计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2020,29(5):69−75 [doi: 10.15888/j.cnki.csa.007421]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于大数据的水生态承载力分析模型①

周晓磊[2], 房　萌[1,2], 刘　枢[3], 姜秋俚[3], 金继鑫[2], 宋春梅[2], 陈　月[2], 王兴刚[2], 毛立爽[2]

[1](中国科学院大学, 北京 100049)
[2](中国科学院 沈阳计算技术研究所, 沈阳 110168)
[3](辽宁省生态环境监测中心, 沈阳 110161)
通讯作者: 房　萌, E-mail: fangmeng17@mails.ucas.ac.cn

摘　要: 随着科学技术的发展, 水文信息数据量发生了巨大的增长, 如何充分利用这些支持决策的大规模数据, 是当前科学家面临一个大问题. 传统的水生态承载力分析计算复杂多样, 涉及的数据种类多样, 扩展性不强, 注重于理论研究和分析, 本文通过研究历史数据, 分析影响水生态承载力的因素, 将数据划分为 3 个指标层, 提出一种基于大数据的水生态承载力分析模型 (ECCBD). 利用 Hadoop 集群的 HDFS 分布式文件系统实现水生态数据的备份存储, 利用 MapReduce 实现海量水生态数据的并行计算. 通过输出值与水生态承载力对比, 判断水资源是否盈余或赤字, 本文提出的方法和模型能够从压力、承载力、弹性力 3 个不同的指标层有效分析水生态环境现况, 对提供水生态保护依据上有重要意义.

关键词: 水资源大数据; Hadoop; MapReduce; 水生态承载力

引用格式: 周晓磊,房萌,刘枢,姜秋俚,金继鑫,宋春梅,陈月,王兴刚,毛立爽.基于大数据的水生态承载力分析模型.计算机系统应用,2020,29(5):69–75. http://www.c-s-a.org.cn/1003-3254/7421.html

Water Ecological Carrying Capacity Analysis Model Based on Big Data

ZHOU Xiao-Lei[2], FANG Meng[1,2], LIU Shu[3], JIANG Qiu-Li[3], JIN Ji-Xin[2], SONG Chun-Mei[2], CHEN Yue[2], WANG Xing-Gang[2], MAO Li-Shuang[2]

[1](University of Chinese Academy of Sciences, Beijing 100049, China)
[2](Shenyang Insititue of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)
[3](Ecological Environment Monitoring Center of Liaoning Province, Shenyang 110161, China)

Abstract: With the development of science and technology, the volume of hydrological information data has increased tremendously, how to make full use of these large-scale data to support decision-making is a big problem for scientists at present. Traditional water ecological carrying capacity analysis and calculation are complex and diverse, involving various types of data, with unsatisfied expansion, and focus on theoretical research and analysis. This work studies historical data, analyzes the factors affecting water ecological carrying capacity, divides the data into three layers, and proposes an analysis model of water Ecological Carrying Capacity based on Big Data (ECCBD). HDFS distributed file system of Hadoop cluster is used to implement the backup and storage of water ecological data, and MapReduce is used to implement the parallel computation of massive water ecological data. By comparing the output value with the water ecological carrying capacity, determining whether the water resources are surplus or deficit, the method and model proposed in this study can effectively analyze the current status of the aquatic environment from three different index

① 基金项目: 国家科技重大专项 (18ZX07601001)
Foundation item: Major Project of National Water Pollution Control and Treatment Science and Technology (18ZX07601001)
收稿时间: 2019-10-07; 修改时间: 2019-10-29, 2019-11-15; 采用时间: 2019-11-29; csa 在线出版时间: 2020-05-07

layers: pressure, bearing capacity, and elasticity, it is of great significance to provide a basis for water ecological protection.

Key words: big data of water resources; Hadoop; MapReduce; water ecological carrying capacity

随着全球气候变暖, 经济和人民物质的提高, 工业、农业、生产生活的用水量增加, 导致对水资源的使用加剧, 水体遭到污染的事件频发[1], 与此同时科学发展的不断进步, 航天遥感在环境监测上的应用及雷达技术支持, 水质环境监测数据的种类和数量正在急剧增长, 监测产生的数据量大多为 GB 或 TB, 有时数据量能够达到 PB. 大数据技术、互联网技术的迅猛发展使大数据存储及分析利用的问题得到了解决[2]. 在过去几年中, 新一代系统被引入大规模数据处理中, 基于 MapReduce 范式的程序[3], 尤其是 Apache Hadoop[4]越来越成为流行, 利用大数据技术对水生态的历史数据进行处理、存储和分析等, 在有效管理水生态数据的同时利用数据的分析结果对水生态环境保护提供有效方案. 近年来, 国内外出现了很多基于大数据技术平台对水资源数据进行分析处理的应用. 例如: Moturi 等[5]研究在处理较小的数据集上大数据系统表现出不理想的速度, 但是在足够大的数据集上实现了合理的加速, 数据集在补充计算节点的数量下, 与正常的数据挖掘和处理相比, 能够减少 30% 的执行时间; Parmar 等[6]将神经网络模型、模糊模型和小波模型相结合对监测站监测的水质参数化学需氧量 (COD) 进行预测. 李辉等[7]提出了在 Hadoop 平台上利用 MapReduce 并发处理水文信息的模型, 利用大数据技术提高了存储执行的效率.

在水生态监测环境中, 水文监测站在网络互通的条件下, 形成网络结构或按河流流向分布, 由于监测站点分布广泛, 传感器可监测水生态、水质、气象等多类数据, 因此随着传感器的监测频率上升, 水生态环境数据大幅度增长, 传统关系型数据库在存储与管理数据中不足以承载如此海量的数据; 文献[8]提出一种基于生态系统服务的生态足迹 (Ecosystem Service-based Ecological Footprint, ESEF) 方法, 在考虑或不考虑水质标准和环境分类的条件下, 采用求平均值法和并集法分析太湖流域上游湖州市的水生态承载力, 但该方法扩展性不强. 基于上述问题, 为解决海量水生态监测数据存储和处理问题以及水生态承载力分析方法的可扩展性, 本文利用 Hadoop 分布式存储以及基于 MapReduce 并行计算提出一种基于大数据的水生态承载力模型, 将影响水生态承载力的数据因素划分为 3 个指标层, 同时结合 HDFS 分布式文件系统来管理存储海量数据, 利用生态足迹法计算水生态足迹和水生态承载力[9]以及利用 MapReduce 的并行计算处理来训练水生态承载力模型, 将生态足迹与生态承载力相比较, 得到生态承载力是否赤字或盈余. 在可持续发展的基础上控制水生态环境在可承载范围内, 完成海量水文数据存储并能完成有效计算.

1 水生态承载力大数据处理架构

1.1 水信息大数据处理总体架构

整体的基于大数据的水文信息处理架构如图 1 所示, 数据源分别由物联网、水生态监控站以及互联网资源构成, 通过远程上传、网络爬虫等完成对数据源的数据采集.

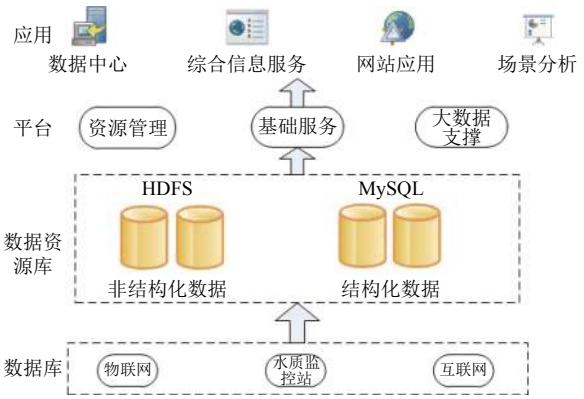

图 1 水文数据中心总体架构

数据资源库是整个系统平台的核心, 为了处理海量数据存储问题, 将其划分为结构化数据库、基础数据库以及非结构化 HDFS 分布式存储数据库等部分. 是经过数据采集及处理后的数据存储部分[10], 实现对水环境空间数据、业务数据与管理数据的管理. 提供不同物理结构数据资源库, 具有良好扩展性.

水生态环境管理平台应用依据 Java 开发框架为基础, 将数据应用开发、运行时环境与支撑业务开发的

软件封装为资源管理、基础服务以及大数据支撑应用平台, 为水环境监管部门提供业务应用, 全面提升信息化应用水平, 增强信息资源共享服务力度.

1.2 大数据采集与处理

目前, 大量的水资源、水环境以及水生态数据不断的由监测站点生成并通过物联网技术上传至远程服务中心处. 在数据仓库的归档过程中, 其中最大的数据集就是关于水质、水位和水流量的实时数据信息. 水质信息在不同站点处实时监测水质的水温、溶解氧、pH、电导率、氨氮离子、余氯等参数, 这些数据集形成海量的水生态信息. 由于水质信息数据中的 pH 值、污水处理厂的排放浓度等具有时空连续性, 因此在进行监测水质数据是否超标实现污染超标预警预报等服务时, 对水生态监测数据的存储和计算的要求提升. 为实现终端用户以交互式的丰富方式分析大量后端数据, 使用 Hadoop 技术实现对大量水生态数据的存储与计算功能, 通过 Hadoop 分布式文件系统将接口应用程序连接到后端 HBase, 通过 MapReduce 框架用户可以轻松的分析和可视化水生态大数据. 如图 2 所示.

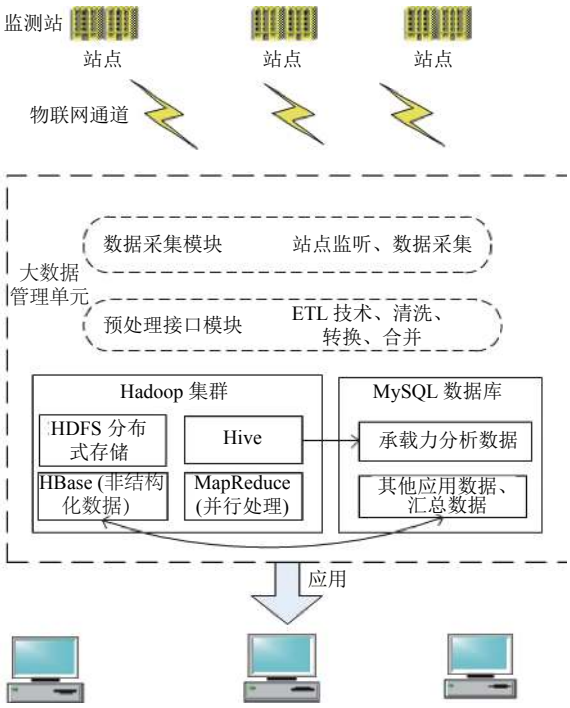

图 2 Hadoop 采集处理架构图

使用高性能的 Linux 集群构建配置, 安装 Yarn、HBase 等 Hadoop 生态大数据技术. 基于 Hadoop 作为水生态数据存储和计算的基础, 主要包括数据采集、数据预处理、数据存储和数据加载 4 个模块. 数据采集单元负责水生态数据监测站点的数据采集任务, 监测站点自动对水质、水量等信息进行监控, 并通过互联网技术将数据信息传送至数据预处理模块; 数据预处理接收到的数据大部分的时候时“肮脏”的数据, 极易遭到噪声的影响导致数据丢失、数据出现冲突等, 采用 ETL (Extra、Transform、Load) 技术将数据采集单元收集到的数据进行清洗转换, 完成数据加载入库操作.

数据存储模块主要是将经过 ETL 技术处理后的数据存放在 HDFS 分布式文件系统、HBase 结构化分布式存储系统中, 这种分布式文件系统提高了数据的容错能力以及提高了可用性, 提供了对文件的操作和存储的支持, 在必要时能够实时读写并随机访问, 数据被分成块, 以分布式的方式存储在 HDFS 集群中存在的节点上, 解决了数据库的单点性能极限. 既能提供较高的可扩展性, 又能达到并发访问的能力.

分布式计算框架 MapReduce 帮助我们在 Hadoop 平台上完成分布式的计算编程, 完成对水生态数据的分类处理, Sqoop 工具能够完成 Hadoop 与关系型数据库之间数据的互相转移, 例如通过 Shell 命令将存在于关系型数据库中具有结构规范无需处理的结构化数据备份到 Hadoop (HDFS、HBase) 中, 完成数据的转移, 从而提高获取数据的速度.

1.3 Hadoop 分布式存储管理

Hadoop 通过分布式存储管理的方式, 将数据存储及计算均衡分布在各个集群上, 从而可保证数据的可靠性和高容错性. Hadoop 的核心包括 HDFS (Hadoop Distributed File System), Hadoop 的分布式存储工具, 实现水生态数据的存储; MapReduce 分布式计算框架处理海量水生态数据的运算. HDFS 的存储结构如图 3 所示.

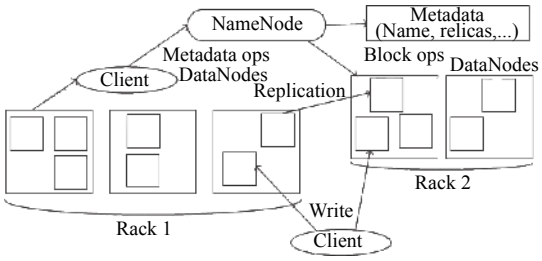

图 3 HDFS 分布式存储结构

与关系型数据库不同, HDFS 将集群中的工作节点均匀分布, 在存储上文件分块存储, 通过设置参数 (dfs.

blocksize) 来确定块的大小, 每个数据块的大小默认设置为 64 MB, 当经过处理的水生态数据上传至 HDFS 时, 可以被切割成不同的块分别存放在不同的 DataNode 上, 完成对水生态数据的备份存储, 使得存储更高效并具有高容错性[7]. Mastar/slave 结构是 HDFS 的架构模式, 其一个 Master (NameNode) 节点包含若干个 slave (DataNode). NameNode 会记录文件分块存储在 DataNode 上的位置信息, 由 dfs. name. dir 指定元数据 (文件的名称、副本系数, Block 存储的 NameNode) 的存储位置, 负责客户端的请求响应. DataNode 负责存储 Block, 在 NameNode 的调度下完成数据库的创建、删除和复制; 根据设置的时间间隔定期向 NameNode报告本身以及所有 Block 的信息.

HDFS 架构中还包含一个辅助 NameNode: Secondary NameNode, 它相当于 NameNode 的助手节点, 负责 fsimage (镜像文件) 备份以及将 edits(日志文件) 与镜像定期合并, 帮助减小 edit logs 的大小, 减轻 NameNode 重新启动时的压力, 使 NameNode 保持文件系统最新的元数据. 当系统发生突发事件的时候, 可以保存最新的改动.

2 水生态承载力分析模型

2.1 水生态承载力影响因素

水生态环境具备弹性力的特点, 能够在一定程度上进行自我恢复. 对水生态承载力的计算主要从环境、生活、资源多个方面进行综合分析, 能够体现数据的多元性、动态性以及分析结果的客观性, 为人们对水生态破坏控制在可以恢复的范围内, 即水污染、水资源利用控制在水生态环境自我恢复能力中, 能够最大化的利用水资源, 净化对水体造成的污染. 本文通过研究分析水环境、水资源与水生态方面的数据, 分别进行分类统计、比对、分析, 总结归纳得出影响水生态承载力评估的主要因素主要包括: 水生态压力数据、水资源支撑力层数据以及弹性力数据. 水生态压力数据主要包括人口增长 P1、经济增长 P2、环境污染 P3, 支撑力指层标数据主要包括水资源自身支持 S1、人类支持 S2, 弹性力指标数据主要包括生态因素数据 E1. 如表 1 所示.

2.2 水生态承载力模型

在研究承载力评估方法的过程中, 从多个领域了解到目前计算承载力的方法主要有系统仿真、灰色关联度计算、系统统计学等方法[11−13]. 生态足迹是从可持续发展的理念出发, 黄林楠等[9]提出了一种水资源生态足迹计算方法, 本文基于生态足迹法, 参考王文国等[14]对生态足迹计算相关参数的修正, 分别计算水生态足迹以及水生态承载力.

$$W_{EC} = Nw_{ec} = 0.88N\lambda_w Q_w \gamma_w / P_w \tag{1}$$

式 (1) 中所述的λ_w为区域水资源产量因子; Q_w为该区域水资源总量; N为人口总数, 人; γ_w为水资源均衡因子, ghm^2/km^2; P_w为区域水资源平均产能 m^3/km^2. 本文在水生态足迹计算过程中, 参考文献[15]在辽宁省水资源生态足迹中的研究, 其在全球部分国家数据统计中, 选定 WWF 确定的均衡因子, 在中国定义范围下, 各区域的水资源产量因子进行生态承载力计算, 并得到中国的单位面积产水量也就是水资源平均产能为 29.46×10^2 m^3/hm^2.

$$W_{EF} = N\gamma_w(W_{\mathrm{i}}/P_{\mathrm{w}}) \tag{2}$$

$$W_{EF} = W_{EFl} + W_{EFp} + W_{EFc} \tag{3}$$

其中, W_{EF}为水生态环境总生态足迹, 其分别由生活用水生态足迹W_{EFl}、生产用水生态足迹W_{EFp}和生态用水生态足迹W_{EFc}组成. $W_{i(i=l,p,c)}$为各项用水消耗量.

表 1 水生态承载力数据

一级指标层	二级指标层	三级指标层
压力层数据	人类活动数据	人口数量 (万人)
		居民生活用水量 (m^3)
		农业灌溉用水量 (m^3)
		工业生产用水量 (m^3)
	经济增长数据	万元 GDP 耗水量 (m^3/万元)
		GDP 年增长率 (%)
	环境污染数据	氨氮排放量 (t/ m^3)
		COD 排放量 (t/ m^3)
承载力层数据	水资源自身	人均水资源量 (m^3/人)
	人类支持	城市污水处理率 (%)
		工业用水重复率 (%)
弹性力层数据	生态数据	森林覆盖率 (%)
		年降水量 (mm)

根据表 1 所示的压力层数据、承载力层数据和弹性力层数据的需求, 设计包含 3 层的水生态承载力分析模型, 通过输出值不断调整模型的权重以及误差, 如图 4 所示.

输入层有 3 个节点, 分别为水资源压力层数据 P、水生态承载力层数据 S 和水生态弹性力层数据 E. 隐

藏层节点的个数通过式 (4) 进行计算, 得出节点的个数范围为[3, 12], 在本模型中选取 6 个节点, 输出层有一个节点, 为 EF 评估值.

$$L = \sqrt{n+m}+a \tag{4}$$

$$\Delta E = W_{EC} - W_{EF} \tag{5}$$

将流域流经区域地区的水生态足迹与生态承载力相比较, 就会得到水生态资源环境是否为生态赤字或者生态盈余, 如式 (5) 所示, 若$\Delta E > 0$, 水生态环境呈现盈余, 说明该区域水生态供给充足, 水资源可持续发展利用. 若$\Delta E < 0$, 水生态环境赤字, 水资源的供给大于自身可以提供的生态环境支撑, 容易对环境过度使用, 对水生态环境环境造成破坏.

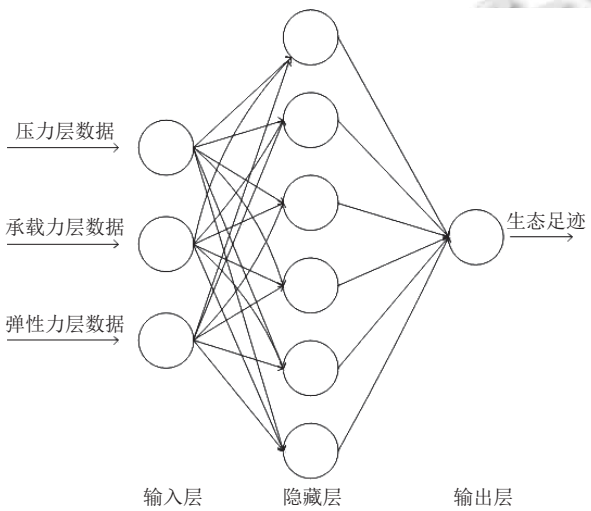

图 4 水生态承载力模型

2.3 水生态承载力模型 MapReduce 并行计算

本文模型实现采用传统的反向传播算法, 并且参考文献[16]中的一些思想.

MapReduce 是一种编程模型适用于大规模数据处理的相关实现. 开发者只需实现 mapper 函数和 reducer 函数就定义好了 MapReduce 作业, mapper 函数初始键/值对 (key-value), 根据初始键/值对计算产生中间键/值对. MapReduce 框架会将产生的中间键值对中键相同的值传递给一个 reduce 函数. Reducer 函数接受一个键以及一组值, 将这组值进行合并产生一组规模更小的值[17], 具体的操作如下:

将训练集分配到多个节点, 执行多个 mapper 任务, 每一个 mapper 接收一个训练项, 然后使用训练项计算出模型中权重的更新值, 并将产生的中间键/值对, 形如 (key=权重; value=更新值) 暂存于本地系统文件; 然后执行多个 reducer 任务, 每个 reducer 收集一个权重的更新值, 并计算更新值的平均值, 然后将计算所得的平均值作为权重的更新值; 更新模型中所有的权重的值. 重复执行 mapper-reducer 任务直到达到预期的精度.

3 实验应用

3.1 实验环境

实验环境基于 Hadoop 大数据平台, 实验应用采用 Java 做为编程语言, JDK 版本为 1.8.0_181, 采用分布式搭建大数据环境, 选择 5 台 PC 搭建, 其中一台作为 Master(NameNode), 其余 4 台作为 slave(DataNode). 环境信息及配置如表 2 和表 3 所示.

表 2 节点配置

属性	值
OS	Centos -6.10
Hadoop	2.6.0
Java	1.8.0_181
内存	1 GB
硬盘	30 GB

表 3 Hadoop 参数信息

Hadoop 参数	值
dfs.block.size	64 MB
dfs.replication	2
dfs.heartbeat.interval	3 s

3.2 实验数据

本文选取辽河流域 2012~2018 年的水生态监测数据以及人口数据、GDP 增长基础数据作为研究对象. 数据分别来源于辽宁省环境监测站监测数据、辽宁省沈阳市、盘锦市、鞍山市、营口市、铁岭市的年水资源公报以及辽宁省统计年鉴. 选取 2012~2017 年的数据作为训练集, 2018 年的数据作为测试集.

3.3 生态承载力结果分析

分析 2012~2017 年辽河流域流经区域的水生态承载力是否符合生态发展的规律, 是否呈现可持续发展状态. 选取人口数量、生活用水、农业灌溉、工业用水、GDP 增长率、万元 GDP 平均耗水量作为生态压力层数据; 选取人均水资源量、城市污水处理量、工业用水重复率作为承载力层数据; 选取年降水量、森林覆盖面积作为弹性力层数据. 对 2012–2017 年间的辽河流域流经区域的数据通过 EF 计算方法对生态足

迹进行计算, 通过 30 组输入数据以及 EF 所得数据对水生态承载力模型进行训练.

将 2018 年辽河流域流经地区的压力层数据 P、支撑力层数据 S、弹性力层数据 E 作为输入层节点数据输入到水生态承载力模型中, 计算出 2018 年的辽河流域流经区域的水生态足迹的评估值, 如表 4 所示.

表 4 水生态承载力盈余/赤字

流经区域	EF 评估值 (10^6hm^2/人)	ECC(10^6hm^2/人)	盈余/赤字
沈阳	4.3	1.45	赤字
盘锦	2.12	0.29	赤字
鞍山	1.57	1.97	盈余
营口	1.27	0.43	赤字
铁岭	1.55	0.86	赤字

将 ECC 所得值与生态足迹评估值进行比较, 通过分析, 沈阳、铁岭、盘锦、营口的生态承载力值均小于生态足迹评估值, 处于水生态环境赤字状况, 尤其是盘锦地区生态承载力达到最低值, 说明对水资源环境过度使用; 鞍山生态承载力值大于生态足迹评估值, 处于生态盈余情况, 如图 5 所示. 通过分析以往数据显示, 鞍山水资源总量较往年相比减少了 30%, 但相较于其他地区多, 水生态环境处于可持续发展状态.

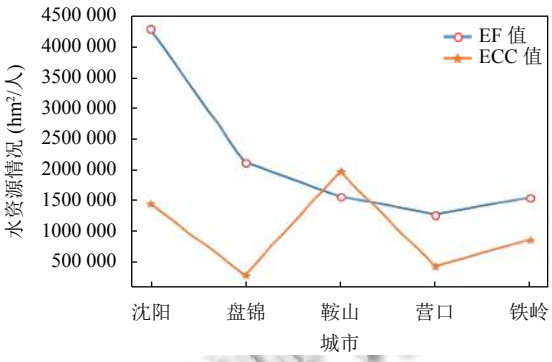

图 5 生态承载力分析图

3.4 Hadoop 集群吞吐能力分析

为了能够验证利用 Hadoop 集群存储处理海量数据方面比单机数据处理技术上能够有更优越的表现[18], 本实验在选取不同的数据量在 Local 单机模式以及 Full-Distributed Mode 集群模式下进行运行时间测试, 在这两种情况下都使用“清洗”后的数据进行规则计算. 数据表包括 3 列: 唯一标识符、监测时间、监测值. 用“清洗”后的数据统计时间段内, 监测物的超标次数.

从表 5 可以看出, 在数据量较小的情况下单机运行的时间更短, 处理数据的效率更高, 而数据在超过 3 GB 之后, Hadoop 集群的运行时间更短, 且稳定运行, 时间跨度不是很大.

表 5 Hadoop 集群与单机运行时间

数据量 (GB)	Hadoop 集群运行时间 (s)	单机运行时间 (s)
1	210	40
3	290	183
5	310	462
8	355	580
15	420	1120

4 结论与展望

本文对水生态环境承载力的分析从现实生态环境出发, 提出基于大数据的水生态承载力分析模型, 利用大数据技术对水资源、水生态数据处理分析, 以及增加生态足迹计算的数据多样性, 通过生态承载力分析模型输出值与生态承载力相比较, 得出水生态环境当前发展情况是否赤字或盈余. 应用案例表明, 在增加数据多样性的同时能够通过水生态承载力模型对生态足迹做出准确的分析, 减少了数据进行各类公式计算的过程, 提高了工作的效率并丰富了数据来源的多样性. 基于大数据的水生态承载力模型加深了对历史数据的分析与挖掘, 在未来科学和技术的发展下, 以及数据资源库资源的不断完善, 能够对水生态环境承载力做出更准确的分析结果.

参考文献

1 李安增, 王宁, 王常权, 等. 大数据技术在环境信息中的应用. 计算机系统应用, 2015, 24(1): 60–64. [doi: 10.3969/j.issn.1003-3254.2015.01.010]

2 刘卫萍, 王宁, 周晓磊, 等. 数据融合技术在环境监测领域的应用. 计算机系统应用, 2016, 25(6): 88–93. [doi: 10.15888/j.cnki.csa.005202]

3 Yang HC, Dasdan A, Hsiao RL, *et al*. Map-reduce-merge: Simplified relational data processing on large clusters. Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data. Beijing, China. 2007. 1029–1040.

4 Apache Hadoop. http://hadoop.apache.org/, 2014.

5 Moturi CA, Maiyo SK. Use of mapreduce for data mining and data optimization on a web Portal. International Journal of Computer Applications, 2012, 56(7): 39–43. [doi: 10.5120/8906-2945]

6 Parmar KS, Bhardwaj R. River water prediction modeling

using neural networks, fuzzy and wavelet coupled model. Water Resour Manage, 2015, 29(1): 17–33. [doi: 10.1007/s11269-014-0824-7]

7 李辉, 王建文, 叶明雯. 基于 Hadoop 的海量气象水文数据并发处理模型. 计算机应用, 2018, 38(S2): 187–191, 205.

8 焦雯珺, 闵庆文, 李文华, 等. 基于 ESEF 的水生态承载力评估-以太湖流域湖州市为例. 长江流域资源与环境, 2016, 25(1): 147–151. [doi: 10.11870/cjlyzyyhj201601018]

9 黄林楠, 张伟新, 姜翠玲, 等. 水资源生态足迹计算方法. 生态学报, 2008, 28(3): 1279–1286. [doi: 10.3321/j.issn:1000-0933.2008.03.044]

10 叶茂, 夏润亮, 刘颖, 等. 基于大数据的省级水利数据中心体系设计. 计算机与网络, 2018, 44(17): 60–62. [doi: 10.3969/j.issn.1008-1739.2018.17.053]

11 翁异静, 邓群钊, 杜磊, 等. 基于系统仿真的提升赣江流域水生态承载力的方案设计. 环境科学学报, 2015, 35(10): 3353–3366.

12 靳超, 周劲风, 李耀初, 等. 基于系统动力学的海洋生态承载力研究——以惠州市为例. 海洋环境科学, 2017, 36(4): 537–543.

13 沈思祎, 钮尔轩, 孟斌. 基于灰色关联度的辽宁近海海域生态环境承载力评价. 大连海事大学学报, 2017, 43(3): 112–118.

14 王文国, 龚久平, 青鹏, 等. 重庆市水资源生态足迹与生态承载力分析. 生态经济, 2011, (7): 159–162.

15 罗娜. 辽宁省水资源生态足迹动态变化与时间序列预测分析研究[硕士学位论文]. 大连: 辽宁师范大学, 2012.

16 周志华. 机器学习. 北京: 清华大学出版社, 2016.

17 Liu ZQ, Li HY, Miao GS. Mapreduce-based backpropagation neural network over large scale mobile data. Proceedings of the 2010 Sixth International Conference on Natural Computation. Yantai, China. 2010. 1726–1730. [doi: 10.1109/ICNC.2010.5584323]

18 曾理, 王以群. Hadoop 集群和单机数据处理的耗时对比实验. 硅谷, 2009, (19): 55–56.