

多级网格环境下的土地利用统计方法*

刘婷¹ 刘仁义² 张丰² 陈志荣³

(1. 杭州师范大学遥感与地球科学研究院, 浙江 杭州 310036;
2. 浙江省资源与环境信息系统重点实验室, 浙江 杭州 310028;
3. 宁波工程学院理学院, 浙江 宁波 315016)

摘要: 基于土地资源多级网格服务架构 LR-MGSP (Land Resource Multi-Grid Service Platform), 对土地利用统计服务按照原子服务和聚合服务的分类方式进行分级处理, 提出对原子服务采用基于统计基表的增量式密集统计, 对聚合服务采用基于网格资源状态的聚合统计, 从而实现基于多级网格资源的土地利用汇总统计方法。通过与传统土地资源管理系统的性能对比测试, 结果表明: 该方法在统计效率上明显优于传统统计方法, 且随着统计量的增大, 效率优势更加明显。该方法的应用实例也表明, 它能够满足海量土地资源数据高效统计处理的需求, 也为跨机构跨部门处理其他土地资源业务提供了参考借鉴。

关键词: 土地利用; 服务网格; 统计分析

中图分类号: P208 **文献标志码:** A **文章编号:** 0529-6579 (2011) 03-0118-05

Land Utilization Statistics Method based on Multi-grid Environment

LIU Ting¹, LIU Renyi², ZHANG Feng², CHEN Zhirong³

(1. Institute of Remote Sensing and Earth Science, Hangzhou Normal University, Hangzhou 310036, China;
2. Zhejiang Provincial Keylab of GIS, Zhejiang University, Hangzhou 310028, China;
3. College of Science, Ningbo University of Technology, Ningbo 315016, China)

Abstract: Based on multi-service grid platform of land resource (LR-MGSP), a novel land utilization statistics method is presented. The statistics services are classified by grading processed compositions. Incremental statistics based on statistic base table is adopted for atomic services, and aggregate statistics based on the state of grid resource is adopted for composition services. The comparison test results show that this method is superior to traditional approach in statistic efficiency, especially for large statistic data. The prototype system can meet with the needs of efficient process of massive land data, and provide new scheme for inter-agency coordination processing other business of land resource management.

Key words: land utilization; service grid; statistics analysis

网格技术通过统一支配和调度各节点大量闲置的计算资源、数据资源和存储资源共同完成任务计划, 网格节点以提供非平凡服务的方式共享各类资源^[1], 充分利用资源的同时也提高了海量数据密集处理性能。当前, 我国各级土地资源管理部门存在海量土地利用调查数据, 土地信息系统建设冗余, 进行大规模的土地利用数据汇总统计时, 往往

需要调集各地、各级相关数据进行海量存储和密集计算处理。为此, 本文构建土地资源多级网格环境, 调度各级各类网格资源, 实现各级土地管理部门土地利用数据的高效统计。

研究表明, 网格技术对地理空间数据共享访问、业务海量数据处理等方面有极大的优势^[2-5]。但已有的文献研究或仅给出土地资源信息网格架

* 收稿日期: 2010-03-30

基金项目: 国家 863 计划基金资助项目 (2009AA12Z222, 2007AA12Z182); 浙江省优先主题重大社会发展资助项目 (2009C03015-5); 浙江省自然科学基金资助项目 (Y5090130)

作者简介: 刘婷 (1985 年生), 女, 讲师, 博士; 通讯作者: 张丰; E-mail: zfcarnation@zju.edu.cn

构，多停留在数据网络共享和访问层面，能够同时解决空间数据分布式共享和数据高效处理的并不多见。本文在前人研究的基础上，首次尝试将网格技术应用到土地利用统计业务处理领域，结合土地部门省市县三级垂直管理的特点及其现有信息服务建设，构建了面向土地利用统计业务应用的土地资源多级网格管理平台，实现了分布式共享；将土地利用统计服务进行分级分类处理，并提出基于统计基表的增量式密集统计和基于网格资源状态的聚合统计方法，实现省市县多级土地利用数据的快速统计。

1 土地资源多级网格服务架构

随着 WSRF (Web Services Resource Framework) 网格体系架构的提出，“网格服务”实际上是 Web 服务与有状态资源的组合。因而，本系统将“网格服务”作为 SOA (Service Oriented Architecture) 的组成构件来搭建土地资源多级网格服务平台 (LR-MGSP)。该平台架构由数据资源层、支撑平台和服务组件层、服务层以及应用层组成，如图 1 所示。四层架构中每层均包含省市县三级元素，形成省市县多级网格架构。

数据资源层包含了各种土地资源，包括多级土地资源数据及其业务应用遗留子系统，它们是土地资源管理部门现有信息化资产的核心价值所在，也正是当前急需扩大共享能力的土地资源主体对象。支撑平台与服务组件层是对各种土地资源的服务化封装，为上层服务访问底层资源提供了统一的操作接口，通过网格平台的中间件构建用户所需的网格

服务组件，借助于网络安全基础设施 GSI (Grid Security Infrastructure) 的引入，保证各种资源得到最大限度的安全访问。服务层聚集了各种土地资源数据服务和功能服务，既有可独立完成简单操作的原子服务，也有可利用本层及其下层服务实现较复杂功能的聚合服务，此外还包括连接各服务的注册发现服务以及服务监控和调度服务。客户端应用层则针对统计业务需要，包括各级地类汇总应用及土地资源多级网格服务门户，为土地资源管理用户提供了可视化的服务交互界面。

2 基于土地资源多级网格的土地利用汇总统计方法

采用多级网格服务管理架构，可以松散耦合地连接各级土地资源，又可以集中使用计算资源，降低平台数据访问的负载量。基于土地资源多级网格的土地利用汇总统计方法将土地利用统计服务进行分级处理，分为原子统计服务与聚合统计服务；实现方法上，在原子级统计处理阶段采用基于统计基表的增量式密集统计技术，聚合服务端则基于网格资源状态进行聚合统计。

2.1 土地利用统计服务分级

网格计算任务调度的目标就是要对用户提交的任务实现最优调度，并实现高效的总体吞吐率^[6]。平台将面向用户业务需求的任务进行再划分，分割成子任务的形式，交给系统中的各级网格节点进行处理。由于子任务的相对独立性和网格节点的分布式部署，可以利用网格节点的计算资源和服务资源，达到网格节点高效并行处理状态。任务的执行通过调用服务的方式来实现，统计服务分级结构如图 2 所示。

直接面向客户端应用的是 I 级聚合统计服务，它一般部署在中心网格节点上，如省级网格节点。聚合服务根据自身实现需要，对请求处理的任务进行划分，分为原子任务或者非原子任务。原子任务表示该任务不可再继续分解，是任务处理的最小单元；非原子任务对应由下一级聚合统计服务实现，该级聚合服务若能直接处理，则将处理结果返回上级聚合服务；否则，继续分解该任务，直至所有的子任务处理完成。划分的任务借由网格调度器调度空闲的网格节点进行并行处理。每一网格节点处理完成后即通知上级节点，并将返回计算结果，由上级网格节点进行协同处理和数据并发控制，最后由中心网格节点汇总数据处理结果返给客户端。网格服务的调度策略依据地理空间特征与服务质量 QoS

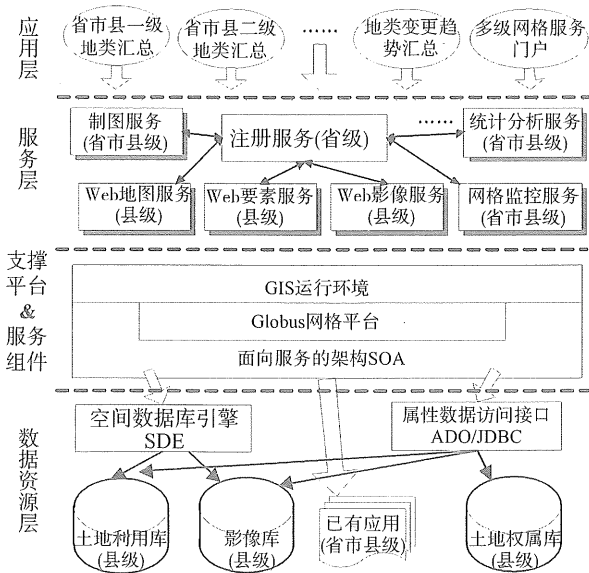


图 1 LR-MGSP 架构图

Fig. 1 LR-MGSP architecture

(Quality of Service) 等多个参数综合考虑。具体调用哪个网格节点的原子服务,对客户来说是透明的。

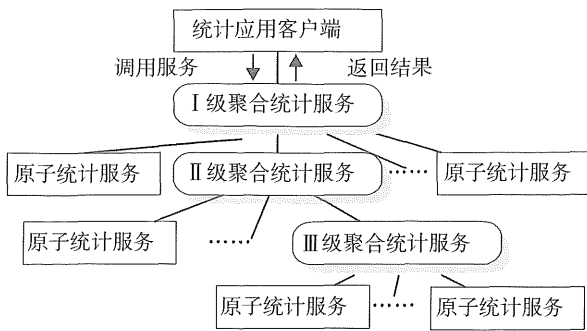


图 2 统计服务分级结构图

Fig. 2 Structure of statistic service hierarchy

2.1.1 原子统计服务 原子统计服务是进行土地利用汇总统计操作过程中最小的独立单元。服务提供者对应土地利用时空数据库最小组织和管理部 门,现行土地利用管理制度下一般为县级土地管理部 门,也可将数据存放权限下放到乡镇级土地资源 管理所,由其提供原子统计服务。进行全国土地调 查数据汇总时,一般先由县(市)级单位提供该 县(市)的土地利用统计信息,在县(市)级网 格节点中部署的土地利用调查统计服务就是原子 服务,负责统计该县(市)域范围内的土地利用现 状数据。

2.1.2 聚合统计服务 网格聚合服务是由多个 服务聚合起来共同完成某一操作的服务。根据聚合 的程度,可以分为顶层的 I 级聚合服务、II 级聚合 服务、……、底层 N 级聚合服务。服务提供者对 应原子服务分析结果数据的上级接收单位,在 LR- MGSP 平台中,即原子服务所在节点的上一层次节 点。仍旧以土地利用调查数据汇总为例,一般来 说,地(市)级以及省级节点中部署的是网格聚 合统计服务。

聚合统计服务按照 OGC (Open Geospatial Consortium) 服务不透明链^[7]调用的方式,异步调用 下级网格节点的土地利用统计服务,最后进行统计 结果汇总。当进行土地利用省级汇总时,省级节点 服务器中的 I 级聚合统计服务将调用辖区范围内市 级节点的 II 级聚合统计服务,而市级节点的 II 级聚 合统计服务又将调用其管辖范围内县级节点中的 原子统计服务或 III 级聚合统计服务或原子统计服 务,依次类推。

2.2. 基于多级网格资源的土地利用统计

2.1.1 基于统计基表的增量式密集统计 土地利

用时空数据库设计采用基于过程的动态时空数据模 型^[8],不同阶段的土地利用要素分别存储在要素 类基表(F表)、要素类新增表(A表)以及要素 类删除表(D表)中。在要素类表之间,提供一个 关系表(R表)建立三者之间的联系,A表与 基表F表图层表结构大致相似,D表记录删除的 要素,表之间通过唯一的要素标识符(FID)进行 关联。

为了便于说明,这里以土地利用专题要素地类 面积统计为例,说明本研究密集统计的实现方法:

1) 土地调查周期元年进行要素的基表统计。 基表统计结果以网格资源的形式对外提供访问。网 格服务的有状态资源特征用来缓存统计结果,平台 首次进行地类面积统计时,获得统计结果并将其作 为网格统计服务资源属性记录其状态,便于后续统 计使用。

2) 土地利用变更发生后进行变更要素的增量 统计。当土地利用要素发生变更时,增加或修改后 的要素记录在 A 表中,删除或修改前的要素则存 储在 D 表中,要素 FID 为 F 表、D 表和 A 表建立 关联;统计变更后的地类面积时,只需对该统计时 间之前 A 表和 D 表中的记录进行统计,计算出 土地利用要素的修改量,值为正数则表示该地类面 积增加了,反之,则是该地类面积减少了。

3) 土地变更调查年基于统计基表的增量式密 集统计。首先从网格资源属性状态记录中提取基表 统计结果,然后根据统计时间进行增量统计,最后 将两项统计数据进行加减得到该时期的地类统计结 果。

2.2.2 基于网格资源状态的聚合统计 原子统计 服务采用前述基于统计基表的增量式密集统计方法 实现,而聚合统计服务则采用基于网格资源状态的 聚合统计方法。该方法的核心思想是利用网格服务 的资源属性进行状态保存。一个统计处理业务的典 型实现过程如图 3 所示。

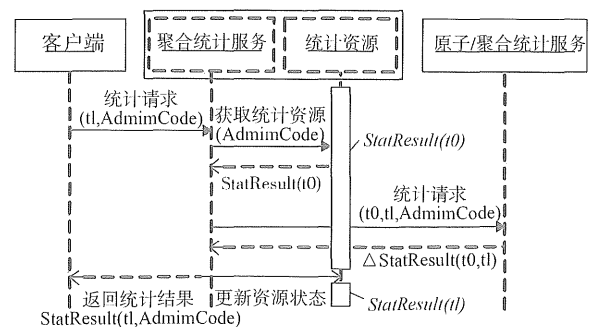


图 3 统计处理业务的实现流程

Fig. 3 Processing flow of statistic business

首先，客户端应用程序调用位于中心网格节点的 I 级聚合服务。逻辑组成上，聚合服务可以看成由聚合统计 Web 服务与统计资源组成。聚合统计服务首先获取统计资源状态，若统计时间等条件符合用户需求，直接将统计资源结果返回给服务调用端。否则，在已有统计资源基础上，重新组织增量统计任务，调度网格节点并请求部署其上的原子统计服务或低级聚合统计服务，生成增量统计结果。最后，将该返回结果与原有资源状态进行聚合，得到新的统计资源取值，进行资源状态的更新，并统计结果等资源属性返回给客户端应用。

3 土地利用多级统计实例及性能对比测试

为了验证本研究方法的合理性和高效性，本文对基于 LR-MGSP 平台的土地利用现状统计响应时间进行了性能对比测试，对比测试对象选用目前在浙江省国土资源厅使用的基于 C/S 架构的土地利用现状更新调查系统 ZJLandUse（以下简称传统土地资源管理系统）。测试结果证明基于 LR-MGSP 平台设计的土地利用汇总统计方法高效，能满足各级土地管理部门统计业务需要。

3.1 网格节点部署及数据准备

根据前述的土地资源多级网格服务架构设计，本测试采用 6 台相同配置的 PC 机（CPU，Core 2 Quad Q8200 2.33G；内存，4G）以及相应的网络设备模拟搭建多级网格环境，节点设计涵盖省市县各级网格节点。数据用例来源于浙江省部分市县的土地利用数据，通过空间数据引擎 ArcSDE 存储在对象关系数据库 Oracle10g 中。

3.2 土地利用多级统计应用实例

本文所述研究方法在浙江省土地利用现状管理系统（GridLandMgr）中得到了实现，应用于统计汇总模块的设计。该模块包括农村土地利用现状一级分类面积汇总、一级分类按权属汇总、二级分类面积汇总以及土地利用变更趋势分析。系统依据用户选定的统计区域进行县级、市级、省级土地利用现状统计。图 4 是土地利用现状管理系统的统计主页面。

3.3 土地利用地类面积专题统计效率对比测试

传统土地资源管理系统集中管理所辖县市空间数据。当进行省市土地利用地类面积汇总时，以县级空间数据为单位，各县市数据排队等待逐一处理，多个县市数据的处理时间与县市数量呈正比增长关系；而本系统则是并发进行多个县市的处理，在最长的单县市处理时间基础上，增加数据汇集时

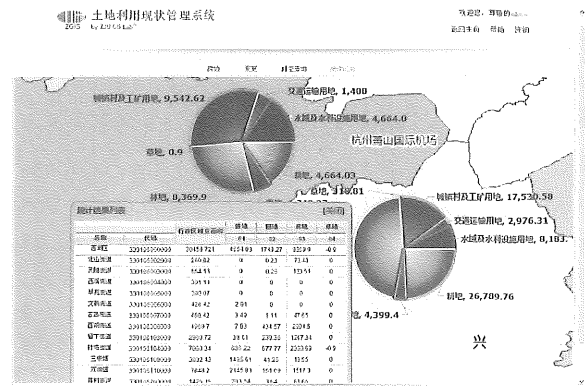


图 4 土地利用现状统计页面

Fig. 4 Page of land use statistics

间，这对密集型统计处理来说是可以忽略的，因而实现了高效处理。本测试以一级地类专题统计和二级地类专题统计为例，分首次测试和再次测试计算地类面积汇总耗时。一级地类专题统计对应测试指标分别为 StatTime11 和 StatTime12，ZJLandUse 系统相应的汇总耗时为 StatTime1_t。二级地类专题统计对应测试指标分别为 StatTime21 和 StatTime22，ZJLandUse 系统相应的汇总耗时为 StatTime2_t。多用户多次测试求平均后的测试结果见表 1，统计效率趋势见图 5 和 6。

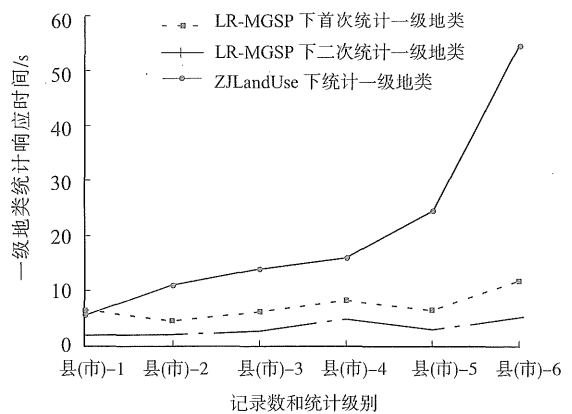


图 5 LR-MGSP 与 ZJLandUse 一级地类面积专题统计效率折线图

Fig. 5 Efficiency line charts of level 1 land-use data area statistics with LR-MGSP and ZJLandUse

土地利用一、二级地类面积汇总效率测试结果表明：

1) 单个县地类汇总时，可以忽略网络传输因素，基于统计基表的增量式汇总算法效率高于传统系统 ZJLandUse 的统计效率。

表1 LR-MGSP与ZJLandUse一级地类专题统计耗时结果比较

Table 1 Time-consuming comparison of level 1 land-use data special statistics with LR-MGSP and ZJLandUse

级别	行政区划编号	统计单元/个	记录数/行	StatTime11	StatTime12	StatTime1_t	ΔE_{Stat1}^*	StatTime21	StatTime22	StatTime2_t	ΔE_{Stat2}^*
				s		%	s		%		
县级	县(市)-1	1	108 031	6.266	1.657	5.5	-12.22	2.344	1.625	13	454.61
县级	县(市)-2	1	382 687	4.485	1.765	11	145.26	4.437	1.719	23	418.37
县级	县(市)-3	1	438 073	6.047	2.391	14	131.52	5.187	1.656	26.5	410.89
市级	地(市)-4	3	928 791	8.063	4.562	16	98.44	8.031	4.609	21.5	167.71
市级	地(市)-5	3	1 248 234	6.25	2.625	24.5	292.00	5.016	2.484	38	657.58
省级	省-6	10	3 392 463	11.594	5.141	55	374.38	9.828	5.172	122.5	1146.44

*注: ΔE_{Stat1} 为本方法一级地类统计效率相对 ZJLandUse 方法提高的百分比, 计算公式为: $\Delta E_{Stat1} = (\text{StatTime1}_t / \text{StatTime11} - 1) * 100\%$ 。 ΔE_{Stat2} 为本方法二级地类统计效率相对 ZJLandUse 方法提高的百分比, 计算公式为: $\Delta E_{Stat2} = (\text{StatTime2}_t / \text{StatTime21} - 1) * 100\%$ 。

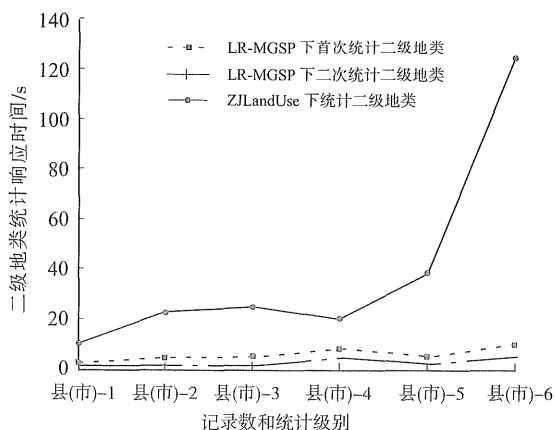


图6 LR-MGSP与ZJLandUse二级地类面积专题统计效率折线图

Fig.6 Efficiency line charts of level 2 land-use data area statistics with LR-MGSP and ZJLandUse

2) 当数据量随县级单位数的增加而增大时, 基于服务分级、任务并行处理策略的 LR-MGSP 系统对数据记录量的敏感度很低。

3) 系统进行二次统计时, 基于网格资源属性缓存统计结果的 LR-MGSP 系统统计时间大大缩短, 统计效率明显增大。

4 结论

本研究针对土地调查数据分布特点及各级土地利用统计汇总要求, 在土地资源多级网格服务管理平台架构的基础上提出任务分解、服务分级的高效统计机制, 结合基于统计基表的增量式密集统计技术与基于网格资源状态的聚合统计计算, 给出了本研究基于土地资源多级网格的土地利用汇总统计方法。通过与传统土地资源管理系统性能对比测试, 证明该统计方法性能优越, 能满足高效统计海量土

地利用数据的目的。方法承载平台基于技术标准和行业标准开发, 具有很好的扩展性和开放性, 便于其他土地资源业务处理集成。

参考文献:

- [1] FOSTER I, KESSELMAN C. The Grid 2: Blueprint for a New Computing Infrastructure [M]. Morgan Kaufmann Publishers Inc, 2003.
- [2] DI L, CHEN A, YANG W, et al. The Development of A Geospatial Data Grid by Integrating OGC Web Services with Globus-based Grid Technology[J]. Concurrency and Computation: Practice and Experience. 2008, 20(14): 1617 - 1635.
- [3] 李德仁, 宾洪超, 邵振峰. 国土资源网格化管理与服务系统的设计与实现[J]. 武汉大学学报: 信息科学版, 2008, 33(1): 1 - 6.
- [4] 金江军, 潘懋. 网络技术及其在国土资源信息化中的应用探讨[J]. 国土资源信息化, 2005(1): 6 - 9.
- [5] 岳海斌, 周庆俊, 陈鸿志, 等. 网络技术在国土资源信息化中的应用探讨[C]//中国土地学会等. 土地信息技术的创新与土地科学技术发展——2006年中国土地学会学术年会论文集. 重庆: 地质出版社, 2007: 72 - 76.
- [6] 陈晶, 潘勋. 基于双重服务质量约束的网格资源调度策略[J]. 计算机集成制造系统, 2008, 14(8): 1571 - 1578.
- [7] ISO/TC211. Geographic information - Services [S/OL]. (2001-09-14) [2003-01-18]. http://portal.opengeospatial.org/files/?artifact_id=1221,
- [8] 张丰, 刘仁义, 刘南, 等. 一种基于过程的动态时空数据模型[J]. 中山大学学报: 自然科学版, 2008, 47(2): 123 - 126.