

## 高性能地学计算进展\*

薛勇<sup>1</sup> 万伟<sup>1,2</sup> 艾建文<sup>1,2</sup>

(1. 遥感科学国家重点实验室,中国科学院遥感应用研究所,北京 100101; 2. 中国科学院研究生院,北京 100049)

**摘要:**地学计算是用计算机解决复杂的空间问题的艺术和科学。计算科学使用计算机去研究科学问题,高性能地学计算(HPGC)是高性能计算科学到地理科学的应用。它将高性能计算资源应用于各种类型的地理科学数据、信息和模型,来解决地球科学问题。它使用计算科学理念发展了专业学科理论、算法、体系结构、系统、支撑工具和基础设施。高性能地学计算的应用领域主要包括空间数据分析、动态建模、仿真、时空动力学、可视化、虚拟现实、采用非传统数据采集与分析技术的应用。本文介绍了高性能地学计算的发展、基本模式和分类。分析了当前国内外的研究现状,从关键技术等方面,重点阐述了具有代表性地学网格计算应用项目。最后总结了高性能地学计算未来的两大发展趋势。

**关键词:**地学计算;网格;高性能;机群;空间信息网格;遥感

**中图分类号:**TP393.4

**文献标识码:**A

## High Performance Geocomputation Developments\*

XUE Yong<sup>1</sup> WAN Wei<sup>1,2</sup> AI Jianwen<sup>1,2</sup>

(1. State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Applications,

Chinese Academy of Sciences, Beijing 100101;

2. Graduate University of Chinese Academy of Sciences, Beijing 100049)

**Abstract:** Geocomputation is an art and science using computer to solve complicated spatial problems. Computational science involves the use of computers to study scientific problems. High Performance Geo-Computation (HPGC) is the application of high performance computational science to Earth sciences. It applies high-performance computational resources to various types of Earth science data, information, and models for solving Earth science problems. It develops discipline-specific theories, algorithms, architectures, systems, supporting tools, and infrastructure within the overall context of computational science. The areas of application for HPGC include, but are not limited to, spatial data analysis, dynamic modeling, simulation, space-time dynamics, visualization, virtual reality, applications employing non-conventional data clustering and analysis techniques. In this paper, we introduced the current developments in HPGC, the basic modes and taxonomy of it. Then we analyzed the current developments in HPGC. Especially, from the key technologies, the typical applications of projects have been addressed in more details. At the end, we concluded two major future trends in this filed.

**Key words:** geocomputation; grid; high performance (HP); computer cluster; spatial information grid (SIG); remote sensing

### 1 引言

地学计算被视为计算科学在地理和地球(“geo”)中各类型问题的应用范例。地学计算技术的发展基础为高性能计算中的计算技术、算法和例证。事实上,它包括四个前沿技术<sup>[1]</sup>:

(1)地理信息系统(GIS),遥感(RS)和全球定位系统(GNSS):产生空间的数据;

(2)人工智能(AI)和计算智能(CI):提供中心工具;

(3)高性能计算(HPC):提供能力;

(4)无线电通讯和英特网:提供连接。

高性能地学计算是地学计算的新发展,同时也是地学与高性能计算技术结合的产物。传统的高性能计算(High Performance Computing)是采用基于向量或并行处理器为基础的超级计算机硬件对大容量数据与对那些复杂而又不能用其他常规计算手段来处理的现实世界(包括地理的时间、空间世界)问题所实施的计算。随着计算技术的发展,现代高

性能计算包括了新的计算技术,如集群、分布式、网格、普适计算等。正是因为有以这种高性能为基础的地学计算,地理环境中全球性或大区域性以时空演变为特征的地理现象(数值天气预报、全球气候变化、地理信息系统和遥感定量解译和图像解译等)的模拟才有可能真正实现;原有的一些受计算机条件限制而难以模拟的模型才能得以完善、改进和运行;进而开始探索地理环境,认知新的概念。

在地理现象研究中,高性能并行计算的应用分为计算密集型(computing intensive)、数据密集型(data intensive)和网络密集型(network intensive)三大类。像数值天气预报这样的大数据处理属计算密集型的高性能计算应用;而像GIS的可视化计算应是数据密集型的高性能计算;现在正在研究发展中的地理信息系统(GIS)、全球定位系统(GPS)和遥感(RS)的集成技术和应用将是计算数据和网络三种密集型高性能计算的综合,其真正意义上的实现还需作很大努力。

虽然近十年来,用于遥感影像处理的并行计算系统绝大多数是同构平台(由一致的计算单元组成)<sup>[2-4]</sup>,但是,HPC结构的发展趋势却是异构平台。这种平台整合了高度异构的计算资源,更加适合解决数据密集型计算问题<sup>[5,6]</sup>。异构平台系统并非最初刻意设计,而是长期计算机技术进化和市场销售和发展的结果。在地学应用里,异构的机群资源能够实现非常高的综合性能。异构机群系统表现出来的计算资

\* 基金项目:本文得到自然科学基金项目(40471091,40671142)、中国科学院知识创新工程重要方向性项目(KZCX2-YW-313)、科技部973计划“陆表生态环境要素主被动遥感协同反演理论与方法”项目(2007CB714407)的支持。

源的蔓延、迁移和动态扩展能力,引发了后来的大规模分布式计算,即网格计算概念的提出。

遥感数据处理算法通常选择由廉价 CPU 处理机群组成的并行系统作为平台,但这些系统规模庞大而且相对高的成本,难以适合于观测平台板载的遥感数据处理情况。特殊的微型高性能计算应用,需要低成本、低重量、低功耗的计算单元,在减轻有效载荷重量的同时能够获得实时的处理能力。一个具有重要意义的发展是可编程器件的出现,包括现场可编程阵列(FPGA)<sup>[7]</sup>、嵌入式信号处理系统(DSP)、以及图像处理单元(GPUs)<sup>[8]</sup>。这些技术解决了平台上板载的要求和实时性要求之间的矛盾。地学应用对计算能力有前所未有的需求,同样也能从这些技术中找到答案。比起集群或者网格,它们紧凑的硬件结构、小规模解决方案和相对低廉的成本是其明显优势。

常见的高性能地学实现方式有以下几类:第一类应用是网络时代出现的专用信息服务器。PC 的普及和 Internet 服务的发展推动了大量分布式的低端服务器的普及。与传统的高性能计算不同,这种由海量廉价服务器组成的机群通过协作调度的方式工作,仍然可以获得很高的计算力。第二类高性能计算应用是局域网的工作站和并行机群,对于科学研究者来说更为常见。在卫星遥感影像地面系统中对海量栅格数据作并行处理,大都基于这种模式<sup>[9]</sup>。第三类是更为传统的超级计算机方式(supercomputer)。计算机本身有多个 CPU,通过 SMP(Symmetrical Multi-Processing)的方式进行并行处理。过去的十几年间,主要是这种方式被用于地学和环境有关的计算任务中,比如数值气候模拟、亚表面的地理学模型等等。最后一类是地学网格计算,我们在下一章详细介绍。

## 2 地学网格计算

网格计算在地学方面的应用指的是结合多种空间信息数据源(例如空间、大气、和地面)、模型(例如预报和评估)。整合广域范围的计算资源、数据资源,通过协同计算和处理,从中提取有用的信息和知识,以便为环境监测和自然资源管理提供服务。Xue 等<sup>[10]</sup>对地学网格在遥感领域应用作了初步的系统研究,认为基于网格计算的 HPGC 系统将很快成为具有代表性的技术,被用于分析高维遥感数据或者其它领域。地学网格能够对海量地学数据进行高效、实时的分析处理;能够实现应用的互连和异构资源(高性能计算机、海量存储、软件系统)共享,提高地学资源利用;集成现有各种地学数据,构建新的先进地学数据系统;跨地域的大规模的地学网格应用,涉及多个异地部门,能提供远程访问与一站式、无障碍服务;地学网格具有动态性,随着地学应用系统的业务需求不断变化,其运行管理策略和使用模式也在不断变化。

地学网格关键技术有别于一般网格关键技术,指的是在网格基础设施上实现地学应用的重要技术,包括三个层面:网格的地学化、地学资源的网格化、地学应用的网格化<sup>[11]</sup>。

网格的地学化,就是对于通用的网格中间件和相关的第

三方组件进行改造,使得它们可以适应于地学领域的特点。比如数据管理方面,通用网格的数据传输协议和复制协议就需要改造,才能适应空间数据的复杂性和海量分布的特点。

地学资源的网格化,就是对于各种地理空间数据、模型、算法进行网格化改造,实现网格环境的封装、部署和注册。使用者可以在网格上搜索、访问到这些资源。地学应用的网格化是地学网格的一个重要方面。

地学网格应用实际就是在网格基础设施平台上对一系列网格资源调用的实现以及资源调用之间的上下文关系。地学应用的网格开发包含:建立应用描述机制、建立应用开发平台、建立可视化界面系统等,这些工具为构建大型的地学网格应用系统提供了必要的基础。

具体说来,包括以下关键技术:

(1) 计算模式。在网格环境中,如何实现地学应用的计算模式是应用的关键。涉及到任务分解的模式和方法。一些传统的地学问题如状态模型和统计分析可以用并行的方式来求解,甚至可以采用像 SETI@HOME 那样使用高度分布的任务并行模式(<http://setiathome.berkeley.edu>)。另一些问题如洪水、暴雪、飓风等自然灾害的监测,由于实时性需要则借助计算中心提供高性能计算服务的方式<sup>[12]</sup>。并行计算的基本实现方式是把串行计算分解为并行子任务,每个子任务在不同的处理机上执行<sup>[13]</sup>。在理想的情况下子任务是完全独立的,但实际上会涉及到处理机间通讯和交互的问题。常用的并行方式分为功能并行、数据和流水线并行,这三种方式一般对空间数据来说都适用。但对于任意边的矢量或图来说,数据域分解的方法有时会有困难。显然,完全的并行化对大多数算法来说都过于理想化,并且采用何种的并行模式不能一概而论,要根据具体应用改变并行策略,有时甚至使用混合的方式。例如栅格图像处理增强、遥感图像的辐射定标、蒙特卡罗(Monte Carlo)模拟和元胞自动机等模型和算法<sup>[14]</sup>,数据并行是完全可行的,而对于更为复杂的地学模型,如卫星气溶胶定量遥感反演、城市化模型等就不能依赖简单的并行模式。

(2) 资源管理。除了一般网格的存储管理、资源监控和发现外,地学网格的资源管理更侧重数据库、算法库和模型库的管理。

(3) 任务调度。任务管理服务包括任务决策、资源选择和任务调度负载管理等技术。地学网格需要实现负载均衡的数据和任务调度。因为地学问题往往既是计算密集型问题,又是数据密集型问题,处理流程是一个混合负载的工作流<sup>[15]</sup>。一般网格没有涉及到在海量数据计算环境中的数据任务管理和负载均衡调度问题。数据传输如果依赖简单传输策略将成为系统瓶颈。一般考虑利用文件副本服务器,数据提前存放策略,任务调度管理算法等技术,实现节点间负载均衡。

(4) 许多网格地学应用本身就是一个工作流实例,由于地学应用需要灵活的数据传输管理和服务调用过程,需要工作流来管理服务程。

(5) 用户访问与可视化技术。地学网格入口的优势在于用户可以通过简单、唯一、有形象表现的入口访问高性能设备,屏蔽底层的计算复杂性。由于地学数据具有时空、多维、多样等特点,需要可视化技术支持甚至是 3D 虚拟技术的支持,存在着对海量数据的显示、分析并能够交互操作的要求。包括数据的可视化、结果的可视化、处理的可视化和资源的可视化。

(6) 现有地学模型、算法、工具的集成。面向服务的概念和网络的复杂技术体系一般用于企业级的网格建设,通常只能被系统开发人员安装配置,由专业管理人员维护,而难以满足一般用户的开发和使用。地学网格基于服务集成的架构,以软件资源重用的思路,简化了网格系统的管理,把网格、地学软件、地学模型以及 HPC 结合在一起。这种构造设计方法实现的地学网格应用具有可灵活性、稳定性和扩展性,能实现按需计算和用户定制,能推广到其它领域的网格应用中去,形成一个通用的、合作的和高性能的虚拟处理平台。

### 3 地学网格发展现状

下面列举有代表性的地学科学相关网格项目。具有先进水平的地学领域网格项目主要分布在美国和欧洲。按照网格应用分为基础网格、数据网格、计算网格、网格地学应用。数据网格管理各种空间数据库资源,计算网格偏重地学模型在互连的高性能计算中心上计算。网格地学应用是目前地学网格的主要构建模式,在基础网格平台组件上开发地学研究应用,结合数据、模型、工具功能,使地学研究者能够在先进的计算基础设施上进行地学处理和分析。

#### 3.1 基础网格项目

EGEE (The Enabling Grids for E-Science) 项目 (<http://www.eu-egee.org>) 最初由欧盟资助,并由欧洲、美洲、东亚的 30 多个国家的超过 90 个机构参与。发展横跨欧洲的科学网格设施,提供 24 小时不间断科学服务。EGEE 目的是加强同波罗的海和拉美地区科技合作交流,其领域包括天体物理、生物医学、化学、地球科学、冶金、金融和多媒体在内的广泛应用领域。

EGEE 的子项目已开发地学应用,解决地球科学研究的关键需要,例如时空元数据管理\实时性需求、专用数据模型和数据同化等。欧空局 ESA 同时发起了由网格界和地学界联合参与的各种工作组研讨会和出版组织。

(1) EOGeo: 使用可持续对地观测\地理空间信息和通讯技术 (EOGeo ICTs), 应用于社会组织活动和居民健康公共卫生环境。

(2) GEOS: 对地观测委员会任务小组 CEOS (Committee on Earth Observation Satellites) 目的是调研网格技术为满足 CEOS 项目应用需求的可能。共享有效使用网格技术的经验。

(3) ESA 网格和 e-collaboration 工作组。ESA 定期组织工作讨论组,致力于评估网格和 e-collaboration 在地学科学

领域的应用现状。UK e-Science Program 英国的 e-Science 项目<sup>[16]</sup>由 UK Research Councils 发起的合作支持。旨在使用海量数据集、万亿次级的计算资源和高性能的可视化,有效地提高科学研究效率。e-Science 的应用实施是依托于七个研究领域的研究委员会,下面再分设研究中心,利用 e-Science 提供的网格协作环境进行应用研究。地学领域项目属于国家环境研究委员会,开展的代表性项目包括:

1) 气候预测 (Climateprediction.net)。这是目前最大规模的基于分布式超级计算的数值天气预报实验 (<http://www.climateprediction.net>)。全球范围互联网用户下载气候模型中间件,通过后台运行的方式贡献闲置的计算力,Climateprediction.net 使用 e-Science 提供的网格软件聚合这些资源。

2) 分子层环境研究 (eMineral) (<http://www.eminerals.org>)。例如污染扩散、疾病的传播、放射性污染等等环境问题,分子层的计算机模拟是理解空间传播规律的有效手段。此项目联合模拟实验科学家、应用开发者和计算机科学家,利用网格基础设施,研究环境问题的分子级模拟。

3) 网格支持的地球系统模型 (GENIE) (<http://www.genie.ac.uk>)。项目目的是开发一个基于网格的计算平台,可以灵活地把各种组件嵌入和组合成地球系统模型,在网格环境里运行这些模型,得到分析结果,共享分布的数据。地学研究者用虚拟组织的方式进行协作研究。

#### 3.2 数据网格

欧盟主持的数据网格项目和 WP9 组件 (<http://styx.esrin.esa.it/grid>)。DataGrid 是首个大规模国际网格项目,在一套网格基础设施上同时建立多个按照应用划分的虚拟组织(高能物理、生物和对地观测)的项目。目标在于构建下一代先进计算基础设施,提供从 Terabyte 量级到 Petabytes 量级,横跨不同学科的大规模数据库资源的共享,以及对数据的计算和分析服务。DataGrid 主要针对 CERN 的高能物理应用,解决海量数据的分解存贮和处理问题,同时扩展到其他应用,如对地观测和生物研究。DataGrid 的任务划分成为 12 个工作包 WP (Work Package), WP9 的目标是定义数据网格中对地观测的技术和应用需求,设计相应的网格中间件,构建对地观测数据网格的原型。WP9 的直接目的是在 DataGrid 基本设施上建设面向对地观测的扩展应用。WP9 项目计划研究制定了 5 个工作任务:定义对地观测科学应用的需求、修改和增加相应的中间件、对地观测数据网格的应用交互界面开发、使用该系统进行 Ozone/Climate 天气变化试验研究、利用该系统进行对地观测全面应用原型研究。

DataGrid 后续项目 DataGrid 项目的成功成果(包括技术和基础软件)被用于日后的后续项目 (EGEE)。第二个 2 年期项目开始于 2006 年 4 月,包括:超过 32 个国家的 90 个合作伙伴,形成了 13 个委员会;网格基础设施跨越了 200 个节点 39 个国家;包含了 24 小时可提供的 20000 个 CPUs;超过 5Petabytes 数据存储;维持负载每天 20000 个作业;海量数据传输大于 1.5 gigabytes/s。

NASA 数据网格项目集成了多种平台,用来对 NASA 的各地的卫星数据池进行管理和提供访问服务<sup>[17]</sup>。Integration Information Power Grid (IPG)以及地学应用扩展是此类的代表。它是一个科研和应用网格,服务于空间科学和其他 NASA 任务。IPG 正在扩展空间地理方面的网格计算子项目。为了实现大规模的科学计算和工程以及满足其他广泛的地理分布、数据密集型计算任务,IPG 正在积极地发展,目标是智能化整合 NASA 各个数据中心的数据集,建立地球观测数据档案、元数据档案、统一访问接口、二级数据访问目录。目前,IPG 已将位于美国的哥达德飞行中心、兰利研究中心、加州喷气推进实验室等数个数据基地连接了起来。

### 3.3 计算网格

I-WAY 是 Globus<sup>[18]</sup>的简化系统,它是用于气象卫星的实时图像处理系统。在这个应用中,卫星中的数据下载后,进入一个远程超级计算机进行云层检测处理,然后再由另外一个图形处理机进行气象图绘制。这些处理均实现在地理上分布的多台机器上。I-WAY 成功验证了 Globus (Globus Ubiquitous Supercomputing Test-bed Organization) 系统一些基本构件和机制。

NESS Grid (Network for Earthquake Simulation Grid) 是由 NeesGrid 组织和 NSF 共同资助的项目<sup>[19]</sup>,目的在于研制并实现网格化的地震模型应用。NeesGrid 的后台是一个系统集成环境,用于整个 NEES 项目,联系着全美的地震领域研究者,提供最先进的计算资源和研究资源,使得合作、协同分析和发布等成为可能。同时提供了物理和数值模型库,为研究者提供了一个全面、准确的建模环境。

ESG (Earth System Grid) 网格由美国能源部主持,目的是利用网格技术将分布式计算技术、大规模数据分析服务等技术结合起来,为天气分析研究、气候变化研究和全球气候模拟等提供无缝的高性能处理环境 (<https://www.earthsystemgrid.org>)。ESG 的研究分为两个阶段,分别为 ESG I 和 ESG II。目前已经有三大模型产品供使用:共享用户模型 CCSM (Community Client System Model)、并行气候模型 PCM (Parallel Climate Model)、科学数据处理和可视化软件 SDPVS (Scientific Data Processing and Visualization Software)。CCSM 是一个全球高性能气候模拟模型,可以提供地球上过去、现在和今后未来的气候模拟计算。PCM 模型主要提供一个连结多个实验室的大规模环境下的气候协作处理模型,其中包括路易斯安娜国家实验室的海洋模型和 Naval Postgraduate School (NPG) 的海冰模型。ESG 实验床连结了美国多个国家实验室和研究中心,为下一代气候研究提供了一个强有力的高性能处理环境。

### 3.4 网格地学应用

ESA G-POD (Grid Processing on Demand) 是欧洲空间局 (ESA) 主持的对地观测网格应用。目的是提供一个虚拟计算环境,把用户的计算处理和数据集成做到按需定制计算;为科研人员提供一个虚拟研究环境,共享工具和函数,中间数据结果等,用户只需要集中精力于自己的算法。G-

POD 网格环境中数据的处理应用能被无缝地加入,通过网格管理动态计算资源,可以灵活地构建获取、处理数据、生成结果的应用。目前,G-POD 使用 Globus (LCG2.6, gLite) 中间件,管理超过 200 个工作节点,70TB 的在线卫星数据。G-POD 数据节点接口可以访问到 ENVISAT、MSG 气象卫星存档数据,覆盖全欧的 MODIS NRT 数据产品,以及某些 NASA 的数据节点。G-POD 的在线处理算法模型包括 IDL、Matlab、BEAT、BEAM、CQFD、编译器,以及公用代码,另外还有数据查询和预览工具。G-POD 的 web portal 为用户提供下列功能:按照时空参数选择卫星数据、任务定义、提交和状态监测、定制结果发布界面。

G-POD 的成功应用有:(1) 产品方面, MERIS 卫星 LEVEL3 级产品和年平均产品,每天产生 400 m 分辨率的 ASAR 数据南极拼图, Meris 数据 9 km 分辨率全球月拼图;(2) 按需计算方面,提供处理 MERIS 全分辨率影像的 Geo-toolbox 算法工具箱,快速提取 MERIS L2 产品像素作统计分析的 Aeromeric 工具;(3) 协作研究方面,提供了 Envisat/MERIS 数据和 ERS-2 GOME 数据在云和水汽测量方面的综合验证和交叉比较实验的交互应用原型等等。

GEON (Geosciences Network GEON) 项目 (<http://www.geongrid.org/>) 由 NSF (Information Technology Research ITR) 资助,多家机构合作并允许自由参与。目的是开发支持一体化地球科学研究的网格环境。GEON 的基础计算设施为地学研究者提供地域分布的资源,包括数据、工具、计算和可视化服务。GEON 技术上解决了由于跨多个地学子领域的数据和工具带来的异构性。最初的 GEON 示范应用包含了一系列的科研主题:(1) 3D 地质特征的重力模型,例如使用岩石和重力数据以及其他的地质数据来模拟深成岩体;(2) 通过激光雷达 LiDAR 数据集、地震和裂缝数据、地球动力学模型的整合,来研究活动构造学;(3) 通过整合地质学、地球物理、地球演化、结构数据和模型,来研究岩石圈层结构和属性。GEON 分布式网格系统是基于 SOA 面向服务结构,集中了先进计算技术支持的支持搜索、语义数据库、多学科数据的可视化和 4D 地球科学数据管理。GEON 网格环境具有访问高性能计算资源的能力,不但支持数据密集型的分析处理,而且还支持计算密集型的模型运算。GEON 的 Portal 为用户提供了一个方便的、基于 Web 的资源访问入口。GEON 的核心底层设施被设计成具有通用性,不但适用于地学应用,也可以扩展到其它领域的网格应用中去。

GISolve (TeraGrid GIScience Gateway) 项目 (<http://www.teragrid.org>) 本质是一个网格 portal,提供访问 TeraGrid 的资源进行地理信息分析的应用服务,帮助研究者或一般用户能用上超级计算资源运行地理模型。GISolve 展示了在网格基础设施上开发地学应用的范例,目前的应用集中在地理信息科学 (GIS Geographic Information Science) 在生态环境科学、交通、公共卫生和经济方面的应用。GISolve 沿用了 OGSA 网格服务体系,提出一个三层的服务模型并开发了自己的 GIS Service 中间应用层。实现了域分解、任务调度、作业调度、数

据传输、 workflow 管理等服务功能。应用层服务的实现是依赖 Globus 和 Condor 等, 整合于 TeraGrid 系统内的网格底层服务。

### 3.5 国内地学网格项目

国内地学网格项目主要包括 ChinaGrid (<http://www.chinagrid.edu.cn>) 和空间信息网格 SIG (Spatial Information Grid) 等。

ChinaGrid 目标是充分利用中国教育科研网 (CERNET) 基础设施和丰富的各类资源, 建设一个先进实用网格, 已发展到 20 所高校的规模, 聚合计算能力达到 15 万亿次, 存储容量达 150 TB。第一期成功部署生物信息、图像处理、海量信息处理、大学课程在线、流体力学网格和数字博物馆等 6 类应用。ChinaGrid 的成员节点和 CERNET 在物理上连接, 网格公共中间件平台是 CGSP (ChinaGrid Support Platform)。ChinaGrid 遥感图像网格应用子项目具有几个方面的功能: 图像处理算法的网格化和服务化; 基于神经网络和决策树的遥感图像高精度并行监督分类算法; 应用于网格环境的遥感图像处理并行函数库的建设和集成。

SIG 空间信息网格是一种汇集和共享地理上分布的海量空间信息资源, 对其进行一体化组织与处理, 具有按需服务能力、强大的空间数据管理和信息处理能力的空间信息基础设施。国家 863 计划“高性能计算机及其核心软件”重大专项于 2002 年 4 月正式启动, 专项研究重点放在网格技术的开发和应用。在 863 的信息获取与处理技术主题中, 专家组设立了空间信息网格专题, 这是我国第一个国家级的应用网格研究计划, 课题包括空间信息网格框架体系和关键支撑技术和基于空间信息网格框架的城市空间信息应用服务系统。

## 4 遥感信息处理网格节点与高性能遥感定量反演

由中国科学院遥感应用研究所远程通讯地学处理课题组所开发的“遥感研究与信息服务节点 RSIN” (<http://www.tgp.ac.cn>), 是网格计算在遥感信息处理和高性能遥感定量反演的一个应用示范。RSIN 为空间信息网格 (SIG) 网格环境下以 MODIS 卫星数据为对象的遥感信息处理与信息发布应用示范节点。借助于该节点提供的计算和模型服务, 可按需处理用户提交的数据, 可发布的信息有: 大气参数信息, 如大气水汽、气溶胶光学厚度; 地表信息, 如地表反射率, 反照率; 植被信息, 如植被指数、叶面积指数; 地表温度、土地覆盖和土壤含水量。

RSIN 的后续项目是以大气气溶胶遥感定量反演应用为背景研究开发了网格支持的高性能气溶胶遥感反演系统, 旨在运用网格聚集计算能力的特点将所有空闲的 PC 资源组织成一个虚拟的后台计算系统, 解决遥感反演数据处理与复杂计算过程中遇到的, 需要数据密集型计算和计算密集型计算整合的科学或技术问题。针对现有网格系统在处理遥感反演流程问题上的不足和局限, 在高吞吐量计算网格平台 Condor 上构建分布式的应用框架, 适用于执行现有多种遥感反

演算法。采用开放式、层次化框架, 使用现有的网格中间件快速搭建遥感应应用。主要内嵌网格 workflow 管理、负载均衡的反演模型分解和任务, 网格资源管理、模型管理等功能。反演模块采用自主开发的气溶胶反演模型。目前已经在系统上实现了全国区域的气溶胶快速定量遥感反演应用, 建立了一套从数据处理、反演计算到数据产品加工的全国气溶胶反演流程, 以及数字制图规范。

## 5 高性能地学计算的未来

高性能地学计算的未来和高性能计算技术密切相关, 地理学的大部分研究内容, 在高性能计算的时代都将成为可能。随着地学计算理论、方法、模型、算法的不断完善, 高性能地学计算技术的发展和普及会把地理科学推向一个全新的水平。计算技术对高性能地学计算的影响将呈现以下趋势:

(1) 超级计算机和网格计算两种实现形态共存。国外高端计算系统今后的开发热点是计算速度为十万亿次每秒左右的系统, 中期目标是百万亿次秒, 长期目标是千万亿次秒甚至更高<sup>[20]</sup>。高端计算应用的多样性导致了高端计算实现形态的多样性。网格计算作为无缝集成的协同计算环境, 又称虚拟计算, 会得到进一步重视。未来高端计算系统是超级计算机群的重要计算资源, 网格技术将是获取高端计算能力的优良环境, 二者将互为补充。

(2) 微客户端和巨型服务器端的模式。数据处理向板载的嵌入式高性能处理单元发展。传感器信息获取技术已经从过去的单一化渐渐向集成化、微型化和网络化方向发展。例如, 传感器网络由随机分布的集成传感器、数据处理单元和通信模块的微小节点通过自组织的方式构成。美国加州大学 Berkeley 分校的 Smart Dust (<http://robotics.eecs.berkeley.edu/~pister/SmartDust/>) 采用光作为通信介质, 可以像尘埃一样悬浮在空中。微型传感器、传感器网络的出现为随机性的研究数据获取提供了便利。下一个阶段新的计算模式刚刚新起, 以一端是并行化高性能计算, 另一端是瘦客户端甚至是微客户端 (Nano-client) 为代表的极端的计算模式。在服务器端, 将演化为服务架构的高性能计算模式, 包括超级计算机、大规模并行机群等。在客户端, 将向着带计算功能的微型化单元的方向发展, 包括手持电脑、无线网络、无线传感器网络等多形式的移动计算模式。这是服务器-瘦客户端是分布式系统发展到极限的结果。这种计算的类型和优势及他们在地学领域的可能应用也将被逐步考虑。

## 参考文献

- [1] Xue Y, Sheng X, Jayaram N. Guest editorial for the special issue: geo-computation [J]. Future Generation Computer System, 2004, 20: 1097 ~ 1099
- [2] Kalluri S, Zhang Z, Jaja J, et al. Characterizing land surface anisotropy from AVHRR data at a global scale using high performance computing [J]. International Journal of Remote Sensing, 2001, 22 (11): 2171 ~ 2191

- [3] Chalermwat P. High performance automatic image registration for remote sensing[D]. Computer Department, Virginia: George Mason University, 1999
- [4] Plaza A, Valencia D, Plaza J, et al. Commodity cluster-based parallel processing of hyperspectral imagery[J]. Journal of Parallel Distributed Computing, 2006, 66: 345 ~ 358
- [5] Lastovetsky A L. Parallel computing on heterogeneous networks[M]. Hoboken, NJ: Wiley-Interscience, 2003
- [6] Plaza A J. Heterogeneous parallel computing in remote sensing applications: current trends and future perspectives[C]. 2006 IEEE International Conference on Cluster Computing, Barcelona, Spain; IEEE Computer Society, 2006, 1 ~ 10
- [7] Valencia D, Plaza A. FPGA - based hyperspectral data compression using spectral unmixing and the pixel purity index algorithm[M]. Lecture Notes in Computer Science, 2006, 3991: 888 ~ 891
- [8] Setoain J, Tenllado C, Prieto M, et al. Parallel hyperspectral image processing on commodity graphics hardware[C]. The 2006 International Conference Workshops on Parallel Processing Columbus, OH, US; IEEE Computer Society, 2006: 465 ~ 472
- [9] Cannataro M. Clusters and grids for distributed and parallel knowledge discovery[M]. Lecture Notes in Computer Science, 2000, 1823: 708 ~ 716
- [10] Xue Y, Wang J, Wang Y, et al. Preliminary study of grid computing for remotely sensed information[J]. International Journal of Remote Sensing, 2005, 26(1): 3613 ~ 3630
- [11] 李国庆. 空间信息应用网格技术研究. 博士论文[D]. 中科院地理科学院资源研究所中科院卫星地面站, 2004
- [12] Phillips R D, Watson L T, Wynne R H. Hybrid image classification and parameter selection using a shared memory parallel algorithm[J]. Computers & Geosciences, 2007, 33(7): 875 ~ 897
- [13] Klimeck G, Oyafuso F, McAuley M, et al. Near real - time parallel image processing using cluster computers[J]. CA, USA: Pasadena, 2003
- [14] Clarke K C, Gaydos L. Loose coupling a cellular automaton model and GIS: long-term growth prediction for San Francisco and Washington/Baltimore[J]. International Journal of Geographical Information Science, 1998, 12(7): 699 ~ 714
- [15] Xue Y, Wan W, Li Y, et al. A data intensive scientific computing framework for quantitative retrieval of geophysical parameters using satellite data[C]. IEEE Computer, 2008, 41(4) (in press)
- [16] Hey T, Trefethen A E. UK e-cience programme: next generation grid applications[J]. International Journal of High Performance Computing Applications, 2004, 18(3): 285 ~ 291
- [17] Di L P, Chen A J, Yang W L, et al. The integration of grid technology with OGC web services (OWS) in NWGIS for NASA EOS data[C]. HPDC12 & GGF8, Seattle, USA, 2003
- [18] Foster I, Kesselman C. Globus: A metacomputing infrastructure toolkit[J]. Supercomputer Applications, 1997, 11(2): 115 ~ 128
- [19] Billie S, Finholt T A. NEESGrid: A distributed collaboratory for advanced earthquake engineering experiment and simulation[C]. The 13th World Conference on Earthquake Engineering Vancouver, B C, Canada, 2004
- [20] 韩冀中, 韩承德. 高性能计算技术的发展现状与趋势[J]. 中国工程科学, 2000, 2(1): 85 ~ 86

### 作者简介

薛 勇(XUE Yong, 1965-), 英国 Dundee 大学博士, 中国科学院遥感应用研究所研究员, 博士生导师, 英国皇家特许物理学家, 英国皇家物理学会专业会员, 英国遥感与摄影测量协会专业成员, IEEE 高级会员。国际计算科学大会地学计算分会主席(2002 - 2008)。主要研究方向为定量遥感与高性能地学计算, 远程通讯地学处理(Tele-geoprocessing), 数字地球。已发表论文 100 余篇(SCI 收录 50 余篇)。

万 伟(WAN Wei, 1979-), 中国科学院遥感应用研究所在读博士研究生, 研究方向是定量遥感与高性能地学计算。

艾建文(AI Jianwen, 1975-), 中国科学院遥感应用研究所在读博士研究生, 研究方向是定量遥感与高性能地学计算。

(责任编辑: 高利丹)