

基于云 GIS 的地理国情统计分析方法探讨

钱亮

湖北创易行工程项目管理有限公司

DOI:10.32629/gmsm.v2i1.76

[摘要] 本文以基于云 GIS 的地理国情统计分析方法探讨为重点进行阐述,从地理国情地表覆盖数据统计分析方法,地理国情统计分析的并行处理算法设计,云 GIS 背景下地理国情统计分析并行实现机制等几方面进行深入探讨,旨意为相关性研究提供一定的数据参考。

[关键词] 云 GIS; 地理国情; 统计分析

引言

当前,我国是社会经济持续发展的特殊时期,同时还是我国国情迅速发展、复杂变化的关键性时期。很多地理国情的信息与之对应科学决策的制定以及科学管理方法的约束发挥着重要作用。相关工作人员根据目前地理国情统计的方式决定服务质量以及服务内容,借助网络服务和高标准的计算技术,可以成立具有灵活性、高效性的总结国情统计分析方法,进而在云 GIS 背景下构建具有弹性可扩展的统计分析服务体系^[1]。

1 地理国情地表覆盖数据统计分析方法

1.1 完善的直接统计法

为了更全面的体现出更新后的实际地理位置以及变化量,能够正确使用 ArcGIS 工具箱中分析功能的叠加指令,将两期数据充分的进行叠加,将两整期数据中 CC 码提取出来,接着再做出相应的统计。相关数据显示,共同变化的位置大约有 8000 处。总面积之和大约为 45000000 平方米,大约占我国地表总面积的 2.3%,这种方法与直接统计法的结果有明显差异^[2]。根据对直接统计法进行的完善,一方面有效解决了统计结果偏大的问题,并且能准确的找到更改的实际地理位置。另一方面此统计结果实际上是各项专题数据实现全面性统计分析的必要数据,最终所统计的必要数据还能够称为具有变化性的必要数据,变化率就是整个数据变化的均值。

1.2 分组统计方法

分组统计方法实质上是根据我国地理国情各个级别的单元,通过图表数量以及面积为统计指标实施的方法,从而得出节点准备前后几期的数据与图表以及面积的变化情况,然而还能够根据数据的实际需求,将统计的分组设为二级地类或者更详细的三级地类。通过对这种统计方法的使用,能够全面反射出各个一级地类的图表变化以及面积变化实际情况,同时应该具体统计出其整体变化率,但是由于其整体统计结果比较分散,特别是如果将分组的基本单元归属为二级地类或者是三级地类时,没有明确的具有统计性的指标代表总体积的变化,造成描述结果比较繁琐。因此这种方法只能用于各级地类进行的分类统计,或者各个政府主管部门在其管辖范围内所涉及的重点地类分类统计应用。

1.3 专题统计方法

专题统计方法,就是在变化必要数据的前提下,结合各个行业以及政府部门和各个管理部门的需求,定制符合自身发展需求的统计指标,同时还应该对最基础的三级地类做出独立统计,这种方式也是地理国情普查数据应用的比较广泛的一种统计方法。并且结合统计的方向可以分为正向统计方向与逆向统计方向。正向统计方向实际上是将核准前数据中某种地类数据作为统计依据,比如相关工作人员挖掘地表面积减少的主要原因,必须先整合总结出原始地表更新形成地类形式,或者可以提取核准前数据中地表的计划图,接着再进行详细的统计,统计过程依然按照基础统计单位为主。

2 地理国情统计分析的并行处理算法设计

云计算方式是大数据背景下地理国情对于大数据统计分析的最具有实际可操作性的处理手段,为了有效降低原始单机所承载的规模较大的普查数据存储以及处理负荷,并且保证符合大量数据准确性高的格式数据可以在普通单机上做出高效性存储与处理。相关研究实际上基于 Hadoop 分布式对架构进行处理,通过云环境向多层次地理国情统计分析出具有高效性并行处理模式。最终利用 HDFS 及 HBase 技术完成与数据普查相关的云储存。或者利用 MapReduce 编程方式将单机所执行的算法完成并行处理,在分布式集成环境影响下必须高效完成对我国地理国情的统计分析并行性处理。

2.1 在 MapReduce 的空间分析的支持下并行相关算法做出设计

空间叠加分析方式其实是根据地理国情的统计分析过程中重复使用条件下的基础空间分析做出的服务,能够通过输入两个或者两个以上的云服务数据是按照一定拓扑叠加实现运算并进行规划以及实际操作,进而获得原始数据集所收获的综合数据集。

2.2 在 MapReduce 的基本指标影响下所做出并行算法设计
基础数量、面积、长度、表面积、长度等基础指标统计服务是地理国情统计分析指标计算方式中具有一定支撑性服务,并且能够为后续的研究提供一定研究单元以及内容参考,大多数研究区域所需要的数据应该按照对应的基本指标算法做出计算与统计,进而有效发挥其指标算法的作用。

3 云 GIS 背景下地理国情统计分析并行实现机制

受云背景的影响可知,我国地理国情统计分析并行架构同时具备规模大、深度大、宽度高、处理时间短、对应设备较为普通化等特点。随之形势的不断改进,相关工作人员设计了符合我国实际情况的地理国情统计分析多层次服务机制,目前能够将单机处理方式进行延伸,进而形成围绕 HDFS、HBase、MapReduce 等多种云技术创设的分布式集群中多个节点服务器共同进行处理的方式。进而根据对 HBase 数据库的应用能够高效的读取统计评价所需要的普查数据,同时还应该在 MapReduce 计算结构中做好对应的数据处理以及统计指标计算的汇总工作,最终进行相应的算法设计,进而实现对地理国情统计分析服务的并行处理,对相关生态格局进行分析可知,详细说明地理国情统计分析的并行处理实现机制^[3]。

第一,将任务管理服务方式进行完善,最终选择所需要的研究单位以及研究指标等相关参数,并且将空间格局评价任务交由控制台进行处理,同时还应该在任务数据库中建设任务标记,成立任务执行制度。

第二,完成分布型数据库的访问服务,通过研究区域、研究对象、研究指标等必要条件在全面对 HBase 数据库中的相关信息进行筛查。如果所研究的单元是行政区域的单元,就不必再选取研究单元数据的信息,只要根据普查数据中行政区划划分码属性就能够提取所需要的数据信息。如果将研究单元归属于非行政区划单元时,应该提前选取研究数据的独立性数据,同时再利用研究单元数据中的空间影响力度,对分布式数据库进行访问服务,接着在利用筛查的方式提取所需要的数据信息,进而合理利用数据集的裁剪服务,实现数据并行管理策略,并且对数据进行深入裁切,进而获得所需研究单元对应的普查数据。

第三,调整总面积统计服务项目,根据研究对象进行分析,在分布式集群中做出数据并行处理方案,进而对数据对象所需面积指标进行并行性处理。第四,对长度统计方案进行服务,按照道路的分层次标准,做出数据处理,并且对各种道路长度指标进行统计,也凭借计算得出各项研究单元中的道路密度指标具体值。第五,实现数据库的操作性服务,筛选

空间格局的一级指标与整体景观格局,设计出资源多样性、资源循环使用、资源质量保障等二级指标计算模型,根据计算可以得出各项二级指标的均值,从而更全面的得到各个研究单元的空间格局以及指标数。第六,执行数据库操作系统的服务,并且将计算结果记入在任务结果中,在并行处理架构背景下的地理国情统计分析处理模型如图。

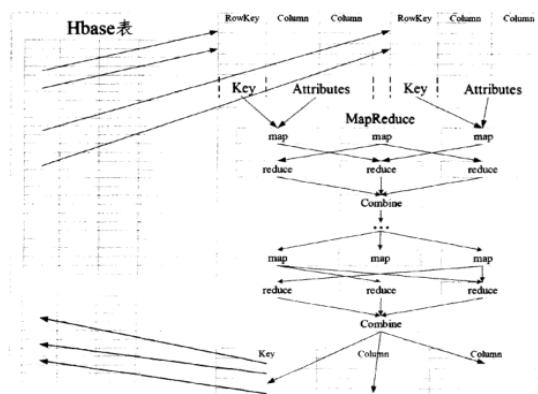


图 3-6 地理国情统计分析并行处理模型

4 结束语

总之,地理国情统计分析在数据信息体系数量丰富、信息多样化、实效性强的地理国情普查数据的影响下,必须适当降低现有单机承载的巨大数据存储以及处理压力方式,按照数据密集以及计算密集的地理国情统计分析方式,实现普通计算机高效的存储以及处理工作。利用各项针对性方案将现有单机以及传统统计分析中长久性运行的统计分析方式通过在分布式集群背景下实现并行性处理,进而完成对地理国情的统计分析方法的全面性实施。

[参考文献]

- [1]张琴琴,杨柏林.地理国情普查试点成果综合分析[J].测绘科学,2016,41(4):83-88.
- [2]白亚辉,邢立宾.应用 GIS 分析方法的地理国情变化检测研究[J].科技资讯,2017,15(23):4-5.
- [3]张俊宇,梁寿强.遥感与 GIS 技术在地理国情监测中的应用与研究[J].低碳世界,2017,(03):85-86.