

基于 CORBA 和 Web Services 的异构数据源集成研究^①

Research of Inhomogeneous Data Source Integration Based on CORBA and Web Services

刘高嵩 龙 军 (中南大学 信息科学与工程学院 湖南 长沙 410075)

摘要: 为解决信息孤岛问题, 提出一个基于 CORBA 和 Web Services 的异构数据源集成的解决方案。讨论了基于 CORBA 和 Web Services 封装数据源, 详细说明了异构数据源集成框架, 并介绍了该解决方案的一个应用。

关键词: 异构数据源 数据集成 CORBA Web Services

随着计算机网络的普及, 解决信息孤岛、实现数据资源的共享已经成为一个热门话题。传统的数据集成技术, 已无法适应人们获取更多更新数据的需要^[1]。人们要求数据集成系统不仅能集成数据库系统中的数据, 而且能集成非数据库系统中的数据; 不仅能集成同构数据, 而且能集成异构数据; 不仅能集成已有数据源中的数据, 而且能集成随时加入的新数据源中的数据, 即实现数据源的“即插即用”。

分布在网络上数据源具有分布性和异构性^[2], 所谓分布性是指各数据源分布于不同的物理位置, 依靠网络传输数据; 所谓异构性主要表现为不同的数据表达形式、不同的数据语义和数据源不同的使用环境。本文提出的异构数据源集成方案是: 基于分布式对象/ Web Services 技术, 将数据源包装为分布式数据构件, 将分布式数据构件的操作封装成 Web 服务, 从而屏蔽数据源的异构性, 实现分布在 Internet 上的异构数据源的逻辑集成。其主要特点是:

(1) 在异构数据源的集成环境中, 用户每次的数据访问, 都是借助 Web 服务, 最终通过分布式数据构件, 从数据源直接获取数据, 这样, 用户所得到的永远是最新的信息。

(2) 每当数据源的数据和结构变更时, 只需更新

分布式数据构件的相关操作, 而访问数据源的应用系统不需要做任何改动, 从而保证了数据源的自治性, 实现了系统间的低耦合。

(3) 数据源实现了“即插即用”。成员数据源通过对分布式数据构件实现了数据异构和物理分布这两方面的透明化, 新的数据源只要被封装, 便能随时集成。

(4) 基于 Web Services 技术, 突破了分布对象技术在实现中所受到的防火墙约束, 为分布在 Internet 环境中数据源的集成提供了有力的支持。

1 异构数据源的封装

基于分布式构件的集成和基于 WEB 服务的数据集成是当前常用的数据源集成的技术。基于分布式构件的集成, 基于分布式构件模型, 如 CORBA、DCOM, 构建异构数据源的集成系统, 这类系统的特点是执行效率高, 并有分布计算规范提供的丰富的底层服务支持, 如 CORBA 规范提供了包括命名服务、交易服务、事务服务、安全服务和生命周期服务等底层服务。其局限性是: ①服务客户端和服务提供者之间必须是就紧耦合, 即要求为同类的基本结构, 如 CORBA 分布式构件模型要求客户端和服务端运行同一 ORB 产品; ②受到防火墙的约束, 限制了数据源集成系统向 Inte-

① 基金项目: 国家自然科学基金(60873081); 高等学校博士学科点专项科研基金(20060533084)
收稿时间: 2008-11-20

rnet 的扩展。

基于 WEB 服务的数据集成是在 Web Services 的框架下,使用一组 Web Services 协议,构建异构数据源的集成系统。对每个数据源构建其 Web 服务,然后通过 WSDL 定义 Web Services 的服务描述向服务中心注册,数据服务请求者使用查找操作,在服务中心检索服务描述,然后使用服务描述与服务提供者进行绑定,并调用 Web Services。此类系统的主要优势为:① Web Services 通过 XML 数据表示和 SOAP 消息交换,使得 Web 客户和服务器实现了完全的松散耦合,解决了如 DCOM 要求所有应用节点必须是 Windows 系统,CORBA 则要求整个分布式应用必须使用同一种 ORB 产品的问题;② 网络协议基于 HTTP、SMTP 等,可穿越防火墙。其不足之处是当连接逻辑和数据位于单个平台上时(传统的紧耦合的分布式体系结构的应用领域,如 CORBA、DCOM 和 EJB),Web Services 的执行效率很低^[3]。

由此可见,基于分布式构件集成和基于 WEB 服务的数据集成在使用上具有一致性,在技术上具有互补性,将两种技术有机结合,实现异构数据源的集成,可以充分利用两者的技术优势。因此,我们封装数据源的策略是,首先将数据源包装成分布式对象,然后将分布式对象操作封装为 Web 服务,其结果既可利用分布式对象技术的性能优势,又能充分发挥 Web Services 技术的互操作性。

1.1 封装数据源的分布式数据构件

CORBA、COM/DCOM 以及 EJB 是当前最具有代表性的分布式构件模型,而 CORBA 作为分布式对象的主要标准,自从 1991 年 OMG 组织推出其第一个版本以来,规范逐渐成熟。因为其独立于网络协议、编程语言和硬件平台的特点,CORBA 被广泛地运用于通信、电子商务、航天航空等诸多应用领域中,逐渐成为分布式应用最具生命力的技术平台^[4]。与其它模型比较,CORBA 实现了平台无关性,而 COM/DCOM 基于微软的开发平台;CORBA 将程序设计语言无关性作为的重要设计原则,而 EJB 依赖于 JAVA 语言^[5]。因此,为完全屏蔽数据源的异构性,基于 CORBA 模型,将数据源包装为分布式对象,并将其称为分布式数据构件。

分布式数据构件模型可用三元组{OID_{ptr}, A_{str}, O_{set}}描述。其中,OID_{ptr}:对象 ID,以序列方式获得

全局唯一 ID; A_{str}:对象别名,与数据源名相同; O_{set}:操作集,即分布式数据对象的接口。

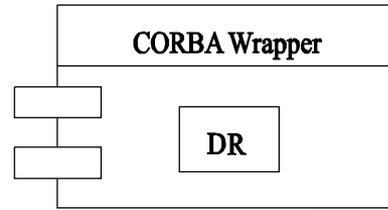


图 1 分布式数据服务构件结构图

图 1 描述了作为数据源包装器的分布式数据构件的结构,CORBA 包装完成从数据访问到 CORBA 对象的映射,即将数据源的访问操作作为 CORBA 对象的接口,对外发布。对于同类型的数据源,根据各自的初始化参数,生成不同的 CORBA 对象实例。

1.2 基于 Web Services 的封装

CORBA 作为一个有代表性分布式计算的技术的平台,由于具有强耦合的特点,并且在实现中受到防火墙的约束,使得它们更适合于在局域网的范围内进行应用系统的开发^[6]。Web Services 基于 Web 计算模式,提出了新的、面向服务的体系结构,提供便于在互联网范围内进行服务注册、发现和调用的手段,在互联网上方便地进行发布、发现和互操作。为更好地支持 Internet 环境下的异构数据源集成,我们将分布式数据对象的操作封装成 Web 服务,利用 Web Service 的互操作,采用 SOAP 通信,实现 Internet 环境下的异构数据源的交互。

将 CORBA 对象部署为 Web 服务,在定义服务时,需要将 CORBA 对象的接口描述转换为 Web Services 的接口描述,即将 CORBA 对象的 IDL 文件映射成为 Web 服务对应的 WSDL 文件;而当运行服务时,同时由于 Web 服务客户端不能直接访问基于 CORBA 的服务实现端,需要一个桥接系统,将 SOAP 请求中的参数转换成对 CORBA 对象调用需要的参数类型,定位和调用 CORBA 对象,并将响应结果封装成 SOAP 消息返回给客户端。

2 异构数据源集成框架

如何利用 CORBA 和 Web Services 体系提供的服务,将封装的数据源集成为一个全局的数据服务体系,实现应用系统对每个分布的异构数据源的透明访问,

这是异构数据源集成需要解决的问题。

出于应用需要和实现复杂性的考虑，我们把异构数据源的集成框架分成两个层次：Intranet 环境下的集成和 Internet 环境下的集成。

2.1 Intranet 环境下的数据源集成框架

封装异构数据源的分布式数据构件，是具备的 CORBA 接口分布式对象，在企业内部的局域网中，数据服务请求者与数据服务提供者之间，基于 CORBA 对象接口和 ORB 软件总线，完成数据交互，以充分利用 CORBA 的性能优势；另一方面，局域网内包装数据源的 CORBA 对象，封装成 Web 服务，将服务描述信息以 WSDL 文档形式在服务注册代理注册，并对外发布。局域网外的数据服务请求者，可通过 UDDI 查找服务注册代理中的服务描述，调用 web 服务完成数据访问。因此，在企业局域网内部，Intranet 集成框架应支持的基于 CORBA 体系结构的数据交互；而对外，Intranet 集成框架接受远程的 SOAP 消息，最后通过相应的 CORBA 服务，完成数据服务请求。

图 2 描述了 Intranet 环境下集成框架的层次结构。其中，框架底层支持企业内部基于软件总线 ORB 的数据源的交互；框架顶层，即数据请求处理层，接收来自 Internet 范围内远程的数据服务请求，并返回 SOAP 应答；桥接系统 CORBA Provider 执行 SOAP 域到 IIOP 域的请求翻译，负责将 SOAP 请求分发给相应的 CORBA 对象，并将 IIOP 应答转换为 SOAP 应答。另外，由于分布式数据构件只具有处理数据源内部事务的能力，对于跨数据源的全局事务，所以需要设立事务控制器，以支持数据源间的事务处理。

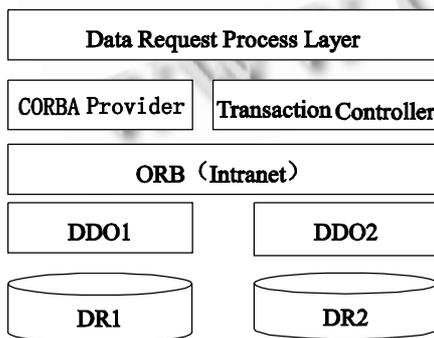


图 2 Intranet 环境下的集成框架

2.2 Internet 环境下的数据源集成框架

Internet 环境下异构数据源的集成，以 Intranet

集成框架作为数据节点，构建 Internet 范围内的统一框架，以实现因特网上异构数据源的分布访问。Internet 集成框架由三个部分组成：数据访问客户端、数据请求处理层和 Intranet 集成框架层，其拓扑结构如图 3 所示。数据访问客户端通过数据请求处理层提出数据请求，因此，数据请求处理层实质上充当服务请求者的角色，而各个 Intranet 数据集成框架层充当服务提供者的角色。

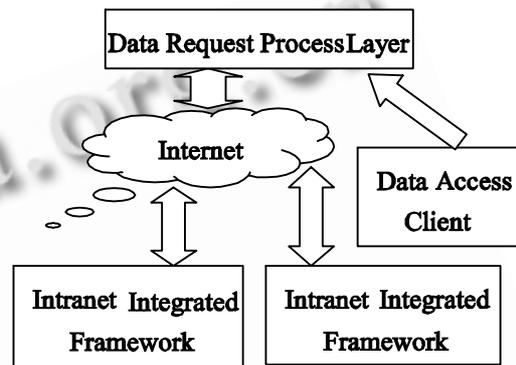


图 3 Internet 集成框拓扑结构

(1) 数据访问客户端

Web 应用系统中的业务逻辑层充当数据访问客户端角色。Web 服务器(IIS、Resin、Apache 等)接收来自浏览器的数据请求，将请求转发给业务逻辑层(Servlet、CS 类等)，业务逻辑层在 Internet 上查找全局数据服务代理，请求执行相应的数据操作。

(2) 数据请求处理层

一方面，作为客户端数据请求的统一处理接口，数据请求处理层接受客户端的数据请求，并检查其数据操作权限，数据请求处理因涉及到从整体到局部的映射[7]，所以，需将客户端的全局数据请求分发给相应的 Intranet 集成层的 web 服务，同时收集各个 web 服务的执行结果，组装成最终的结果返回给客户端，另一方面，数据请求处理层作为 web 服务的管理者，实现对 Web 服务的注册管理、服务发布和服务发现等功能。

数据请求处理层是 Internet 集成框架中的核心部分，其处理能力是决定 Internet 集成框架性能的关键。其对数据请求的处理过程如下：

① 数据请求处理层接收到数据访问请求后，进行权限检查，确定客户是否有操作特定数据的权限，若

无权限,则终止处理过程。

②分解数据请求,将全局数据请求转化为一个或多个局部数据请求,并将局部数据请求加入相应的数据请求队列。

③执行相应 Intranet 集成层中 web 服务,最终通过分布式数据构件完成局部数据请求,并把执行结果返回给数据请求处理层。

④收集并组装所有回局部数据请求的执行结果,返回给数据请求客户端。

(3) Intranet 集成层

Intranet 集成层是 Internet 集成框架的数据节点,为 Internet 集成框架中数据请求处理层提供数据服务。它集成局域网范围内的多个分布式数据源为一个整体,以数据元的身份参与 Internet 集成框架这一更高层次的集成。

3 应用

专家信息共享平台是 Internet 环境下异构数据源集成的典型应用,其目标是实现全国范围内科技项目评审专家信息的集成和共享。

在科技项目的评审过程中,专家是评审的主体,也是实现评审客观、公正的关键。目前全国评审专家信息资源重复建设、各自独立,各省市都拥有本省的评审专家信息库,而国家科技奖励工作办公室和各部委分别拥有全国范围内和本行业内的评审专家信息库。此外,不少省市的科技管理部门内部也有多个的专家信息库,如科技厅下设的奖励办、基金办、计划等处处室都有各自独立的专家信息库。专家信息共享平台在集成省市科技管理部门内部各专家信息库的基础上,实现全国范围内的各省市、部委的专家信息共享。

专家信息共享平台由门户网站、管理服务器、数据节点等几部分组成。

门户网站 门户网站是专家信息共享平台对外的窗口,是客户端信息请求服务的接入点。门户网站建

立在国家科技奖励工作办公室服务器集群上。

管理服务器 管理服务器是 Internet 数据源集成的控制中心,实现 Web 服务管理、权限管理、任务调度分配和数据操作结果的合成。利用中心数据库对频繁检索的数据进行缓存,并负责各数据节点间专家信息数据的一致性和完整性。

数据节点 数据节点是专家信息数据源,是各省市、部委所拥有的所有专家信息库的集成,数据节点中的各数据源封装成 CORBA 对象,并将 CORBA 对象的数据操作封装为 Web 服务,在管理服务器上注册,并对外发布。

用户通过门户网站提交数据访问请求,由管理服务器验证用户身份和权限,进行任务调度,执行相应 Intranet 集成层中 web 服务,通过访问数据节点,完成局部数据请求,最后由管理服务器合成操作结果,返回给用户。

参考文献

- 1 王宁,陈滢,俞本权,等.一个基于 CORBA 的异构数据源集成系统的设计.软件学报,1998,9(5):378-382.
- 2 Hasselbring W. Information System Integration. Communications of the ACM, 2000,43(6):33-38.
- 3 张驰,吴健,胡正国,等. CORBA 与 Web 服务的比较与集成. 计算机工程与设计, 2005,26(8):2213-2215.
- 4 Inoski V, Corba S. Integrating Diverse Applications Within Distributed Heterogeneous Environments. IEEE Communications Magazine, 1997,35(2).
- 5 郑红,李师贤.可重用的分布式软件构件模型分析. 计算机工程与应用, 2002,38(15):68-71.
- 6 谷清范,吴介一,张飒兵.基于 CORBA/Web Service 的 PDM 图档管理系统. 计算机工程, 2006,32(16):23-25.
- 7 王宁,王能斌.异构数据源集成系统查询分解和优化的实现. 软件学报, 2000,11(2):222-228.