

基于 EJB 和 XQuery 的异构数据集成技术研究^①

Research heterogeneous data Integration base on EJB and XQuery

杨 云 赵正文 孙华林 (西南石油大学 计算机科学学院 四川 成都 610500)

摘 要:企业信息化的逐步深入,企业内部及企业之间存在着大量的异构信息的交互和共享,Java 与可扩展的标记语言(XML)技术为实现网络环境下异构信息的交互提供了解决方案。提出应用 Java 的 EJB 和 XML 的查询语言 XQuery 技术建立适用于信息化的面向异构数据集成的框架结构,并对所涉及到的关键技术和实现方法进行了研究。

关键词:数据集成 异构数据 Java EJB xml XQuery

1 引言

企业信息化的逐步深入,在企业内部以及企业间必然存在着大量异构信息的交互和共享问题。对于一个具有多种应用系统的企业来说,各系统中数据的交换与信息的集成是至关重要的,只有实现了各种信息系统的集成,才能真正实现企业的信息化。数据信息系统的交换与集成,是一个企业信息化的基础,没有对企业现有的各种应用信息系统的集成,企业中的各种数据信息系统,只能是各种自动化和信息化的“孤岛”,不能共享各系统的数据,不能对企业的整体业务流程进行同意的管理,特别是对于设计制造业来说,更是如此。同时企业必然要与外部企业进行数据的交换和共享,企业之间数据的交换和共享也需要一种工具来帮助不同企业间的数据交换,对于企业之间和分布式企业的数据交换和共享,国际标准组织 W3C 推出了可扩展的标记语言(extensible markup language, xml),它是一种精简的 SGML(Standard generalized markup language)语言,这种描述 Web 页面的内容和直接 Web 数据的通用方法,用于实现网络上信息的共享和交互,基于 XML 的信息集成,已成为国内外学者研究的热点。

2 EJB 和 XQuery 技术

EJB(Enterprise Java Bean)是 J2EE 平台服务器端

组件模型,可用于创建分层次、跨平台的分布式应用,提供了对事务、安全以及连续等的支持。与 OL E/ COM 的 .NET 集成平台只限于 Windows 操作系统相比,J2EE 是一个标准化和开放性的平台,能够在不同操作系统间移植,并且提供了系统集成所需的全部技术和中间件解决方案。

XQuery 是一种灵活的查询语言,还在不断的修订和完善。作为一种新型的查询语言,XQuery 汲取了多种查询语言的优点,适用于各种类型的 XML 数据源的查询,功能强大,简洁灵活且易于实现。而且,XQuery 还具有从多种数据库中检索信息的特点,能对各种数据和文档进行查询。XQuery 构建在 XPath 规范之上,其最大的特征就是 FLWR 表达式。通过它所支持的各种表达式,它的查询可以有各种不同的形式。各种 XQuery 表达式可以完全嵌套,也支持子查询。目前,Oracle、IBM、Microsoft 等都已经各自的产品中提供了对 XQuery 规范的支持。

在数据集成中,一个关键的问题就是屏蔽不同数据源在结构上存在的差异。采用 XQuery 将关系数据库中的数据转换为 XML 数据就非常方便地实现了数据源异构统一化。

3 EJB 和 XQuery 异构应用系统集成框架

异构应用系统是指在不同的硬件的体系结构和操

^① 本课题为西南石油大学研究生院第五批创新基金资助课题(CXJJ27082)

作系统平台 (WINDOWS, UNIX, LINUX) 以及网络数据库系统上运行的各种信息系统与辅助工具 (如 CAD/CAM/CAPP 等)。信息系统集成的含义是指将基于信息技术的资源及应用 (计算机硬件, 接口及机器) 聚集成一个协同工作的整体, 继承包含功能交互, 信息共享以及数据通信三个方面的管理与控制。从信息系统继承的角度来看, 在应用系统之间的数据交换与集成有三中基本的方式, 即应用封装, 接口集成和紧密集成。

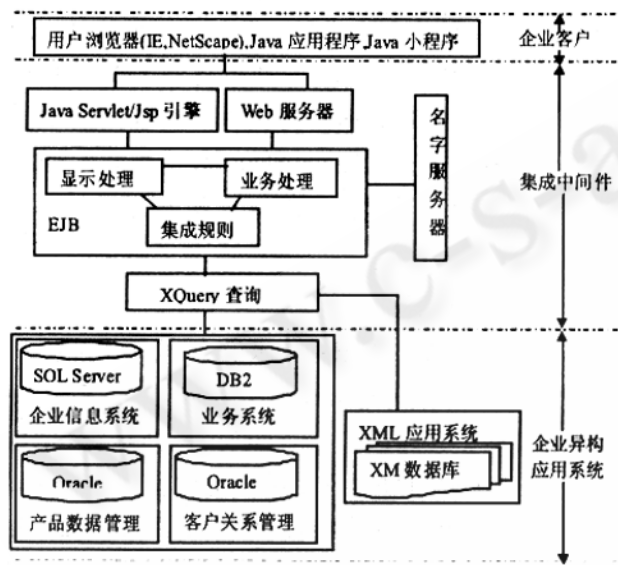


图 1 企业异构系统集成框架

集成平台与集成框架技术是 20 世纪 80 年代末 90 年代出, 随着企业信息集成与系统集成复杂度的提高, 采用常规的集成方法已难以适应集成的要求, 于是作为先进的应用集成工具, 集成平台与集成框架技术应运而生。目前企业应用集成有几种不同的技术规范, 大都采用面向对象和分布式技术, 如分布式计算环境, 面向对象分布式处理, COBAR 规范和 OLE/COM, .NET 技术和 J2EE 技术规范以及产品交换标准 STEP 和数据交换标准 XML 等。

本文提出的异构应用系统集成框架采用的是面向对象的 EJB 和 XQuery 技术。其集成框架结构如图 1 所示。整个集成框架采用三层体系结构, 从上到下分别是: 企业用户, 企业集成中间件和企业异构应用系统。

(1) 企业拥护是集成框架的顶层, 该层是企业异

构应用系统数据对企业用户的表示。企业用户提出请求, 通过中间层传递到底层的异构应用系统, 数据的返回也借助中间层集成中间件将数据返回给用户。在企业客户层的应用工具是各种浏览器及 JAVA 应用程序和 JAVA Applet 小程序。

(2) 集成中间件是集成中的中间层。它是该集成框架的核心, 它是由, Java Servlet/Jsp 引擎, web 服务器, EJB, 名字目录服务器, XQuery 等组成。他们各自的功能分别是: (1) Web 服务器接收浏览器的请求, 同时将请求转发给, Java Servlet/Jsp 引擎, Java Servlet/Jsp 引擎处理, Java Servlet/Jsp 程序; (2) 这些程序从 EJB 中获取数据, 同时将请求与响应数据文件的名字存入目录服务器, 以便检索。(3) 名字目录服务器同时管理企业的拥护和客户及企业的各种资源。XQuery 负责将各种异构应用系统的数据转换成可以交互的格式, 即 XML 格式的数据。这样有关数据交互的应用程序开发回变得非常简单。

EJB 和 XQuery 是该层的核心, EJB 处理业务规则和表示规则, 实现在异构应用系统间的交互和共享工作。同时, XQuery 技术使数据可以有多种多样的表现形式。在分布式环境下个体的情况千差万别, 数据的记录, 存储方式也是各有千秋, 为了更好地进行数据的共享, 本文提出了用 XQuery 技术把数据表示成与系统, 与应用无关的形式。

(3) 企业异构应用系统。该层是各种企业异构应用系统, 其存储采用不同的存储结构。这些应用系统的存储单元很可能是异构的, 如图 1 所示, 它们有 MS SQL SERVER, ORACLE, DB2 等主流的商业数据库系统, 还有 XML 数据库, 由这些异构的数据管理系统支撑应用系统。

本文提出的这个新的异构应用系统数据集成框架, 采用的 EJB 和 XQuery 技术, 把 EJB 和 XQuery 技术结合起来可以创建具有高度平台独立性的分布式数据交换和处理系统。

4 框架结构的实现关键技术

上面提出的异构应用系统数据集成框架的实现过程主要有三个方面的关键技术需要解决, 即基于关系数据库的 XQuery 查询, 基于异构数据源 XQuery 集成规则, 基于 EJB 和 XQuery 数据集成的实现。

4.1 基于关系数据库的 XQuery 查询

XQuery 是一种功能强大的数据查询语言,它能够从 XML 文档中选择并抽取复杂的模式,进而把查询结果重构成用户需要的新的 XML 文件结构。它极有可能成为“XML 中的 SQL”,因为 SQL 语言是对建立在关系代数基础上的数据表进行操作的,而 XQuery

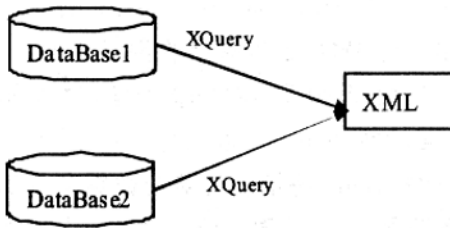


图 2 使用 XQuery 查询异构数据源

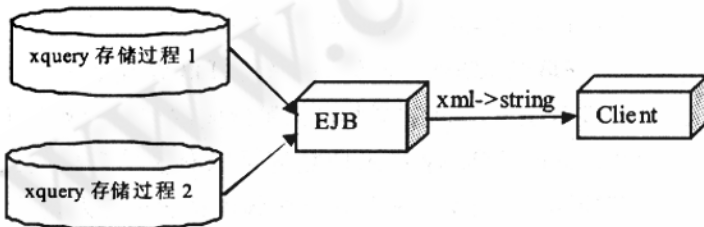


图 3 EJB 调用异构数据源存储过程

则是对建立在 XQuery 数据模型上的 XML 数据进行查询的语言,并且它还是一种书写格式自由的语言。目前在 10g R2 版本中第一个在商业上包含了对 W3C XML Query 标准的支持。XQuery 还未最终完成,然而,Oracle 最大的数据库产品竞争对手,IBM 和微软,都在自己的数据库产品中都提供了 XML 支持。在不久的将来这些主流数据库都将包含对 W3C XML Query 标准的支持,基于关系数据库的 XQuery 查询将会同 SQL 查询一样普遍应用于关系数据库的操作中。

4.2 基于异构数据源 XQuery 集成规则

XQuery 能够从 XML 文档中选择并抽取复杂的模式,进而把查询结果重构成用户需要的新的 XML 文件结构。不同结构的数据源经 XQuery 语言查询都可以转换为 XML 格式数据,而 XML 具有自描述性,跨平台性等特性。只需要在此对返回的 XML 数据进行操作即可。我们描述了一个使用 XQuery 查询异构数据源并生成 XML 数据,如图 2。

4.3 基于 EJB 和 XQuery 数据集成的实现

为了在 EJB 中能对 XQuery 查询的数据结果更好的进行处理,我们选择了以存储过程的方式来进行操作,主要有两个方面的原因:第一,直接在 EJB 中使用 XQuery 查询,返回的 xmltype 数据类型在 EJB 中不便于直接进行操作;第二,在 EJB 的 find 方法中需要带相应的返回参数,直接使用 XQuery 查询使得 find 方法难以实现。因此在这里,主要操作通过 EJBFindByPrimaryKey(String docID), EJBFindByXML(String docID) 两个 EJB 的 find 方法来实现,EJBFindByPrimaryKey() 是实体 bean 中必须实现的方法,EJBFindByXML(String docID) 方法是自定义的 find 方法,它通过返回 EJBFindByPrimaryKey() 的实例对数据进行操作。EJBFindByXM() 调用存储过程,并将返回的 xmltype 数据转换为 string 型数据。EJB 调用异构数据源存储过程如图 3。

5 结束语

对于企业的信息化来说,异构应用系统之间数据的共享和交互是十分常见的。EJB(Java)的平台无关,“一次编写,到处运行”的特点;XQuery(xml)文档的结构化和自解释性以及与应用无关的特性,它们是“基于 EJB 和 XQuery 的异构数据集成”的基础。本文对实现该集成框架的关键技术进行了分析,说明此集成框架的有效性。

参考文献

- 1 D. Chamberlin: XQuery: An XML query language, 2002. <http://www.research.ibm.com>.
- 2 UsingXMLinOracleDatabaseApplicationsExchanging BusinessDataAmongApplications, <http://www.oracle.com/technology/tech/xml>, 1999.
- 3 Matsuoka S, Yonezawa A. Analysis of inheritance anomaly in object - oriented concurrent programming languages. In: Agha G.
- 4 孙会峰、许德刚,基于 J2EE 组件的数据集成平台设计,电脑开发与应用,郑州,河南工业大学,2006.
- 5 李军怀、周明全、耿国华、张景西,XML 在异构数据集成中的应用研究,计算机应用,北京大学,2002.