

企业数据网格平台模型的研究与实现^①

The Study of Enterprise Data Grid Platform

侯勇 于炯 吐尔根 (乌鲁木齐新疆大学信息科学与工程学院 830046)

李 瑗 (石家庄学院计算机科学与技术系 050035)

摘要:数据网格作为网格客体分类中的最基础层,将企业内部大量业务不能互联互通、资源不能共享、数据不能联动的异构应用系统整和在一起。通过利用基于 OGSA 体系结构的最新网格开发工具 GT4 构建一个企业数据网格平台(Enterprise Data Grid Platform, EDGP)来集成现有应用系统,将各应用系统的数据封装为服务,实现不同应用系统数据库、数据中心之间数据交互和联动。该文首先论述了 EDGP 的结构及部署机制,接着对服务动态集成机制、系统安全机制、非标准数据与标准数据转换机制,及数据联动机制等关键技术进行了研究。

关键词:数据网格 OGSA 体系结构 网格服务

1 引言

传统企业内部存在着大量基于数据库技术的应用系统。企业各部门在业务上相互关联,但运行于各部门的应用系统是使用不同工具开发、运行于不同的运行平台、数据格式也各不一致,导致部门之间数据不能共享,信息不能联动,在互联网上形成了一个信息孤岛和资源孤岛,最终使得企业无法高效办公,实时做出决策。若废弃现有应用系统,在统一环境中根据企业需求重新开发应用系统,则将耗费大量人力、物力,在短期内也无法完成,同时也造成了巨大浪费。

针对企业面临的上述问题,充分利用网格技术,实现互连网上所有资源(计算资源、存储资源、通信资源、软件资源、数据资源、信息资源、知识资源等)的全面共享与协同工作,将企业内部网整合成一台巨大的超级计算机,为用户提供“即连即用”式的服务^[1],以成为企业的迫切需要。文章介绍了一个面向企业的数据网格平台,即企业数据网格平台(Enterprise Data Grid Platform, EDGP)。从技术的角度来说,此平台具有友好性和开放性,把复杂的应用程序移植到服务器端,使得“胖”客户端转换为“瘦”客户端,方便了系统的使用、升级和维护。从企业需求和企业信息化管理角度来说,EDGP 采用 B/S 模式,方便实现企业不同部门的技术人员、管理人员及时了解当前企业实时信息,保

证各部门间高效协同工作。

本文提出了 EDGP 的部署机制和系统结构,并对实现 EDGP 的核心技术,如服务动态集成机制、系统安全机制、非标准数据与标准数据转换机制,及数据联动机制等关键部分作了详细描述。

2 EDGP 部署机制

Web service^[2,3]是一种分布式计算技术,一个可以被 URI(Uniform Resource Identifier)识别的软件应用。Web service 的实现类似一个远程调用 RPC。服务方首先提供一个 Web service 应用,接着服务使用方可以在网络的任何位置调用这个 Web service,好像调用本地的 API 函数。所以 Web service 可以在不关心服务功能如何实现的情况下把不同平台开发的不同类型的功能块集成在一起,提供相互之间的互操作。基于上述原因,Globus 联盟和 IBM 迅速推出了基于该协议规范的 Web service 资源框架 WSRF (Web service Resource Framework),正式将 Web service 列为基于 OGSA (Open Grid Service Architecture)体系结构的新一代网格服务的核心,并推出 GT4 (Globus Toolkit 4),将原有的 GridService 演变成 Web service 和资源文档两部分,实现了网格和 Web service 的融合,加快了科研界和

^① 基金项目:国家自然科学基金(编号:60563002);新疆自治区高校科研重点项目(XJEDU2004I03)

工业界的接轨^[4]。

GT4^[5]是 Globus 联盟推出的最新遵循 WSRF 规范的网格开发工具, EDGP 实现利用 GT4 搭建一个数据网格平台,在该平台上建立一个数据中心,将各应用系统中需要交互的数据提取出来,统一格式存放于数

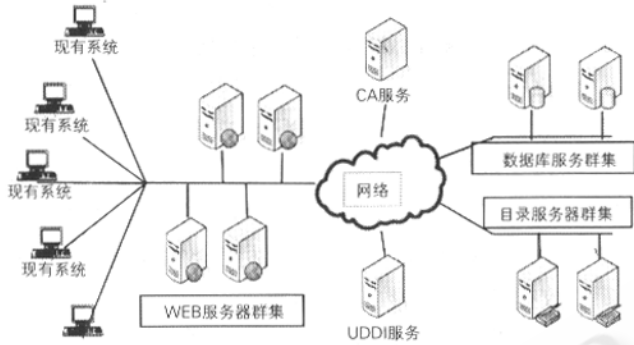


图 1 EDGP 部署机制图

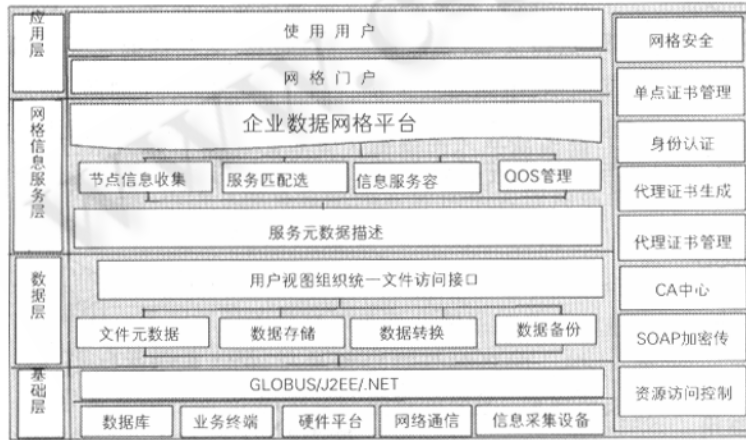


图 2 EDGP 体系结构图

据中心,作为应用系统的数据备份。在现有应用系统的服务器上部署 Web service,为集成现有系统和共享现有系统中的数据资源提供接口。在 Web 服务器上部署各种 Web service,为在 EDGP 中实现数据联动以及提供 EDGP 用户简单易用的数据网格接口。通过目录服务器(是数据中心各种数据资源的索引)查找所需访问数据资源的位置。使用 CA 服务器实现服务访问的安全管理,包括认证和授权。使用 UDDI 服务器实现 Web service 服务的动态集成。EDGP 部署机制如图 1 所示。

3 EDGP 系统结构

参照基于 OGSA 的网格体系结构^[6], EDGP 将集成的系统结构分为基础层、数据层、网格信息服务层和应用层。下层为上层提供服务,上层调用下层提供的服务,同时实现底层数据之间的联动,EDGP 体系结构如图 2 所示。

(1) 基础层包括各种异构的数据库资源、硬件平台、信息采集设备、网络通信设备、计算机等各种软硬件资源;基于数据网格平台的运行环境包括 Globus 平台、J2EE 平台和 .NET 平台等。

(2) 数据层是 EDGP 的核心层,作为上层应用的数据基础。数据层主要功能是从基础层抽取信息元数据,并将元数据转换为统一的标准格式,接着将数据层平台提供的数据进行聚合封装为服务,最后将这些服务进行收集并对外发布。

(3) 网格信息服务层由企业数据网格平台和一些网格应用组成。网格信息服务层主要功能是为网格应用和运行提供一个全局的、抽象的资源视图,使 EDGP 用户能够透明的访问企业网格环境上的服务资源。

(4) 网格应用由网络门户和使用用户构成。网络门户是对数据操作的入口,向用户提供注册、处理、数据查询、数据统计、数据挖掘等功能。

网络安全基础设施(GSI)分布在不同的层中,验证操作权限,保证整个系统的安全。

4 关键技术

4.1 EDGP 服务动态集成机制

EDGP 采用 UDDI 技术实现 Web service 的动态集成。基于 OGSA 体系结构的网格将一切资源都看作服务,因此应用系统的数据可看作是被 Web service 封装后形成的服务。UDDI(Universal Description, Discovery and Integration)指服务的统一描述、发现和集成^[7,8]。由于 EDGP 中内含大量 Web service,所以使用一台 UDDI 服务器,作为注册、发布、查找 Web service,实现 Web service 的统一化,标准化。

EDGP 实现以编码形式在 UDDI 服务器上查找所需 Web service 的过程,即使用客户端 SDK,通过

API 连接 UDDI 服务器,将编写好的 Web service 发布到 UDDI 上。如果 Web service 出现变动,应用程序会很好地重新恢复,不需在程序中修改代码,只需维护 UDDI 服务器上的 Web service。另外,新 Web service 可以自由加入,现有 Web service 可以自由撤消,实现各种服务的“即插即用”,即实现了新应用软件只要部署了 Web service 就可以加入到 EDGP 中,现有的应用软件只要禁用 Web service 就可以撤离 EDGP,方便软件的升级换代和新老软件的交替。服务动态集成机制如图 3 所示。

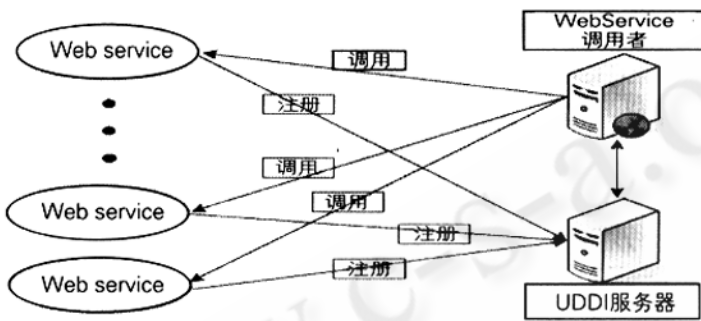


图 3 服务动态集成机制图

使用 UDDI 进行 Web service 动态集成的步骤:

- (1) Web service 在 UDDI 服务器上注册;
- (2) 应用程序在 UDDI 上查询要调用的 Web service;
- (3) UDDI 服务器返回给应用程序此 Web service 的地址;
- (4) 应用程序根据 UDDI 服务器返回的地址调用 Web service;
- (5) Web service 执行数据封装的功能,与 EDGP 客户端数据库进行交互;
- (6) 应用程序接收 Web service 提取到的数据。

4.2 系统安全机制

EDGP 采用 CA 认证方式实现系统安全机制。CA 中心为每个合法用户颁发包含其操作权限的数字证书^[9],用户利用 CA 中心所颁发的数字证书访问 EDGP 门户,EDGP 门户通过 CA 中心对用户数字证书进行验证,根据用户的操作权限提供相应的操作界面,实现资源的安全调用与用户操作的不可抵赖性。CA 系统信息安全管理流程如图 4 所示。

- (1) CA 中心给所有服务器发布私钥服务器证书;
- (2) 用户使用私钥去 CA 中心验证,通过 CA 公钥验证无误后,返回到用户;

(3) 用户使用私钥在 UDDI 上查询所调用 Web service 地址,UDDI 服务器用私钥进行验证;

(4) 验证通过后 UDDI 给出 Web service 地址,用户调用 Web service 去查询相应的信息;

(5) Web service 调用完毕后,系统自动断开连接;如果需要再次调用相应的信息,返回第 3 步进行操作。

4.3 非标准数据与标准数据转换机制

企业各部门业务流中包含大量数据。基于部门间业务流数据共享的需求,EDGP 要求分析企业各业务流程中包含的数据,梳理出其中的“权威数据”。所谓“权威数据”是指属于某个部门业务的原始数据,该数据被别的业务所共享,一旦发生数据重复等问题就以数据共享源的数据为标准。

非标准数据与标准数据之间转换的核心在于完成业务数据库表和标准数据库表之间的字段对应规则配置。中心数据以标准的数据格式存储数据,各个异构数据源之间的数据通过中心数据库的“对照表”进行交互,通过查寻“对照表”交互数据之间的对应关系,根据对应关系将表中的字段一一对应,最后根据需求进行添删改查操作。例如,人员管理系统与财务系统都有人员表,但两表的人员名称字段分

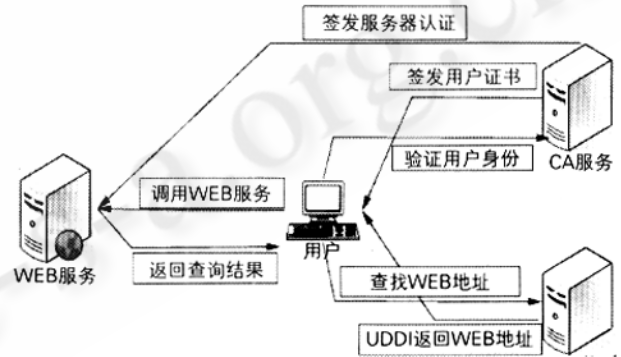


图 4 CA 系统信息安全管理流程图

别为“姓名”和“职员”,当两表进行数据交互时通过对照表查出标准字段名称为“Name”,而且人员管理系统与财务系统在人员表中有“姓名”和“职员”字段,它们关系是对应的,随即可以进行两种数据源之间的数据转换。非标准数据与标准数据之间的转换机制如图 5 所示。

4.4 数据联动机制

EDGP 数据中心存放着应用系统之间需交互数据的备份。当应用系统的数据发生改变时应及时更新数据中心的数据,但由于应用系统逻辑中没有本机数据

与数据中心数据联动的功能,也没有与 Web service 交互的功能,因此数据的联动只能依靠 Web service 的定时抓取来完成。由于一次数据联动可能会形成大量数据在网络中的流动,因此数据联动不宜太频繁,可以根据实际情况设定执行联动的具体时间。

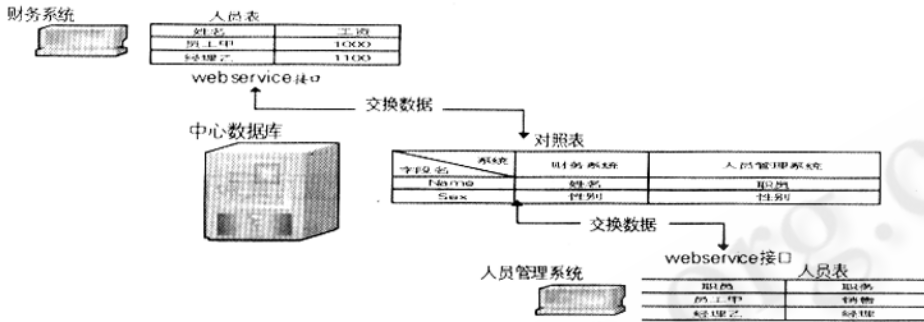


图 5 非标准数据与标准数据之间转换机制图

应用系统与数据中心之间数据联动时,首先由 Web service 根据数据索引与应用系统数据库的内容确定需要更新的内容,接着将这些数据上传到数据中心更新相应数据,同时刷新数据索引,最后将需要与其它系统联动的数据项写入数据中心的数据缓冲池,等待适当的时机,用这些数据去替换相关应用系统数据库中的数据项。同时部署在应用系统端的 Web service 也会定时查看数据中心数据缓冲池,及时用数据中心缓冲池中的数据项更新应用系统数据库的内容,并通知相关人员,达到数据联动的目的。其中,数据索引是指在应用系统端设置数据索引,记录上次数据更新时,应用系统端数据库中数据的情况,并与当前数据库中数据进行比较,确定已经发生变化的数据,使数据的更新仅针对这些数据进行,减少网络中数据的传输量。在数据中心设置数据缓冲池,是为了临时存放联动的数据项,在适当的时机将这些数据通过调用部署在应用系统端的 Web service 写入应用系统的数据库中。

5 结束语

该文介绍了根据企业内部业务的运行机制,充分利用网格技术,设计的企业数据网格平台 EDGP。保持原有不同时期、使用不同工具开发的客户端系统的情况下,将这些应用程序集成到企业数据网格平台中,实现了数据联动,统一了数据格式,便于上级部门对下级部门业务情况的实时了解,大大降低了办公费用,且

提高了办公效率。EDGP 是一种可扩展的网格体系结构,具有全局统一的信息访问接口,可自由加入新的系统,撤消现有系统,同时,该系统提供了可靠的安全保证。EDGP 已很好的应用于某大型企业,应用效果良好。

参考文献

1. I. Foster, C. Kesselman, J. Nick eds. The physiology of the grid: An open grid services architecture for distributed systems integration [EB]. <http://www.globus.org/research/papers/ogsa.pdf>, 2004.10.
2. Tim Banks. Web Services Resource Framework (WSRF) - Primer V1. 2 [EB]. <http://docs.oasis-open.org/wsrp/wsrp-primer-1.2-primer-cd-02.pdf>, 2006.5.
3. Borja Sotomayor. The Globus Toolkit 4 Programmer's Tutorial [EB]. http://gdp.globus.org/gt4-tutorial/download/progtutorial-pdf_0.2.1.tar.gz, 2005.6.
4. Karl Czajkowski, Donald F Ferguson, Ian Foster eds. From Open Grid Services Infrastructure to WSResource Framework: Refactoring Evolution [EB]. http://www.globus.org/wsrp/specs/ogsi_to_wsrp_1.0.pdf. 2004.5.
5. I. Foster, C. Kesselman. 网格计算. 第二版[M], 北京: 机械工业出版社, 2005.4: 30-326 I. Foster. A globus Primer. The Open Grid Services Architecture, Version 1.0 [EB]. <http://www.gridforum.org/documents/GWD-1-E/GFD-1.030.pdf>. 2005.6.
6. I. Foster, C. Kesselman, S Tuecke. The anatomy of the grid: Enabling scalable virtual organizations [J]. International Journal of Supercomputer Applications, 2001; 15(3): 200~222.