

谢榕,刘亚文,李翔翔. 大数据环境下卫星对地观测数据集成系统的关键技术[J]. 地球科学进展, 2015, 30(8): 855-862. doi: 10.11867/j.issn.1001-8166.2015.08.0855. [Xie Rong, Liu Yawen, Li Xiangxiang. Key technologies of Earth observation satellite data integration system under big data environment[J]. Advances in Earth Science, 2015, 30(8): 855-862. doi: 10.11867/j.issn.1001-8166.2015.08.0855.]

# 大数据环境下卫星对地观测数据集成系统的关键技术\*

谢榕<sup>1</sup>,刘亚文<sup>2</sup>,李翔翔<sup>3</sup>

(1. 武汉大学国际软件学院,湖北 武汉 430079; 2. 武汉大学遥感信息工程学院,湖北 武汉 430079; 3. 航天恒星科技有限公司,北京 100086)

**摘要:** 建立卫星对地观测数据集成系统是遥感卫星数据信息资源有效管理与应用的重要手段。从我国对地观测重大需求以及前沿科学问题入手,提出大数据环境下卫星对地观测数据集成系统建立中亟待解决的关键技术,包括大容量异构对地观测数据集成的语义技术、基于网格的遥感图像快速处理技术、遥感大数据深度分析技术、多数据中心协同处理及云平台技术,为实现集成卫星图像、地面观测数据和模拟模型的元数据管理、几何精度纠正和卫星数据质量评价、海量卫星图像数据的空间分析与知识发现、分布式高性能卫星图像数据管理和归档等基本功能,为解决海量卫星数据分布式存储与计算、数据集成与互操作、空间数据分析与地学知识发现提供新思路、新技术与新方法。

**关键词:** 卫星观测数据集成系统; 语义技术; 网格计算; 大数据深度分析; 云平台

中图分类号: P208 文献标志码: A 文章编号: 1001-8166(2015)08-0855-08

## 1 引言

世界各国相继发射各类遥感卫星,全球对地观测高分时代到来,卫星遥感对地观测范围越来越广泛,覆盖水与能量循环、陆地、海洋等多方面<sup>[1]</sup>,并由此呈现出“三高”特点,即高空间分辨率、高时间分辨率和高光谱分辨率<sup>[2]</sup>。在经济全球化和网络信息技术飞速发展的背景下,快速获取、高效处理和有效应用对地观测数据成为世界各国的共同需求<sup>[3]</sup>。面临卫星遥感对地面观测所形成的地球空间海量信息流,一方面如何对卫星图像、地面观测数据以及各种模拟模型等多源、异构、不同平台的数据

进行有效集成,另一方面如何将其中未经加工的数据转变成可供用户直接利用的有价值信息<sup>[4,5]</sup>,建立卫星对地观测数据集成系统正成为遥感卫星数据信息资源有效管理与应用的重要手段<sup>[6]</sup>。

### 1.1 对地观测基础科学问题

在对地观测基础科学问题<sup>[2,7]</sup>中,卫星图像数据与地面观测数据的有效集成方法是其中主要问题之一。从认知科学来看,研究很大程度上沿着“信号—数据—信息—知识”的数字化信息化过程展开,具有显著的信息学基本特征。同时,从知识工程角度来看,对地观测活动从海量遥感数据的获取、存储、处理,到地学知识分析、发现、分发,也体现出

\* 收稿日期: 2015-04-07; 修回日期: 2015-07-01.

\* 基金项目: 中央高校基本科研业务费专项资金项目“面向卫星对地观测数据集成系统的大数据应用关键技术”(编号: 2042014k0297); 中国航天科技集团公司卫星应用研究院创新基金项目“卫星观测数据集成系统的建立”(编号: 2014\_CXJJ-YG\_02)资助。

作者简介: 谢榕(1968-),女,福建泉州人,教授,主要从事时空数据挖掘研究. E-mail: xierong@whu.edu.cn

“信号传输—数据获取—信息处理—知识呈现”的数据处理生命周期。因此,除地球科学方法论以外,认知理论、信息学、知识工程方法论也对地观测的研究与应用起着十分重要的指导意义与推进作用。

## 1.2 大数据下对地观测研究的技术难点和瓶颈

大数据<sup>[8-11]</sup>的出现给对地观测研究带来了全新挑战、发展机遇和解决思路<sup>[12]</sup>。超大规模数据、海量数据、大数据可看作是三代数据管理技术的标签<sup>[13]</sup>。“超大规模”数据研究重点是数据模型、事务管理、索引技术与查询优化等方面。“海量数据”主要考虑各种非结构化数据有效管理、多数据源集成等问题。而“大数据”的技术难点和瓶颈<sup>[14,15]</sup>归纳起来主要表现为以下 3 个方面,需要研究更有效、更实用的大数据管理与分析技术。

(1) 数据量的膨胀。遥感平台技术的进步使得获取的观测数据量大幅度增大。以 ZY3 卫星为例,在轨工作期间每天获取的观测数据量可高达 10 TB 以上。同样,当前大量传感器部署在卫星、飞机等飞行平台上,未来 10 年全球天、空、地部署的数百万计传感器每天获取的观测数据将超过 10 PB,呈现出显著的大数据“4V”特征,即海量数据规模(Volume)、多样数据类型(Variety)、动态数据体系(Velocity)和巨大数据价值(Value)。从对地观测过程的信息学特征可以清晰地看到此领域大数据的发展趋势,需要解决对地观测数据量增大所带来的遥感大数据存储与处理问题<sup>[12]</sup>。

(2) 数据深度分析需求的增长。遥感技术发展初期,专业人员通过目视判读对信息进行解译及修正。当观测数据量较小时,传统数据挖掘手段已成功解决一定地学知识的定量查询与分析以及单一内容应用服务,但它们不能满足日益增大的观测数据量和日益复杂的应用模式需求。在数据维度和规模不断增大、信息提取精度不断提高的情况下,层级深度与复杂性也随之增大。在大数据深度分析(Deep Analysis)的挑战下,传统数据挖掘技术的扩展性遇到了前所未有的困难<sup>[16-18]</sup>,对 PB 级以上的大数据分析还需要研究新的方法<sup>[19]</sup>,依赖于计算机处理的数据深度分析新模式的出现以及以数据为核心的知识发现方法来解决分析较大地理尺度上的复杂地学问题。

(3) 多源数据的高度异构与分散自治。对地观测数据集成的最终目标是建立能使用户直接获取有价值信息的卫星集成数据管理与归档系统。信息化与网络化的飞速发展与深入应用,遍布全球的多个

数据观测中心都积聚了巨大的海量数据,它们高度异构、分散自治和动态更新。集成系统的建设涉及到众多领域、众多部门,目前这些子系统、组件和服务被独立开发和部署,协调困难,难以实现卫星遥感信息资源在领域之间和部门之间的交流与共享。如何提供一个高效的数据管理与信息整合的途径或平台,支持分布式环境下这些大规模数据的逻辑关联表达、语义集成、协同综合管理以及共享归档成为迫切需要和亟待解决的难题。

## 1.3 国际卫星标准计划

国际标准组织 ISO/TC211、OGC(Open GIS Consortium)等正在制定一系列地理信息元数据、卫星图像相关的标准计划<sup>[20-21]</sup>来有效地管理地理信息,如 ISO 制定了 19115 地理信息元数据(Geographic Information-Metadata)、19115-2 地理信息元数据 Part II: 影像和栅格数据的扩展(Metadata Extensions for Imagery and Gridded Data)、19130 地理信息影像与栅格数据的传感器数据模型(Sensor Data Model for Imagery and Gridded Data)、19139 地理信息元数据-XML 模式实现(Metadata-XML Schema Implementation)等;OGC 制定了 Topic 15 图像使用服务(Image Exploitation Service)、Topic 16 图像坐标转换服务(Image Coordinate Transformation Service)以及地球影像(Earth Imagery Case)等。然而这些标准并不能完全满足卫星数据集成系统开发的需要,具体表现在:

(1) 卫星数据模型尚缺乏表达卫星图像数据的地理定位、地理参考等语义信息,特别是缺乏一个统一的卫星图像元数据模型来描述和集成卫星图像、地面观测和仿真模拟等数据。

(2) 缺乏卫星影像与地面栅格之间的空间对应关系的定义,不能满足地面坐标转换的几何纠正服务功能和精度的要求。

(3) 缺乏标准化卫星集成数据仓库和知识库,以及集数据集成、空间查询与知识发现为一体的卫星数据分析功能,不能满足为用户自动提取有用信息的应用服务。

## 1.4 我国卫星应用现状及主要问题

在我国,卫星遥感信息已成为不可或缺的战略资源和经济资源,未来 5~10 年自主遥感卫星数据将呈爆炸性增长,遥感卫星应用产业蕴藏着巨大的跨越式发展机遇<sup>[22]</sup>。我国国家重大专项项目“高分辨率对地观测系统”<sup>[23]</sup>把实施目标确定为全面提升我国自主获取高分辨率观测数据的能力,加快我国

空间信息应用体系的建设,推动卫星及应用技术的跨越发展。然而,国内对地观测系统建立以及对地观测集成标准化建设的严重不足也直接影响了自主遥感卫星数据的开放利用,存在卫星数据“不能用、不好用、不会用、得不到”等现实问题<sup>[24]</sup>,迫切需要对多源卫星和航空遥感数据、不同平台卫星遥感数据之间、卫星数据与地面观测数据之间进行整合,建立完整的卫星应用技术支撑体系。

针对对地观测的重大需求以及前沿科学问题,结合当今计算机科学与技术领域中大数据应用的最新技术,本文提出大数据环境下卫星对地观测数据集成系统建立与应用中亟待解决的关键技术。

## 2 卫星对地观测数据集成系统及其总体技术框架

### 2.1 卫星对地观测数据集成系统的定义与内涵

面临卫星对地面遥感所形成的地球空间海量信息流,对卫星图像数据、地面观测数据以及仿真模型等多源、异构、不同平台的数据进行有效集成,建立能反映地面参数时空变化与信息关联的卫星图像知识库,面向海量卫星数据开发集数据集成、空间查询与知识发现为一体的卫星大数据分析功能,并进行几何精度纠正和卫星数据质量评价,最终形成能使用户直接获取有价值信息的分布式高性能的卫星集成数据管理与归档云平台。

### 2.2 基于大数据技术的系统总体技术框架

提出如图 1 所示的基于大数据技术的卫星观测数据集成系统总体技术框架。卫星观测数据集成系统包括标准化卫星数据集成元数据管理、几何精度纠正和卫星数据质量评价、海量卫星遥感图像数据分析与知识发现、分布式高性能卫星图像数据管理与归档云平台这些基本功能。在系统建立过程中贯穿运用大数据技术体系(包括大数据存储与管理、大数据计算模式、大数据分析挖掘),其中关键实现技术包括大容量异构对地观测数据集成的语义技术、基于网格的遥感图像快速处理技术、卫星遥感大数据深度分析与地学知识发现技术、基于共享知识库的多数据中心协同处理及云平台技术。充分利用国际标准组织 ISO/TC211、OGC 等一系列卫星数据国际标准计划以及国内标准,对它们进行扩展与集成,使得通过建立共同的标准和协议联接全球分布式网络化卫星遥感数据库,确保卫星数据与服务的兼容性和互用性,使所有的卫星遥感数据都可以通过网络实现规范化共享与应用。

### 3 大容量异构对地观测数据集成的语义技术

对卫星图像、地面观测数据、模型模拟结果等异构数据源数据以及不同数据归档系统之间数据进行有效集成,可建立如图2所示的标准化卫星数据集成

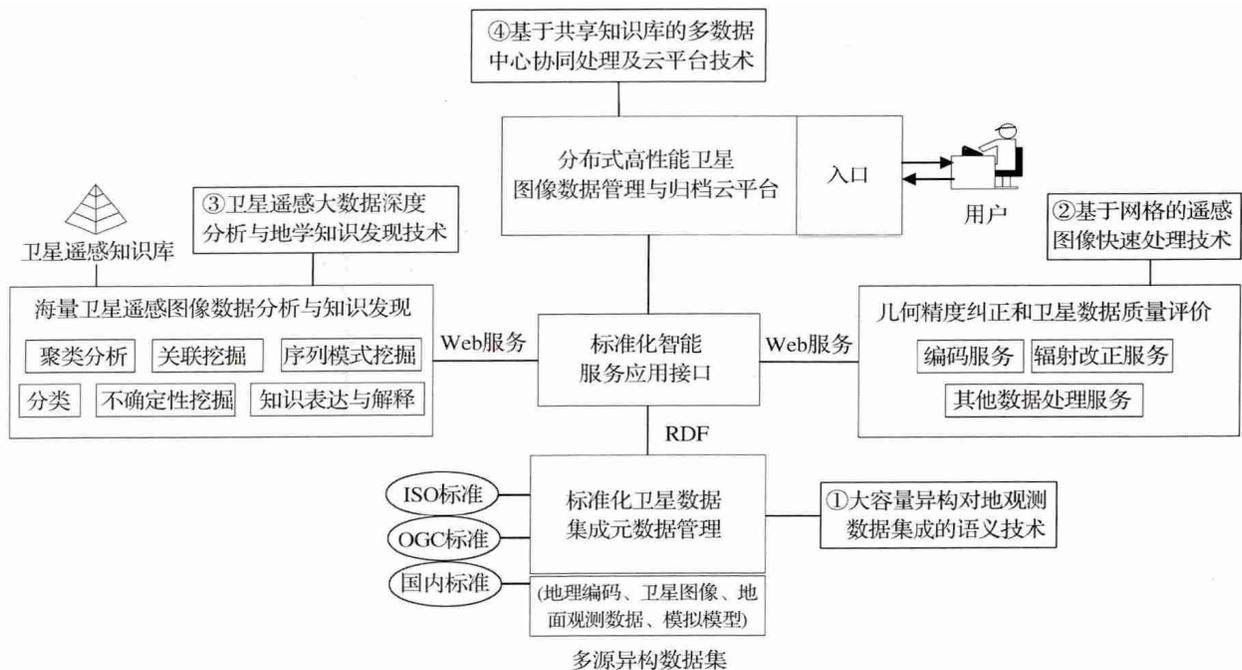


图 1 卫星观测数据集成系统及其大数据技术应用的总体技术框架

Fig. 1 Overall technical framework of satellite data integration system based on big data technologies

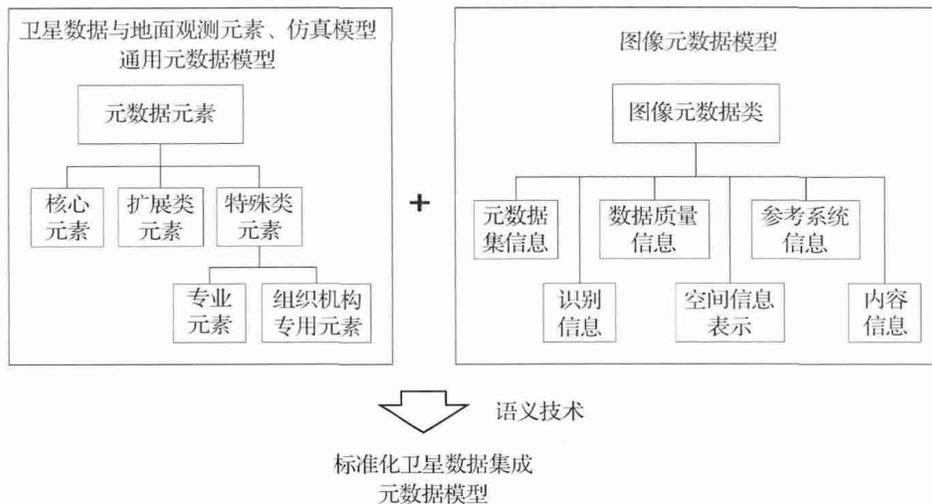


图 2 基于语义技术的大容量异构对地观测数据集成

Fig. 2 Massive heterogeneous Earth observation data integration based on semantic technologies

成元数据模型。基于时空间信息认知模式,通过语义技术,将卫星数据、地面观测数据及仿真模型相集成的通用元数据模型与图像元数据模型进行语义集成,实现对卫星元数据的高效管理。

### 3.1 时空间信息认知模式

从时空间信息的认知机理出发,建立一种形式化数据结构方式来表达概念的内涵和外延以及概念与概念之间的不同层次的抽象关系,同时描述时空数据时间、空间概念的形成、时空概念的结构关系。利用形式化的理论和方法,表达与描述卫星遥感时空数据分析过程中时间、空间概念的形成、时空概念的结构关系,反映遥感数据的时空特点,形成统一框架下“概念—关系”为中心<sup>[25]</sup>的认知模式与语义模型。

### 3.2 卫星遥感语义模型

地理定位信息是定义图像数据地理定位的重要信息,但这些信息在 ISO 19115、19115-2 中没有得到定义,而 ISO 19130 仅支持地理定位和传感器特性。因此,为了把图像数据运用于地理信息,并能有效地描述图像元数据,需要对 ISO 元数据标准 (ISO 19115、19115-2) 进行扩展,并结合 ISO 图像标准 (ISO 19130) 中的地理定位信息和传感器特性。在 ISO 19115、19115-2 和 19130 的基础上,开发卫星数据集成通用元数据模型以及图像元数据模型。通过定义元数据元素公共集、元数据的定义和内在的关联以及元数据的扩展,运用 Protégé 工具对卫星数据的抽象结构和内容进行描述,并将不同来源的异构数据映射成一种规范化形式的本体数据类型,构建

卫星遥感语义模型。

### 3.3 高效卫星元数据管理

大容量异构卫星数据的集成,需要解决卫星数据目录功能复杂性,因此建立标准化卫星数据集成元数据模型,对卫星遥感、地面观测、仿真模型等异构数据源以及不同数据归档系统之间数据进行有效集成。一个完整的元数据模型可包含多个元数据包<sup>[26]</sup>,包括主类定义、支撑类定义和描述图像的扩展类定义和特殊类定义。每个元数据包包含一个或多个元数据实体。元数据实体由一系列元数据元素组成。元数据元素可包括 3 类成分,即核心元素、特殊类元素和扩展类元素。其中,核心元素定义每个数据集应该包含的元数据最小集;扩展类元素则定义描述图像的扩展元数据;特殊类元素包括专业元素和组织机构专用元素。图像元数据类包括元数据集信息、识别信息、数据质量信息、空间信息表示、参考系统信息、内容信息。

在此基础上,采用分布式数据管理架构<sup>[27]</sup>,将元数据分散在多个节点上,以目录为粒度对元数据进行划分,并根据集群负载状况建立目录子树,实现元数据在集群中的合理分布与存储,解决元数据服务器性能瓶颈问题,提高可扩展性。

## 4 基于网格的遥感图像快速处理技术

快速生产标准遥感产品需要开发系统具有大吞吐量的、高精度的以及自动化的数据处理能力。海量遥感数据的高效处理及其标准产品的快速生成属于数据密集型的计算工作,而传统计算模式无法满

足这种遥感图像实时快速处理的应用需求,因此,需要充分利用庞大的网络计算资源,通过集群计算、分布式处理等技术来实现网络化大容量数据处理及多机分布式并行处理。基于网格计算模式,开发基于网格计算的图像处理中间件,以及影像处理算法和数据质量评价,可集中实现对海量遥感图像的快速处理及应用。

#### 4.1 网格计算模式

利用网格技术<sup>[28]</sup>可以把分散于不同地理位置的计算机集中起来组织成一个虚拟超级计算机,为完成数据密集型的计算任务提供高吞吐量、高性能的计算环境;同时充分利用网络上一些闲置资源设备及其处理能力,完成传统计算模式下难以完成的各种大数据量的计算任务,保证卫星数据快速处理以及标准产品生成。

#### 4.2 图像处理中间件

根据用户具体任务以及网格计算资源实际情况,对图像处理任务进行分割管理<sup>[29]</sup>,并将分割任务及其执行任务所需程序和参数提交给中央管理服务服务器,通过中央管理器分发给网格计算资源中各个节点。完成图像处理,再将各个计算节点上作业的计算执行结果返回到中央管理节点服务器进行数据合成。在处理过程中,还需要监控网格平台中计算资源的状况,包括工作状态、闲置状态以及各节点上作业执行情况等。

#### 4.3 影像处理及数据质量评价

由于遥感平台运动、地球自转等因素影响,卫星遥感对地观测所获得的影像会在几何上产生形变、在灰度上产生衰减。为了能从对地观测数据中精确提取所需地理空间信息,必须对获取的遥感影像进行影像处理,因此开发影像处理算法,包括遥感影像精确几何纠正算法和辐射校正算法,同时建立卫星数据质量评价体系。

### 5 卫星遥感大数据深度分析与地学知识发现技术

在数据量增大、信息提取精度高等需求下,从对地观测数据和时空信息中发现地学知识,需要解决海量遥感数据深度分析的新模式问题。目前,以Hadoop-MapReduce为代表的非关系数据分析技术,以其适合非结构化数据处理、大规模并行处理等突出优势,在海量数据存储与分析应用领域中取得了重大进展,已成为大数据分析的主流技术。尽管如此,Hadoop在应用性能等方面仍存在问题,其

编程模型处于较低层次。针对不同数据分析任务,需要开发不同MapReduce程序进行处理,使系统具有高扩展性大数据分析能力。在进行海量卫星遥感数据深度分析与地学知识发现时,可建立反映地面参数时空变化信息及其关联的卫星图像数据仓库,开发基于Hadoop的高扩展性数据分析算法以及统一框架的时空分析具体任务,包括聚类、关联、分类、时序分析、不确定性挖掘以及知识表达与解释。

#### 5.1 卫星图像数据仓库

从应用深度上,可将遥感大数据分析划分为3个层次空间,即:①数据空间。在该空间上利用现有数据库管理系统的查询检索功能,进行基于关键字或字段的信息查询,实现联机事务处理。②聚合空间。从时空数据集中获取满足空间与时间约束的时空对象集合,在该空间上利用聚集运算,并结合多维分析和统计分析,实现联机分析处理,以提供决策参考的统计分析。③影响空间。按照相似性的聚类,发现关联性、相似时序、分类结构以及不确定性知识等。按照这种层次空间建立卫星图像数据仓库,并从数据仓库中发现隐含的有用信息。

#### 5.2 基于Hadoop的高扩展性数据分析算法

针对大数据计算模式和处理环境,利用Hadoop<sup>[30]</sup>,并结合MapReduce编程模型,可采用一种基于数据本地化计算的数据分析技术<sup>[31,32]</sup>,包括数据分布策略和数据挖掘算法2个方面。

数据分布策略将逻辑相关数据聚集存储在相同节点上,对其特定属性进行哈希操作,使得属于同一哈希分区的数据以及属于不同数据表但具有相同哈希分区序号的数据能够集中存放在同一节点上。哈希技术将数据表在机群上进行分布,以分区作为处理单位,分析算法只需执行Map任务,进行分析处理并直接完成结果输出,从而可以避免Reduce操作带来的巨大时间开销,大幅度提升连接查询与分析处理的效率。

然后在MapReduce上开发增量式模式挖掘快速算法,针对对地观测系统在时间粒度、空间尺度、语义层次上的时空聚类、时空关联等特征,通过遥感卫星数据的时空分析,提取面向不同时空知识类型与形式的高层时空模式,建立聚类、关联、时序、分类、不确定性等挖掘为一体的统一模式知识发现体系框架,实现高效的模式分析与挖掘以及知识表达解释。

## 6 基于共享知识库的多数据中心协同处理及云平台技术

对地观测活动的最终目标是以满足用户需求为导向,为不同应用需求的用户提供有价值的卫星数据产品和信息资源。因此,通过建立分布式多中心计算环境,实现海量卫星数据分布式存储与共享,使用户能从不同节点方便地获取所需数据、并能直接获得数据分析与知识发现中价值信息。实现该目标的关键在于建立共享知识库、多中心协同计算以及分布式高性能的卫星图像数据管理与归档。

### 6.1 共享知识库

通过卫星数据管理中心主节点的一级知识目录(即共享知识库)及操作工具,从宏观上引导用户使用所发现的信息。同时通过常驻在各个分节点上的二级知识目录,提供详细信息的线索,使用户能进一步了解信息,确定需要获取的信息内容、获取途径和方法,并支持通过网络传输查询结果。对内部用户,通过知识目录及操作工具,既可查询检索其他站点的信息,也可维护管理自己的知识目录。对外部用户,通过知识目录及其浏览工具发现信息、概略或详细地了解信息,并通过适当途径获取信息。同时信息共享知识实施还应包括提供使用卫星数据服务界面的详细描述。

### 6.2 多中心协同计算

通过构建分布式多中心计算环境,开发算法实现有效地调度计算资源以及跨异构系统高性能计算,将独立的或大量松散绑定的数据处理任务动态地分配给闲置计算资源,实现动态资源调度及任务分配。

### 6.3 分布式高性能卫星遥感信息归档云平台

遥感信息的应用需要为各类需求用户提供一个基础平台。运用云计算模式<sup>[33]</sup>,借助云平台先进的基础架构与管理方式,构建有效的遥感信息公共服务平台,提供权限管理、遥感图像智能化搜索、图像资源浏览、结果获取、订单处理及反馈等基本功能。

利用高性能处理终端集群<sup>[34]</sup>可在遥感数据处理中建立云平台来处理卫星遥感数据,证明云计算模式具有较好的应用潜力。

云平台的构建包括 3 个不同任务,即虚拟化、平台搭建以及服务提供<sup>[35]</sup>。其中,虚拟化是整个云平台构建的基础,在此基础上,平台搭建实现对信息资源以及相应虚拟化资源的调度和管理,服务提供则将相应信息资源转化为服务。采用虚拟化软件(如

Vmware, Virtual PC 等)在虚拟服务器和底层硬件之间建立一个抽象层,然后将卫星遥感相关应用模块迁移到虚拟层上,不同应用模块共享底层硬件计算和存储资源。在建设时,服务器、存储设备以及应用程序等通过虚拟化软件整合成统一资源,动态地给各个应用系统按需分配资源,实现应用的动态迁移。平台搭建具体包括搭建公共云、业务云和支撑云 3 个平台,对相应信息和数据资源等进行整合。其中,公共云将可供公共使用的数据和信息以及其他相应资源放在该平台上;业务云为数据中心内部各个业务部门的相互连通而搭建;支撑云为公众云和业务云的搭建和运行提供资源和技术层面的支持。服务提供使所建云平台向不同的集成系统提供丰富的云端服务。通过分布式高性能卫星遥感信息归档云平台,为用户提供能进入数据分析、知识查询的专用入口。

## 7 结 语

本文提出大数据环境下卫星对地观测数据集成系统建立与应用中亟待解决的关键技术,包括大容量对地观测数据的存储优化技术、基于网格的遥感图像快速处理技术、海量卫星遥感图像数据的深度分析与地学知识发现技术、基于共享知识库的多数据中心协同处理及归档技术。通过这些关键技术建立卫星对地观测数据集成系统,实现集成卫星图像、地面观测数据和模拟模型的元数据管理、几何精度纠正和卫星数据质量评价、海量卫星图像数据的空间分析与知识发现、分布式高性能卫星图像数据管理和归档系统等基本功能,能够为解决海量卫星数据分布式存储与计算、数据集成与互操作、空间数据分析与地学知识发现这些问题提供新思路、新技术与新方法。

### 参考文献(References):

- [1] Wang Yi. The development of the Earth observation system [J]. *Advances in Earth Science*, 2005, 20(9): 980-989. [王毅. 新一代对地观测系统的发展 [J]. 地球科学进展, 2005, 20(9): 980-989.]
- [2] Li Deren, Tong Qingxi, Li Rongxing, et al. Current issues in high-resolution Earth observation technology [J]. *Science in China (Series D)*, 2012, 42(6): 805-813. [李德仁, 童庆禧, 李荣兴, 等. 高分辨率对地观测的若干前沿科学问题 [J]. 中国科学: D 辑, 2012, 42(6): 805-813.]
- [3] Max Craglia, Kees de Bie, Martino Pesaresi, et al. Digital Earth 2020: Towards the vision for the next decade [J]. *International Journal Digital Earth*, 2012, 5: 4-21.

- [4] Li Deren. Opportunities for geomatics [J]. *Geomatics and Information Science of Wuhan University*, 2004, 29(9): 753-756. [李德仁. 地球空间信息学的机遇 [J]. 武汉大学学报: 信息科学版, 2004, 29(9): 753-756.]
- [5] Li Deren, Shen Xin. Intelligent Earth observing system [J]. *Science of Surveying and Mapping*, 2005, 30(4): 9-11. [李德仁, 沈欣. 论智能化对地观测系统 [J]. 测绘科学, 2005, 30(4): 9-11.]
- [6] Rong Xie, Ryosuke Shibasaki. Creating a satellite-data integration system [J]. *GIM International*, 2006, 20(12): 21-23.
- [7] Gong Jianya. Advances in Earth Observation Data Processing and Analysis [M]. Wuhan: Wuhan University Press, 2007. [龚健雅. 对地观测数据处理与分析研究进展 [M]. 武汉: 武汉大学出版社, 2007.]
- [8] James Manyika, Michael Chui, Brad Brown, et al. Big Data: The Next Frontier for Innovation, Competition, and Productivity [R]. McKinsey Global Institute 2011.
- [9] Daniel Sui, Sarah Elwood, Michael Goodchild. Introduction: Volunteered geographic information, the exaflood, and the growing digital divide [M] // *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. New York: Springer, 2013: 1-44.
- [10] Big Data Expert Committee of China Computer Federation. White Paper of Big Data Technologies and Industry Development in China [Z]. Beijing: China Computer Federation, 2013. [中国计算机学会大数据专家委员会. 中国大数据技术与产业发展白皮书 [Z]. 北京: 中国计算机学会, 2013.]
- [11] Yu Xing. The bigdata tool for geochemical study of seabed rocks—PetDB and its application in geoscience [J]. *Advances in Earth Science*, 2014, 29(2): 306-314. [余星. 海底岩石地球化学研究中的“大数据”——PetDB 及其应用 [J]. 地球科学进展, 2014, 29(2): 306-314.]
- [12] Li Guoqing, Wu Yanhui. Scientific research of Earth observation in era of big data [J]. *China Computer Society Newsletter*, 2013, 9(9): 27-31. [李国庆, 邱延辉. 大数据时代的对地观测科学研究 [J]. 中国计算机学会通讯, 2013, 9(9): 27-31.]
- [13] Zhou Xiaofang, Lu Jiaheng, Li Cuiping, et al. Challenges of big data from the perspective of data management [J]. *China Computer Society Newsletter*, 2012, 8(9): 16-20. [周晓方, 陆嘉恒, 李翠平, 等. 从数据管理视角看大数据挑战 [J]. 中国计算机学会通讯, 2012, 8(9): 16-20.]
- [14] Qin Xiongpai, Wang Huiju, Du Xiaoyong, et al. Big data analysis—Competition and symbiosis of RDBMS and MapReduce [J]. *Journal of Software*, 2012, 23(1): 32-45. [覃雄派, 王会举, 杜小勇, 等. 大数据分析——RDBMS 与 MapReduce 的竞争与共生 [J]. 软件学报, 2012, 23(1): 32-45.]
- [15] Yang Bingxin. Brief introduction to Xiangshan science conferences of nos [J]. *China Basic Science*, 2012, 14(4): 22-29. [杨炳忻. 香山科学会议第 420-424 次学术讨论会简述 [J]. 中国基础科学, 2012, 14(4): 22-29.]
- [16] Gong Xueqing, Jin Cheqing, Wang Xiaoling, et al. Data-intensive science and engineering: Requirements and challenges [J]. *Chinese Journal of Computers*, 2012, 35(8): 1 563-1 578. [宫学庆, 金澈清, 王晓玲, 等. 数据密集型科学与工程: 需求和挑战 [J]. 计算机学报, 2012, 35(8): 1 563-1 578.]
- [17] Wang Shan, Wang Huiju, Qin Xiongpai, et al. Architecting big data: Challenges, studies and forecasts [J]. *Chinese Journal of Computers*, 2011, 34(10): 1 741-1 752. [王珊, 王会举, 覃雄派, 等. 架构大数据: 挑战、现状与展望 [J]. 计算机学报, 2011, 34(10): 1 741-1 752.]
- [18] Ma Shuai, Li Jianxin, Hu Chunming. Challenging and thinking of big data science and engineering [J]. *China Computer Society Newsletter*, 2012, 8(9): 22-30. [马帅, 李建欣, 胡春明. 大数据科学与工程挑战与思考 [J]. 中国计算机学会通讯, 2012, 8(9): 22-30.]
- [19] Li Guojie. The scientific value of big data research [J]. *China Computer Society Newsletter*, 2012, 8(9): 8-15. [李国杰. 大数据研究的科学价值 [J]. 中国计算机学会通讯, 2012, 8(9): 8-15.]
- [20] International Organization for Standardization. ISO/TC 211—Geographic information/Geomatics [S/OL]. (2005-07-28) [2015-01-24]. [http://www.iso.org/iso/home/store/catalogue\\_tc/catalogue\\_tc\\_browse.htm?commid=54904](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_tc_browse.htm?commid=54904).
- [21] OGC. The OpenGIS Abstract Specification [S/OL]. (2005-07-28) [2015-01-24]. <http://www.opengeospatial.org/standard/as>.
- [22] Gu Xingfa. Development Status and Prospects of China Earth Observation System [R]. Wuhan: Wuhan University, 2010. [顾行发. 中国空间对地观测体系发展现状与前瞻 [R]. 武汉: 武汉大学, 2010.]
- [23] State Administration of Science, Technology and Industry for National Defense [EB/OL]. China Major Projects of High-Resolution Earth Observation System, 2007. [2015-04-06]. <http://www.sastind.gov.cn/n25770/index.html>. [国家国防科技工业局 [EB/OL]. 中国高分辨率对地观测系统重大专项网, 2007. [2015-04-06]. <http://www.sastind.gov.cn/n25770/index.html>.]
- [24] Gu Xingfa. Thinking of Industry Development of China's Satellite Applications [R]. Wuhan: Wuhan University, 2010. [顾行发. 我国卫星应用产业发展思考 [R]. 武汉: 武汉大学, 2010.]
- [25] Dai Weimin. Techniques and Methods of Semantic Web Information organization [M]. Shanghai: Academia Press, 2008. [戴维民. 语义网信息组织技术与方法 [M]. 上海: 学林出版社, 2008.]
- [26] Rong Xie, Ryosuke Shibasaki. Imagery metadata development based on ISO/TC 211 standards [J]. *CODATA Data Science Journal*, 2007, 6(3): 28-45.
- [27] Hua Yu, Jiang Hong, Zhu Yifeng, et al. SmartStore: A new metadata organization paradigm with semantic-awareness for next-generation file systems [C] // *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 2009: 1-12.
- [28] Zeng Shaobin, Xie Chuanjie, Li Jiaqi, et al. Parallel fusion for

- remote sensing images based on grid services[J]. *Journal of Geo-Information Science* 2010 ,12( 2) : 269-274. [曾少斌,谢传节,李佳琪,等. 基于网格服务的遥感图像并行融合[J]. 地球信息科学学报 2010 ,12( 2) : 269-274. ]
- [29] Li Shengyang , Zhang Aijun , Zhu Chongguang , et al. Fast processing of grid-based remote sensing images[J]. *Computer Engineering* ,2007 ,33( 6) : 35-37. [李盛阳,张爱军,朱重光,等. 基于网格的遥感图像快速处理[J]. 计算机工程 ,2007 ,33( 6) : 35-37. ]
- [30] Dhruba Borthakur. The Hadoop Distributed File System: Architecture and Design [R]. The Apache Software Foundation , 2007.
- [31] Jeffrey Dean , Sanjay Ghemawat. MapReduce: Simplified data processing on large clusters [J]. *Communications of the ACM* , 2008 ,51(1) : 107-113.
- [32] Kyuseok Shim. MapReduce Algorithms for Big Data Analysis , Proceedings of 8th International Workshop [C]. Berlin: Springer-Verlag Heidelberg , 2013: 44-48.
- [33] Thomas Erl , Ricardo Puttini , Zaigham Mahmood. Cloud computing: Concepts , Technology & Architecture [M]. New York: ServiceTech Press , 2013.
- [34] Navid Golpayegani , Milton Halem. Cloud computing for satellite data processing on high end compute clusters [C] // IEEE International Conference on Cloud Computing. Bangalore , India , 2009: 88-92.
- [35] Lu Junjie , Hou Weizhen. Construction of e-government cloud platform oriented information resources integration [J]. *Library Studies* , 2012 ,13: 36-40. [鲁俊杰,侯卫真. 面向信息资源整合的电子政务云平台构建研究[J]. 图书馆学研究 2012 ,13: 36-40. ]

## Key Technologies of Earth Observation Satellite Data Integration System under Big Data Environment

Xie Rong<sup>1</sup> , Liu Yawen<sup>2</sup> , Li Xiangxiang<sup>3</sup>

( 1. *International School of Software , Wuhan University , Wuhan 430079 , China;*

*2. School of Remote Sensing and Information Engineering , Wuhan University , Wuhan 430079 , China;*

*3. Space Star Technology Co. , Ltd. , Beijing 100086 , China)*

**Abstract:** The establishment of satellite earth observation system is an important means for effective management and application of satellite information resources. From significant demands for earth observation in China as well as cutting-edge scientific issues , we propose some key technologies of developing earth observation satellite data integration system under big data environment , including semantic integration of large heterogeneous earth observation data , fast data processing of satellite remote sensing imagery based on grid , in-depth analysis and knowledge discovery of big satellite data , and collaboration processing of multiple data centers and cloud-platform. It is hoped to provide with new technologies and methods for satellite big data management , analysis and archiving.

**Key words:** Satellite observation data integration system; Semantic technologies; Grid computing; Big data deep analysis; Cloud platform.