

Application of Big Data and Cloud Platform Technology in the Study of Poyang Lake Ecological Environment

Wenjing Li, Xiaohua Xu*, Haijing Wang

Jiangxi Institute of Water Sciences, Nanchang Jiangxi
Email: 1015733858@qq.com, *76964225@qq.com

Received: August 12th, 2019; accepted: August 27th, 2019; published: September 3rd, 2019

Abstract

This paper adopts high-tech, cloud computing, cloud platform and other high-tech, with the ecological environment protection and comprehensive management of Poyang Lake as the demand, and aims to improve the modern management level of Poyang Lake, build a cloud platform through Hadoop cluster, and the establishment of Poyang Lake ecological environment information-sharing-service cloud platform system. The establishment of this platform has realized the functions of monitoring and analysis of the ecological environment of Poyang Lake, information sharing services, GIS map services and data maintenance. Through this platform, we can fully grasp the current ecological environment of the Poyang Lake area, and provide support for the monitoring and evaluation measures and supporting government decision-making. It will contribute to the construction and management of Poyang Lake's informatization and promote the development of smart water conservancy in our province.

Keywords

Big Data, Cloud Computing, Cloud Platform, Poyang Lake Ecological Environment, Smart Water Conservancy

大数据与云平台技术在鄱阳湖生态环境研究中的应用

李文晶, 许小华*, 王海菁

江西省水利科学研究院, 江西 南昌
Email: 1015733858@qq.com, *76964225@qq.com

*通讯作者。

摘要

本文采用大数据、云计算、云平台等高新技术，以鄱阳湖生态环境保护与综合管理为需求，以提高鄱阳湖现代化管理水平为目标，通过Hadoop集群的方式搭建云平台，建立鄱阳湖生态环境信息共享服务云平台系统。本平台的建立实现了鄱阳湖生态环境的监测与分析、信息共享服务、GIS地图服务和数据维护等功能。通过本平台能够全面掌握鄱阳湖环湖区生态环境现状，为监测评估措施成效和辅助政府决策提供支持依据。有助于鄱阳湖信息化建设与管理，促进我省智慧水利的发展。

关键词

大数据，云计算，云平台，鄱阳湖生态环境，智慧水利

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着互联网技术的高速发展，从国家战备到各行业层面都大力推进云计算、大数据、物联网、移动互联和人工智能等技术创新和深入应用[1]。近年来，大数据与云计算技术在水利信息化发展中也发挥了重要作用[2]。“鄱阳湖生态经济区”是江西省第一个纳入国家战略的区域性发展规划，也是我国十大生态功能保护区之一。它拥有丰富的鸟类和鱼类等物种资源，并负责调节气候，调节洪水储存和降低污染等各种功能[3][4]。近年来，随着江西省经济社会的发展，人口增长过快，生态问题开始显现，鄱阳湖环湖区生态问题不仅严重影响国民经济和社会的可持续发展，还威胁人民群众的健康。利用大数据、云计算、云平台等技术建立鄱阳湖生态环境信息共享服务云平台。整合鄱阳湖生态资源数据，充分利用云平台将鄱阳湖生态信息建立在云平台上，方便相关人员快速访问鄱阳湖生态变化信息，为鄱阳湖生态环境管理提供科学指导，有利于全面规划鄱阳湖区的建设与保护，其研究成果具有重要应用价值。探索大数据、云平台等技术并优先在鄱阳湖保护和管理中的示范应用也是推动江西省智慧水利建设的重要手段。

2. 大数据技术

大数据自提出以来尚未有统一定义[5]。维基百科提供了大数据定义，因为涉及的数据非常大，无法通过当前的主流软件工具在合理的时间内管理，处理和组织成有用的信息。它不是一个空的概念，其出现对应了数据产生方式的变革。在大数据时代，大数据的核心是提取有价值的数据并进行预测，因此经过分析和处理的数据才能创造价值[6]。大数据技术主要有数据采集、数据存储、数据处理、数据分析、数据挖掘、数据可视化展示等[7]。在本平台数据处理中，数据需要及时更新、存储和分析，因此，大数据技术在本平台建设中占有重要位置。本平台数据处理流程如图1所示。

3. 云计算平台技术

谷歌的CEO埃里克·施密于2006年首次提出了云计算平台[8]。当时人们还不理解云计算的概念，经过近几年的研究和发展，各个领域已经开始运用云计算平台技术。云计算平台是面向用户服务的计算

平台, 通过互联网将基础设施资源、软件资源整合起来, 根据用户实际需求提供服务收取相关费用。本平台采用采用公共化的方式计算资源, 用户不需要关心底层服务, 只需要关注自身应用业务逻辑, 降低了用户开发的难度, 节省了企业开发应用的成本。云平台技术提供的三种服务模式[9]有基础设施作为服务(Infrastructure as a Service, IaaS)、平台作为服务(Platform as a Service, PaaS)和软件作为服务(Software as a service, SaaS)。云平台层次结构如图 2 所示。

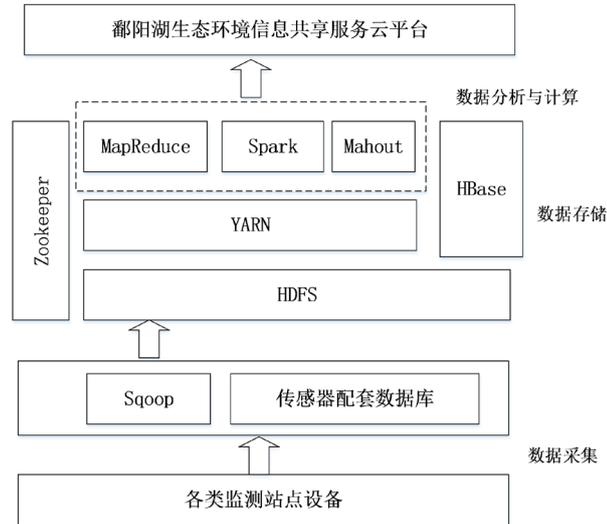


Figure 1. Data processing flow chart

图 1. 数据处理流程图

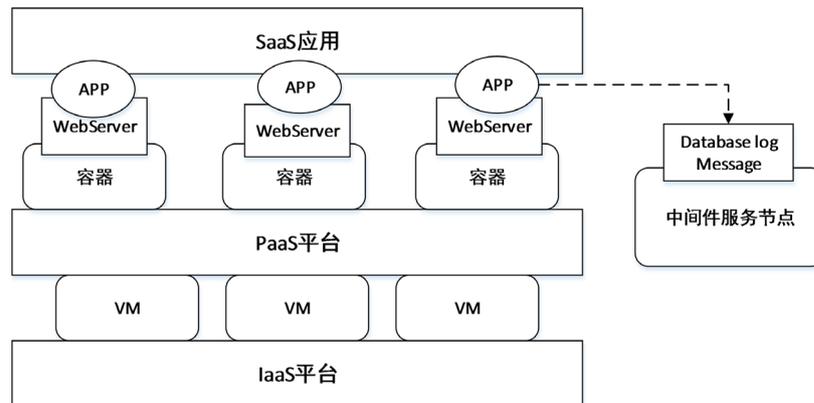


Figure 2. Cloud platform hierarchy diagram

图 2. 云平台层次结构图

从图 2 可知, IaaS 平台即为用户提供物理资源虚拟、网络资源虚拟、存储资源虚拟等的基础设施的服务, 提供最底层的服务。PaaS 平台即为开发者用户提供云端集成的软件开发服务。SaaS 应用即为云平台顶层, 为用户提供各类云端应用。

4. 鄱阳湖生态环境信息共享服务云平台

4.1. 平台架构

鄱阳湖生态环境信息共享服务云平台基于 SOA 架构, 依托云计算、大数据、云平台等技术为支撑,

根据云计算的服务模式构建鄱阳湖基础设施服务层(IaaS)、鄱阳湖大数据层(DaaS)、鄱阳湖云平台服务层(PaaS)和鄱阳湖云应用层(SaaS)四层架构。整体架构如图 3 所示。



Figure 3. Platform architecture diagram
图 3. 平台架构图

鄱阳湖基础设施服务层(IaaS): 为达到“互联互通、资源共享、支撑应用、提高效能、减少成本”的目的, 基于云计算技术, 构建鄱阳湖基础设施云平台, 提供本平台相关物理资源设备支撑, 包括相关服务器、路由器、存储设备、核心交换机等硬件设备。

鄱阳湖大数据层(DaaS): 利用大数据集群、存储、数据库等技术, 构建鄱阳湖数据资源管理平台, 支撑鄱阳湖生态环境各类基础数据、多媒体数据等信息资源的管理, 同时, 采集人口、经济、社会、水位、水生态、水资源等信息, 形成鄱阳湖水生态数据库。

鄱阳湖云平台服务层(PaaS): 基于大数据处理平台, 构建资源共享服务平台, 满足数据资源汇集共享、应用服务, 数据处理等功能, 提供数据处理相关的操作引擎, 提高数据处理效率, 为本平台的应用提供软件支撑环境。

鄱阳湖云应用层(SaaS): 利用大数据 SQL 语言、可视化服务展现等技术, 建立鄱阳湖生态环境信息共享服务云平台。

4.2. 平台环境

鄱阳湖生态环境信息共享服务云平台开发需要三台硬件服务器, 服务器配置为浪潮英信 NF5270M4 8G 2T, 一台服务器作为集群主节点服务器, 另外两台服务器作为集群从节点服务器。开发所需要的软件

有 MyEclipse 10.0、Hadoop3.0、分布式协调器 Zookeeper3.4、jdk1.8、Tomcat8.0、分布式数据库 Hbase 2.1、数据仓库 Hive2.3、数据库 Mysql 5.0 等。

4.3. 平台搭建

鄱阳湖生态环境信息共享服务云平台存储、计算、网络等物理资源均采用国际云计算应用中公认的 VMware、CitrixIBM 等虚拟化产品进行资源分配[10]。根据需求进行弹性的资源分配服务,再结合 Hadoop 的集群技术,实现在成本低投入的情况下获得高性能、较好的灵活性和可靠性[11]。VMware 虚拟机安装和 Hadoop 集群相关配置如下所述。

(1) 虚拟系统安装

在本系统服务器安装 ESXi 服务器,通过虚拟软件虚拟多台主机,并安装相关虚拟化软件,通过 vsSphere Client 客户端来管理资源,并在虚拟计算机上安装部署 Hadoop、Spark、ElasticSearch 等软件。

(2) Hadoop 系统配置

鄱阳湖生态环境信息共享服务云平台采用三台物理服务器、运用虚拟化技术虚拟成 11 台虚拟机来构建 Hadoop 集群环境,并修改相关 Hadoop 配置文件,slaves 配置文件、masters 配置文件等,实现 Hadoop 系统集群。

4.4. 平台数据库设计

通过本平台所需使用的数据库进行分析,本平台包含的数据库有基础信息数据库、水文与水资源数据库、生态环境数据库及系统管理数据库等。鄱阳湖生态环境信息共享服务云平台数据库设计如图 4 所示。

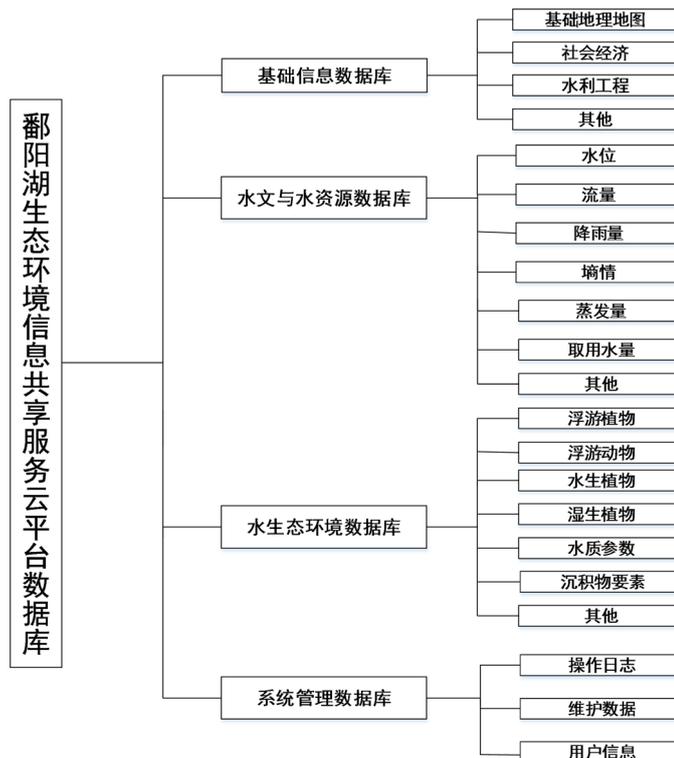


Figure 4. Platform database design

图 4. 平台数据库设计

(1) 基础信息数据库：基础信息数据库存储鄱阳湖基础地理地图、社会经济、水利工程和其他数据。基础地理地图包括数字线画图(DLG)、数字正摄影像图(DOM)、数字高程模型图(DOM)等，为本平台中的 GIS 地图服务提供基础数据支撑。

(2) 水文与水资源数据库：水文与水资源数据库存储与水文与水资源相关的水位、流量、蒸发量、土壤墒情、取用水户用水量等数据，为本平台的信息查询提供服务。

(3) 水生态环境数据库：水生态环境数据库存储水生态环境相关数据，如鄱阳湖区域中湿地、候鸟、钉螺、水土流失、生物、水环境等数据，为本平台的生态指标查询提供服务。

(4) 系统管理数据库：系统管理数据库主要是存储系统后台相关数据，如操作日志、用户信息、维护数据等。

4.5. 平台功能设计与实现

本平台的总体目标是利用云计算，大数据，云平台等技术构建符合统一数据标准的鄱阳湖生态环境信息共享服务平台。平台开发使用 Java 语言，采用 B/S 结构、前后端完全分离的方式设计，后台开发使用 SM (Spring Boot + Mybatis)构建服务，完全遵循 RESTful 规范，前端开发采用 webpack + vue 组件化模式。平台地图服务功能基于 WebGIS 平台二次开发实现。平台设计中充分考虑了系统的性能可靠性和兼容性问题，保证系统中的数据正确性和准确性，支撑每秒 500 人以上的并发操作，图表以及复杂查询能在 5 秒内显示查询结果，简单查询 3 秒内显示查询结果，并且兼容常用的 360、IE、谷歌等浏览器访问。

平台主要功能实现目标：建立鄱阳湖生态大数据，实现鄱阳湖生态数据的有效存储，提供信息查询服务，提供信息操作与管理、信息共享服务等。通过将异构、分布式的信息资源按照统一标准进行整合，实现信息共享。平台具体功能结构如图 5 所示。

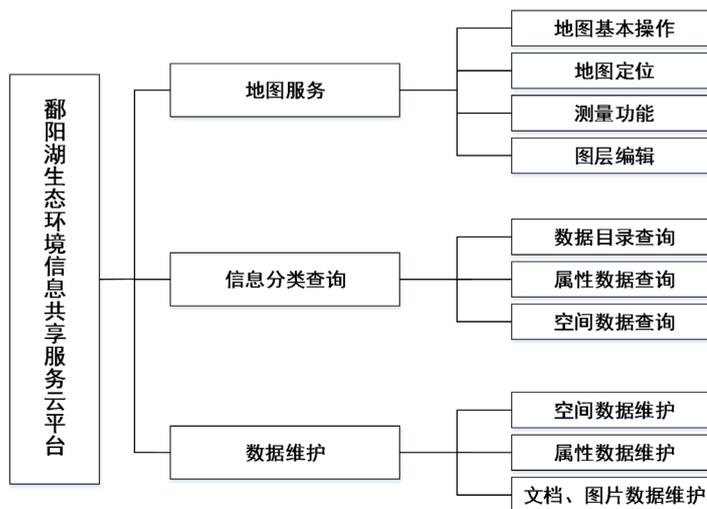


Figure 5. Platform function structure diagram

图 5. 平台功能结构图

(1) GIS 地图服务

实现地图显示，交互，操作，查询等功能。1) 基本地图操作：放大，缩小，平移，完整地图，打印，鹰眼视图等。2) 图层控制：图层的显示控制。所有的地理信息图形是分层管理的，可根据需求选择相关图层内容。3) 图例：提供显示所有资源和对应的图例列表。4) 测量功能：可完成相关图形测量工作。例如：测量点之间的线段长度、矩形面积、不规则图形面积等；5) 地图定位：可以通过对字段进行搜索，

在地图上进行展示相应字段的图标位置。如：行政区域、河流湖泊、乡镇、水利工程、各类监测站点等。

6) 地图查询：通过在地图上展示的图标信息，如监测站点图标、乡镇图标、水利工程图标等，查看具体的图标内容详情信息。

(2) 信息分类查询

可以在目录树中直接选择数据分类查询以指定查询数据类型。文本查询以用户输入文本信息的形式提供查询，并为不同的业务数据类型设置查询选项。如设置数据的时间间隔频率，空间范围，数据精度和其他技术指标来筛选数据。

1) 数据分类查询

数据分类主要分为水文气象监测、水质监测、鸟类监测、渔业监测、水生动植物监测、湿地水生动植物监测、水利工程、区域水质和水草等。数据分类查询展示如图 6 所示。

序号	站点	经度	纬度	风向风力	气压	监测日期	操作
1	栎寄	114.7	26.56667	东南风4级	3500.0	2012-01-08	🔍 📄
2	西安	114.98333	27.1	东南风4级	3455.0	2012-01-08	🔍 📄
3	樊江	115.15	27.55	东南风4级	3422.0	2012-01-08	🔍 📄
4	樟树	115.53333	28.06667	东南风4级	3443.0	2012-01-08	🔍 📄
5	外洲	115.83333	28.63333	东风5级	4433.0	2012-01-08	🔍 📄
6	樊山	115.21667	25.91667	东风5级	2233.0	2012-01-08	🔍 📄
7	郑门牌	115.75	25.23333	东风5级	2222.0	2012-01-08	🔍 📄
8	麻州	115.78333	25.51667	东风5级	1333.0	2012-01-08	🔍 📄
9	半岗江	115.38333	25.31667	东南风4级	3444.0	2012-01-08	🔍 📄
10	1					1900-01-01	🔍 📄
11	宁都	116.01667	26.48333	东南风4级	6443.0	2012-01-08	🔍 📄
12	100					1900-01-01	🔍 📄
13	塘坑	115.66667	26.13333	南风3级	33.0	1900-01-01	🔍 📄
14	11	111.0				1900-01-01	🔍 📄
15	桥下壑	116.05	26.55	东南风5级	34.0	2013-01-08	🔍 📄
16	2					1900-01-01	🔍 📄
17	222	222.0				1900-01-01	🔍 📄
18	3					1900-01-01	🔍 📄
19	4					1900-01-01	🔍 📄
20	5					1900-01-01	🔍 📄

Figure 6. Data classification query display

图 6. 数据分类查询展示

2) 属性查询

通过资源对象的属性字段进行模糊查询，例如通过鄱阳湖区水质监测站点名称，查询相关站点相关的详细信息，并能在地图上进行相关的展示。

3) 空间查询

提供三个功能：空间区域查询、生态资源目录、信息查询。空间区域查询分为矩形查询、多边形查询、固定区域查询和缓冲区查询。矩形查询允许用户拉动地图上的矩形区域框架以查询矩形区域中的生态信息。多边形查询允许用户绘制任意多边形区域查询相关生态资源信息。固定区域查询是供用户选择如城市，区域，乡镇等条件下的生态信息。信息查询允许用户选择指定资源类型的属性信息，并查看列表中资源类型的详细信息，例如查询自然保护区信息列表中特定自然保护区的细节。

4) 生态指标查询

按照三级指标体系从数据库中读取生态指标生成生态指标目录树，用户可以选择查看任意一级的生态指标，显示生态指标数据。

(3) 数据维护

数据维护提供人工方式录入和批量导入数据两种数据维护方式。

1) 空间地理数据维护

空间矢量数据维护涉及线、多边形等，并提供基于 Web 的分布式空间数据编辑。空间数据编辑界面如图 7 所示。



Figure 7. Spatial data editing interface
图 7. 空间数据编辑界面

2) 属性数据维护

为用户提供生态资源数据对象属性数据维护，具有添加、修改、删除数据记录和和数据项的功能。

3) 多媒体数据维护

将平台中相关的文档、视频、音频、图片等进行记录与存储，用户能进行上传、下载、修改和删除等操作。

5. 结语

本文通过对大数据、云计算平台等技术的分析和探讨，对鄱阳湖生态环境的研究，构建了鄱阳湖生态环境信息共享云服务平台。本文从云平台的架构、环境、搭建方法和系统功能阐述了鄱阳湖生态环境信息共享云服务平台。本平台的建立将鄱阳湖生态数据进行共享，能够及时发现鄱阳湖区生态环境变化危机，为决策者提供解决方法。本平台的研究对改善鄱阳湖区生态环境具有重要意义。

基金项目

江西省水利厅科技计划项目(201818ZDKT14)。

参考文献

[1] 柳林, 刘恒伟, 王哲. 云计算在水利信息化建设中的应用探讨[J]. 海河水利, 2012(4): 52-56.

-
- [2] 罗婷予. 智慧城市研究与规划中大数据技术应用[J]. 低碳世界, 2017(32): 178-179.
- [3] 黄宽. 鄱阳湖流域生态数据库和信息共享平台功能设计与实现[D]: [硕士学位论文]. 南昌: 东华理工大学, 2013.
- [4] 许小华, 章重, 雷声, 等. 基于 GIS 的鄱阳湖区生态数据库和信息共享服务平台构建[J]. 江西水利科技, 2015, 41(5): 327-331, 350.
- [5] 刘蔚然, 刘莉娜. 大数据技术[J]. 冶金设备管理与维修, 2014, 32(4): 33-36.
- [6] 张锋军. 大数据技术研究综述[J]. 通信技术, 2014, 47(11): 1240-1248.
- [7] 张颖. 大数据时代高校图书馆信息资源管理的创新与发展[J]. 武汉理工大学学报(社会科学版), 2015, 28(3): 449-452.
- [8] 唐箭. 云计算研究综述[J]. 民族高等教育研究, 2010, 16(2): 15-16.
- [9] 吴国强. 基于虚拟化技术的云计算平台架构研究[J]. 电脑编程技巧与维护, 2017(16): 61-62.
- [10] 单冬红, 郭静博, 赵伟艇. Hadoop 集群作业调度算法优化技术研究[J]. 现代电子技术, 2016, 39(6): 25-29.
- [11] 郑柏恒, 孟文, 易东, 等. 在 Hadoop 集群下的智能电网数据云仓库设计[J]. 制造业自动化, 2014(19): 134-138.

知网检索的两种方式:

1. 打开知网首页: <http://cnki.net/>, 点击页面中“外文资源总库 CNKI SCHOLAR”, 跳转至: <http://scholar.cnki.net/new>, 搜索框内直接输入文章标题, 即可查询;
或点击“高级检索”, 下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询。
2. 通过知网首页 <http://cnki.net/>顶部“旧版入口”进入知网旧版: <http://www.cnki.net/old/>, 左侧选择“国际文献总库”进入, 搜索框直接输入文章标题, 即可查询。

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: csa@hanspub.org