

# 基于科学工作流的空间数据处理技术探讨

柴 胜<sup>1,2</sup>, 周云轩<sup>1,3</sup>

CHAI Sheng<sup>1,2</sup>, ZHOU Yun-xuan<sup>1,3</sup>

1. 吉林大学 地球探测科学与技术学院, 长春 130012

2. 吉林大学 计算机科学与技术学院, 长春 130026

3. 华东师范大学 河口海岸学国家重点实验室, 上海 200062

1. College of Faculty of GeoExploration Science and Technology, Jilin University, Changchun 130012, China

2. College of Computer Science and Technology, Jilin University, Changchun 130026, China

3. State Key Laboratory of Estuarine and Coastal Research, East China Normal University, Shanghai 200062, China

E-mail: chaisheng@jlu.edu.cn

CHAI Sheng, ZHOU Yun-xuan. Survey of geospatial data processing technology based on scientific workflow. Computer Engineering and Applications, 2007, 43(2): 187-189.

**Abstract:** The paper introduces scientific workflow technique related with the geospatial data processing, and contrast with the traditional business workflow technique. Then taking the Kepler project as an example, it explains scientific workflow technique to handle realm in the geospatial data. At last it discusses the problem existed currently and gives the conclusion.

**Key words:** scientific workflow; geospatial data; Web service

**摘 要:** 首先介绍了与地理空间数据处理相关的科学工作流技术, 并与传统商业工作流技术作对比, 然后以 Kepler 项目为例说明科学工作流技术在空间数据处理领域的应用, 最后讨论目前存在的问题并给出了结论。

**关键词:** 科学工作流; 地理空间数据; Web 服务

文章编号: 1002-8331(2007)02-0187-03 文献标识码: A 中图分类号: TP311.52

## 1 简介

目前, 各个 WebGIS 提供商采用各自的技术和方法来构造本企业的 WebGIS, 所使用的技术、数据结构、语义都不尽相同。在网络上的不同节点上分布着许多功能相似但结构异构和语义异构的 WebGIS, 这种分布式系统是独立存在的、封闭的分布式, 难以与其它分布式系统互操作, 只能通过转换数据格式或调用组件外部接口来实现。从技术上讲, 用不同技术构建的组件之间是很难或不能互相调用。因此, 这种紧密耦合的方式很难实现资源的共享和系统的互操作。

面向服务构架(SOA: Service Oriented Architecture)<sup>[1]</sup>可以解决这样的问题。SOA 是一种软件体系架构概念。服务的概念出现在 20 世纪 90 年代初。SOA 最重要的特点是把服务的实现和接口分离。在服务消费者眼中, 服务是一个支持特定请求格式和契约的端点。消费者不用关心服务如何执行其请求。SOA 用发现、绑定和执行模式代替软件组件之间的紧密耦合方式。服务提供者把服务发布在第三方注册中心中, 服务消费者向注册中心请求服务, 注册中心将满足消费者要求的服务端点和服务契约返回给消费者。这种方式改变原来的紧密耦合方式, 容易实现资源的共享和程序的互操作。

Web 服务是实现 SOA 的一种技术集合。Web 服务可以把

运行在通过 Intranet, Extranet 或 Internet 连接的分布式服务器上的应用程序整合在一起, 而且使整合更加方便和安全。实现一个完整的 Web 服务架构需要有一系列的协议规范来支撑。目前开发的 Web 服务的相关标准协议包括服务调用协议 SOAP(Simple Object Access Protocol)、服务描述协议 WSDL<sup>[2]</sup>(Web Services Description Language)和服务发现/集成协议 UDDI(Universal Description, Discovery and Integration)等。

近年来, 随着科学研究数据和知识的指数级增长(例如地理学、生物学、环境学), 如何简单、高效地利用这些数据已经成为科学界面临的最大的挑战。对这些数据的利用和分析本质上可以理解成科学工作流的范畴, 即相应学科数据从分析的第一个步骤流入到下一个步骤, 整个流程使用标准的工作流语言描述。本文以地理学科为例, 探讨利用科学工作流技术进行地理空间数据的处理。首先介绍分布式地理空间数据处理的各种方法, 然后说明科学工作流系统 Kepler<sup>[3]</sup>并与传统商业工作流作比较, 最后利用 Kepler 系统针对特定问题进行建模举例。

## 2 分布式地理空间数据处理

早期, 地理空间数据只能通过物理介质拷贝的方式进行共享。随着网络技术地飞速发展, 通过一些软件工具或者特殊

程序脚本可以自动完成部分空间数据利用和处理工作。但是这种方法仍然耗费人们大量的精力而且效率比较低。

近年来,主要的 GIS 系统软件,例如 ESRI 等,已经扩展了它们的软件系统,提供分布式地理空间数据的管理。主要是利用主流的 DBMS 技术来支持整个网络环境内远程的数据访问和计算。但是这种方法在数据宿主技术、技术复杂性、互操作性等方面代价很大。更重要的是,DBMS<sup>[4]</sup>虽然在利用索引技术进行强大的数据的查询和过滤,但是对空间数据格式转换支持不够,例如数据类型转换、空间数据值转换等。

目前,许多主流的商业数据库系统已经开始逐步具有空间数据转换的功能,但是它们对于非格式化、准格式化数据的数据处理还是不尽人意。因此,空间数据共享不能完全靠纯的基于 SQL 的分布式空间数据库方法。另一方面,虽然 Web GIS 在分布式空间数据处理方面影响巨大,采用客户/服务器架构,而且提供了比较好的客户端可视化能力,但是没有提供适合科学研究的复杂数据查询和分析功能。

一般在客户端显示地理空间数据有两种方案。一种是直接显示相关数据的图片,另一种是用 Java applet 或者 ActiveX 控件方式装载空间数据进行显示。这两种方式都无法支持空间数据的整合和重用。Web 服务基于平台独立的 WSDL 标准并提供了公开的接口,能够在基于 Web 的空间数据处理中应用。随着地理空间数据格式标准的出台,GML<sup>[5]</sup>已经能够在各种数据间达成一个格式转换的标准。这里基于 Web 服务和 GML 技术说明目前的一种开放的体系结构,该结构:

- (1) 对于原有技术实现的各种资源能够有效兼容和利用;
- (2) 支持各种模型以科学工作流方式工作,科学家能够与流程交互或者让流程自动执行;
- (3) 支持所有数据格式,包括非格式化数据、准格式化数据、格式化数据。

结合科学工作流和 Web service 技术能够完全满足上述需求。首先,将地理空间数据和处理过程发布成 Web service,然后在系统仓库中注册它们,最后利用 Kepler 系统将它们组合成科学流并进行访问。下面介绍 Kepler 科学工作流系统,将它作为组合 Web service 和执行空间数据处理的框架。

### 3 科学工作流

#### 3.1 与商业工作流对比

科学工作流的特征与商业工作流有部分重叠。目前已经出现了几个有影响的基于 Web service 的商业工作流标准,例如 BPEL4WS<sup>[6]</sup>、IBM 的 WSFL 和微软的 XLANG。传统的商业工作流面向文件处理、任务管理、流程控制<sup>[7]</sup>,科学工作流主要面向数据流,侧重于密集的数据操作和计算,同时带有数据转换、数据分析和模拟仿真。比较典型的科学工作流系统有 Kepler、Taverna<sup>[8]</sup>、Triana<sup>[9]</sup>等。

面向数据流和面向控制流在模型表示方面有所区别。例如,商业工作流的可视化界面和流程图、状态转换图或者 UML 的活动图相似,侧重在数据流之上的事件以及流程控制。而目前的科学工作流系统的执行模型通常和数据流过程网络(dataflow Process Network, PN)<sup>[10]</sup>相似。

#### 3.2 Kepler

PtolemyII 是一个通过各种方式封装组件的建模和设计工具。Kepler 构造在 PtolemyII<sup>[11]</sup>系统上,是一个科学工作流系统,

能够设计、发现、执行和部署不同科学领域的工作流。科学家可以利用 Kepler 的图形化界面进行注册和发现各种服务和资源,交互的设计和科学工作流。

Kepler 利用内嵌的并行控制和工作调度机制,无缝的将科学工作流的设计、执行、运行时交互、本地和远程数据的访问、本地和远程服务调用组合起来。另外,它继承了 PtolemyII 系统的特点,能够在单个科学工作流中组合不同计算模型。Kepler 的计算单元称为“actor”,通过输入输出端口,每个 actor 都是彼此通信的可重用的组件,多个 actor 连接在一起形成科学工作流。整个工作流在交互调度组件的调度下执行(参见图 1)。

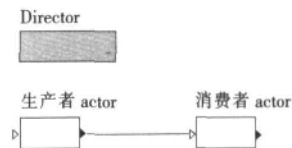


图 1 actor 交互图

Kepler 系统中已经实现了能够访问 Web service 的 actor,在工作流中可以利用相应 actor 作为客户端访问远程分布式服务和资源。工作流执行过程中,执行 Web service 的 actor 相当于 Web service 的代理,该 actor 通过端口连接到其它 actor。基于这样的 actor 组件,任何应用中的功能都可以封装实现为远程 Web 服务,然后成为 Kepler 系统中的组件。

图 2 描述了利用 Kepler 进行基于 Web service 科学实验的整个结构。Kepler 系统中的服务可以是一个分析操作或者是一个基于数据集查询的数据服务。(1)通过发布机制将服务注册到系统仓库中;(2)通过服务发现机制可以查询到已经注册的服务;(3)通过服务组合机制可以组合系统仓库中的各个服务,成为科学工作流;(4)通过工作流执行机制执行组合后的科学工作流。另外,Kepler 系统提供了错误恢复、过程监控、数据日志、科学工作流部署等相关功能。

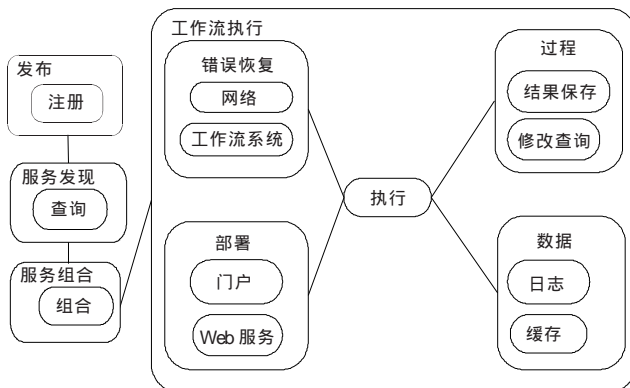


图 2 体系结构

#### 4 基于科学工作流系统的应用举例

基于科学工作流系统的应用举例见图 3。

一般情况下,对外发布的服务和相关的元数据需要注册到系统的仓库中,然后其它使用服务的系统再通过元数据发现服务。使用 Kepler 工作流系统说明服务注册和发现过程,同时描述在该框架内,各个服务可以被组合成科学工作流,然后执行相关的科学任务。Kepler 系统中,经过组合后的科学工作流也是服务,将它们存储起来,可以方便地供其它系统使用。

下面假设一个地理信息门户的检索场景,整个地理空间数据处理过程可以使用 Kepler 建模。用户登录地理信息门户,在

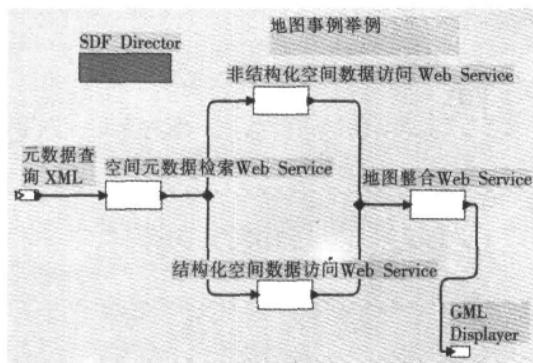


图3 应用实例

门户内输入元数据查询地图, 门户将查询封装成 XML 格式, 提交给空间元数据检索 Web service。该服务可以调用其它 Web service, 例如结构化空间数据访问 Web service, 这些服务可以由远端异构系统提供。最后各个服务返回的结果提交给门户内地图整合 Web service 进行地图融合操作, 该 Web service 返回 GML 格式结果, 传递给 GML Displayer actor 进行显示。这里描述的工作流完全利用 Kepler 科学工作流系统, SDF Director 是整个流程的控制器, Web service 和 GML Displayer 都是 Kepler 中已经实现的 actor。其中 Web service actor 用来调用 Web 服务, GML Displayer actor 用来可视化 GML 格式的数据。各个 Web service 的实现可以有相同或者不同组织提供, 其对应的参数, 例如 WSDL url 可以直接在 Kepler 的 Web service actor 中配置。上述应用只是列举了一个简单的场景, 实际的 Kepler 科学工作流系统实现了很多功能, 并且针对不同领域例如 GIS 有很多扩展。

## 5 结论

本文讨论了基于 Web 服务的科学工作流技术, 并举例说明了科学工作流技术在地理空间数据处理方面的典型应用。Kepler 作为典型的科学工作流系统, 提供了方便的接口注册和发现 Web 服务, 并且将这些服务组合起来进行调用。科学工作流侧重密集的数据操作和计算服务, 完全基于 Web ser-

vice 技术。

目前基于 Web service 的科学工作流系统存在三方数据传输问题。例如假设有 3 个 Web service, 即 A、B 和 C, 并且位于异地。其中 C 调用 A, 得到结果后并将该结果直接传入 B 并调用它。目前 Web service 只能将数据先从 A 拷贝到 C, 然后再从 C 拷贝到 B, 并不支持直接将数据从 A 拷贝到 B。如果 A 返回的数据量大, 将导致系统的效率比较低。

(收稿日期: 2006 年 9 月)

## 参考文献:

- [1] Web Services Roadmap — Guiding the Transition to Web Services and SOA [EB/OL]. <http://www-306.ibm.com/software/solutions/Webservices/pdf/cbdirroadmap2.pdf>.
- [2] Web Services Description Language (WSDL) [EB/OL]. <http://www.w3.org/TR/wsdl>.
- [3] Kepler [EB/OL]. <http://kepler-project.org/>.
- [4] ORACLE Spatial [EB/OL]. <http://www.oracle.com/technology/products/spatial/index.html>.
- [5] 张霞, 李德仁, 朱欣焰. 基于 GML 构建 WebGIS 的研究[J]. 测绘通报, 2003, 10: 4-7.
- [6] BPEL4WS [EB/OL]. <http://www-128.ibm.com/developerworks/library/specification/ws-bpel/>.
- [7] Aalst van der W M P, Hofstede ter A H M, Kiepuszewski B, et al. Workflow patterns[J]. Distributed and Parallel Databases, 2003, 14 (1): 5-51.
- [8] Oinn T M, Addis M, Ferris J, et al. Taverna: a tool for the composition and enactment of bioinformatics workflows[J]. Bioinformatics, 2004, 20 (17): 3045-3054.
- [9] Majithia S, Shields M S, Taylor I J, et al. Triana: a graphical Web service composition and execution toolkit[C]//Proc of the IEEE Intl Conf on Web Services (ICWS), IEEE Computer Society, 2004.
- [10] Lee E A, Parks T M. Dataflow process networks[C]//Proc of the IEEE, 1995, 83 (5): 773-801.
- [11] Ptolemy II [EB/OL]. <http://ptolemy.eecs.berkeley.edu/ptolemyII/>.

(上接 130 页)

断信息“A 点正确, 请检测 B 点”, 根据提示进行下面的操作。该系统能非常准确地对被测功放进行故障定位, 并给出维修方案, 操作简单方便, 可靠性高, 特别适合应急维修的场合。

## 5 结论

基于虚拟仪器开发的本套智能功放测试系统, 采用自动量程切换和霍尔电流传感器实现了高电压宽信号测量和大电流的测量, 同时充分利用虚拟仪器“软件就是仪器”的特点, 开发的软件测试系统克服了传统仪器在数据处理、分析、显示、存储方面的限制, 解决了开发成本高、需要的仪器设备多、操作复杂、精度低、界面不友好等缺陷。同时因为 LabVIEW 是一种图形化编程软件, 因此减少了系统开发时间, 且便于扩展, 大大提高了效率。该设备已经在某单位应用, 且取得了很好的效果, 另外该设计思想也适应于其他类功放设备的智能测试。

(收稿日期: 2006 年 3 月)

## 参考文献:

- [1] 时秋兰, 赵伟, 候国屏. 基于 LabVIEW 环境开发虚拟仪器的几点体会[J]. 电测与仪表, 2001 (12): 29-33.
- [2] 潘戈. 基于虚拟仪器技术的智能仪器的开发[J]. 计算技术与自动化, 2003, 22 (3): 92-94.
- [3] 赵新民, 王祁. 智能仪器设计基础[M]. 哈尔滨: 哈尔滨工业大学出版社, 2002.
- [4] 杨明, 甘欣辉, 张景文. 智能仪器的量程自动转换设计[J]. 国外电子元器件, 2004 (5): 10-12.
- [5] 李泽勇, 王文生. 闭环霍尔电流传感器在车用电源系统中的应用 [EB/OL]. <http://www.china-power.net/dzkw/312/12.htm>.
- [6] Advantech Co, Ltd. PCI-1721 user's manual[S], 2001.
- [7] Advantech Co, Ltd. PCI-1716 user's manual[S], 2001.
- [8] 石博强, 赵德永. LabVIEW6.1 编程技术实用教程[M]. 北京: 中国铁道出版社, 2002.
- [9] 邓炎, 王磊. LabVIEW7.1 测试技术与仪器应用[M]. 北京: 机械工业出版社, 2004.