

数据中心集成开发技术:新一代 GIS 架构技术与开发模式

吴信才^{1,2}

1. 中国地质大学信息工程学院, 湖北武汉 430074

2. 教育部地理信息系统软件开发与应用工程中心, 湖北武汉 430074

摘要: 解决分布式异构多源空间数据共享、不同厂家的功能共享和合作开发是当前 GIS 开发和应用领域的难题。提出了数据中心集成开发模型框架,并基于该框架提出了新一代 GIS 架构技术和开发模式,介绍了数据中心的核心理念——数据仓库技术、功能仓库技术、搭建与配置技术等,同时给出了国土数据中心的应用实例,为基于数据中心大型信息化系统解决方案的实现提供了一种新的模型。

关键词: 数据中心;数据仓库;功能仓库;搭建与配置式开发。

中图分类号: TP311

文章编号: 1000-2383(2009)03-0540-07

收稿日期: 2009-01-03

Datacenter Integration Development Technology: The Next Generation GIS Architecture and Development Model

WU Xin-cai^{1,2}

1. Faculty of Information Engineering, China University of Geosciences, Wuhan 430074, China

2. GIS Software Research and Application Engineering Center, Ministry of Education, Wuhan 430074, China

Abstract: The sharing of distributed heterogeneous multi-source spatial data and the function sharing and cooperative development among different manufacturers are tough problems in the domain of the current GIS application and development. We present a datacenter framework model for integration development, based on which the next generation GIS architecture and development model is proposed. In addition, we introduce the core technologies of datacenter-data warehouse technology, function warehouse technology, architecture and configuration technology, etc. Examples of applying the above technologies by the National Land and Resource Datacenter are presented, which provides a new solution for the implementation of large-scale information systems.

Key words: datacenter; data warehouse; function warehouse; architecture and configuration development.

0 引言

经过几十年的发展, GIS 应用的领域越来越广泛。GIS 应用系统需要访问分布在多个数据源的异构数据,也需要整合诸如文档、数据库属性表等非空间数据。如何在分布式异构环境下对空间数据及非空间数据进行有效的集成管理,成为一个亟待解决

的问题(李德仁等, 2005)。与此同时, GIS 应用系统本身的构架也变得越来越复杂,从处理来自多种数据库的异构数据,到使来自多种 GIS 平台的功能能够协同工作;从 GIS 空间分析,到复杂的领域业务逻辑的实施;从单机单应用程序到分布多服务器集群的运用等;除此之外,用户的需求不断变动,系统的设计不断调整,使系统的开发与维护成为巨大的

基金项目: 国家重点 863 项目(No. 2007AA120503)。

作者简介: 吴信才(1952—),男,博士,教授,博士生导师,主要研究方向为地理信息系统工程与应用,曾荣获“湖北省青年科技精英”、“地矿部跨世纪人才”、“国家有突出贡献的中青年专家”、“全国优秀教师”、“全国优秀科技工作者”、“湖北省劳动模范”、“长江学者首批特聘教授”等荣誉。E-mail: xcwu@263.net

挑战,用户也未真正参与到系统的开发过程中,不能自发地对新的需求做出响应,哪怕看似简单的功能也需要通过需求变更、系统软件版本升级等软件过程才能解决;随着信息化应用和发展,信息系统越来越庞大,以致任何一个软件公司都无法独立完成开发,因此,如何使多家厂商能共同合作开发同一个系统,是急需解决的问题。另一方面,服务组件技术(如 Web Service 技术)逐渐普及,面向服务的系统构架技术的初步形成,使开发具备高可伸缩性(Erl, 2004)。在这种形势下,一种崭新的革命性的技术(即基于“面向服务的 GIS 架构”的“搭建式、插件式、配置式”集成开发技术)的诞生成为可能,笔者称这种技术为“数据中心集成开发技术”,简称为“数据中心”。

1 数据中心基本概念

数据中心是多种概念及思想的综合体,从不同的视角来看,数据中心可以从以下 5 个方面进行理解:

(1)数据中心是各种数据的集成与交换中心,是集基础与应用为一体的综合开发与应用集成平台。

(2)数据中心是一种搭建平台,提供一系列的完整语义的地理数据处理功能,提供对流程、表单、应用程序界面、数据等完整的搭建方案(由一系列的程序模块及数据组成),因而也是关于应用系统的集成设计器和工具箱。

(3)数据中心是一种强兼容性的数据仓库,可以在同一个框架下,把来自不同生产厂商、不同格式、不同标准和分布在不同位置的数据统一在一个系统之下,即实现对分布式多源异构空间数据的管理能力。

(4)数据中心是一个定义完备的功能仓库,支持和管理以多种方式(组件、插件、流程、动态库、程序片断和脚本)提供的功能,并能对这些功能以一种一致的方式进行调用和执行。

(5)数据中心是一种应用集成系统,采用柔性设计理念,系统能够被快捷地搭建出来,且能适应需求的变化迅速做出调整。它根据行业的标准建立起功能与界面相分离的具有高度可定制性的应用系统解决方案,以更容易地适应需求的变动、使用户参与到设计开发过程中,可以使功能模块达到搭建级别的可复用,并实现功能重组。

总之,数据中心同时具备“资源管理器”和“系统开发器”功能。“资源管理器”管理两大资源:(1)数据资源;(2)功能资源。数据资源通过数据仓库来管理;功能资源通过功能仓库来管理。“系统开发器”提供搭建式、插件式和配置式的开发,提供对流程、表单、应用程序界面和数据等完整的搭建方案。数据中心提供适合于多种 GIS 应用领域的应用系统快速构建技术,为多领域应用系统的集成及功能复用提供手段;实现多源异构数据的统一和层次化管理;在统一的框架下实现多个地理信息系统的协调工作;支持应用方案的集成搭建和配置可视化,增强 GIS 应用系统适应需求不断变化的能力,降低 GIS 应用系统的开发难度,为开发地理信息应用系统提供基础支撑。

2 数据中心系统架构

数据中心的架构是“面向服务的系统架构”,面向服务的架构(Service-Oriented Architecture),又称为 SOA 架构(Foster and Kesselrnan, 1997; Mock *et al.*, 2002; Papazoglou and Georgakopoulos, 2003);在这架构下,无数软件制造者可将他们的研制软件功能以“服务”形式提供出来,各功能之间是相互独立的,以一种称为“松耦合”的协议机制来组合。因此理论上系统可以无限扩大,而无须担忧负荷过大。数据中心架构通过目录配置、可视化配置和搭建充分利用功能仓库和数据仓库的管理机制,基于功能仓库中已有的资源和针对具体业务开发相应业务插件,从而配置、搭建出具体业务的应用系统的解决方案,如国土资源应用系统、管网资源应用系统、电信资源应用系统等;通过运行设计的解决方案,实现多源异构数据的统一管理。基于数据仓库、功能仓库和搭建运行平台的数据中心结构如图 1 所示。

按照多层体系结构建立数据中心的总体架构,具体可以分为用户层、框架层、功能插件层和仓库管理层(吴信才和吴亮, 2006; 吴信才等, 2008)。在实际的应用中,随着开发项目领域的扩展,功能插件层是不断被丰富的,并在仓库管理层的构件仓库中被统一管理和维护;仓库管理层除了利用构件仓库管理维护用户开发的插件资源和功能资源外,还利用数据仓库负责访问存放于各分布的服务器、工作站、主机上的数据资源;在框架层,用户基于框架层

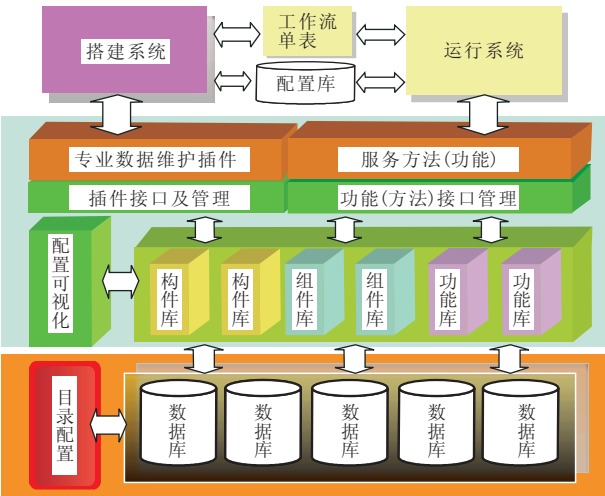


图 1 数据中心架构

Fig. 1 The datacenter architecture

进行搭建、配置式二次开发,得到具体业务的解决方案,并运行;表示层直接面向客户,提供异构数据表现和信息可视化功能,运行于 Windows 系列操作系统。数据中心的分层结构如图 2 所示。

多层结构提供了灵活的系统伸缩性,在框架层、功能插件层、仓库管理层以及表示层之间建立符合国际标准的访问接口。在实际应用部署时,可根据需求扩展系统的某个层面。数据中心采用“框架+可聚合的插件+功能仓库+数据仓库”的模式:数据中心的框架负责提供数据中心逻辑,并装载/卸载插件;插件是针对不同业务系统的特性而言,插件可以集成到框架中,通过专题激活,便可以使用插件功能。

框架看上去没有多大变化;插件应该遵循框架的接口协议。针对已经存在的功能,用户可通过功能仓库进行配置,形成新的插件,所以数据中心的插件是可聚合的。

数据中心的这种架构设计目的是为了支持分布式数据存储,提供集成化开发;提供统一数据管理平台,支持子系统相对独立运行;开发的应用系统适用稳定,能够充分满足业务需求;采用基于 GUID 资源转换和元数据过滤规则形成安全的数据仓库和安全的功能仓库的模式,保障数据的安全性;提供当前最新的搭建式、配置式和插件式二次开发技术,以最快的方式构建应用系统的。

3 数据中心核心技术

数据中心包含搭建平台的搭建能力,负责搭建、管理、维护数据仓库和功能仓库,实现业务应用系统的搭建,实现对多元异构数据的集中统一的管理,实现以开发资源的可复用。

数据中心由 3 个主要的功能集群组成,如图 3 所示:仓库系统、工作空间系统、设计管理与部署系统。仓库系统包括:元数据仓库、目录系统、功能仓库、数据仓库、资源注册与发现系统、权限系统;数据资源与功能资源和驱动体系;工作空间系统包括:地图文档及工程管理、图层及其扩展体系、集成视图和交互体系;设计管理与部署系统包括:设计器、工作流、框架界面、帮助系统、查询检索系统和插件体系。

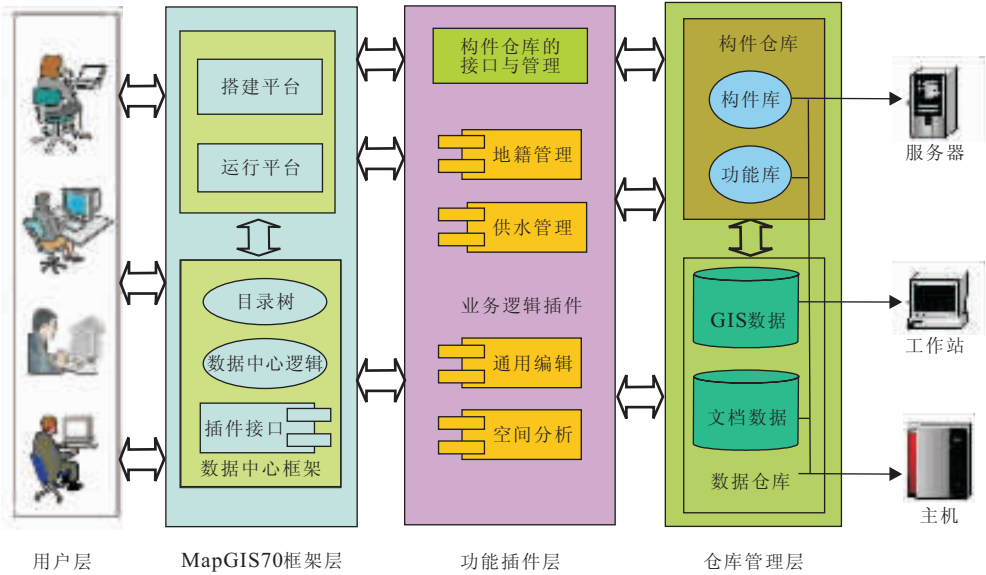


图 2 数据中心的分层结构

Fig. 2 The datacenter stratification structure

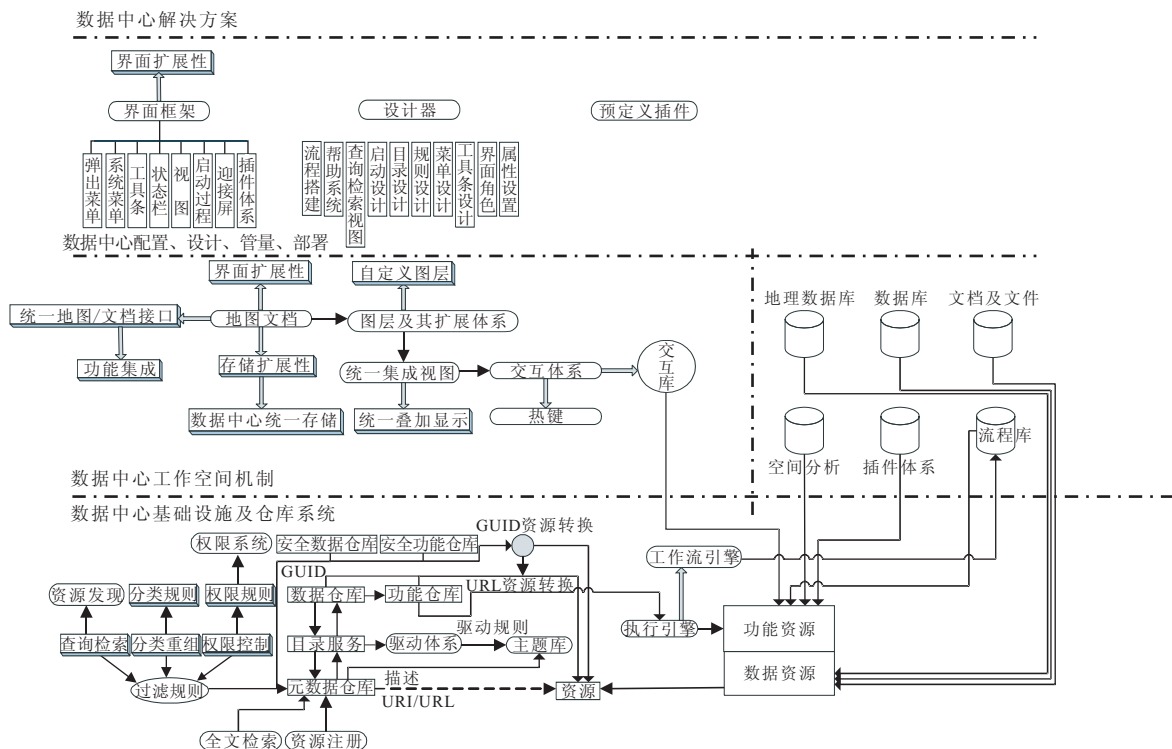


图3 数据中心功能集群

Fig. 3 The functions set of datacenter

数据中心首先应是一个数据仓库,按照一定的主题域进行数据组织,可以存放和管理各种类型的数据(信息)资源和系统自身的信息;其次数据中心是一个服务中心,可以向外提供各种服务.数据中心是采用可定制的目录树结构管理数据,实现层次化的管理.在实际应用中,可根据系统的实际应用主题域进行配置.

3.1 数据仓库技术

数据中心的数据仓库提供异构 GIS 数据的中间件、数据目录管理、数据维护及安全机制,提供数据入库管理的清洗机制和方法、数据的挖掘(提取、分析、检索)等.

数据仓库技术在目录系统上实现对数据的仓库式管理.数据仓库技术无需格式转换即可直接管理主流 GIS 软件的空间数据,通过维护数据位置的描述,根据描述可以得到该数据的位置及类型,通过合适的程序模块去访问该位置的数据.数据仓库技术实现以下目标:

(1)各种数据通过多个目录规则被分门别类的组织起来;(2)提供对各种异构数据的访问机制;(3)提供专门的元数据,部分影响目录系统在界面上的表现及定义数据的界面事件;(4)提供异构数据的目的

录配置和可视化工具;(5)提供数据维护的插件和服务方法;(6)提供查询与检索机制,方便访问用户需要的数据;(7)在数据仓库中的节点,不管是直接定位还是间接搜索出的,都可以直接支持应用.

3.2 功能仓库技术

数据中心的功能仓库是功能仓库的概念模型.它提供异构 GIS 功能管理、功能目录管理,提供功能入库管理的清洗机制和方法,以及功能的查询、合并、挖掘(功能的提取与重构),完善功能管理机制,实现功能仓库管理等.

功能仓库在目录系统上实现对功能的仓库式管理,各种功能通过多个目录规则被分门别类的组织起来;提供对各种异构功能的调用机制;提供专门的元数据部分影响目录系统在界面上的表现及定义功能的界面事件;提供查询与检索机制,在功能仓库中的节点,不管是直接定位的还是间接搜索出的,都可以直接支持应用.

功能仓库将集成大量稳定、可靠的标准组件和 MapGIS 平台的相关 GIS 插件,能够在开发软件系统时直接以拖拽的方式使用.对于用户按照数据中心提供的标准开发针对具体业务的功能,也可以放入功能仓库中,与原有的功能仓库一起形成二次开

发用户自己特有的功能仓库.

3.3 搭建与配置技术

数据中心通过自身提供的系统设计工具——数据中心设计器,构建适合多种 GIS 平台的应用系统,突破面向业务系统的开发特定业务系统及面向问题的开发陈旧开发模式,提倡以配置式二次开发和搭建式二次开发为主要开发模式,配合以少量编码的插件式二次开发(如果功能仓库资源足够丰富,那么用户可以零编码),可以快速地构建多种 GIS 应用领域的应用系统;提倡多领域应用系统的集成管理和功能复用;提倡开发的 GIS 应用系统应具备随需求不断变化的能力;提倡开发应用系统自动产品化的方法;提倡通过新的二次开发技术、搭建工具和配置工具降低 GIS 应用系统开发的难度.基于数据中心提倡的搭建和配置式开发模式具备以下特点:

- (1)响应速度快,软件的质量得到提高;
- (2)使更多的人能够参与实施、降低开发难度;
- (3)强调业务系统和功能的搭建;
- (4)普通人员经过简单培训即可上手,不需要编码;
- (5)使功能可以积累,并能够快速重用;
- (6)系统的部署、移植不过是配置文件的拷贝、覆盖,不需要做其他任何操作;
- (7)更多的精力集中于业务而非技术细节;
- (8)减少每个环节之间的隔阂,尽量使上一个环节的输出能够被下一个环节使用;
- (9)部署更加容易,影响更小、耦合度低、热插拔;
- (10)减少软件的开发量,增加软件的开发质量;
- (11)对业务流程、模型的变化能够当场修改,即时反映;
- (12)可复用性高,原有的功能资源可在数据中心框架下直接调用.

数据中心提供的开发方式注重软件功能的积累

与复用,致力于提供简单易用的开发方式,实现了快速构建行业应用和 GIS 功能的快速搭建,适合于初涉 GIS 专业的开发商、用户.

数据中心基于模型驱动架构,不仅能够快速地开发管理异构数据的应用软件,而且开发的软件系统具有高度的扩展性和可维护性,能够根据用户的需求快速地对软件系统进行调整,降低了软件实施和维护过程中的成本和风险,使软件系统能够充分满足用户的个性化需求,以及实现用户不断发展和完善特有业务领域应用系统的长期目标.

利用数据中心搭建应用系统时,可以充分地利用数据中心提供的搭建配置工具,即可快速、高效地构建各方面的应用系统.数据中心的功能仓库已经具备了各种插件、二次开发组件和流程方法库,并预置了大量的流程资源,这些都支持在数据中心框架下的直接调用;数据中心的框架还支持对应用程序的直接调用;用户可以通过 workflow、表单工具非常方便地搭建功能和界面嵌入到数据中心;数据录入及规则、目录系统负责对异构数据集中统一管理的配置实现;应用系统的系统搭建、配置,如功能仓库资源的调用,数据仓库嵌入数据中心等,最终由数据中心集成设计器来完成,如图 4 所示.

用“搭建”的方式构建应用程序是当今应用软件发展的大趋势.无论是 B/S 结构还是 C/S 结构的应用程序,大致分为以下几个层面:界面表现层、逻辑控制层、功能服务层和数据管理层.

搭建平台力图从界面表现到数据管理各个层面上实现搭建: Visual Form Designer 实现界面表现和逻辑控制的融合; MapGIS Work Flow 实现逻辑控制; MapGIS Function Library 提供基础功能服

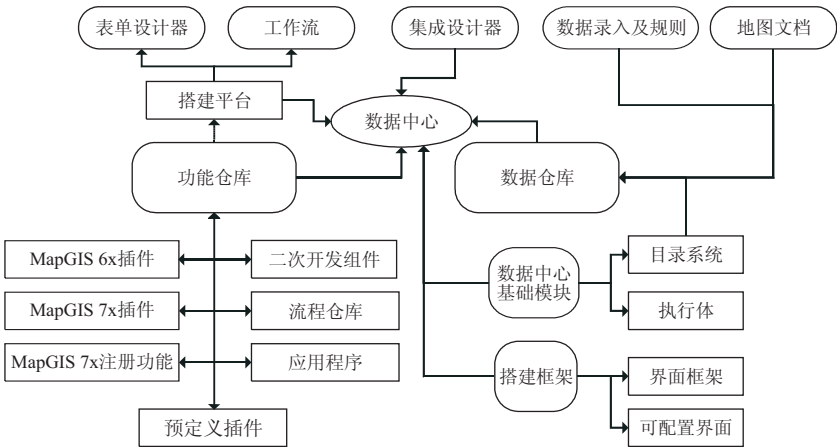


图 4 数据中心搭建流程图

Fig. 4 The datacenter building flow

务,数据管理层由 MapGIS DB Tool 或其他数据库工具来管理,搭建平台从两个方面支持表现搭建的优越性:业务流程的搭建和 GIS 功能的搭建,它提倡像积木一样生产面向具体业务的应用系统;提倡通过改变开发方式从根本上降低软件开发的难度;提倡零编码或少量编码快速构建政务系统,业务流程可灵活定制、灵活调整;提倡软件的使用人员参与到开发过程中。

4 应用实例

MapGIS 7x 数据中心可针对国土资源的数据管理(图 5)、市政管线数据管理和电信数据服务管理等各领域的业务特点,进行基于国土业务应用、市政管线业务应用和电信业务应用的领域分析,从中提取共性需求与功能,采用面向服务的架构思想,设计开发出相应的抽象功能模块,而每个功能模块又由若干基本功能单元构成,如异构数据的视图、GIS 功能、处理数据的工作空间和保障数据安全性的权限管理模块等;而对特定业务领域提供标准的功能模块扩展接口,支持特定业务逻辑的集成,特定业务的功能开发完成后,也可以纳入功能仓库中,成为功能仓库的有机部分,从而实现特定业务功能的可重用性。

数据中心功能模块与功能模块之间的连接是采用一种“松耦合”方式,“松耦合”方式是互联网的最

佳耦合方式(结构灵活、可扩展性强),它受网络环境影响最小,操作采取面向“服务”方式进行,就是把“进行数据存取操作”变为“请求数据存取服务”,“数据存取服务”是所有“服务”的特例,充分体现了“面向服务”的最新设计思想。

数据中心的目录管理是一种动态的可重配置的目录管理,可按“区域”、“专题”、“年度”次序分级配置(图 5),也可按“专题”、“年度”、“区域”次序分级配置,目录分级可达 256 级。

5 结论

在实际创建大型信息化解决方案时,一个解决方案通常包括多个业务领域的应用,产品功能和结构都非常复杂,数据中心基础平台同时支持 C/S 架构和 B/S 架构,能够极大地增加软件系统部署和运行的灵活性,因此,通过数据中心设计的解决方案在运行后得到的应用程序,能够实现一次设计同时拥有 C/S 架构和 B/S 架构软件的能力,极大地降低了软件系统的开发和维护成本。

利用数据中心集成开发平台,不需要编码或少量编码即可实现复杂的应用,可将开发周期缩短 50%~80%,可以做到“一次搭建、处处运行”,同时大大提高系统的可靠性和稳定性,从而提高了系统的开发质量。

再者,数据中心采用基于组件和面向服务的架构体系,注重体系架构的兼容性和集成性,采用门户技术提供个性化和适应性接口;采用工作流技术实现业务处理和协同办公;采用跨平台数据、信息交换技术实现信息共享;采用多媒体文档存储及管理技术实现文档数据库管理;能快速构建应用系统。

数据中心还支持多用户在线、多事务并发等应用模式,对用户提交的请求进行快速反应,针对海量数据(如 TB 级)的多用户/多系统调用等情况,通过建立缓冲池等措施,有效地改善速度,保证应用系统的正常运行。

由于数据中心采用“搭建”式开发模式,抛开了复杂的编码方式,因而维护起来很方便,基本上不需要程序编写人员维护,甚至企业用户自己都可以完成,维护期间无须停止服务器,不耽误应用系统的正常工作。

可以说,数据中心带来了软件开发模式的一场革命性的变革。

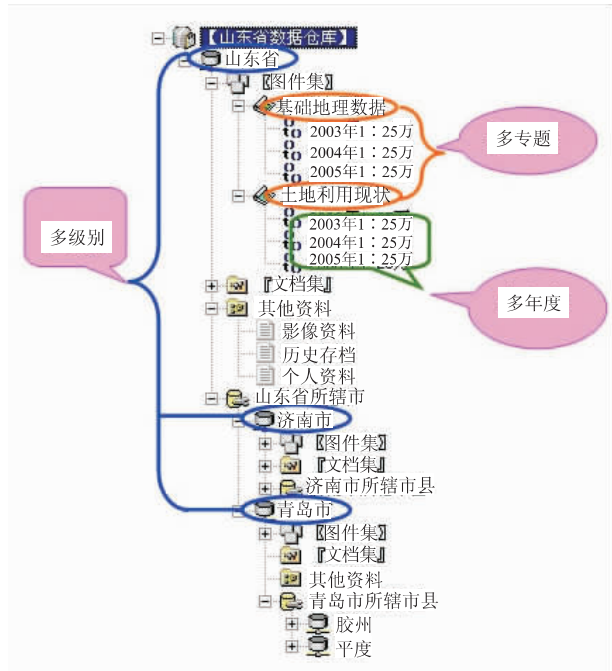


图 5 国土资源数据中心的目录管理实例

Fig. 5 A catalog of a datacenter for land resources application

References

Erl, T. , 2004. Service-oriented architecture: A field guide to integrating XML and Web Services. Prentice Hall PTR, NJ, USA.

Foster, I. , Kesselrnan, C. , 1997. Globus: A metacomputing infrastructure toolkit. *The International Journal of Super-Computer Applications and High Performance Computing*, 11(2): 115—129.

Li, D. R. , Yi, H. R. , Jiang, Z. J. , 2005. Introduction and analysis of grid technology. *Geomatics and Information Science of Wuhan University*, 30(9): 757—761 (in Chinese with English abstract).

Mock, S. , Mueller, K. , Youn, C. , et al. , 2002. A batch script generator Web Service for computational portals. In: Proc. of the 11th IEEE international symp. on high performance distributed computing—HPDC-11. IEEE Computer Society Press, Edinburgh, 422.

Papazoglou, M. P. , Georgakopoulos, D. , 2003. Service-oriented computing. *Communications of the ACM*, 46(10): 25—28.

Wu, X. C. , Wu, L. , 2006. Service-oriented distributed spatial information supporting system. *Earth Science—Journal of China University of Geosciences*, 31(5): 585—589 (in Chinese with English abstract).

Wu, X. C. , Yang, L. , Zhou, S. P. , et al. , 2008. Multimodal supported composite transportation network model. *Geomatics and Information Science of Wuhan University*, 33(4): 342—346 (in Chinese with English abstract).

附中文参考文献

李德仁, 易华蓉, 江志军, 2005. 论网格技术及其与空间信息技术的集成. *武汉大学学报(信息科学版)*, 30(9): 757—761.

吴信才, 吴亮, 2006. 面向服务的分布式空间信息支撑平台. *地球科学——中国地质大学学报*, 31(5): 585—589.

吴信才, 杨林, 周顺平, 等, 2008. 支持多模式的复合交通网络模型研究. *武汉大学学报(信息科学版)*, 33(4): 342—346.

(上接 539 页)

制了自激振荡脉冲射流 PDC 石油钻头. 现在中国地质大学长江钻头公司生产的 PDC 石油钻头已经大面积地采用了自激振荡射流技术和装置, 通过生产使用普遍获得了很好的经济效果; 钻头钻进效率提高 30% 左右, 钻头寿命也有相应的提高, 特别是这种钻头破碎的岩屑比普通钻头破碎的岩屑颗粒粒径大 2~4 倍, 这样就扩大了 PDC 石油钻头的使用范围, 特别是勘探井.

《金刚石、PDC 钻头与工艺学》这本书, 顺应时代潮流, 在知识经济时代, 新知识、新技术日新月异、层出不穷, 一旦停滞不前, 便可能被很快淘汰. 李大佛教授深知此理, 每日都觉是在逆水行舟, 70 多岁的高龄仍然研究不止、笔耕不辍, 在出版《电镀金刚石钻头研究》、《电镀金刚石钻头技术》等专著的基础上, 又推出了这本《金刚石、PDC 钻头与工艺学》. 该书对业内科技创新能力提升具有很好的参考价值.

本书注重理论性与实践性的耦合, 较好地达到

先进技术与市场需求的双向关照. 书中列举了第一手实验数据和大量的实物照片, 主要论述了金刚石地质钻头、PDC 石油钻头及其钻进机理, 介绍了坚硬、致密、“打滑”地层用电镀金刚石钻头、绳索取心电镀金刚石钻头、复合片电镀金刚石钻头及电镀 PDC 石油钻头的研制过程及其使用方法, 钻进性能, 质量测试方法及在国内外推广实验情况, 是作者 30 多年来在金刚石钻头制造方面的理论研究和生产实践的总结.

本书题材新颖, 图文并茂, 实用易懂, 是中国地质大学第一本讲授“金刚石钻头及其工艺学”细节的、新的、不可多得的专业书籍, 为有关专业师生提供了很有参考价值. 本书对于关心地质事业发展、关心钻探技术研究、关心金刚石钻头研制的各类人士而言, 都是一部值得一读的好书. 该书共三篇、23 章、95 节、50 万字, 已于 2008 年 12 月由地质出版社出版.