

# Research and Design of Hydrological Big Data Sharing Platform

Hua Chen<sup>1</sup>, Jian Xu<sup>1</sup>, Zhiyuan Xiao<sup>2</sup>, Jiawei Yang<sup>1</sup>, Jie Chen<sup>1</sup>, Shenglian Guo<sup>1</sup>, Chongyu Xu<sup>1</sup>

<sup>1</sup>State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan Hubei

<sup>2</sup>Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan Hubei

Email: chua@whu.edu.cn, xujianwhu@gmail.com

Received: Dec. 5<sup>th</sup>, 2017; accepted: Dec. 21<sup>st</sup>, 2017; published: Dec. 28<sup>th</sup>, 2017

---

## Abstract

In this paper, the methodology of hydrological big data standardization is discussed upon analyzing on the characteristics of hydrological data. Solutions on data preprocessing, data indexing and highly efficient data reading and writing are also introduced. The mass storage capacity and high speed computing capability of Hadoop are utilized for designing and implementing hydrological big data sharing platform. Accordingly, the platform can be technical support for information sharing and space integration between water conservancy industry and other industries, as well as the interdisciplinary sustainable development.

## Keywords

Big Data, Hydrological Data Sharing, HDFS, HBase

---

# 水文大数据共享平台研究与设计

陈 华<sup>1</sup>, 徐 坚<sup>1</sup>, 肖志远<sup>2</sup>, 杨家伟<sup>1</sup>, 陈 杰<sup>1</sup>, 郭生练<sup>1</sup>, 许崇育<sup>1</sup>

<sup>1</sup>武汉大学水资源与水电工程科学国家重点实验室, 湖北 武汉

<sup>2</sup>长江水利委员会水文局, 湖北 武汉

Email: chua@whu.edu.cn, xujianwhu@gmail.com

收稿日期: 2017年12月5日; 录用日期: 2017年12月21日; 发布日期: 2017年12月28日

---

## 摘 要

本文根据水文数据的特点探讨了水文大数据标准化方法, 探索数据预处理、数据索引、数据高效存储等水文大  
作者简介: 陈华(1977-), 男, 福建人。武汉大学教授、博导。研究方向为流域水文模拟、气候变化对水文水资源的影响研究、水利信息化。

文章引用: 陈华, 徐坚, 肖志远, 杨家伟, 陈杰, 郭生练, 许崇育. 水文大数据共享平台研究与设计[J]. 水资源研究, 2018, 7(1): 10-18. DOI: 10.12677/jwrr.2018.71002

数据共享平台关键技术；利用Hadoop对多源异构数据的海量存储能力及高速计算能力，研究基于MapReduce的水文大数据分布式数据处理模型，设计和实现水文大数据共享平台，为水利及跨行业跨部门的信息共享、空间集成，以及跨学科的可持续发展研究提供技术支撑。

## 关键词

大数据，水文数据共享，分布式文件系统，分布式数据库

Copyright © 2018 by authors 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

数据共享作为科技创新的原动力之一，已受到科学家乃至社会公众的广泛关注。美国的科学数据共享工作始于 20 世纪 80 年代，1990 年，美国航空航天局(NASA)着手建设“分布式最活跃数据档案中心群(DAACs, Distributed Active Archive Centers)”，DAACs 由 9 个数据中心构成，水循环数据中心 MSFC 也在其中。1999 年，DAACs 为了扩大其服务的领域，建立起 DAAC 联盟(DAACs Alliance)，全球水文研究中心也加入到这个联盟。在欧洲，欧洲环境总署(EEA, European Environment Agency)负责研究水利科学数据共享方面的工作，欧洲水主题中心(ETC WTR, European Topic Center on Water)协助 EEA 的日常工作，发布相关水信息[1]。

2001 年，我国开始实施科学数据共享工程，并启动了基础性工作专项资金项目“气象资料共享系统建设”，标志着我国科学数据共享迈入了试点建设阶段[2]。2002 年底，科技部又启动了 5 个科学数据中心，即测绘科学数据共享服务中心、水文信息共享服务中心、地震科学数据共享中心、林业科学数据中心、农业科学数据中心。当前，我国水文数据共享虽然取得了一定进展，但总体上还存在很多问题，可归纳为以下几点：1) 共享机制与标准不够健全。机制与标准是水文大数据共享的基础，目前水文信息汇集方面已有成熟的标准，但水文大数据交换与共享方面的标准还很少；2) 同一机构多类数据存储分散。例如，在某县市水利局中，可能同时存在实时雨水情数据库、基础水文数据库、地下水项目数据库、山洪灾害防治项目数据库和国家水资源能力监控数据库等，这样造成数据的分散存储，不利于后期数据库的扩充和整合，也不利于海量水文数据的分析；3) 不同机构数据存储冗余。例如，水利部水文局与各流域水文局、各省水文局里存储了相当一部分相似的数据。数据的冗余存储导致信息的不一致性，阻碍了数据共享；4) 数据存储能力有限。采用传统方式存储海量水文数据对硬件配置要求非常高，即便硬件性能的发展非常快，也赶不上数据增长的需求。如何实现资源和计算能力的分布式共享以及如何应对当前数据量高速增长的势头，也是水文数据管理、数据处理领域亟待解决的问题。

大数据时代的信息运行模式正在改变人们的工作方式和学习方式，为未来的世界认知和交流方式提供全新的运行环境。国家“十三五”规划建议提出：“实施国家大数据战略，推进数据资源开放共享”。大数据已经成为一种重要的国家战略资源，拥有数据才能拥有话语权。近几年国内外针对水利大数据的研究刚刚起步，取得了一些研究成果，主要集中在水利大数据特征和内涵、水利大数据存储和共享、水利大数据挖掘等方面[3]。莫荣强等 2013 年构造了水利大数据中心基础架构，并实现了基于混合云的水利大数据存储和管理[4]；Ai 和 Yue 2014 年构建了多源异构数据动态配置的水利大数据平台[5]；冯吉平等 2014 年将大数据技术应用于送料流域水环境管理[6]；Chalh 等在 2015 年构建了一套基于大数据的开放水资源管理平台[7]；Li 等 2016 年基于 POT 的大数据分析方法来预测洪水[8]；孙欣欣 2016 年采用大数据分析技术研究城市突发水涝灾害[9]。

水利部 2017 年印发的《关于推进水利大数据发展的指导意见》中，要求整合现有各种基础数据、动态监测

和监督数据, 汇聚其他行业数据和社会监督数据, 为全面推行河长制提供有效的数据支撑, 因此, 加快水利大数据相关研究迫在眉睫。本文根据水文数据的特点探讨水文大数据标准化方法和水文大数据共享平台关键技术, 并在此基础上设计和实现水文大数据共享平台系统。

## 2. 水文数据特性

水文数据是指水文的实测资料, 即通过水文测验所收集的各种水文要素的原始记录, 广义的水文数据还应包括水文年鉴、水文统计值、水文图集及水文调查资料等。仅水利行业与水文数据相关的数据库就包括了实时雨水情数据库[10]、基础水文数据库[11]和水质数据库[12], 几乎覆盖了水文部门的主要业务数据。水文数据因其特性具有总量大、种类多、增速快和价值高的特点:

**数据总量大:** 水文数据为一系列的时间序列数据, 且覆盖全国各大流域及行政区划的大、中、小河流流经的区域, 以测站为单位进行数据采集, 具有海量数据的特性。

**数据种类多:** 目前有实时雨水情数据库、基础水文数据库和水质数据库三类数据库覆盖了水文部门的主要业务数据。其中实时雨水情数据包含存储降水、蒸发、河道、水库、闸坝、泵站、潮汐、沙情、冰情、地下水、墒情、特殊水情、水文预报等 13 大类数据; 基础水文数据包括降水、蒸发及蒸发辅助项目、水位、流量(水量)、泥沙、水温、冰凌、潮汐、测站属性以及其他等十类数据; 水质数据库包括了大气降水、地表水以及地下水等各类水体中的水质监测信息。除此之外水文大数据还应包含地理信息数据、遥感数据和社会经济指标等多源异构数据, 具有数据种类多的特性。

**数据增速快:** 早在 2007 年, 全国分布在 7 大流域、31 个省和 125 个水情分中心的基本测站已达到数万[13], 实时水文信息, 非汛期每天至少更新 1 次, 汛期每天更新 4~8 次, 部分站点甚至每半小时更新一次, 数据更新频率高, 仅每天更新的水情数据可达到数十万条。

**数据价值高:** 水文数据作为水文行业资源性产品, 其价值是隐藏而非显性的, 随着水文数据规模的扩增, 通过数据挖掘和知识发现, 其潜在价值将会越来越大; 水文学科是地学的一个重要分支, 水文数据是研究区域或全球水循环过程的重要基础, 因此其数据的完整性和代表性在水文科学研究中具有重要意义。

目前水文数据主要是以表格的形式存储于诸如 Oracle、SQL Server 等关系型数据库中。水文测站的更新频率非常高, 几乎每小时都有实时数据更新, 产生数据表过大, 检索速度慢, 甚至超过数据库容量上限等问题。常用的解决手段有两种: 一是将历史数据单独存储备份, 但产生检索困难的问题; 二是将中间数据舍弃, 仅保留测站日数据, 但又有历史水文信息缺失的问题。这样一来对水文数据后续的分析和应用产生了重大影响。

大数据是指对数据规模大、结构复杂度高、关联度强的数据集进行处理与应用的信息技术。领域内对于大数据的概念众说纷纭, 未能统一, 主要通过“4V”特性来表征, 即: 巨量性(Volume), 存储空间大, 计算量大; 多样性(Variety), 来源多, 种类多, 格式多; 时效性(Velocity), 增长速度快, 处理速度要求快且高时效; 实用性(Value), 数据中包含着有实用价值的信息。大数据技术的更新换代可谓日新月异, 相关技术可成为数据获取、存储、处理分析或可视化的有效手段。因此, 水文数据的存储模式由传统关系型数据库向大数据转变是水利行业顺应时代发展的必然趋势。

## 3. 水文大数据标准化设计

### 3.1. 水文大数据分类

针对大数据本身和各个行业的大数据模型或标准研究开始较晚, 直到 2015 年才发布了首个大数据国际标准《ITU-T Y.3600 基于云计算的大数据需求与能力标准》, 回答了大数据的定义、特征、生态系统、功能需求等基本问题。水文行业内大数据模型或标准至今仍未出现, 而水文大数据共享机制研究的核心是水文大数据标准化。

水文是自然界中水的变化、运动等的各种现象，一般指自然界水的时空分布、变化规律。因此水文数据必备的元素包括时间、空间和业务信息。目前我国水文行业数据种类繁多，主要包括实时雨水情数据、基础水文数据和水质数据等，除此之外还包括水文测验数据、水质采样数据、水资源数据库、山洪灾害监测数据等。水文数据所涉及的特征主要包括降水、蒸发、径流(水位、流量、流速)、泥沙含量、水质、土壤含水量等。根据目前水文行业数据的特点，监测数据需要长期进行，具有海量大数据特征。本文选取了 Apache 开源软件基金会发布的 HDFS 和 HBase 存储水文大数据集，对各类水文海量数据集进行分析、归类和总结，实现水文大数据集存储。

### 3.2. 水文大数据集存储

水文大数据集是指更新十分频繁，且数据量为海量数据，包括结构化数据和非结构化数据两类。其中结构化数据主要包括目前存储于关系型数据库中的水文业务数据，如降雨量表、河道水清表和水库水情表等，此类数据存储于 Hbase 中；而非结构化数据是指其字段长度可变，且每个字段的记录又可由可重复或不可重复的字段构成的数据，可处理的数据包括文本、图像、声音和视频等，非结构化水文大数据集主要包含各类报告、实景图片、实景音频视频等数据，此类数据可直接存储于 HDFS 中。

HDFS (Hadoop Distributed File System)是被设计成适合运行在通用硬件上的分布式文件系统，是一个能提供高吞吐量数据访问的高度容错性系统，具有高容错性、适合大数据处理、流式文件写入和可部署于低廉的分布式硬件系统之上等特点。HBase 是构建在 HDFS 上的分布式列存储非关系型数据库模型，具备存储结构化和半结构化数据的优势，支持随机读写超大规模数据集，具备高可靠性、高性能、列存储、可伸缩、实时读写的特性。

## 4. 水文大数据共享平台设计

### 4.1. 平台架构

水文大数据共享平台建设内容可分为基础设施、数据资源、信息组织平台和保障环境。其中：基础设施建设包括硬件、网络和机房的建设；数据资源建设是数据共享平台建设的重要内容之一，包括数据资源的规划和组织，信息资源目录、元数据体系的建立，若干主题和专用数据库的建设；信息组织平台建设包括目录服务、信息资源门户服务、信息组织与管理服务和信息交换等部分，实现数据共享平台的信息共享与综合开发等支撑功能；保障环境建设包括安全、容灾备份、技术标准、日常管理制度等的建设，其框架示意图如图 1 所示。

### 4.2. 平台功能

水文大数据共享平台提供了资源管理、基础服务、应用支撑等不同层面的服务，建立了一个合理、开放和标准、统一的应用支撑平台，为跨地域、部门、平台的不同应用系统，不同数据之间的互联互通，以及实现“松耦合”结构的应用和数据集成奠定基础。平台主要功能应用包括：数据管理、目录服务、数据服务、扩展服务和系统管理等功能，其功能结构图如图 2 所示。

1) 数据管理：数据管理包括数据汇交、数据存储和数据更新等三类的功能。数据汇交包含数据的审查、发布和管理功能；数据存储提供对数据实体的入库功能，包括数据抽取、转换与装载；数据更新功能对数据进行实时更新。

2) 数据服务：数据服务是在目录服务基础上的深入，主要是通过全文检索、动态图形检索等技术对各种空间、非空间数据以及结构化、非结构化数据提供浏览、查询和下载等多种功能。

3) 目录服务：目录服务是以水文大数据索引为核心查询服务，根据水文元数据标准为依据，通过水文大数据索引将信息以动态分类的形式展现给用户。

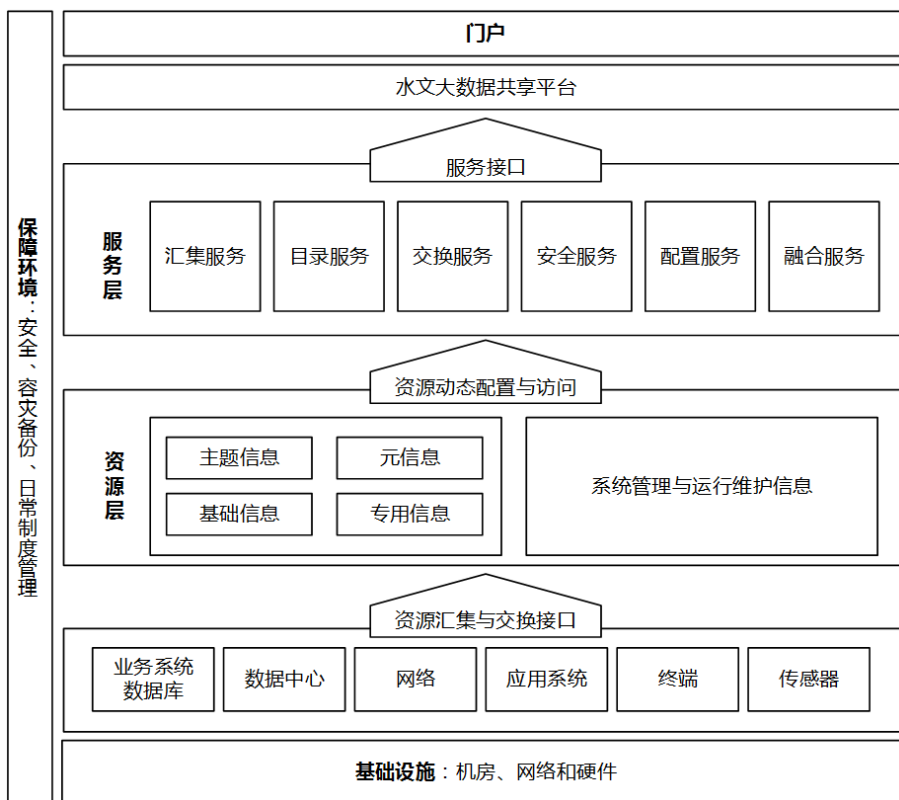


Figure 1. Framework of hydrological big data platform  
图 1. 水文大数据平台架构

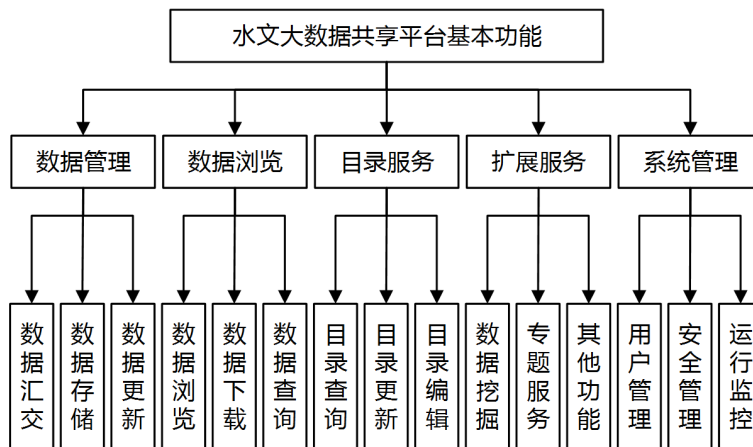


Figure 2. Function diagram of hydrological big data platform  
图 2. 水文大数据平台功能图

- 4) 扩展服务：扩展服务是一种高级服务，提供数据挖掘、专题服务和其他服务等功能。
- 5) 系统管理：提供与水文大数据共享平台运行管理相关的功能，如用户管理、安全管理和运行监控等。

### 5. 水文大数据共享平台实现

水文大数据共享平台采用分布式的部署，需充分发挥各主机的性能，其实现主要包括数据服务器的物理部署、应用服务器的物理部署和开发语言框架与实现三部分。



## 5.1. 数据服务器的物理部署

本平台采用 8 台主机组成的分布式物理部署测试环境, 分别命名为 master 和 slaver01-slaver07, 在这些机器中分别安装 HDFS、HBase、MapReduce2、Spark2、YARN 和 Zookeeper 等, 集群部署如图 3 和表 1 所示。

所有节点启动后, 设置负载均衡的工作模式, 将各节点配置为 HDFS、HBase、MapReduce2、Spark2、YARN 和 Zookeeper 的相应功能节点, 部署结构如表 1 所示。

若在本平台在今后的使用中需扩充性能, 直接在本网段局域网内接入新的主机, 修改配置文件即可, 无需重启集群中所有机器。

## 5.2. 应用服务器的物理部署

应用层服务器采用 Tomcat 发布网站, 并采用 Nginx 反向代理技术提高网站性能, 从而应对海量水文数据下载的并发请求。应用层主要结构如图 4 所示, Nginx VIP1 和 Nginx VIP2 作为服务发布的主节点, 安装 Nginx 反向代理, 共享同一个虚拟网络地址对外提供水文数据服务; application1-application5 为应用服务器, 均安装 Tomcat 来发布网站服务器程序。

具体实现为: 在 Nginx VIP1 和 Nginx VIP2 上分别部署 Nginx 消息分发器, 分别将前端业务消息负载均衡转发到 application1-application5 的 Tomcat 上; 此外在 Nginx VIP1 和 Nginx VIP2 上分别部署虚拟网络地址管理软件, 配置两台主机的共享虚拟网络地址, 将 Nginx VIP1 设置为主节点, Nginx VIP2 设置为备用节点, 当 Nginx VIP1 发生故障时, 将虚拟 IP 地址切换到 Nginx VIP2 上。在每一台应用服务器中安装 Tomcat, 并部署相同的网

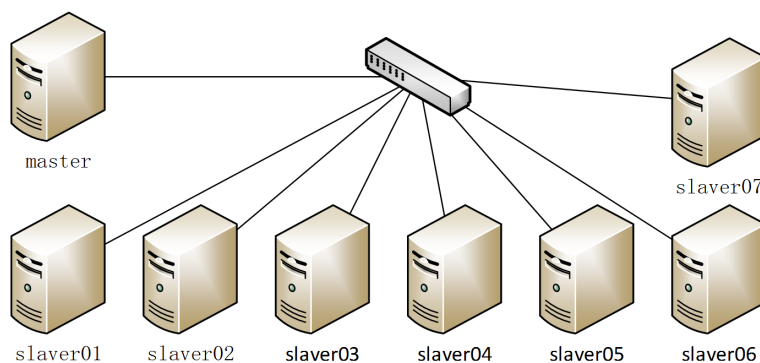


Figure 3. Deployment of data server cluster

图 3. 数据服务器集群部署结构

Table 1. Assignment of Hadoop nodes

表 1. Hadoop 平台的节点分配

主机名	master	slaver07	slaver01-slaver06
HDFS 节点分配	NameNode	SNameNode	
	DataNode	DataNode	DataNode
HBase 节点分配	Hmaster		
	RegionServer	RegionServer	RegionServer
Spark 节点分配	Driver		
	Worker	Worker	Worker
MapReduce 节点分配	JobTracker		
	TaskTracker	TaskTracker	TaskTracker
YARN 节点分配	ResourceManager	NodeManager	NodeManager
Zookeeper 节点分配	Server	Client	Client

站客户端程序，同时连接数据服务器集群，获取数据。

### 5.3. 开发语言框架与实现

水文大数据共享平台实现了数据管理、数据浏览、目录服务、扩展服务和系统服务等功能。后台主要采用 Spring MVC 框架和 JAVA 语言，前端主要采用 Bootstrap 框架和 Javascript 语言开发。

#### 1) Spring MVC

Spring MVC 框架通过实现 Model-View-Controller 模式来很好地将数据、业务与视图分离，管理程序开发中众多对象，其主要思路如图 5 所示。其中模型(Model)是指数据模型，即所要展示的数据，包括数据和行为，提供了模型数据查询和模型数据状态更新等功能。视图(View)负责进行模型展示，也就是与最终客户直接交互的界面。控制器(Controller)通过接收用户请求，委托给模型进行处理，处理完毕后将返回的模型数据返回给视图，由视图负责展示。这样的实现将模型和视图分离，将控制器中的逻辑控制与功能处理分离，将控制器中的视图选择和具体视图技术分离，从而使视图层只包含展现的 web 页面内容，不包含其他逻辑功能，保证整体架构的精炼和明确。

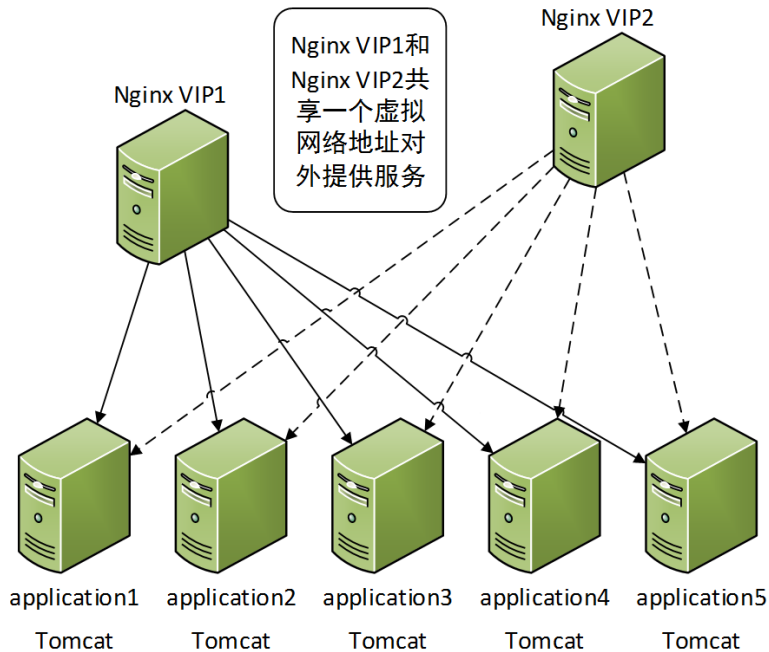


Figure 4. Deployment of application server cluster

图 4. 应用层集群部署结构

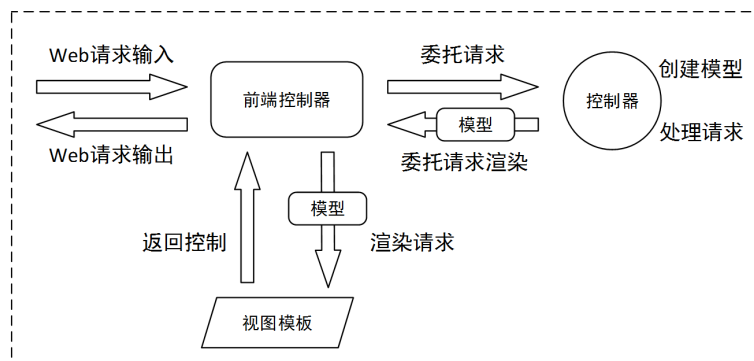


Figure 5. Design idea of Spring MVC

图 5. Spring MVC 设计思想

## 2) Bootstrap

Bootstrap 是基于 HTML、CSS 和 JavaScript 的框架，可以快速搭建漂亮的前端框架，并且可开发全响应式网页，可自适应于各种客户端和浏览设备，如台式电脑、笔记本电脑、平板电脑和手机等，为开发人员创建接口提供了一个简洁统一的解决方案，包含功能强大的内置组件，易于定制，并且完全开源。

## 6. 结语

本文根据水文数据的特点探讨了水文大数据标准化方法，介绍了数据预处理、数据索引、数据高效存储等水文大数据共享平台关键技术，并在此基础上设计和实现了水文大数据共享平台，为水利及跨行业跨部门的信息共享、空间集成，以及跨学科的可持续发展研究提供技术支撑。

随着水文数据量的不断增大，数据类型不断扩展，各行业用户需求不断增多，多源异构海量水文信息的集成共享日趋重要，水文大数据共享平台实现了水文大数据资源的集中统一、安全可靠的存储和管理，构建了统一的信息接入标准和共享机制，建立了业务支撑环境和数据交换共享体系，为水文行业大数据共享集成提供了规范化、标准化的模式。随着水文大数据研究的不断深入，更多问题需要解决，笔者认为可从以下几个方面入手。

### 1) 大数据共享在生产环境中的应用

在实际生产中，参与数据共享的机构包括中央、流域、省级、地市的水文和防汛抗旱部门，大型水利工程管理单位等，都需要做到监测资料的实时共享。如何在实际生产环境中应用水文大数据共享平台也是将来需要研究的方向之一。

水文数据是国家重要的战略数据资源，因此水文大数据共享平台在实际生产环境中必须考虑数据共享安全设计。结合我国当前政务网布局结构和平台部署的网络环境，研究合适的数据共享安全机制是水文大数据共享平台在实际生产环境中应用的研究方向之一。

### 2) 数据服务接口的研究

通过服务接口的方式对外发布水文大数据是目前最理想的数据共享方式。如何定义服务功能，如何划分服务功能边界，如何合理地建立数据服务相互之间的关系，都是需要解决的问题，均可作为水文大数据共享平台近期研究方向。

### 3) 多源异构大数据融合

水文学及水资源学科与大气科学、环境科学、生态科学、信息科学、经济学、管理学、社会学和遥测技术等关系日益密切，学科交叉发展现象明显，并逐渐拓展为对广泛的资源、环境与生态问题研究的支撑[14]。如此趋势之下，水文大数据平台研究需要融合更多来源的异构数据，研究其特性和数据之间的相互联系，探讨多学科数据融合的大数据存储模式。

### 4) 水文大数据机器学习

建立水文大数据平台不仅是为了数据查询与共享，更重要的是为机器学习提供平台基础。大数据机器学习包括回归、聚类、降维、朴素贝叶斯、决策树、神经网络、支持向量机、深度学习等算法。Foster 和 Tom 提出了基于数据的科学和其与数据驱动决策支持和大数据之间的关系[15]。大数据机器学习是通过对数据的归纳、挖掘和深度学习发现水文规律，而不是传统水文学通过演绎反映水文过程，是另外一种深入大自然、了解大自然规律可行的方式。所以对于水文大数据平台更深入的研究需从机器学习算法入手，扩充水文大数据平台的应用范围，从而解决更多的水文学问题。

## 基金项目

国家自然科学基金重点项目(51539009)。



## 参考文献 (References)

- [1] 蔡佳男, 耿庆斋. 水利科学数据共享汇交体系探索与构建[J]. 中国水利水电科学研究院学报, 2006, 4(1): 31-35.  
CAI Jianan, GENG Qingzhai. Study on concept and data collection system of scientific data sharing of water resources. Journal of China Institute of Water Resources and Hydropower Research, 2006, 4(1): 31-35. (in Chinese)
- [2] 郭亚曦. 我国气象科学数据共享系统建设与服务[J]. 中国科技资源导刊, 2008, 40(2): 14-18.  
GUO Yaxi. Meteorological science data sharing system construction and service in China. China Science & Technology Resources Review, 2008, 40(2): 14-18. (in Chinese)
- [3] 陈军飞, 邓梦华, 王慧敏. 水利大数据研究综述[J]. 水科学进展, 2017, 28(4): 622-631.  
CHEN Junfei, DENG Menghua and WANG Huimin. A review of water resources big data. Advances in Water Science, 2017, 28(4): 622-631. (in Chinese)
- [4] 莫荣强, 艾萍, 吴礼福, 等. 一种支持大数据的水利数据中心基础框架[J]. 水利信息化, 2013(3): 16-20.  
MO Rongqiang, AI Ping, Wu Lifu, et al. A fundamental frame of water resources data center supporting big data. Water Resources Informatization, 2013(3): 16-20. (in Chinese)
- [5] AI, P., YUE, Z. X. A framework for processing water resources big data and application. Applied Mechanics and Materials, 2014: 3-8.
- [6] 冯吉平, 陈微, 官滂, 等. 大数据技术在松辽流域水环境管理中的应用展望[J]. 水利发展研究, 2014, 14(9): 63-65.  
FENG Jiping, CHEN Wei, GUAN Di, et al. Application prospect of big data technology in water environment management of Songliao River Basin. Water Resources Development Research, 2014, 14(9): 63-65. (in Chinese)
- [7] CHALH, R., BAKKOURY, Z., OUAZAR, D., et al. Big data open platform for water resources management. 2015 International Conference on Cloud Technologies and Applications (CloudTech), 2015: 1-8.
- [8] LI, D., GUO, S. and YIN, J. Big data analysis based on POT method for design flood prediction. 2016 IEEE International Conference on Big Data Analysis (ICBDA), 2016: 1-5.
- [9] 孙欣欣. 城市突发水涝灾害大数据技术分析研究[J]. 科技通报, 2016, 32(4): 196-201.  
SUN Xinxin. Study on the big data technology of the city unexpected flood disaster. Bulletin of Science and Technology, 2016, 32(4): 196-201. (in Chinese)
- [10] 中华人民共和国水利部. SL323-2011 实时雨水情数据库表结构与标识符[S], 2011.  
The Ministry of Water Resources of the People's Republic of China. Structure and identifier for real-time hydrological information database, 2011. (in Chinese)
- [11] 中华人民共和国水利部. SL324-2013 基础水文数据库表结构与标识符标准[S], 2013.  
The Ministry of Water Resources of the People's Republic of China. Standard for structure and identifier in fundamental hydrological database, 2013. (in Chinese)
- [12] 中华人民共和国水利部. SL325-2014 水质数据库表结构及标识符[S], 2014.  
The Ministry of Water Resources of the People's Republic of China. Structure and identifier for water quality database, 2014.
- [13] 李洁. 我国水利信息化建设现状及趋势[J]. 城市建设理论研究: 电子版, 2013(18): 2095-2104.  
LI Jie. The current situation and tendency of the construction of Chinese water conservation informatization. Urban Construction Theory Research, 2013(18): 2095-2104. (in Chinese)
- [14] 贾仰文, 王浩, 彭辉. 水文学及水资源学科发展动态[J]. 中国水利水电科学研究院学报, 2009, 7(2): 81-88.  
JIA Yangwen, WANG Hao and PENG Hui. A review on international development trend in the field of hydrology and water resources. Journal of China Institute of Water Resources and Hydropower Research, 2009, 7(2): 81-88. (in Chinese)
- [15] PROVOST, F., FAWCETT, T. Data science and its relationship to big data and data-driven decision making. Big Data, 2013, 1(1): 51-59. <https://doi.org/10.1089/big.2013.1508>