

# 基于 G/S 模式的空间信息云服务架构研究

苗放<sup>1,2,3</sup>, 姚丹丹<sup>1</sup>, 杨文晖<sup>1</sup>

(1. 成都理工大学 地球探测与信息技术教育部重点实验室, 成都 610059;

2. 成都理工大学 地质灾害防治与地质环境保护国家重点实验室, 成都 610059;

3. 成都大学 信息科学与技术学院, 成都 610106)

**摘要:** 针对多系统间的数据交换、共享问题, 尤其在大数据时代, 海量的数据增长给数据的管理带来的挑战, 提出了一种基于 G/S (General-Browser/Service-Cloud, G/S) 模式的空间信息云服务架构——面向数据的架构 (data oriented architecture, DOA)。以数据为核心, 以数据标识为主线, 通过数据注册中心和数据交换规范 XXML (specific industry markup language), 对海量异构数据进行存储、计算和管理, 实现多系统间的数据共享、访问和协同; 通过在地质灾害监测预警方面的应用, 证明了该架构作为信息技术的系统构建方法在大数据时代独特的优势。

**关键词:** 大数据; 云服务; 数据管理; 面向数据的架构; 信息技术

## Research on Spatial Information Cloud Service Architecture Based on G/S Model

Miao Fang<sup>1,2,3</sup>, Yao Dandan<sup>1</sup>, Yang Wenhui<sup>1</sup>

(1. Key Lab of Earth Exploration & Information Techniques of Ministry of Education,

Chengdu University of Technology, Chengdu 610059, China; 2. State Key Laboratory of Geohazard Prevention and

Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, China; 3. Information Science and

Technology College, Chengdu College, Chengdu 610106, China)

**Abstract:** Specific to data exchange and sharing problems between multisystems, especially in the era of big data, growth of vast amounts of data brought great challenges to data management, proposed one spatial information service cloud architecture based on G/S model, data oriented architecture. In this architecture data as the core, data identification as the main line, through the data registry center and data exchange specification XXML, huge amounts of heterogeneous data are stored, computed and management, and implement the data sharing, access and synergy between multisystems. Application in geological disaster monitoring and warning proved that as a method of system establishing of information technology the architecture have unique advantages in the era of big data.

**Keywords:** big data; cloud service; data management; data oriented architecture; information technology

## 0 引言

随着社会的发展和科学技术的进步, 数据量剧增, 已经进入大数据时代, 这些都极大的影响着计算机系统的软硬件设计与应用。工信部发布的物联网“十二五”规划上, 把信息处理技术作为 4 项关键技术创新工程之一被提出来<sup>[1]</sup>。大数据管理面临着诸多的挑战, 例如能耗、性能等。大数据还需要解决 IT 技术架构的挑战, 存储能力的增长远远赶不上数据的增长<sup>[2-3]</sup>。百度每天大约要处理几十 PB 数据; Facebook 注册用户超过 10 亿, 每天生成 300 TB 以上的日志数据; 淘宝网每天交易数千万笔, 产生约 20 TB 数据<sup>[4]</sup>。如何建立信息资源采集、处理、交换、共享、运营和服务的机制和规程, 实现分布在各种系统中的信息资源的有效采集、交换、共享和应用, 最

终达到数据的统一管理, 是大数据时代数据管理及服务的核心任务<sup>[5-9]</sup>。

## 1 G/S 模式概述

### 1.1 G/S 模式定义

苗放教授在第 303 次香山科学会议上首次提出新型空间信息网络访问 (G/S) 模式<sup>[10]</sup>, 即地学浏览器/分布式空间数据服务器群模式, 是一种基于互联网的网状空间信息服务模式。随着研究的深入发展, 又演变为地理信息浏览器/服务云模式。目前的 G/S 模式指的是通用浏览器/服务云工作模式, G 代表具有时空增强的通用浏览器, 是各类面向人类的信息终端; S 代表以数据为核心构成的服务云。

### 1.2 G/S 模式框架

G/S 模式采用以超地理信息组织交换标准——HGML 为基础的、互联网环境下的分布式网络结构, 具有两层或多层结构。G/S 模式的架构体系包括地学浏览器、超地理标记语言 HGML 和分布式空间数据服务器群。

C/S 模式、B/S 模式和 G/S 模式对比如表 1 所示。

收稿日期: 2014-10-13; 修回日期: 2015-03-19。

基金项目: 国家自然科学基金项目 (61071121); 成都市经信委科技专项项目 (201102153)。

作者简介: 苗放 (1958-), 男, 北京人, 教授, 博士生导师, 主要从事计算机技术及应用、空间信息技术及应用方向的研究。

杨文晖 (1969-), 女, 上海人, 副教授, 主要从事计算机科学与技术方向的研究。

表 1 C/S、B/S 和 G/S 模式比

特点	C/S 模式	B/S 模式	G/S 模式
空间分析	客户端进行	服务器运行	云端运行
效率限制	计算处理能力受到客户端 CPU 的限制	效率随客户端连接的增多而急剧降低甚至瘫痪	弹性
通用性	客户端通用性差	增强了客户端的通用性	各种客户端都通用
结果展现	客户端	浏览器	个性化、多样

## 2 G/S 模式与面向数据的体系结构的关系

### 2.1 二者的关联

面向数据的体系结构 (data oriented architecture, DOA)<sup>[11]</sup> 是一种用于构建大型复杂系统的分布式系统架构, 以数据为核心, 以数据标识为主线, 通过数据注册中心和基于 XML 的数据交换规范 XXML, 对海量异构数据进行存储、管理和计算。

通用浏览器/服务云 (General-Browser/ Service-Cloud, G/S) 模式, 是基于 DOA 架构的网络信息访问与服务模式, 以 HGML (hyper geographic markup language, 超地理标记语言) 为核心, 以基于 XML 的数据注册与数据交换机制为主线, 通过云服务下的海量数据管理及客户端聚合服务机制, 有效解决存储管理、服务响应、网络带宽、信息表达、操作控制、技术和分析等瓶颈问题。

DOA 研究以数据为中心的系统构建机制, 包括数据的获取, 传输, 存储, 管理, 调度, 权限控制, 数据的价值挖掘, 数据检索与服务等机制的研究。DOA 采用通用浏览器/服务云 G/S 工作模式, 二者的关系如图 1 所示。

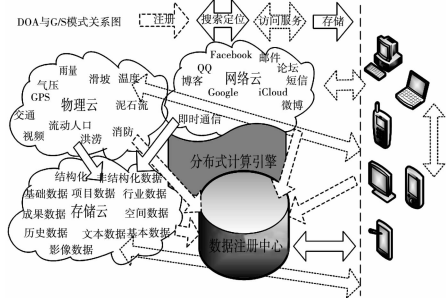


图 1 DOA 与 G/S 模式架构示意图

### 2.2 DOA 数据流程

首先来自“存储云”、“物理采集云”、“网络采集云”或客户端的数据在数据注册中心进行注册。然后用户通过向数据注册中心请求有关元数据来搜索和定位数据服务, 获得查询结果。用户根据结果对云端或其他客户端发起数据访问请求; 对方在收到请求后为用户提供数据服务。最后来自“物理采集云”和“网络采集云”的实时数据被持久化到“存储云”中, 成为历史/归档数据。

### 2.3 架构特性

DOA 架构以数据为核心, 以数据标识为主线。符合当前“大数据”在存储、管理和计算等方面的需求。数据标识 (data identifier, DI) 体系解决了海量异构数据的标识和定位问题。通过数据注册中心对海量异构数据进行管理, 独立于各种硬件平台和软件系统, 极大地降低异构系统间接口的耦合度,

能运用到多种不同的环境中, 具有较高的普适性。采用基于 XML 的数据交换规范对海量异构数据进行管理, 适用于多种数据类型。架构中的数据注册中心和分布式计算引擎能够对存储资源和计算资源进行统一管理实现了存储和计算的有机结合。

## 3 面向数据的体系结构 DOA

### 3.1 大数据服务云

云计算的核心在于其服务模式, 大数据服务云是一种新型的 DaaS 云计算服务模式, 通过对海量异构数据进行统一的存储、管理和计算, 为用户提供按需的实时/归档数据支撑服务。主要由存储云、“物理”云和“网络”云三部分构成。

存储云由多种存储系统组成, 主要用于存储各种归档、历史、离线的数据, 为数据服务和计算分析提供数据支撑。“物理”采集云通过部署在各处的多个传感器采集现实世界的物理特征数据, 将这些数据持久化到存储云中, 为事件发展状态分析和辅助决策分析等深度分析计算提供支持。“网络”采集云对微博、论坛、社交网络等数据进行实时分析, 提取地理位置等相关信息, 根据相关标识信息过滤与主题无关的数据, 快速挖掘网络中出现的热点数据, 将其以服务的方式提供给用户, 同时将相关数据持久化到存储云中。

以上“三朵云”通过数据注册中心进行管理和调度, 实现了对动态数据和静态数据的统一管理, 基于 DOA 架构的 DaaS 既能够为使用 SOA 架构、IaaS、PaaS 或 SaaS 的平台级用户提供大数据存储与计算支持, 也可以按照个人用户需求提供小型数据服务, 是一种具有发展潜力的云计算服务模式。

### 3.2 XXML 与数据注册中心

#### 3.2.1 大数据交换与共享的纽带 XXML

特定行业标记语言 (specific industry markup language, XXML) 是一系列基于 XML 的应用服务层面数据交换规范的统称, 根据各行业领域相关数据类型和特征进行设计, 其语法定义具有很强的针对性。XXML 采用 XML Schema 标准, 可定义文档中的元素与属性, 并可定义子元素与子元素数目, 以及子元素是否为空或者可否包含文本, 能够定义元素和属性的数据类型以及默认值和固定值。

特定行业领域 XXML 的建立流程如图 2 所示。目前 XXML 几个应用领域有: 应急管理标记语言 EMMML (emergency management markup language), 地质灾害标记语言 GHML (geo-hazard markup language), 国土资源标记语言 NLRML (national land resource markup language), 社交网络标记语言 SNML (society network markup language)。

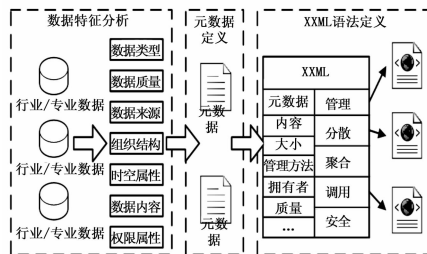


图 2 XXML 的建立流程图

#### 3.2.2 数据注册中心 DRC

数据注册中心在 DOA 中处于核心地位, 对异构系统中海量数据的管理和应用发挥着至关重要的作用。

1) 提供统一的数据注册与管理平台。DRC 用于解决分散数据集中管理的问题。通过以 XXML 为载体的数据交换规范实现分散数据的集中管理和维护, 实现海量异构数据的统一管理 with 访问。

2) 提供可靠的数据查询定位机制。DRC 为数据访问者提供数据查询定位机制, 保证客户端快速查询数据, 同时准确定位, 空间信息资源的高效访问。

数据注册中心由应用接口层、服务层和信息存储层构成, 如图 3 所示。

1) 应用接口层。为两端提供了访问数据注册中心的途径, 包括客户端接口和云端接口。

2) 服务层。提供元数据的管理和应用服务, 包括信息注册管理服务 and 目录检索服务。信息注册管理服务负责元数据的注册、更新、删除等功能; 目录检索服务向客户端用户提供信息查询支持。作为中间层, 服务层实现数据注册中心的业务逻辑。

3) 信息存储层。提供元数据的存储服务, 包括元数据存储库和元数据发布库。

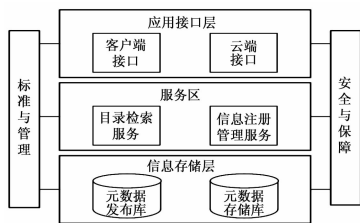


图 3 数据注册中心架构模型

### 3.3 分布式计算引擎

在面向大数据的分布式计算中, 参与计算的数据量往往大到 TB 甚至 PB 级。将计算数据分发到计算节点的过程往往会涉及到大规模的数据迁移与传输, 造成较为严重的网络阻塞、传输时延和存储空间开销。为了解决这些问题, 分布式计算引擎采用了多种不同的构建策略, 将属于不同集群的计算节点整合起来, 提高整体计算资源的利用率。

#### 3.3.1 分布式计算引擎的构建策略

1) 计算节点与存储节点融合。DOA 架构的大数据服务云中存在大量的数据服务器, 这些服务器重要用于存储数据资源, 提供数据服务, 其计算能力得不到充分的利用, 计算框架部署在这些服务器上, 再通过分布式计算引擎进行统一调度, 能够将空闲的计算资源充分利用起来, 提高服务器资源的利用率。

2) 云端计算与客户端计算融合。分布式计算引擎能够把适合在客户端进行计算的任务分发给客户端执行, 充分利用客户端的计算能力, 在一定程度上减轻服务器节点的计算压力。

3) 多种计算框架融合。DOA 架构中可以部署多种现有的分布式计算框架, 包括用于离线计算的 Hadoop/MapReduce, Spark 等, 和用于实时计算的 Storm 计算框架等。这些计算框架也通过分布式计算引擎进行统一的调度与管理。

4) 实时/离线计算融合。DOA 架构的分布式计算引擎能够同时进行实时计算和离线计算。将离线计算框架部署到存储云中, 将实时计算框架部署到“物理”采集云和“网络”采集云中, 既保证了计算能够在最靠近数据资源的节点上完成, 又实现了实时计算资源和离线计算资源的分离。

#### 3.3.2 分布式计算引擎的工作模式

分布式计算引擎能够将客户端提交的计算任务分发到合适的计算节点, 由各节点分别进行计算之后, 再将结果缓存到云端, 返回给客户端。

1) 计算任务提交, 客户端将计算任务以 XXML 的格式提交到分布式计算引擎, 提交的内容主要包括计算方法和计算数据的元信息。

2) 计算任务分发, 分布式计算引擎按照计算方法和计算数据元信息找到最合适的计算节点, 将计算任务分发给这些节点。

3) 结果数据缓存, 计算节点在分配到任务后, 按照计算方法进行计算, 并将计算结果缓存到本地或云端的其他节点上。

4) 结果数据返回, 当计算全部完成后, 将缓存在节点上的结果数据返回给客户端。

### 3.4 多终端通用浏览器

多终端通用浏览器是时一空增强的通用客户端, 采用请求一聚合的工作机制, 能够按照用户需求提供个性化服务, 其关键技术包括:

1) 时一空可视化技术。通用浏览器通过 HGML 进行可视化表达, 将具有时间属性和空间地理特征的海量数据以三维可视化的方式呈现给用户, 使用户获得数据基于时间序列和空间位置的直观感受, 极大地提高了用户体验。

2) 动态服务聚合技术。通过动态聚合服务技术, 通用浏览器能够根据用户的需求自主的从云端或其他客户端获得服务所需的各种数据, 并将这些数据以服务的方式呈现给用户, 实现客户端服务聚合。

3) 多终端协同技术。以基于数据注册中心的数据共享为基础, 对客户端实时操作数据进行注册, 其他终端可以通过 XXML 将这些数据同步到本地, 实现多终端的协同显示与操作, 为基于通用浏览器的多用户会商提供支持。

4) 智能缓存技术。通用浏览器通过基于用户行为统计分析的智能预缓存模型实现对用户兴趣点数据的提前缓存, 降低客户端的数据获取延迟, 提高用户体验。

## 4 实例验证及分析

以地质灾害监测预警平台为例, 验证 G/S 模式和 DOA 的先进性和实用性。该平台采用地学浏览器/服务云的架构, 通过数据注册中心, 利用 GHML 进行统一的数据交换, 实现对多种地质灾害信息的统一管理。

地质灾害监测管理平台中对灾害点的管理及地形三维展示如图 4 所示。

用户通过浏览器客户端可以查看三维展示的地质环境如坡度、坡向, 实时查询地灾勘测数据 (滑坡、泥石流等)。服务云从数据层面对地灾信息进行综合管理, 采用面向信息汇聚的数据调度机制, 是数据存储和服务发布端。分布式服务器群是基础设施, 实现对数据的分布式管理。

在地质灾害监测管理平台中, 涉及的数据来源于各业务部门、行业部门。从地质灾害监测管理平台的功能应用出发, 对比 SOA 与 DOA, 可以看出 DOA 在解决大数据的存储、共享、交换等问题上具有独特优势。

(下转第 1733 页)

型电站的 6/10 kV 配电网正在新起, 尤其是一些箱体变电站,

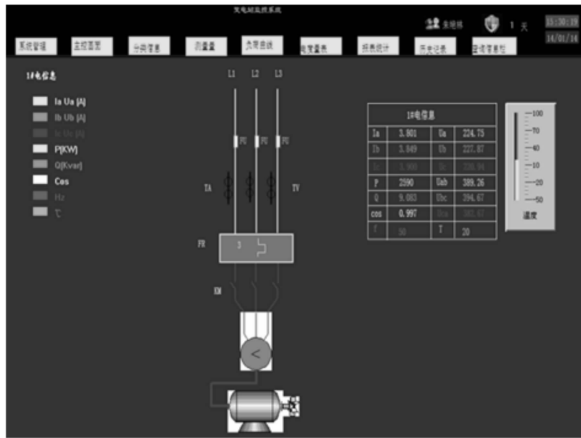


图 5 调试运行监测数据画面

(上接第 1730 页)



图 4 地质灾害三维展示及灾害点管理

表 2 DOA 和 SOA 的特征对比

	DOA	SOA
面向对象	数据	服务
描述方式	数据标识	服务文档
访问机制	注册—请求—访问	代理—发现—请求—服务
系统连接方式	数据注册中心、数据交换平台	接口、契约、协议
交换媒介	HGML	XML
平台相关性	无关	无关
适用数据类型	任何数据类型	封装的服务
对各类数据的管理	统一、一致	缺乏
对云计算和云服务的支持	能	不能
对客户个性化支持	容易	繁琐
对客户协同的支持	能	不能
网络服务模式	G/S、B/S	B/S
系统构建特点	构建简单、维护容易、扩展方便	系统依赖于功能和服务、修改和扩展繁琐
多系统开发复杂度	线性增长	指数级增长
适用范围	需要进行数据交换的复杂系统环境	需要进行服务提供的异构系

在线监测技术是电网正常稳定运行的保证<sup>[3]</sup>。基于组态王与 RS485 配电网监测系统减少了值班人员巡视频次, 做出了故障预警, 减少了故障发生, 实时准确记录配电网运行状况, 对配电网监测自动化具有很好的使用价值。

参考文献:

[1] 安丽娟. 配电网电能质量在线监测系统研究 [J]. 科技信息, 2012 (28): 266-266.  
 [2] 韦 军. 组态王和 RS-485 总线在配电监控系统中的应用 [J]. 机电信息, 2011 (14): 41-42.  
 [3] 高军丽, 王宝珠. 基于 RS-485 总线的智能分布式测控系统设计 [J]. 计算机与现代化, 2012 (3): 151-154.  
 [4] 张 伟. 基于 Modbus 现场总线技术的智能配电系统设计与实现 [D]. 南京: 南京邮电大学, 2012.  
 [5] 黄海波, 何智强, 万 勋. 高电压设备在线监测技术在智能变电站的应用 [J]. 湖南电力, 2013 (A01): 52-55.

5 结束语

大数据管理面临着诸多的挑战, 例如能耗、性能等。还需要解决 IT 技术架构的挑战, 因为存储能力的增长远远赶不上数据的增长。本文提出的面向数据的架构, 是一种基于 G/S 模式的空间信息云服务架构, 以数据为核心、以数据标识为主线, 能够实现多系统间的数据交换和共享, 并在地质灾害监测预警方面得到应用。该架构具有独特的优势, 打破了传统的以业务为核心的思路, 围绕数据展开, 解决了多系统间的数据管理问题。该架构在其他应用领域的推广是我们下一步研究的方向。

参考文献:

[1] 刘菊花. 中国工信部发布《“十二五”物联网发展规划》[EB/OL]. [http://news2.xinhuanet.com/fortune/2012-02/14/c\\_111523632.htm](http://news2.xinhuanet.com/fortune/2012-02/14/c_111523632.htm), 2012-02-14.  
 [2] 陈 军, 卢涵宇, 姚丹丹. 基于处理时间的网络地图云服务调度算法 [J]. 计算机测量与控制, 2013, 21 (7): 1987-1989.  
 [3] 孟小峰, 慈 祥. 大数据管理: 概念、技术与挑战 [J]. 计算机研究与发展, 2013, 50 (1): 146-169.  
 [4] 吴春林. 海量数据的挑战 [EB/OL]. [http://blog.sina.com.cn/s/blog\\_54bdf32501019tkp.html](http://blog.sina.com.cn/s/blog_54bdf32501019tkp.html), 2013 (3): 27.  
 [5] 李 磊. 分布式计算引擎的集群管理及负载均衡策略研究 [D]. 成都: 电子科技大学, 2011.  
 [6] O'Driscoll A, Daugelaite J, Roy D, Sleator. 'Big data', Hadoop and Cloud Computing in Genomics [J]. Journal of Biomedical Informatics, 2013, 4 (6): 774-781.  
 [7] Tang H, Zhou X, Han Y N. Next-generation broadband networks evolution of content-centric [J]. Information and Communications Technologies, 2011 (4): 38-43  
 [8] 程学旗, 王元卓. 大数据计算的技术体系与引擎系统 [J]. 高科技与产业化, 2013, 20 (4): 62-65.  
 [9] 陈宇达, 汪新庆, 刘艳梅. 地质灾害数据共享平台下共享和互操作的研究 [J]. 环境科学与管理, 2007, 32 (6): 185-189.  
 [10] 郭曦榕, 苗 放, 王华军, 等. 空间信息 G/S 网络访问模式体系架构初探 [J]. 计算机应用与软件, 2009, 26 (10): 72-74.  
 [11] chenmuzheng\_10. DOA (面向数据的体系结构) [EL/OL]. <http://baike.baidu.com/subview/649092/12822804.htm>.