

面向能力培养的 Hadoop 课程“四维一体” 教学改革与实践

李天阳, 詹金珍, 韩立锋

西安明德理工学院, 陕西 西安 710018

DOI: 10.61369/ETR.2026140041

摘 要 : 针对 Hadoop 课程教学中存在的知识体系庞杂、实践环境搭建困难、课程间衔接不畅、思政融入生硬等问题, 本文提出“四维一体”的教学改革框架。从知识体系重构、实践项目设计、实验平台建设、课程思政融入四个维度, 构建以能力培养为核心的 Hadoop 课程教学体系。通过“三阶递进”项目化教学、基于开源技术的统一实验平台、思政元素的有机融入, 实现学生从知识习得能力迁移的转变。教学实践表明, 该模式有效提升了学生的工程实践能力和综合素养。

关 键 词 : Hadoop; 课程建设; 教学改革; 开源技术; 课程思政

Hadoop Course Oriented Towards Competency Development: 'Four-Dimensional Integrated' Teaching Reform and Practice

Li Tianyang, Zhan Jinzhen, Han Lifeng

Xi'an Mingde Institute of Technology, Xi'an, Shaanxi 710018

Abstract : To address issues in Hadoop course teaching, such as complex knowledge systems, difficulty in setting up practice environments, poor course connectivity, and awkward integration of ideological education, this paper proposes a "four-dimensional integrated" teaching reform framework. From the four dimensions of knowledge system reconstruction, practical project design, experimental platform construction, and integration of ideological education, a Hadoop course teaching system centred on capability development is constructed. Through a "three-stage progressive" project-based teaching approach, a unified experimental platform based on open-source technology, and the organic integration of ideological elements, students transition from knowledge acquisition to capability application. Teaching practice shows that this model effectively enhances students' engineering practice abilities and overall quality.

Keywords : Hadoop; curriculum development; teaching reform; open-source technology; ideological and political elements

引言

大数据时代, 以 Hadoop 为代表的开源分布式技术已成为数据存储与处理的事实标准^[1]。随着数据科学与大数据技术专业在全国范围内的广泛开设, Hadoop 作为专业核心课程, 其教学质量直接关系到大数据人才培养的成效^[2]。然而, 当前 Hadoop 课程教学面临多重困境: 知识体系庞杂导致学生“只见树木不见森林”^[3]; 实验环境搭建复杂成为教学“拦路虎”^[4]; 课程间知识点割裂难以形成系统认知^[5]; 课程思政融入生硬影响育人效果^[6]。

为了解决上述问题, 让学生掌握 Hadoop 生态体系, 提出“四维一体”的 Hadoop 课程教学改革框架, 以能力培养为核心, 从知识体系、实践项目、实验平台、课程思政四个维度协同推进, 并通过教学实践验证改革效果。

一、Hadoop 课程教学的现实困境

(一) 知识体系庞杂, 学生难以建立系统认知

Hadoop 技术生态体系涵盖 HDFS、MapReduce、YARN、Hive、HBase、ZooKeeper 等众多组件, 各组件间既有明确分

工又存在复杂关联。传统教学往往按组件逐一讲解, 学生虽然了解了各个部件的功能, 却难以形成对大数据处理全流程的系统认知^[3]。理解分布式系统理论基础是掌握大数据技术的关键^[4], 但多数课程对分布式系统基本概念讲解不足, 导致学生对 Hadoop 设计原理理解不深。

（二）实验环境搭建复杂，制约实践教学开展

Hadoop 实验环境搭建涉及 Linux 系统、虚拟机软件、JDK 及多个开源组件的安装配置。据调查，大部分学生在环境搭建环节遇到困难，部分学生甚至因环境问题放弃后续学习。单机伪分布式环境虽能降低门槛，但与真实分布式环境存在差距，难以让学生体验分布式系统的完整特性^[3]。各课程间实验环境相互割裂，学生每门课都需要重复搭建环境，造成时间和精力浪费^[5]。

（三）课程间衔接不畅，知识碎片化严重

大数据专业课程体系往往按照“总-分-总”思路设计^[5]，但在实际教学中，“分”课程之间缺乏有机衔接。Hadoop 课程与数据采集、NoSQL 数据库、数据可视化等课程各自为战，学生难以将不同课程知识整合应用于完整的大数据项目^[3]。林子雨指出，课程间需要做到“知识分割不重复”和“知识传承不断层”^[2]，但现实中这一目标远未实现。

（四）课程思政融入生硬，价值引领效果不佳

当前课程思政在 Hadoop 教学中存在两种倾向：一是生搬硬套，将思政内容强行植入技术教学；二是蜻蜓点水，仅在课程开头或结尾简单提及^[6]。如何将数据主权、技术伦理、工匠精神等思政元素自然融入 Hadoop 技术教学，实现“盐溶于水”的育人效果，仍是亟待解决的问题^[6]。

二、“四维一体”教学改革框架设计

（一）知识体系重构：从组件罗列到系统思维

1. 以大数据生命周期为主线整合知识

以“数据采集→存储→处理→分析→应用”为主线重构 Hadoop 知识体系。将 Hadoop 生态组件按其生命周期中的角色重新组织：Flume 负责数据采集，HDFS 负责存储，MapReduce/Spark 负责批处理，YARN 负责资源调度，Hive 负责数据查询分析，HBase 负责实时读写。通过主线串联，合理构建课程体系^[7-9]，帮助学生建立全流程技术视图。

2. 强化分布式系统理论基础

在 Hadoop 课程中融入分布式系统基础理论教学，包括：分布式系统模型（节点状态、通信语义）、副本机制、一致性协议等。通过理论讲解与 Hadoop 内核源码实现对照的方式，让学生不仅会用 Hadoop，更能理解其设计与实现原理，更能提升学生自主学习能力。

3. 融入技术演进对比

将 Hadoop 2.x 与 3.x 新特性进行对比，引入分层存储、高性能 I/O、纠删码等前沿内容，让学生在对比中理解技术演进逻辑^[3]。同时，设置丰富的基于 Hadoop 集群的实验项目，帮助学生建立技术选型能力。

表 1 Hadoop 课程实验项目

项目名称	实验类型
Hadoop 全分布式安装	验证型
HDFS 操作与基础编程	设计性

MapReduce 编程	设计性
Hadoop HA 集群搭建与配置	设计性
Flume 组件部署与应用	设计性

（二）实践项目设计：“三阶递进”项目化教学

1. 基础验证层：核心知识点单项实验

围绕 HDFS API 操作、MapReduce 编程、YARN 资源调度、HBase 表操作等核心知识点，设计 8-10 个单项实验。实验设计遵循“讲解-演示-演练”闭环^[3]，每个实验包含明确的目标、步骤指导和验证标准。例如，MapReduce 编程实验从 WordCount 入门，逐步扩展到倒排索引、数据去重等经典案例。

2. 综合设计层：中等复杂度场景项目

以真实场景为载体，设计 3-4 个中等复杂度项目，要求学生 2-3 人小组合作完成。项目需覆盖数据采集到应用的全流程：

（1）电商用户行为分析：使用 Flume 采集用户日志，HDFS 存储，MapReduce/Hive 进行购买行为分析，结果可视化展示。

（2）气象数据温度统计：使用 Kafka 传输气象站数据，Spark Streaming 实时统计，HBase 存储结果。

（3）交通流量预测：基于公开交通数据集，使用 Hive 进行数据清洗，Spark MLlib 进行流量预测建模。

3. 创新探索层：开放式团队项目

开放行业真实数据集（如电商交易、交通流量、能源消耗等），学生自主选题组成 3-5 人团队，完成从需求分析、方案设计、系统实现到成果展示的全流程项目^[5]。鼓励优秀作品参加学科竞赛或申报大创项目^[10]。此阶段引入企业导师参与指导和评审，增强项目的工程实践性^[6]。

（三）实验平台建设：基于开源技术的统一平台

1. “1+3”集群架构设计

采用“1 个主节点+3 个从节点”的虚拟机集群架构。主节点配置 4G 内存+2 核 CPU，运行 NameNode、ResourceManager 等管理进程；从节点各配置 2G 内存+1 核 CPU，运行 DataNode、NodeManager 等计算存储进程。该架构对宿主机要求不高（16G 内存+8 核 CPU 即可满足），可在普通机房电脑或学生个人电脑上部署。

2. 统一技术栈选型

基于开源社区构建统一实验平台，技术选型覆盖大数据整个生命周期的各阶段：

（1）数据收集：Flume、Kafka

（2）数据处理：Pandas、Numpy（预处理）、MapReduce、Spark

（3）数据存储：HDFS、HBase、Redis、MongoDB

（4）数据分析：Hive、Spark SQL、Spark MLlib

（5）数据应用：Matplotlib、ECharts、Flask

3. 容器化封装与版本管理

使用 Docker 容器技术将配置好的实验环境封装为镜像，学生可通过一条命令快速拉起完整集群^[5]。同时，为不同课程（如数据采集、NoSQL 数据库、数据可视化）提供对应镜像版本，实现课

程间环境平滑过渡，避免重复配置。

（四）课程思政融入：技术逻辑与价值引领的统一

1. 思政元素的系统挖掘

从 Hadoop 技术特性中挖掘思政映射点：

（1）数据主权与国家安全：在 HDFS 权限管理、数据备份教学中，引申至国家数据主权、关键信息基础设施安全。

（2）技术自立与自主创新：对比 Hadoop 与国产大数据平台（如 MaxCompute、OceanBase）的发展，激发学生科技报国情怀。

（3）团队协作与工匠精神：在项目开发中，强调代码规范、文档撰写、团队配合，培养精益求精的工程素养。

（4）数据伦理与隐私保护：在数据采集和处理实验中，讨论用户隐私边界、数据脱敏技术、算法公平性等伦理议题。

2. “三阶融入”教学设计

构建“知识点—映射点—融入点”三阶转化模型^[6]：

（1）课前引导：教学辅助平台发布预习材料，包含与知识点相关的思政案例（如“某公司数据泄露事件与 HDFS 权限设置的关系”）。

（2）课中渗透：在技术讲解中自然穿插思政讨论，如在

讲 MapReduce “分而治之”思想时，引申至“集体主义与协作精神”。

（3）课后深化：项目实践中设置伦理审查环节，要求学生从数据安全、隐私保护角度反思自己的设计方案。

三、结语

针对 Hadoop 课程教学中存在的问题，本文提出“四维一体”教学改革框架，从知识体系、实践项目、实验平台、课程思政四个维度协同推进。通过以大数据生命周期为主线重构知识体系、设计“三阶递进”项目化教学、构建基于开源技术的统一实验平台、实现技术逻辑与价值引领相统一的思政融入，有效提升了学生的工程实践能力和综合素养。

教学实践证明，这一改革思路具有较强的可操作性和推广价值。后续研究将从以下方面深化：一是建设跨课程的贯通式项目群，实现 Hadoop 课程与数据仓库、数据可视化等课程的深度联动；二是引入更多企业真实案例，增强项目的工程实践性；三是探索基于学习行为数据分析的个性化教学支持，进一步提升教学效果。

参考文献

- [1] 云仲伦. 中国信息通信研究院发布《大数据白皮书（2022年）》报告[J]. 科技中国, 2023, (04): 105.
- [2] 林子雨. 大数据技术原理与应用课程建设经验分享[J]. 大数据, 2018, 4 (06): 29-37.
- [3] 杨阳, 郝玉婷, 陶丽, 王英. 大数据时代背景下数据分析类课程教学改革与实践探索[J]. 高等工程教育研究, 2023, (05): 54-59+116.
- [4] 李卫强, 王继强, 马先珍. 基于 OBE 的大数据课程体系建设[J]. 计算机教育, 2021, (07): 139-142+147.
- [5] 蔡培茂, 张申勇. 基于开源技术的应用型大数据课程实验建设[J]. 计算机教育, 2025, (11): 275-282.
- [6] 李方东. 数据科学与大数据技术专业实验实践教学探析[J]. 长春大学学报, 2021, 31 (02): 88-91.
- [7] 聂沛, 罗凌云, 何啸峰. 新工科背景下大数据专业课程体系建设研究与探索——以南华大学为例[J]. 教育教学论坛, 2025, (40): 106-109.
- [8] 张玉宏, 蒋玉英, 王贵财. 产教融合背景下的大数据专业课程体系建设——以《数据分析与可视化》教材建设为例[J]. 工业和信息化教育, 2024, (05): 90-94.
- [9] 王国仁, 金福生, 刘驰, 王树良. 面向国际化的数据科学与大数据专业课程体系建设[J]. 中国大学教学, 2018, (12): 43-45+51.
- [10] 王英瑛. 竞赛赋能视角下的实践教学改革研究——以数据科学与大数据技术专业为例[J]. 山东工商学院学报, 2025, 39 (04): 118-128.