证可用性。新的基于 Internet 的 Client/Server 系统必须具备原来的 Client/Server 数据库系统的多数功能,否则,这种集成是无意义的。(2) 最小改变。增加基于 Internet 的使用方式不应造成对原代码的大量修改,原有的客户和服务程序应该能够正常使用。(3) 不修改现有浏览器的使用方式。

本文介绍了我们利用 WWW 视图,在 Internet 上的信息组织结构和基于脚本描述的改造工具方面的工作,讨论了数据库系统与 Internet 集成的有关问题和主要途径。Internet 的普及无疑为数据库应用系统提供了新的广阔应用前景。

参考文献

- 1 TCPIMS Development Group. System Specification for the TC Project Information Management System. Tsinghua Development Team, Vienna, May 1996
- 2 周立柱等. 面向市场分析预测的多媒体智能数据库系统. 软件学报, 1996, 7(增刊), 462~468
(Zhou Li-zhu et al. A multimedia intelligent database system for market analysis and forecast. Journal of Software, 1996, 7 (supplement); 462~468)

Approaches of Database Information Dissemination on Internet

ZHAO Hong-biao ZHANG Su-qin ZHOU Li-zhu

(Department of Computer Science and Technology Tsinghua University Beijing 100084)

Abstract When Internet is becoming the popular highway for information exchange, the domain of applications and the types of client software are diversifying. Thus the integration of database with Internet is becoming an imminent research issue as most data in computer are managed by DBMS and the amount of data managed by DBMS is huge. But the modern DBMSs do not support connection to Internet. And the applications based on those DBMSs will become legacy system in the future. In this paper, the authors introduce the design of database information dissemination on Internet in TCPIMS project and describe the information model based on WWW-View and the rebuilding tools based on script input file.

Key words Database, web client, WWW-view, script file, HTML document.