

浅析大数据技术在测绘地理信息中的应用

孙安利¹, 丁希杰², 党元军¹

(1. 自然资源部第一地形测量队 陕西西安 710054; 2. 北京城建勘测设计研究院有限责任公司 北京 100101)

Application of Big Data Technology in Surveying, Mapping, and Geo-Information

SUN Anli, DING Xijie, DANG Yuanjun

摘要: 对于传统专业技术手段在当前测绘地理信息数据生产和行业服务中整体工作效率和服务能力不足的问题, 大数据技术无疑是重要的解决手段。对现有关于大数据技术应用方面的资料进行研究与分析, 总结大数据技术在数据存储、数据管理、数据处理及分析挖掘等方面的优势, 通过基于大数据技术平台的生产应用模式和传统生产模式相关试验结论的对比分析发现, 大数据技术的应用能大幅提升测绘地理信息生产效率与运算能力, 解决传统技术手段存在的问题, 说明大数据技术在测绘地理信息行业中具有广阔的应用前景。

关键词: 测绘地理信息; 大数据技术; 时空数据处理; 数据挖掘; 数据管理

Keywords: Surveying, Mapping, and Geo-Information; Big Data Technology; Tempo-Spatial Data Processing; Data Mining; Data Management

中图法分类号: P208

1 引言

当今社会科技迅猛发展, 信息获取手段越来越多, 信息量越来越大, 信息爆炸在推动社会进步的同时也带来新的问题: 如何更加有效地存储、管理和利用海量数据, 以满足更高的需求。大数据技术应运而生, 有效解决了这一问题, 也为众多领域提供了更广阔的发展空间。测绘地理信息作为经济建设和社会发展的基础性行业之一, 大数据技术的应用对其发展有着重要意义, 本文就大数据技术在测绘地理信息数据的存储、管理、处理、挖掘等几个方面的具体应用进行相关探讨与分析。

2 测绘地理信息的发展现状与面临的问题

近20年间, 3S技术的快速发展推动了传统测绘向测绘地理信息产业的转化, 测绘生产主体模式发生了根本性变化: 大量外业测量工作被内业地理信息处理所取代, 测绘地理信息成果向智能化应用转变, 服务范围也扩展到更多领域, 并融入到人们的

日常生活中。随着行业的发展, 传统专业技术手段在整体工作效率和服务能力等方面逐渐显现出不足, 主要表现在以下几个方面:

1) 测绘地理信息由人工采集到各类传感器快速获取, 由静态空间数据到动态时空信息, 尤其是物联网时代的到来, 数据获取手段和产生方式的变革导致数据量与日俱增。尽管计算机硬件存储设备在不断提升, 但远远无法满足地理信息前沿科技对数据存储和管理能力的需求。

2) 高分卫星遥感、无人机倾斜摄影、激光雷达等技术的发展, 测绘地理信息数据质量不断提升、数据种类日趋多样化, 大量的人工作业模式逐步被自动化处理模式所取代, 计算机处理能力与高质量数据出现不协调现象。

3) 信息化社会对各类信息产品的需求出现多样化, 测绘地理信息服务范围不断扩大, 边缘模糊甚至出现跨界现象, 传统的测绘成果和测绘技术已不能满足社会的需要。

4) 测绘地理信息数据的快速产生和大量聚集, 要求其表达能力更强且更具全局性, 急需提高数据

收稿日期: 2020-03-27

第一作者简介: 孙安利, 工程师, 硕士, 主要从事测绘工程项目生产和管理工作。

实时处理分析和信息挖掘的能力。

因此,在新的时代背景下,测绘地理信息行业面临诸多的考验。

3 大数据技术在测绘地理信息行业中的应用

大数据技术是对海量数据进行存储、管理、计算、统计和分析的一系列处理手段,处理的数据量通常是TB级甚至更大量级的数据,这是传统数据处理手段无法完成的。大数据涉及分布式计算、高并发处理、高可用处理、集群、实时性计算等技术,为解决测绘地理信息行业面临的发展瓶颈找到了出路。近几年,测绘行业的专家对有关问题相继开展了大量的研究^[1-8],本文主要对这些研究成果进行一些概述和粗浅的分析。

3.1 地理信息数据的存储和管理

目前,地理信息数据类型不仅包括各类图形、数据库等结构化数据,还包括文档、表格、像片、视频等非结构化数据。

面对海量的多源异构地理信息数据,可以构建大数据处理平台,实现数据的有效存储和管理。以Hadoop分布式处理平台为例,首先,利用HDFS构建数据管理集群,数据管理集群节点的数量可根据现有数据量规模和近期需求状况进行配置,集群中所有节点协同工作,负责数据的存储和计算。其次,利用Hbase数据工厂对大量半结构化、非结构化数据进行处理,生成结构化数据存储到Hadoop的数据库中。再次,布置Hive数据仓库,对结构化数据进行管理入库。通过Hive、Hbase的联合使用,实现结构化与非结构化2类地理信息数据的集群式存储与管理。最后,利用MapReduce进行分布式计算,通过Hadoop集群中的并行计算来实现大量数据的同步处理,由此形成一个基于Hadoop框架,包括文件系统(HDFS)、数据库(Hive、Hbase)、数据处理(MapReduce)等功能的完整的地理信息档案大数据处理平台^[1]。

Hadoop架构和HDFS的存储管理模式如图1所示。

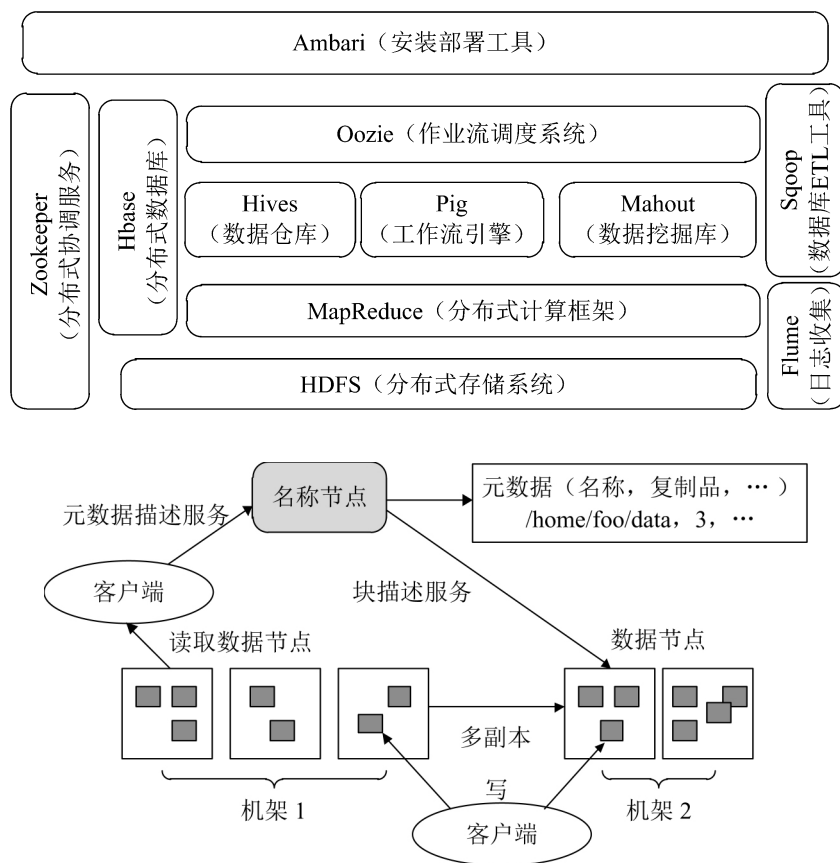


图1 Hadoop架构和HDFS的存储管理模式

Fig. 1 Hadoop Architecture and Mode of HDFS Storage Management

分布式文件系统的存储管理模式较传统的数据存储管理模式有明显的优势:

1) 硬件方面。分布式文件系统的计算机集群规模可大可小,系统兼容性和可扩展性强,可根据用户需求增减节点,且兼容廉价的硬件设备,满足动态使用需求,同时使已有资源得到充分利用。

2) 数据管理和应用方面。通过数据切块处理,可支持大规模文件的存储且易于备份;通过名称节点提供元数据服务,实现数据的快速查找和使用。

3) 系统的冗余副本机制。能在确保数据安全性的同时具备高容错性,即某一副本发生问题或机器发生故障时,其他副本和节点仍可保证提供正常服务。此外,在高并发访问数据时,冗余副本机制还可加快数据传输速度,便于检查数据错误,保证数据可靠性。

3.2 测绘地理信息时空数据的运算与处理

计算机处理能力逐渐成为测绘地理信息生产率的主要决定因素。大数据分布式处理技术,具有针对大规模数据的快速批量处理、流数据的实时流计算、图结构数据的图计算处理以及全局数据查询分析等能力,实际应用中通过搭建运行环境和部署专业处理软件可并行处理海量的、多种结构的地理信息数据,且工作效率高。杨宏山等^[2]采用多源、多时相、多分辨率、多区域影像数据先后开展影像处理、数据入库及镶嵌数据集构建等试验。其中,数据源为资源3号卫星影像,面积约270 000 km²,采用PixelGrid-SAT软件开展DEM、DOM、DSM成果制作;在超融合技术的分布式存储云主机环境下,基于影像密集匹配技术进行DSM提取、正射纠正、影像融合、影像匀色等处理,全工序处理平均每景耗时约510 min;当并发100个任务时,全工序处理平均每景耗时5~6 min;陕西省域范围内205 600 km²高达10 TB的地理国情普查影像数据的入库存储及镶嵌数据集的构建仅用时8~12 h,较传统方式效率提高了8倍以上。

笔者工作团队利用地理国情监测、三调影像一张图、倾斜影像建模等项目数据开展试验,以OceanStor 9000 V5为大数据存储平台,通过Fusion-Sphere搭建10台虚拟机建立试验环境,虚拟机配置4核中央处理器,内存为8 GB,操作系统为Linux。将多个地理信息专业软件部署到大数据应用平台进行并行处理,通过与之前单个8 GB内存4核处理器

图站处理相同数据量的效率对比发现:同样使用地理国情质检软件,在单台图站上需运行8~12 h,在大数据应用平台缩减至0.5~1 h,效率提升10多倍,解决了三调影像一张图大数据量在图站上卡顿慢、无法编辑等问题;倾斜影像建模处理与约15台图站同步处理速度相当。

可见,大数据技术的应用能大幅度提升地理信息数据的生产效率。

3.3 大数据技术在地理信息数据挖掘中的应用

在以数据为驱动的全新科学研究时代,从海量数据中挖掘隐藏价值信息成为一个重要的发展方向。大数据技术是地理信息数据挖掘的有力工具,利用大数据技术对海量的不同结构的地理时空数据进行处理分析,可得到对事物较为准确的认知和判断,为科学决策提供依据。例如,国内外一些专家利用大数据技术对全球夜光遥感影像进行了丰富多样的分析^[3-8]。在GDP估算方面,由于夜光能够客观地反映区域的繁荣程度,因此能解决统计数据中存在的误差以及空间信息量不足等问题。基于计量经济学模型,结合GDP统计数据 and 夜光影像,可以修正不同区域GDP以及GDP增长率。另外,武汉大学李熙等^[8]对全球169个国家1992—2010年的夜光影像进行时空数据挖掘,发现战争爆发往往导致夜光减少,夜光波动较大的国家发生战争的概率较高。科罗拉多州立大学的Witmer等^[7]利用夜光影像对车臣战争和格鲁吉亚武装冲突进行了评估,较好地反映了居民迁徙等情况。

由此可见,时空数据中往往隐藏着很多有价值的信息,而大数据技术对数据与事物相关性的分析方法,能降低分析难度,跨越因果得到极为丰富而准确的结果,是挖掘地理信息数据背后隐含的一个有效途径。

3.4 大数据背景下的测绘地理信息未来发展前景

在大数据时代的背景下,测绘地理信息行业将实现转型升级,数据获取将从空天地专用传感器扩展到物联网中无数个非专用传感器,显著提高数据获取能力。地理信息的用户群体不断地从专业用户扩大到大众用户,并通过网上众包方式,产生大量的自发地理信息,丰富时空信息资源。另一方面,人们能以前所未有的速度获得多维动态数据来描述和研究地球上的各种实体和人类活动,实现基于时空动态数据的感知、分析、认知和变化监测,在人类社会

可持续发展中发挥越来越大的作用。时空大数据管理分析功能的应用如图2所示。

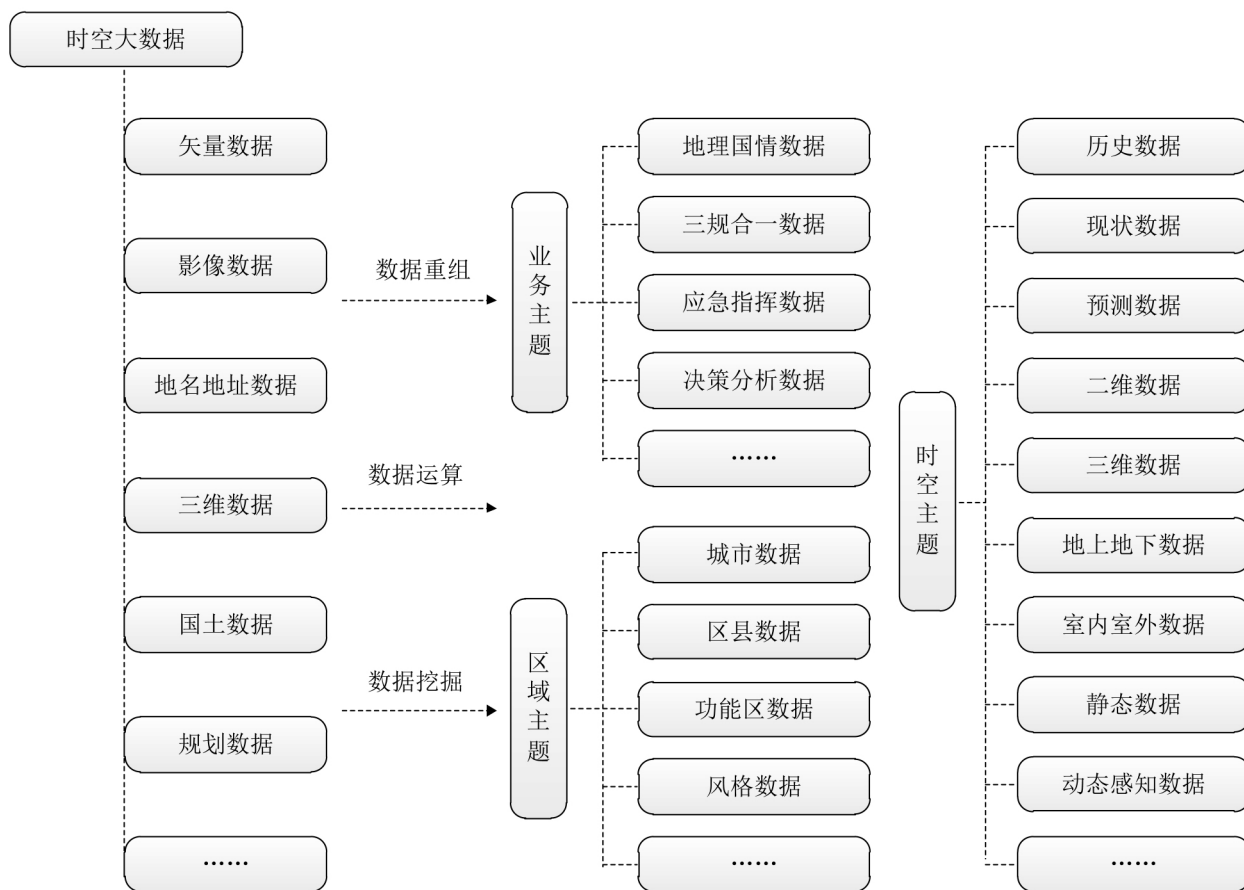


图2 时空大数据管理分析功能应用

Fig.2 Application of Big Tempo-Spatial Data Management and Analysis

4 结束语

综上,大数据技术不仅能解决测绘地理信息中海量数据的存储、管理问题,提高数据处理效率,增强分析和服务能力,未来在地理信息行业中具有更广阔的应用空间。因此,加快推进大数据技术的应用,在提高工作效率和服务能力的基础上,开展深入研究并不断创新,构建大数据地理信息时空大平台,为灾害预警、环境治理、资源保护、社会管理等方方面面提供强有力的技术支持,助推社会快速发展。

参考文献

- [1] 胡瑛. 大数据处理技术在地理信息档案管理中的应用[J]. 测绘通报, 2016(9): 112-114.
- [2] 杨宏山, 闫正龙, 白穆. 陕西时空大数据资源体系构建关键问题研究[J]. 测绘科学, 2019, 44(12): 184-188.
- [3] 李德仁, 马军, 邵振峰. 论时空大数据及其应用[J]. 卫星应用, 2015(9): 7-11.
- [4] Henderson J V, Storeygard A, Weil D N. Measuring Economic Growth from Outer Space[J]. The American Economic Review, 2012(2): 994-1028.
- [5] Chen X, Nordhaus W D. Using Luminosity Data as A Proxy for Economic Statistics[J]. Proceedings of the National Academy of Sciences of the United States of America, 2011, 108(21): 8589-8594.
- [6] Agnew J, Gillespie T W, Gonzalez J et al. Baghdad Nights: Evaluating the US Military 'Surge' Using Nighttime Light Signatures[J]. Environment and Planning A, 2008, 40(10): 2285-2295.
- [7] Witmer F D W, O'Loughlin J. Detecting the Effects of Wars in the Caucasus Regions of Russia and Georgia Using Radiometrically Normalized DMSP-OLS Nighttime Lights Imagery[J]. GIScience & Remote Sensing, 2011, 48(4): 478-500.
- [8] Li X, Chen F, Chen X. Satellite-Observed Nighttime Light Variation as Evidence for Global Armed Conflicts[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2013, 6(5): 2302-2315.