

地质资料数据中心基础设施建设研究

孔昭煜^{1,2}, 李晨阳^{1,2}, 张像源³, 吴轩^{1,2}

(1. 中国地质调查局发展研究中心, 北京 100037; 2. 全国地质资料馆, 北京 100037;
3. 天津市地质调查研究院, 天津 300191)

摘要: 在信息技术高速发展的今天,云计算、大数据技术的不断成熟和普及,使得地质资料数据规模急剧增长。将海量的地质资料数据更加有效便捷的,为国民经济建设和社会发展,提供基础地质信息资料数据服务支撑,同时面向社会提供公益性社会化服务,地质资料数据中心基础设施建设是重要的基础工作保障之一。本文通过对地质资料数据中心的基础设施建设进行研究,对地质资料数据中心基础环境、基础架构和管理体系进行综合研究,对地质资料管理相关行业的基础设施建设和规划起到示范作用,同时为地质资料信息化向大数据时代转型奠定必备的基础环境支撑。

关键词: 地质资料; 数据中心; 信息化; 基础设施建设

中图分类号: G252.8 **文献标识码:** A **文章编号:** 1004-4051(2018)04-0057-06

Research on infrastructure construction of geological archives data center

KONG Zhaoyu^{1,2}, LI Chenyang^{1,2}, ZHANG Xiangyuan³, WU Xuan^{1,2}

(1. Development and Research Center, China Geological Survey, Beijing 100037, China;
2. National Geological Archives of China, Beijing 100037, China;
3. Tianjin Geology Investigation & Research Institute, Tianjin 300191, China)

Abstract: With the rapid development of information technology, the continuous maturity and popularity of cloud computing and big data technologies have led to a dramatic increase in the scale of geological data. To provide more efficient and convenient geological archives, provide support for basic geological information and data services for national economic development and social development, and provide public welfare social services to the public. The construction of geological archives and data center infrastructure is an important basic work. In this paper, through the research on the infrastructure construction of the geological archives data center, a comprehensive study of the basic environment, infrastructure, and management system of the geological archives data center is conducted to demonstrate the infrastructure construction and planning of the industries related to geological archives management. The informatization of geological archives has laid a necessary foundation for the transition to the era of big data.

Keywords: geological archives; data center; informatization; infrastructure construction

0 引言

地质资料成果数据是地质工作者辛勤的劳动和智慧结晶,在具有资料、档案和科研属性的同时,也具获取成本高、难以重复获取、再利用价值高等特点。地质资料工作在地质工作环节中起到承上启下

的重要作用,即是上一轮地质工作的终结,又是下一轮地质工作起始的理论依据,也是地质工作成果社会化服务的重要展示窗口。

现阶段我国各级馆藏机构中普遍存在着对地质资料管理工作主体业务整体规划薄弱,缺乏成体系

收稿日期: 2018-02-25 责任编辑: 赵奎涛

基金项目: 中国地质调查局地质调查项目“全国地质资料汇聚与数据整理”资助(编号:121201004000150018)

第一作者简介: 孔昭煜(1981—),男,北京人,工程师,主要从事地质资料信息化基础设施建设与研究工作, E-mail: kzaoyu@mail. cgs. gov. cn。

通信作者简介: 吴轩(1981—),男,高级工程师,从事地质资料信息研究工作, E-mail: wxuan@mail. cgs. gov. cn。

引用格式: 孔昭煜,李晨阳,张像源,等.地质资料数据中心基础设施建设研究[J].中国矿业,2018,27(4):57-62.

的系统支撑,对纸介质和电子地质资料数据没有进行精细化管理,在地质资料利用方面没有发挥数字资源优势,同时也缺乏对地质资料进一步的开发利用。在大数据、云计算、超融合和人工智能等新兴信息技术不断成熟和广泛应用的当今,数字化的地质调查成果资料数据的种类和占用量呈爆发式增加,因此现有的IT基础设施中的计算资源、网络资源和存储资源等基础设施面临极大的运行压力和发展瓶颈。开展地质资料数据中心基础设施建设,利用信息技术对传统的工作流程进行梳理和升级改造,是地质资料管理工作向新时代数字地质资料馆藏馆转型的有效措施,同时也是地质调查部门响应社会公众对地质调查成果资料数据的开发利用迫切需求而推出的积极举措。全国地质资料馆已经开展并完成地质资料数据中心建设,利用信息技术全面推进地质资料业务全流程向数字化管理改造和升级,汇聚全国地质数字资源、集各学科各专业地质数据之大成、提供公益权威、开放稳定、持续发展的地质信息服务^[1-4]。

全国地质资料馆开展地质资料数据中心基础设施建设,主要由基础物理环境、网络环境、计算能力、存储备份系统、运维安全体系和标准制度所组成,完成了地质资料数据中心的基础设施支撑保障能力建设,实现其先进、高效、稳定、可靠的地质资料数据中心的有效运转。本文通过对地质资料数据中心基础设施建设研究和总结,以先进实用的技术,低成本地开展建设地质资料数据中心基础设施,并提出一套完整切实有效,具有较强可操作性的指导建议,有效地带动和促进地质资料信息化建设工作^[5]。

1 地质资料数据中心建设原则与功能

1.1 建设原则

1.1.1 需求指导原则

地质资料数据中心基础设施的软硬件支撑系统平台建设规划,需要对业务发展、数据规模发展、信息技术发展、保密安全和网络安全等多层次的不同实际需求进行调研,结合已有的工作现状进行规范、细致的规划和设计。以实际的服务需求、数据增长量、安全和基础设施资源占用需求为主导,集中统一建设、按需分配高效利用资源。

1.1.2 标准化建设原则

在开展地质资料数据中心基础设施规划和设计过程中,需严格遵循国际、国内和行业有关的标准和规范。规范术语表达方式,规范系统设计方案、规范数据库建设和规范实施方案,加强地质资料数据中心基础设施建设工作严谨、一致和规范,符合国际惯

用方法和流程。

1.1.3 成熟与先进性原则

地质资料数据中心基础设施在总体架构规划设计时,采用的基础环境、网络、计算能力、存储备份系统和运维监控系统等系统的核心技术,需在现有成熟的产品中选择,避免自行研发投入的成本高于现有产品价格。同时在选型中需要遵循符合国际和国家标准,并在国内外重大类似系统工程中成功运用的,能够代表主流技术发展方向,引领未来技术发展的两项核心原则^[4]。

1.1.4 安全保障原则

由于全国地质资料馆馆藏地质调查成果资料数据中含有部分涉密数据和敏感数据,因此,在基础设施建设规划设计中,需要对物理安全、供电安全、消防安全、系统安全、网络安全和数字资源安全等方面进行严格的规范化设计,以国家安全相关标准和规范为核心设计原则^[4,6]。

1.2 功能组成

地质资料数据中心基础设施建设主要满足全国地质资料馆主要业务和功能,分别建立了物理隔离工作网络和互联网络。其中物理隔离工作网络的建设规划是参照涉密网进行设计和建设的,满足日常数据交换、数据加工生产和到馆服务查询等支撑保障需求;互联网络主要保障面向社会公众提供公开的地质资料成果数据的在线查询和使用^[4,7]。

1.2.1 物理隔离工作区

物理隔离工作区主要由数据交换、数据生产、基础网络、基础系统、核心存储系统和核心备份系统几大功能组成,实现了全国地质资料馆内部业务数字化全面升级,同时利用信息化技术打造数字地质资料馆一体化数据管理系统,切实有效的推动地质资料信息化建设。

1.2.2 互联网区

互联网区主要由基础网络、基础系统、应用系统、私有云平台、核心存储系统和核心备份系统等功能组成,实现高效、即时、全面、精准的面向社会提供具有权威、丰富、科学的地质调查成果资料数据服务平台。同时利用现有成熟的信息技术,完成构建地质资料数据中心基础设施平台,为未来发展的人工智能和机器学习奠定必要的基础资源环境。

2 地质资料数据中心建设

2.1 地质资料数据中心机房环境基础设施建设

目前全国地质资料馆与中国地质调查局中央机房合并使用,该机房的基礎环境建设和基本运行由相关部门负责。在建设地质资料数据中心机房规划

和建设应严格遵循消防、供电、空调系统、综合布线和机房建设等有关标准和规范^[7]。

2.1.1 机房的基础环境建设

地质资料数据中心基础设施机房在现有的机房环境的基础上,完成核心功能区的进一步细化,物理隔离工作区和互联网区严格规划相应部署区域,减少存在的安全隐患。同时在现有的防尘、防水、防静电、防鼠、防结露、保温和阻燃的基本条件下,建立独立缓冲区和维修区,进一步降低防尘和防静电安全隐患,保障其运行安全平稳。

2.1.2 供电系统

地质资料数据中心基础设施机房的建设理念是绿色机房,在电能消耗上力争控制符合节能要求的范围之内。其供电方式由双路市电支撑双路 UPS 不间断电源,为机房每台机柜提供双路供电,切实保障设备运行安全可靠。

2.1.3 空调新风系统

数据中心基础设施机房由高密度、高集成率的网络设备、服务器设备和存储等设备组成,因此机房的环境温度直接影响各设备的正常运行。在温度过高的环境下设备容易出现宕机和损坏,导致的直接和间接损失较大。目前中央机房完成了行间距空调的补充和机柜背板散热系统建设等改造工程,有效的控制和降低了机房环境温度,提高了设备运行的稳定性和降低设备损坏率。

2.1.4 消防和防雷系统

数据中心机房在有关部门的规划和建设下,顺利通过消防部门的认定和验收。同时在原有的基础上对

新增机柜进行防静电、防雷击等功能的完善和优化。

2.1.5 综合布线系统

综合布线系统是地质资料数据中心的基础环境保障必备条件,同时依据有关规定需要对强弱电进行严格区分和距离管控。结合中心机房规划,调整为上走线形式,改善机房下送风存在的风道不畅和风量不足的问题。规范施工和标识清晰有助于故障的排除和维护。

2.2 地质资料数据中心网络系统建设

在建设地质资料数据中心基础设施网络系统过程中,根据全国地质资料馆主要业务特点和数据的重要程度差别,在地质资料数据中心建设和规划中需要建设两套且物理隔离的工作网络,保障数据中心的主要功能能够高效有序运行,在此基础上进行进一步细化分别建立工作网络、存储备份系统网络和设备管理网络^[7-8]。

2.2.1 物理隔离工作网

物理隔离工作网主要承担了全国地质资料馆内部的主要业务需求,并提供基础网络环境保障。主要包括利用三层网络交换结构构建核心层、汇聚层和接入层,实现各部门之间的业务流和数据流转稳定畅通;提供主要业务的数据生产加工基础环境保障;建设内部核心存储备份系统网络,保障数据流使用的光纤通信网络具有高效高速的网络环境;利用老旧设备建设网络基础设备的管理网络。各司其职的网络结构,避免了全业务混合导致单套网络吞吐和运行压力的增加,利用较少的经费投入切实有效的提高了网络吞吐能力和运行能力。图 1 为工作网拓扑图。

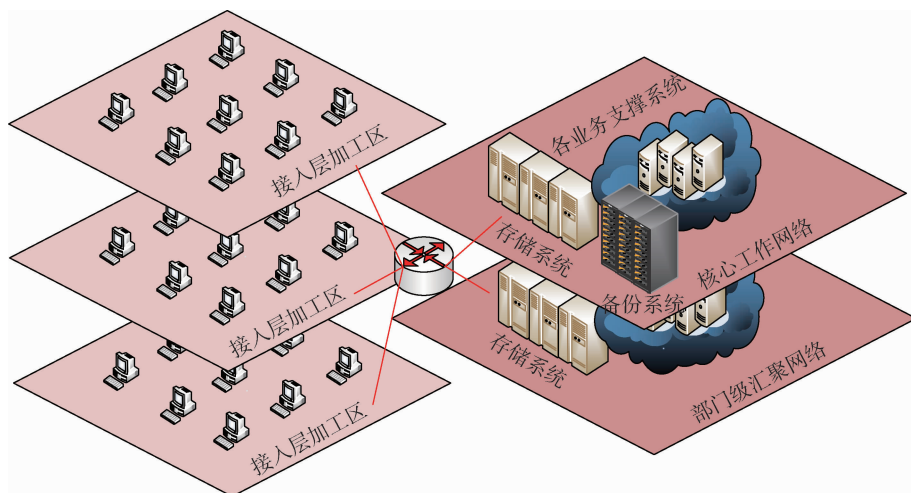


图 1 地质资料数据中心工作网拓扑图

2.2.2 互联网

互联网区主要承担了地质资料数据中心门户网

站、MapGIS 发布集群、ArcGIS 发布门户、目录数据发布系统、虚拟展馆、数据出版门户等服务系统,用

于面向社会公众,提供专业、权威、全面的地质调查成果资料数据在线公开化服务支撑平台。该区域网络主要有互联网连接区、存储备份系统网络和设备管理网络等功能组成。由于互联网链路由中央机房管理部门承担,因此仅需在现有基础上进行使用的规划和部署。存储备份系统由于对数据读写速度要求高,因此利用光纤技术完成构建光纤网络。利用老旧设备构建基础设施设备管理网络。优化清理业务需求,建立最为优化的网络结构,充分提升地质资料数据中心高效快速的社会化服务能力和稳定安全的服务环境保障。图 2 为互联网拓扑图。

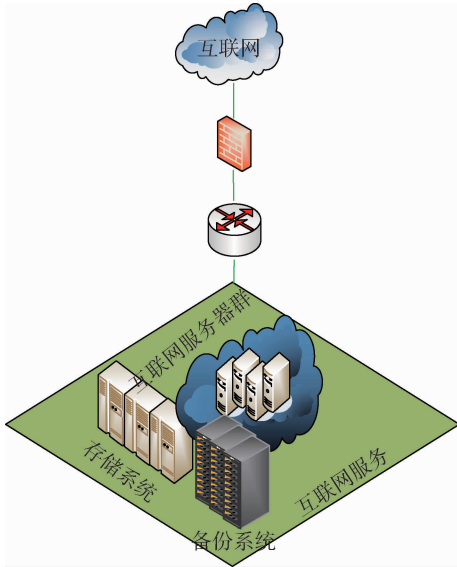


图 2 地质资料数据中心互联网拓扑图

2.3 地质资料数据中心服务器架构建设

随着信息技术不断发展,地质资料数据中心的核心理算能力目前全部由 PC 服务器承担。在有限的硬件资源下,需要系统的规划建设,实现地质资料数据中心基础运算扩展能力更加灵活,更加高效的利用 PC 服务器设备性能,有效降低采购、能耗和运营的投入成本。在对服务器进行整体架构设计中,以支撑服务的功能和系统为需求,结合虚拟化、私有云技术,将服务器整体架构分为核心基础区、私有云基础区和服务应用支撑区。根据系统对硬件需求特点,充分发挥和利用硬件资源,切实有效提高设备的使用率^[9-10]。

1) 核心基础区。该区域主要由普通 PC 服务器组成,主要承担地质资料数据中心核心底层支撑,保障服务应用支撑区各系统所需必备环境支撑能力。该区部署的系统主要包括:AD 集群、基础 SQL 集群和门户网站系统平台等核心基础类系统。

2) 私有云基础区。利用现有成熟架构,完成搭

建私有云系统,有效的保障了服务应用支撑区所需要的服务器资源。目前私有云系统通过 2 年的运行,实现了门户网站系统、数据发布集群系统、办公系统和服务支撑系统等各类应用系统对服务器的需求保障。

3) 服务应用支撑区。在私有云系统的支撑下,该区目前实现地质资料数据中心门户网站、ArcGIS 发布门户、MapGIS 发布集群、OA 系统、国家地质虚拟展馆、中国地质数据出版系统,统一身份认证系统、运行监控系统、邮件系统和部分服务应用系统。

2.4 地质资料数据中心存储备份系统建设

2.4.1 地质资料数据中心存储系统

存储系统是地质资料数据中心的核心理算平台,由于物理隔离工作网和互联网的应用场景和数据类型不同,因此两套存储系统的偏重有所差异,前者由于是存放全国地质资料馆的馆藏数据,因此注重的是存储系统的容量;后者的重点工作是提供社会化在线服务,因此注重的存储系统的使用效率。

物理隔离工作网络中的核心存储系统由 1 套 Isilon 网络存储系统和 1 套 FDS2210 分布式存储系统组成。截至 2017 年底,全国地质资料馆馆藏数据单套已经达到 195.6 TB,共计 187 187 610 档电子文件^[11]。随着新的地质资料成果汇交的有关要求规定,预计 2018 年单套馆藏数据增长量有望突破 100 TB,直接造成了存储系统的运行压力。充分发挥分布式存储系统的性能,利用闪盘和存储虚拟化技术,对存储硬件融合在一个虚拟池中,将“冷、温、热”数据根据其访问读写频率特点进行分层,充分发挥闪盘的高速读写能力,同时建立独立的存储系统的元数据层,提升数据查询的速度和效率。有效保障核心存储系统的使用效率和安全性。图 3 为物理隔离工作网存储系统。

互联网存储系统主要由 2 套 NSC 2600 SAN 存储系统和 1 套 IBM V3700 组成,其中 NSC 2600 主要承担了私有云核心存储和发布集群存储应用,V3700 承担部分归档和其它应用存储应用。利用闪盘结合分层技术,完成对私有云支撑的存储系统进行分层,大幅度提高了虚拟服务器的使用速度。以成熟的技术自行完成私有云建设,实现低成本建设私有云系统。图 4 为互联网存储系统。

2.4.2 地质资料数据中心备份系统

目前备份系统物理隔离工作网由昆腾 I6000 大型带库提供支撑,互联网由惠普 4048 小型带库提供支撑。如何提升物理隔离工作网中的馆藏数据安全,提高互联网应用系统和发布数据安全是备份系统需要保障的重点工作^[12-13]。

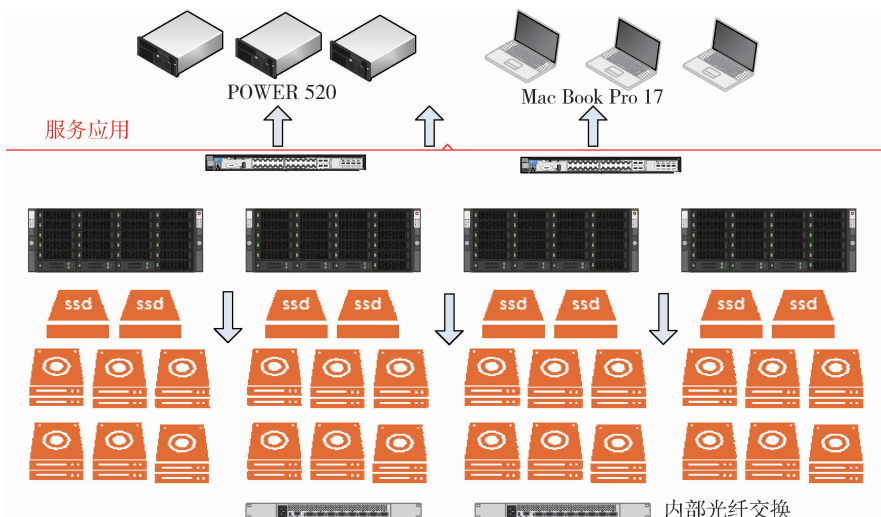


图 3 地质资料数据中心物理隔离工作网存储系统

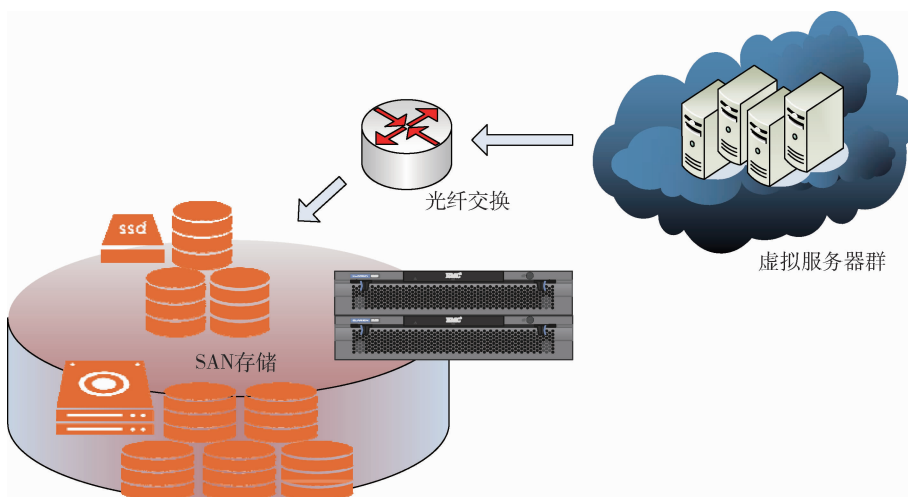


图 4 地质资料数据中心互联网存储系统

在物理隔离工作网中由于数据量大,其数据涵盖种类多因此需要制定详细的备份策略开展备份工作。目前工作网中数据量大且小文件多,因此每季度进行一次完整备份,为弥补备份完整备份周期内的数据变化,每月进行差异化备份,确保数据的高安全可靠,同时依据部分数据特点进行存档备份,用于部分废弃数据的存放,对回溯性查找提供必备的数据源。

在互联网区主要需要进行发布数据备份,同时结合相关软件实现对私有云中的虚拟服务器进行备份。目前发布数据每月进行完整备份,每周进行差异化备份,虚拟服务器每月进行完整备份,根据系统调整情况即时开展差异备份工作。

2.5 地质资料数据中心安全体系建设

地质资料数据中心基础设施建设的安全体系建设,需要应对物理安全、网络安全、数据安全、系统安

全和应用安全等 5 方面进行建设。只有保障基础设施的安全运行才能有效保障数据中心的安全可靠,主要遵循隔离、控制和保护三个基本原则^[14-16]。

1) 物理安全。严格管理物理区域的访问,合理规划隔离网络和互联网区域,通过距离和必备防护措施做到符合安全要求。建立专用维修区域和缓冲区域,降低人为原因造成的次生事故的发生,同时减少物理环境的灰尘和静电隐患。根据相关要求规划和推进涉密机房的建设工作,切实提高数据安全保障能力。

2) 网络安全。利用主流成熟的网络防护软硬件,提高入侵检测和防病毒等基础防护系统,同时依据有关要求,对网络连路和网络设备进行安全防护和安全加固,增加相应的检测手段和防护措施,有效提高互联网安全防护,保障地质资料数据中心安全可靠。

3) 数据安全。地质资料数据中心的核⼼是数据,在保障数据安全⽅⾯需要开展两⽅⾯⼯作:①数据存放和管理安全,其主要包括对数据的校验和验证⼯作,确保数据是完整可信的;②数据的使用安全,主要包括数据内容是否符合保密规定,使用过程中使用者的是否具有使用数据的权限。通过技术⼿段对数据的完整性、可用性、可控性、可追查性、机密性⽅⾯的严格管理和控制,保证了数据的安全性。

4) 系统安全。根据有关要求,定期开展漏洞扫描对系统进行安全威胁性评估,定期开展安全巡检,根据巡检结果提出修补和加固建议进⾏完善,安装和部署防病毒系统,保障服务器各系统不受非法软件和病毒的侵袭。

5) 应用安全。利⽤⾃⾏建设的 CA(数字证书)和购买的商业 CA 相结合,建⽴一套适用于地质资料数据中心基础设施环境的数字证书注册审批机制,实现通过身份认证对数据服务进⾏有效管控,通过访问控制权管理提⾼相应的安全管理机制。

2.6 地质资料数据中心标准制度建设

地质资料数据中心在规划和建设之初,需要对其总体定位、整体规划、业务规划、数据规划和安全规划等有关需求的详细分析和确定。在建设 and 运行维护过程中始终遵循标准先⾏的基本⼯作原则。在整体架构规划时制定详尽的基础设施建设标准和基础系统支撑标准,有效的实现服务接⼝标准统⼀的基础设施环境平台,为服务应用平台奠定一套良好的基础环境支撑。同时对数据格式和存储结构制定最优标准,便于后期的数据挖掘和标准化接⼝服务。在使用和维护过程中,依据标准实现使用的规范化和运行维护的规范化,有效提⾼基础设施对应用系统的标准化的全开放支撑,更加有效的降低信息系统孤岛效应。同时推动数据共享,提⾼数字挖掘能⼒和服务能⼒^[3,15]。

3 结语

地质资料数据中心基础设施建设⼯作,是实现传统地质资料馆业务向全流程信息化转变的⼀个重要契机,有效的提⾼地质资料数据的社会化服务支撑能⼒,同时也响应国家推进的数据开放共享等⼯作。地质资料数据中心历经 5 年建设已初见成效,在物理隔离⼯作网重点完成了 PB 级的存储备份系统建设,有效的保障了馆藏地质调查成果资料数据的安全可靠,并实现全国地质资料馆业务全部数字化迁移;在互联网利⽤ PC 服务器和入门级存储系统完成私有云平台建设,实现数字地质资料馆

门户、发布集群、辅助系统和备用系统及测试系统的有效支撑能⼒。通过对基础设施持建设和完善,有效的保障了地质资料数据中心各系统和功能的游⾏运转。

通过建设地质资料数据中心基础设施,有效改善全国地质资料馆的信息化⼯作环境和支撑能⼒,同时积累的建设经验是建设省级和行业级分中心的重要技术⼿段。通过该项⼯作不断积累有关经验,对省级和行业有关馆藏机构具有极⾼的借鉴和指导作⽤。在有力的保障地质资料数据中心的稳定、可靠、⾼效和安全运行的同时,基础设施建设也是有效的提⾼生产率的重要⼿段之⼀。建设和运行一套⼀流的基础设施系统平台,是⼀流的地质资料数据中心的核⼼基础,也是地质资料⾼效利⽤的核⼼保障⼿段之⼀,更是新一代地质资料⼯作者核⼼⼯作⽬标^[3-4,10]。

参考文献

- [1] 连建,颜世强,王黔驹.地质资料信息化建设趋势、问题及对策[J].中国国土资源经济,2016(1):28-32.
- [2] 卜小平,张翠光,赵亚利.全国地质资料数据中心建设方案初探[J].国土资源科技管理,2007(6):127-131.
- [3] 孔昭煜,李晨阳,范麟,等.大数据时代地质资料数字资源管理保障体系研究[J].中国矿业,2017,26(4):73-77
- [4] 李晨阳,吴轩,王新春,等.国家数字地质资料馆技术系统设计[M].北京:地质出版社,2016:188-193.
- [5] 黄少芳,刘晓涛.大数据时代地质资料信息化发展探讨[J].中国国土资源经济,2013(10):67-69.
- [6] 武允文,管锡文.数据中心规划设计方法论[J].中外建筑,2017(5):87-89.
- [7] 覃建国.大型数据中心机房基础设施的发展趋势[J].电信工程技术与标准化,2015(11):1-6.
- [8] 颜世强,张惠,王黔驹,等.地质资料服务产品基本模式构建[J].中国矿业,2015,24(10):164-167.
- [9] 高学正,李晓蕾.地质资料网络服务产品及服务价值研究[J].中国矿业,2017,26(4):64-68.
- [10] 孔昭煜,李晨阳,贾丽琼.大数据时代下地质资料数据安全保障的思考[J].中国矿业,2017,26(S1):43-46.
- [11] 齐钊宇,吴轩,商云涛,等.地质资料属性及其社会化服务价值研究[J].中国矿业,2016,25(S2):85-88.
- [12] 魏祥麟,陈鸣,范建华,等.数据中心网络的体系结构[J].软件学报,2013(2):295-316.
- [13] 尤克勤,王毅,沈昱.浅谈数据中心数据备份管理[J].中国金融电脑,2017(7):37-41.
- [14] 元海平.强化保密意识做好保密⼯作[J].中国科技信息,2011(4):200-201.
- [15] 王露,庄青.基于大数据技术的数据中心建设规划[J].信息化研究,2017(2):5-11.
- [16] 冯国礼,李蓉,王晔.浅析数据中心网络安全防护与设计要求[J].信息系统工程,2017(3):132-134.