

网络服务下的数据访问与集成规范架构

王 威,刘卫东,宋佳兴

(清华大学 计算机科学与技术系,北京 100084)

(Wwei98@mails.tsinghua.edu.cn)

摘 要:在开放网格服务体系下的数据访问与集成(OGSA-DAI)项目和网络服务下的集成资源框架(WSRF)规范的背景下,介绍了网络服务下的数据访问与集成(WS-DAI)提出的一组 Web Service 数据访问与集成的通用数据接口规范,能扩展支持各种类型的数据资源。分析了 WS-DAI 的规范架构,详细研究了核心抽象模型并给出应用实例。最后讨论了此规范与 WSRF 之间的映射。

关键词:数据网格;WS-DAI;WSRF;OGSA-DAI

中图分类号: TP393 **文献标识码:** A

Research on the specification framework and key technology of WS-DAI

WANG Wei, LIU Wei-dong, SONG Jia-xing

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

Abstract: With the background of OGSA-DAI and WSRF, a Web Service data access and integration specification for a collection of generic data interfaces presented by WS-DAI was introduced. The specification can be extended to support specific kind of data resources. The framework of the specification was analyzed. The paper also studied on the core of the abstract model in detail and an application example was presented. Finally the mapping to WSRF was discussed.

Key words: data grid; WS-DAI; WSRF; OGSA-DAI

0 引言

大规模、分布异构的数据资源在广域网格应用中起着越来越重要的作用,而 Web Service 对于网络资源集的描述、调用和协作提供了设备中立、平台无关的实现。全球网格论坛(GGF)的开放网格服务体系(Open Grid Services Architecture, OGSA)及相关规范定义了一致的接口,通过 Web Service 访问网格基础结构构件。OGSA-DAI 项目的发展和 WSRF 规范的提出使得在这样相互融合的背景下,Web 与 grid 双方都可以从协调一致的支持数据资源及其管理系统的 Web Service 接口中获益^[1]。

WS-DAI(Web Services-Data Access and Integration)工作组(即 GGF 的 DAIS-WG)在相关背景的研究支持下,提出了一组 Web Service 数据访问与集成的通用数据接口规范。这些规范既能应用于 Web Service 环境,也可以作为网格架构中的一部分,扩展了 OGSA 体系应用于更广泛的领域,发挥了更重要的作用。

1 相关背景

1.1 OGSA-DAI

OGSA-DAI^[2]是一个通过网格支持对分离的数据资源实现数据访问和集成的中间件开发项目,其目的是在 OGSA 的基础上开发一个网格服务,实现数据访问和集成目标。

OGSA-DAI 用网格服务的形式为用户提供数据访问和管理服务。网格数据服务工厂(GDSF)是一个永久的服务,它根据客户的请求创建网格数据服务(GDS)。一个 GDSF 能创建

不同类型的 GDS,满足不同客户的要求。客户可以请求 GDSF 确定 GDS 可用的配置,请求 GDSF 创建最能满足用户需求的 GDS。GDS 通过网格数据服务端口类型提供数据访问和集成等功能。用户通过 GDS 执行文档(GDS-Perform)指定需要该服务完成的操作。GDSF 创建时给定 GDS 一个生存期限,过期后该 GDS 就不可用了。

1.2 WSRF

开放网格服务基础设施(OGSI)提供了基于开放标准的网格工具,重造了网格服务,通过 Web Service 定义语言(WSDL)和其他可用 XML 工具将网格服务转变成 Web Service。Web Service 资源框架(WSRF)^[3]是 OGSI 的重构和发展,利用新的 Web Service 标准,基本保留了 OGSI 中的所有功能,同时更改了一些语法,还在其表示中采用了不同的技术。Web Service 通知(WSN)为 Web Service 提供基于消息发布和预定能力。WSRF 和 WSN 都是建立在已存在的 Web Service 定义和技术基础上的,实现了网格计算、系统管理和 Web Service 的统一。

WSRF 由六种 Web Service 规范组成:1) WS-ResourceLifetime:定义了 WS-Resource 的资源破坏机制,其中包括允许请求方立即或使用基于时间的安排好的资源终止机制来破坏资源。2) WS-ResourceProperties:定义了 WS-Resource 的类型和如何与 Web Service 的接口描述相关联,以及检索、更改或删除 WS-Resource 属性的消息交换过程。3) WS-Notification:通过基于主题的发布/订阅模式定义事件订阅和通知机制。4) WS-RenewableReferences:定义了端点变为无效时,对需要检索端点引用更新版本的策略信息的常规

收稿日期:2005-04-26;修订日期:2005-06-24

基金项目:国家自然科学基金资助项目(90412012;90104002);国家 973 规划项目(1999032707)

作者简介:王威(1979-),男,安徽黄山人,硕士研究生,主要研究方向:数据网格、分布式信息系统;刘卫东(1968-),男,副教授,主要研究方向:分布式信息系统、数据库、计算机网络;宋佳兴(1974-),男,讲师,主要研究方向:分布式信息系统、数据库、计算机网络。

WS-Addressing 端点引用。5) WS-ServiceGroup: 定义了通过异质引用集合访问 Web Service 的接口。6) WS-BaseFaults: 为在 Web Service 消息交换过程中返回的错误信息定义了基本的 XML 错误类型。

WS-Resource 是一个有状态资源和它与之交互的 Web Service 的组合。即: Web Service + 有状态资源 = WS-Resource^[4]。WSRF 通过定义“Web Service 资源法”(WS-Resource approach), 在 Web Service 的上下文中实现对状态的建模和管理。这种方法的内容包括: 在已经建立起来的 Web Service 标准上下文中发现和观察状态资源, 并为之交互的能力, 以及将服务和作用于该服务的状态资源的分离。

WSRF 规范仍然在发展修订之中。2005 年 1 月发布的 Globus Toolkit 4 (GT4), 实现了 WSRF 和 WSN 标准。GT4 提供 API 来构建有状态的 Web Service, 其目标是建立分布式异构计算环境。所有知名的 GT3 协议都被重新设计为可以使用 WSRF。并且 GT4 也在其中增添了一些新的 Web Service 组件。

2 WS-DAI 规范架构

2.1 WS-DAI 概览

OGSA-DAI 与 DAIS 相符合, 也成为 DAIS 网络数据服务推荐标准的第一个参考实现。WSRF 的目标则是通过创建“状态”概念以及处理状态的方法使得 Web Service 应用到网络架构中。数据访问在多种类型的网络应用扮演着一个重要角色。数据访问即数据的检索、处理和插入, 数据可能以不同的格式和底层结构存储。网络数据访问既需要一种灵活的架构来处理对数据资源的数据请求, 又能整合到 OGSA 定义的整体网络架构中去。

在这些背景与要求的驱动下, WS-DAI^[5] 提供了一组与 Web Service 一样可灵活使用的通用网络数据访问接口集规范, 接口根据提供的支持分类描述数据描述、数据访问、数据工厂与数据管理。

WS-DAI 的目标是提供数据服务的接口和属性架构, 描述各种特殊接口。这里描述的架构可以扩展定义各种特殊类型数据的访问, 如已经完成的关系数据库访问 (WS-DAIR)^[6] 和 XML 格式数据访问 (WS-DAIX)^[7]。其余特殊格式的数据访问规范可以通过本文档定义的基本接口集来扩展, 如文件类型数据访问, 在 WS-DAIF 草案基础上修改扩充, 以及对对象数据库, 将在 GGF14 上完成提交。图 1 是 WS-DAI 的整体架构概览, 并列出了参考引用的其他规范。

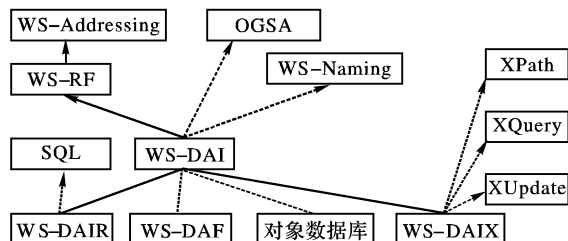


图 1 WS-DAI 整体架构概览

WS-DAI 根据基本的数据访问和数据工厂接口规定了数据服务以及数据服务实现的数据描述属性。WS-DAI 并不定义新的查询语言和数据模型, 因此数据访问接口根据现有的数据服务提供访问的底层数据资源所支持的语言接口来描述。数据管理则参考了 Web Service 接口和存储数据的数据资源间的关系的管理。WS-DAI 把数据服务的抽象模型和操作表示分开, 并使用下面的定义与概念来描述模型。

2.2 基本定义

数据资源: 任何系统, 能作为数据的来源和汇集。例如: 关系数据库、XML 结构存储库, 文件系统, 对象数据库, 传感器网络等。现阶段期望网络中的数据一般仍然是可用现有技术来管理的, 如关系数据库和文件系统等。一个现有的数据资源为用户提供访问存储数据的机制。DAIS 规范提供了一种通用的面向数据资源处理的服务, 并以支持集成到基于 OGSA 的网络的方式列出这些机制。

数据服务: 数据服务就是一个实现了一个或多个 DAIS 规定接口的 Web Service, 提供到数据资源的访问。WS-DAI 规范提供了基于 Web Service 的数据访问架构, 使得现有的数据访问技术和其他相关规范在数据资源中仍然可用并必需。

使用者: 使用者是利用数据服务提供接口来访问数据资源的一个应用。一个数据服务也可以扮演另一个数据服务的使用者。

这三个是最核心的概念。WS-DAI 还给出了数据集 (Data Set)、抽象名称 (Abstract Name) 与地址 (Address) 等的定义。

2.3 数据服务模型与接口

WS-DAI 的重点是定义基本的数据服务接口与数据服务模型。假定使用者与数据服务之间以及数据服务与数据资源之间是一种多对多的关系。如图 2 所示。

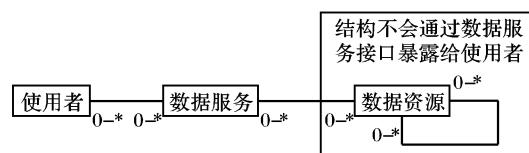


图 2 WS-DAI 数据服务模型

数据服务为使用者提供访问数据资源的接口。数据资源可拥有任意复杂性。例如, 一个联邦数据库或关系数据库。数据资源的这种复杂性结构不会暴露给使用者, 并且可在数据服务提供接口的范围与语义内操作。

接口参考消息集和 XML 结构, 描述使用者和数据服务相互作用的方式。这里并不授权接口如何组成服务; 被提议的接口可用于与其他接口分离或连接。接口的可行组合最初将遵循数据访问已制定的模式。

图 3 中的数据服务为一个 XMLCollection 数据服务提供了 XPathAccess 和 XUpdateAccess 接口, XMLCollection 数据服务关联于一个 XML 数据库。

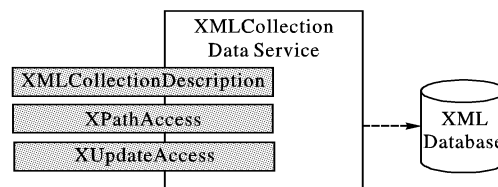


图 3 XMLCollection 数据服务接口组成

3 WS-DAI 核心抽象模型

WS-DAI 的核心抽象模型包括数据描述、数据访问与数据工厂。

3.1 数据描述

数据描述包含了描述数据资源属性的 XML 结构。这里描述的属性不必设置为工厂模式的一部分, 但必须显现在所有的实现数据描述的数据服务中。基于此规范的 DAIS 实现可以进一步扩展需要的属性。

3.2 数据访问

数据访问把直接访问和改写数据服务表现数据的消息连

同描述访问消息行为的属性聚集在一起。WS-DAI 中定义的数据访问行为属性必须显现在所有数据服务中。基于此规范的 DAIS 实现可以按需要扩展属性和消息列表,必须定义一个或多个访问消息。

DAIS 定义了支持允许数据集传递进入或从数据服务中检索的消息数据访问接口,描述了从数据服务提供直接数据访问的消息。同时也描述了数据服务的数据访问消息的行为属性,如: Readable、Writeable、ConcurrentAccess、TransactionInitiation、TransactionIsolation、Sensitivity、DataAccessPropertyType 等,作为实现工厂模式的消息交换的一部分。

直接数据访问意味着使用者能期待对数据服务的请求返回一个直接的响应,包含请求的结果数据。例如,给数据服务传递一个 XPathQuery 消息,将得到一个包含一组 XML 分段的响应消息。

间接数据访问意味着使用者并不期待结果作为对数据服务的请求的一个响应。访问数据的请求将由数据服务和数据资源处理,结果对于使用者是间接可用的,通常通过不同的数据服务和接口提供。例如,给数据服务传递一个 SQL 查询,将引起到另一个数据服务的引用(也可能是另一个数据资源),这一引用是由允许访问原始查询结果产生的。这也允许了结果保存在服务一方,最小化不必要的数据移动。当结果可用时使用者可以使用新的引用来检索结果或允许此后的过程发生在服务端。

这种模型也能在数据增加时使用。例如,当数据服务表示文件系统中的目录时,间接数据访问能用来表示数据可加入的新文件。

3.3 数据工厂

数据工厂定义的消息实现了工厂模式,消息在数据资源和数据服务之间建立了新的关系。用这种办法,数据服务可以用于表示一个查询结果或者作为被插入数据资源的保存数据的地方。数据工厂描述了指示数据服务必须接收工厂消息的行为属性。工厂模式可包含新数据资源的建立,还可以包含 Web Service 的配置部署。

工厂模式允许数据服务和数据资源间的新关系以数据服务的消息交换为结果。这种来自另一个服务—资源关系或对同一数据资源提供可选视图的性能带来了概念上相关的服务—资源关系的集合。可以参看下面的例子。

3.4 应用实例

如图 4 中的数据库数据服务提供一个 SQLFactory 接口。SQLExecuteFactory 操作用于构建 SQL 响应数据服务。服务提供了来自相对于关系数据库的一个 SQL 表达式所得的 RowSet 结果的访问,这里假定 SQL 表达式中包含一个 SELECT 语句。RowSet 能以表的形式存储在关系数据库里或从数据库里分离出来,但这里重要的区别是数据经数据服务以行集合的形式表示出来,而数据服务不实现 SQLAccessPortType。作为 SQL 响应的代替,数据服务提供 SQLResponseAccess 操作集,允许 RowSet 被检索,但不提供提交 SQL 表达式的功能。

4 与 WSRF 的映射

WSRF 是一组 Web Service 规范的集合,描述了作为表示有状态的资源和 Web Service 之间关系方法的 WS-Resource 结构。DAIS 属性及消息在映射到 WSRF 时,会用到以下一些规范: WS-I (包括 SOAP 和 WSDL)、WS-Addressing、WS-ResourceProperties、WS-ResourceLifetime 等^[3]。

以数据服务为例,在映射情境中,一个数据服务就是一个

使用 WSRF 来允许访问使用 DAIS 规定接口的数据资源的 Web Service。WS-Addressing End Point References (EPR) 实现了隐含的标识数据服务—数据资源关系的资源模式 (WS-Resource)。例如,一个关系数据库可引用 EPR1 使用一组 SQL 访问消息通过一个数据服务来访问。作为选择,同一个关系数据库可引用 EPR2 使用一组 XML 访问消息通过一个数据服务来访问。EPR 可通过注册器提供给使用者或由使用者发现。注册在注册器的 EPR 将与数据服务所特有的属性所关联,包括名称以及其他的描述信息。

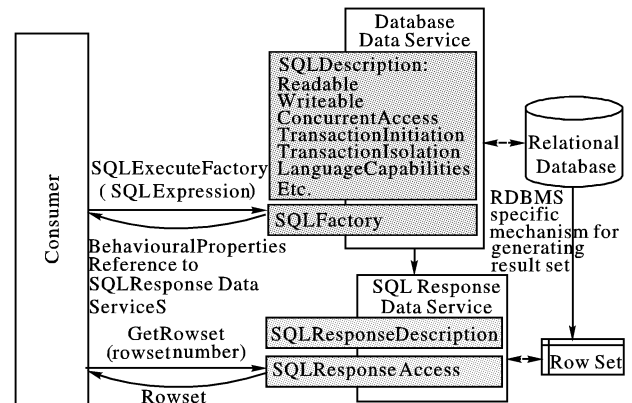


图4 数据服务应用实例

多种不同的 EPR 可通过不同接口提供到同一数据资源的访问,也可通过同一数据服务提供到不同数据资源的访问。为了从现有的 EPR 获得一个新的 EPR,或通过一个新的接口访问一个现有的数据资源,使用者必须重回注册器或调用一个工厂操作。

数据服务实现了一组消息和信息属性。为了处理消息并为信息属性赋值,数据服务能够与数据资源相互作用。属性被表示为资源属性 [WS-ResourceProperties], 因此在服务的 WSDL 文档中静态定义了给定的数据服务的属性结构。为了与 WS-ResourceProperties 规范中概述的方法相一致,所有适于数据服务的属性元素将被聚集到一个唯一的资源属性类型里,依次与数据服务的端口类型相关联。实现中定义的数据访问消息将在 WSDL 1.1 定义中显现,作为数据服务端口类型的操作。来自基本规范和任何实现规范的属性被一起组合为一个 XML schema 类型,是开发数据服务所特有的。这个类型的一个元素被暴露作为数据服务的资源属性结构。未来的属性组允许与任何工厂消息进行传递。

WS-DAI (甚至 WSRF) 依然是发展中的规范。进一步留待解决的问题包括:依赖的相关规范的标准化,资源分辨,会话与生命期,访问控制与安全性等。

参考文献:

- [1] FOSTER I, KESSELMAN C. The Grid 2: Blueprint for a New Computing Infrastructure [M]. Morgan-Kaufman, 2003.
- [2] <http://www.ogsadai.org.uk/> [EB/OL], 2005.
- [3] <http://www.globus.org/wsrfl/> [EB/OL], 2005.
- [4] <http://www-128.ibm.com/developerworks/cn/> [EB/OL], 2005.
- [5] ANTONIOLETTI M, ATKINSON M, KRAUSE A *et al.* Web Services Data Access and Integration (WS-DAI) [R/OL]. <http://forge.gridforum.org/projects/dais-wg>. 2005.
- [6] ANTONIOLETTI M, COLLINS B, KRAUSE A *et al.* Web Services Data Access and Integration—The Relational Realisation (WS-DAIR) [R/OL]. <http://forge.gridforum.org/projects/dais-wg>. 2005.
- [7] ANTONIOLETTI M, HASTINGS S, KRAUSE A *et al.* Web Services Data Access and Integration—The XML Realization (WS-DAIX) [R/OL]. <http://forge.gridforum.org/projects/dais-wg>. 2005.