

基于网格技术的政府信息门户研究

杜文华

(中南民族大学管理学院 武汉 430074)

摘 要 提出了基于网格的政府信息门户的总体构架,分析了其服务结构、设计原则,阐述了服务之间的数据发布/订阅方式,以及 OGSA-DAI 的工作流程。

关键词 网格技术 政府信息门户 OGSA

1 政府信息门户

电子政务是政府在国民经济和社会信息化的背景下,以提高政府办公效率,改善决策和投资环境为目标,将政府的信息发布、管理、服务、沟通功能向因特网上迁移的系统解决方案。目前我国电子政务建设的主要特点是:政府上网和建立纵向的业务应用系统。一方面,国家投入了大量资金,各级政府机构建设电子政务的积极性都很高,政府网站数量快速增加;但另一方面,政府网站的社会应用水平却提不上去,条块分割、信息孤岛的弊端难以解决。这严重制约了电子政务的社会化应用,对未来电子政务的顺利发展将造成极为不利的影响。

突破这种困境的出路在于大力发展政府信息门户(Government Information Portal,简称 GIP)。政府信息门户是一个将用户的所有应用和数据集成到一个信息管理平台上,并以统一的界面提供给用户,藉此可以快速地建立针对用户、内部职员以及合作伙伴的一致信息服务。也就是变“人找信息”为“信息找人”的智能化信息服务平台。随着用户逐步开始应用各种软件系统,组织内的信息会高速增长,这些数据通常都被分布存储在组织内部的数据库、主机、文件服务器上,因此建立跨平台的数据集成信息系统是提高电子政务系统效能的重要手段。

由于各地市的政府信息一般是分布存储的,信息交换频繁,要建立跨部门的应用系统可采用同一数据库的大集中方式进行统一管理,但这种方式成本高且灵活性差。

2 网格技术

近年来出现的网格(Grid)技术为政府信息门户的建立提供了新的思路。网格是一个集成的计算与资源环境,或者说是一个计算资源池。网格能够充分吸纳各种计算资源,并将转化成一种随处可得的、可靠的、标准的、经济的计算能力。除了各种类型的计算机,这里的计算资源还包括网络通信能力、数据资料、仪器设备、甚至是人等各种相关的资源。

2.1 开放网格服务体系结构

开放网格服务体系结构(Open Grid Services Architecture,简称 OGSA)是全球网格论坛于2002年2月通

过的重要标准建议,是继五层沙漏结构之后最重要,也是目前最新的一种网格体系结构。

OGSA 最突出的思想就是以服务为中心。OGSA 定义的网格服务的概念是:网格服务是 Web 服务的扩展,该服务提供了一组接口,使得它可以支持服务发现、动态服务创建、生命周期管理、通知机制等。这里的服务是指具有特定功能的网络化实体。在 OGSA 框架中,将一切都抽象为服务,包括计算机、程序、数据、仪器设备等。这种观念十分有利于灵活、一致、动态的共享机制的实现,使得分布式的网格系统有了标准的接口和行为。开放网格体系结构着眼于构建动态的虚拟组织来达到数据和计算资源的共享。

2.2 网格服务

OGSA 为建设基于网格的应用定义了一个公共和标准的体系结构。OGSA 的核心是定义了网格服务的概念。但 OGSA 并不涉及到技术细节和规范,这些细节和规范是由开发网格服务基础设施(OGSI)定义的。OGSI 是一个正式的技术规格说明书,它用于实现 OGSA 中所定义的包括网格服务在内的各种概念。

根据 OGSI 规范,一个网格服务是一个符合 OGSI 标准的 Web 服务,通过 WSDL 来表现自己的接口、行为和扩展。因此,这种意义上的网格服务可以被视为一种满足网格环境下需求的标准 Web 服务。

基于网格服务的概念,OGSA 将整个网格看作是网格服务的集合,但是这个集合不是一成不变的,而是可以扩展的。这反映了网格的动态特性。网格服务通过定义接口来完成不同的功能,服务数据是关于网格实例的信息,因此网格服务可以简单地表示为:网格服务=接口/行为+服务数据。

OGSI 规范中定义的标准接口构成了 OGSI 的核心。接口主要包括网格服务(Grid Service)、通知(Notification)、句柄映射(Handle Map)、服务工厂(Factory)和注册中心(Registry)接口。

2.3 Globus Toolkit 3(GT3)

Globus 项目是世界上最具有影响的网格项目之一。Globus Toolkit 是该项目的重要实践成果,它是一种基于社团的、开放结构、开放源码的服务的集合,也是支持网格和网格的软件库。目前推出了基于 OGSI 的实现

3.9版。该工具包解决了网络安全、信息发现、资源管理、数据管理、通信、错误检测等问题。Globus Toolkit已经应用在世界大多数的网格项目中。

GT3的核心结构是基于OGSI的基本模块和协议,它的设计目标是使OGSI技术易于使用、复用,在开发新的网格应用时易于扩展。GT3允许开发可移植的与OGSI兼容的网格服务而不用了解具体的协议和传输绑定。

GT3.2作为OGSA的标准实现,目前已经成为建立大型网格系统的首选。

3 基于网格的政府信息门户系统设计

3.1 政府信息门户系统概述

政府信息门户系统由基础部分、信息门户、基础服务平台和门户服务平台四个部分组成。其中网格基础提供通信、安全、系统和网络管理、分发管理等功能,这一部分属于比较底层的部分,开发时一般不予考虑。建立在基础部分之上,政府信息门户系统由基础服务平台和门户服务平台和信息门户几个部分组成。下面重点研究基础服务平台和门户服务平台建设中的关键部分,通过对这部分的具体研究与分析,进而推广到整个政府信息门户系统中。

3.2 服务结构分析

由于网格环境中所有的组件都是虚拟的(即对相同接口实现不同的封装),因此,通过提供一组相对统一的核心接口,所有的网格服务都基于这些接口实现,就可很容易地构造出具有层次结构的、更高级别的服务,如图1所示。这些服务可以跨越不同的抽象层次,以一种统一的方式来看待。我们的信息门户就是这样一个高级服务,它提供一个统一的入口来访问系统中各样具体的服务。

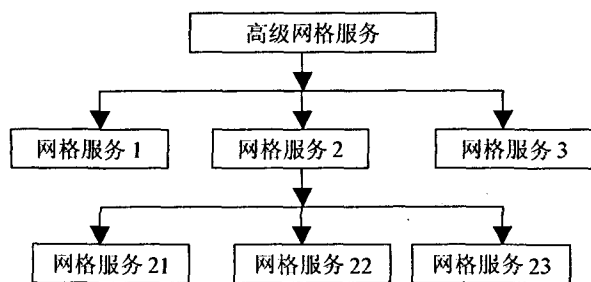


图1 服务结构示意图

3.3 设计原则

参考国内外在设计政府信息门户体系结构的经验,分析网格环境下电子政务信息门户的特点,并结合我国的实际情况,政府信息门户系统的设计原则如下:

(1)面向服务的软件体系结构

网格以网格服务为基本单位,它以Web服务为基础。这是一种以服务为核心的体系结构。这里的服务可以包括各种计算资料、存储资源、网络、程序、数据库等。在网格中一切都是服务。在网格环境下我们作为服务提供者,编写软件所要做的工作就是编写一个个具有基本功能的网格服务并发布出去,以供调用。客户作

为服务使用者,负责发现相应的网格服务,并加以控制流程,使之能够通过组合完成自己需要的功能。

(2)能最大程度地集成遗留系统

网格不是一个全新的系统,它基于现有的技术:Internet技术、对象技术和Web服务技术,并对现有的系统进行了扩展,它不需要一种新的编程语言。它没有什么本质性的新技术,只是对原有的技术加以优化和重新组合使用。

3.4 服务之间的数据发布/订阅

网格的基本单元称为网格服务。网格服务之间通信不仅使用传统的RPC为接口,而且采用更为灵活的消息传递机制——发布/订阅机制。这种机制以速算、可靠、异步、松散耦合、与语言无关及平台无关的方式在网格服务之间传递消息,从而保证软件系统之间的灵活集成和良好的互操作。GT3中通过通知(Notification)接口实现发布/订阅机制,它支持服务到服务的传输,也支持第三方的传输如Java消息服务等。GT3中一个典型的通知从源到目的地的传递过程如图2:

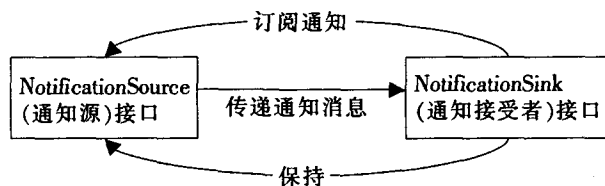


图2 通知传送过程

第一步是消息的接收方用自己的GSH(Grid Service Handle)在通知源接口上激活订阅操作,第二步是通知消息流从“源”发向“目的地”,不能够持续的获得通知,还需要接受者周期性的给通知源存活消息,让通知源知道它对接收通知感兴趣。

3.5 数据访问集成

网格数据库是对现有数据库的网格化,基于开放网格服务体系结构提供网格数据库服务,使网格用户或其他网格服务可通过网格数据库服务访问网格中的各种异构数据库,从而达到数据资源的调试共享和协同处理,对数据资源的访问更加透明、高效、可靠,网格数据处理的能力更强,更好地满足更广泛的数据处理需求。OGSA-DAI项目致力于建造网格访问和集成来自不再的孤立数据源的中间件。OGSA-DAI符合基于OGSA的网格标准,并在Globus Toolkit上进行开发。OGSA-DAI的结构是一种工具箱,它具有一些扩展点,可供开发人员扩展其功能,以适应自身的特定需求。OGSA-DAI的体系结构如图3:

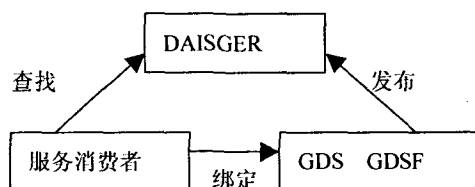


图3 OGSA-DAI的体系结构

论元数据格式及其关系和转化问题

肖述清

(贵州大学图书馆 贵阳 550003)

摘 要 介绍了元数据定义和四种元数据格式,重点探讨这四种元数据格式的关系及转化问题,预测元数据的发展前景。

关键词 元数据 Dublin Core XML RDF MARC

数字图书馆是数字技术搜集、存储和组织信息,并通过计算机网络查询和检索的一种现代化信息系统,其目的在于通过因特网大幅度提高采集、存储、组织数字信息的方法,使其用于搜索、存储、处理和利用,满足各界对信息日益增加的需求,为用户提供更全面的信息系统和更周全的服务。图书馆在传统的文献型信息资源组织中,依据各种分类法进行科学分类与建立各类目录体系,以保证用户可以在图书馆里找到所需的信息。伴随着信息技术的发展,网络信息资源不断增长、信息类型不断丰富、信息传递的方式不断改变,图书馆急需创造全新的网络信息资源组织规则,来适应数字图书馆的发展。

1 何谓元数据

元数据(Metadata)即“关于数据的数据”,是描述任何 Internet 数据和资源,促进 Internet 信息资源的组织和发现的数据。从定义可以看出,在 Metadata 及其所描述的信息资源之间存在着某种必然的联系。例如,图书馆的书目记录就是 Metadata 数据元素的集合,这些元素通过索书号与图书馆馆藏中的文献建立连接。Metadata 具有与书目卡片相似的作用,利用元数据格式描述一个信息对象的内容和位置,以便在网络上可以方便地查找。

2 四种元数据格式

(1)Dublin Core.1995年3月由 OCLC 和 NCSA 共同主持在 Ohio 的 Dublin 召开第一届元数据研讨会,经过对

Metadata 的讨论,终于产生了都柏林核心元素集,简称为都柏林核心(DC)。DC 具有简练、易于理解、可扩展、能与其他元数据格式进行桥接等特性,这些优点使它成为了一个良好的网络信息资源描述的元数据格式。此后又经过6次元数据研讨会增补修订,使得 DC 在结构和功能上逐步完善。DC 适合于广泛网络信息资源的标识。

(2)XML。XML 是可扩展置标语言,是一种基于 SGML 的简单灵活的语言。XML 本身不仅仅是一个标记语言,还是一个元语言,允许用户设计自己的标记语言。XML 实际上定义了一个简单的 SGML 语言集,用现在在 HTML 中可能的方式实现普通 SGML 文件的使用、接受和处理。作为标记语言的元语言,XML 可以运用到广泛的领域,MCF、RDF、Webcollectionsd 等都使用 XML 作为基本的句法。XML 能够运行在任何平台和操作系统上。它的可扩展性表现在它并不像 HTML 那样有固定的格式,设计 XML 的目的就是使得用户能够在 WWW 上使用 SGML。

(3)RDF。RDF 即资源描述框架,是一个使用 XML 语法来表示的资料模型,用来描述 Web 资源的特性,及资源与资源之间的关系。它是为试图解决许多元数据的互操作性和兼容性而制定出来的,为元数据在 Web 上的各种应用提供一个基础结构,使应用程序之间能够在 Web 上交换元数据,以促进网络资源的自动化处理。它符合多种需要,提供的框架,包括范围之广,足以使它能容纳得下所有领域的应用。

从图3中可以看出,其体系结构与 Web Service 很相似,都是查找、绑定、发布机制。

OGSA-DAI 逻辑上可以认为是一些协同工作的网格服务。这些网格服务在实际处理数据的系统和用户之间充当代理的角色。OGSA-DAI 的流程如下:

(1)客户端或服务调用网格服务组注册服务(DAISGR),来得到网格数据服务工厂(GDSF)的句柄;

(2)调用 GDSF,创建一个网格数据服务实例(GDS),并返回这个 GDS 的句柄;

(3)调用这个 GDS,并提交一个 XML 格式请求,GDS 执行这个请求,并通过 JDBC 和后端的数据库进行交互,然后把结果返回给客户端。

目前 OGSA-DAI 已经在英国国家科学网格、天文

网络计划、美国国家虚拟天文台等项目中得到了应用。

参考文献

- 1 林林. 政府信息门户(GIP)在电子政务中的应用. 情报科学,2003(7)
- 2 王波,童维勤. 网格服务通知机制探讨. 计算机工程与设计,2004(9)
- 3 余向东. 国内外电子政务发展浅析. 信息技术与标准化,2003(3)
- 4 刘建新,阎保平. OGSA-DAI 体系结构及其关键技术研究. 计算机应用,2004(11)
- 5 刘昕鹏,罗英伟,汪小林,许卓群. 基于 XML 的 WebGIS 通信协议的设计与实现. 软件学报,2004(6)

(责任编辑:赵日珑)