

面向新工科的大数据专业实验室 建设模式创新与实践^{*}

匡慈维 薛睿^{**} 苏婷 王婷

哈尔滨工业大学（深圳）实验与创新实践教育中心（分析测试中心），深圳 518055

摘 要 针对大数据实验室建设中实践资源匮乏、数据单一、实验室开放欠缺等问题，本研究基于新工科人才培养需求，构建了“课程实验-实践项目-实践平台”三位一体的实践教学体系；设计了学科交叉融合、知识递进式的实验课程体系；实施了体系化的产教融合实践项目教学；集成了功能完备、操作便捷的智能化实践平台；并优化了实验室开放管理机制。实践表明，该方案有效提升大数据实践教学质量，为新工科人才培养提供有力支撑，具有重要实践意义和推广价值。

关键字 大数据实验室，一体化建设，实践项目教学，开放共享

Innovation and Practice in the Construction Model of Big Data Laboratories for Emerging Engineering Education

Kuang Ciwei Xue Rui Su Ting Wang Ting

Education Center of Experiments and Innovations Analytical and Testing Center
Harbin Institute of Technology, ShenZhen
ShenZhen 518100, China;

Abstract—To address critical challenges in big data laboratory construction—including insufficient practical resources, homogeneous data sources, and limited laboratory accessibility—this study established a tripartite practical teaching system integrating "course experiments, practical projects, and practical platforms" in alignment with talent development goals for emerging engineering education. Key initiatives include: designed an interdisciplinary and progressively structured experimental curriculum; implemented systematic project-based learning through industry collaboration; developed an intelligent and user-friendly platform with comprehensive functionalities; and optimized open-access management mechanisms for laboratory resources. Experimental results demonstrate that the proposed approach significantly enhances the quality of big data practical instruction, provides strong support for cultivating qualified talents in emerging engineering disciplines, and offers substantial practical value with broad applicability.

Keywords—Big data laboratory, integrated construction, practical project teaching, open sharing

1 引 言

新工科专业建设的核心在于实现人才培养与产业发展的深度融合，旨在系统提升学生洞察问题、解决问题的能力以及创新能力。以数据科学与大数据技术专业为例，为实现这一目标，必须构建符合新工科内涵要求的人才培养体系^[1-2]，强化创新与实践教学环节，践行“从实践中来，到实践中去”的教育理念。实践性不仅是该专业的显著特征，更是其人才培养的内在要求^[3-4]。因此，推进与专业实践能力培养相匹配的实验室建设，成为推动大数据教育高质量发展的重要保障。近几年，各高校在大数据实验室建设方面进行了诸多探索，也取得了一定的建设成果^[5-8]，但仍存在项

目式、案例式实验教学资源开发不足，难以体现大数据技术的工程实践特性；实验数据集规模有限、类型单一，缺乏真实业务场景下的高质量数据，难以支撑复杂的数据分析与建模实践；实验室资源开放程度不够，跨专业、跨院系的资源共享机制不完善等问题。为了满足大数据专业的实践教学需求，改进以上不足的问题，本研究在建设大数据专业本科教学实验室方面做了丰富的探索和实践，来满足社会及企业对大数据人才需求。

2 大数据教学实验室的建设方案研究

大数据本科教学实验室建设以新工科人才培养需求为导向^[9-10]，坚持软硬件协同建设原则，打造课程体系、实验平台和管理机制协同发展的新型教学模式。通过系统规划实验教学体系、优化教学内容、整合教学资源并完善创新生态系统，有效推进了产教融合与科教融通，为新工科人才培养提供了有力支撑。其建

^{*} **基金资助**：本文得到广东省教育科学规划课题：面向粤港澳大湾区开源人才培养的创新实践教学体系探索与建设（2022GXJK095）项目基金的资助。

^{**} **通讯作者**：薛睿 xuerui@hit.edu.cn

设思路是以学生发展为中心,以能力培养为主线,紧密对接大数据专业人才培养目标,构建“课程实验夯实基础-实践项目强化能力-实践平台培育创新”的递进式三位一体实践教学体系,有力促进新工科人才培养目标的实现。

3 具体举措

3.1 以市场需求为导向,设计学科交叉融合、知识递进式的实验课程体系

数据科学与大数据技术专业作为新工科建设的典型代表,深度融合数学、统计学与计算机科学的学科优势,致力于培养具备完整大数据理论体系知识架构和系统开发能力的复合型人才。该专业毕业生可在IT产业、金融科技、教育科研等大数据应用领域发挥核心作用。当前就业市场对大数据人才的需求呈现多元化特征,主要涵盖大数据采集与处理、系统运维、智能分析、平台研发以及机器学习/深度学习算法开发等专业技术岗位。特别是随着BAT等科技巨头在大数据应用领域的业务扩张,行业对具备跨学科背景的大数

据专业人才需求呈现持续增长态势,这对高校的人才培养体系提出了更高的要求。

大数据领域的核心技术研究主要围绕算法底层逻辑与工程实现范式展开,先用数学和统计学去分析大数据应用底层的业务场景,再用计算机的技术手段做编程应用。通过系统分析行业人才需求特征与学科交叉属性,本研究构建了大数据专业人才的六大核心能力模型:

- (1) 数学维度:科学思维能力和数学建模能力;
- (2) 统计学维度:逻辑推理能力和数据分析能力;
- (3) 计算机维度:数据采集与处理能力、数据挖掘与应用能力。

基于此能力模型,本研究创新性地设计了“三域融合、五阶递进”的实验课程体系:在学科维度上,实现数学-统计学-计算机科学的深度交叉;在课程架构上,设置包含基础实验、专业核心实验、专业选修实验、综合实践和创新实践共五层递进式培养路径。如图1所示,该实验课程体系涵盖数学、统计学与计算机科学三大领域,并分为五个递进层次。

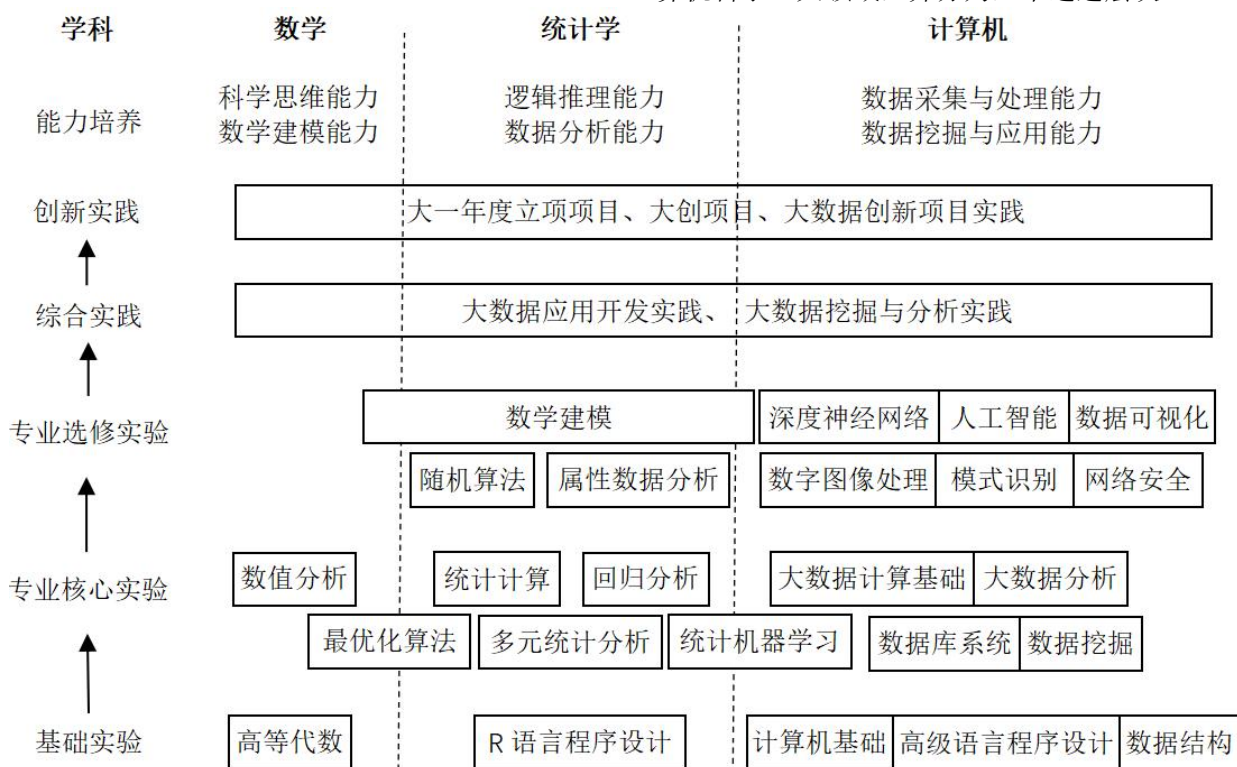


图1 大数据专业结合六维能力的“三域融合、五阶递进”的实验课程体系

在实验教学体系设计中,课程设置充分考虑学科交叉优势:如最优化方法课程融合数学与统计学理论;统计机器学习课程整合统计学方法与计算机技术;数学建模、大数据应用开发实践及大数据创新项目等综合实践课程,则通过项目驱动方式实现数学基础、统

计分析与计算机技术的深度融合。此外,在课程结构设置上渐进式培养路径:基础实验课程聚焦数据理论基础、编程语言以及环境部署运维等基础能力培养,为后续专业课程与实践环节奠定坚实基础;专业实验课程(包含核心课和选修课)通过整合各课程核心知

识点,设计验证型与设计型实验内容,使学生系统掌握大数据开发流程中的关键技术;综合实践课程基于前期理论知识与专业技术,构建综合性实验项目,引导学生运用大数据技术解决多层次问题,深化对核心技术的理解并提升工程实践能力;创新实践课程则以实际应用课题或竞赛项目为载体,面向各年级开放,结合行业真实场景,培养学生综合运用大数据技术解决复杂工程问题的创新能力。本体系通过精心设计的难度梯度曲线,确保学生专业能力呈指数型提升,最终实现人才能力与市场需求精准对接,有效满足新工科背景下复合型人才培养要求。

3.2 推行校企协同育人模式,实施体系化的产教融合实践项目教学

在新工科建设与产业数字化转型双重驱动下,本研究基于 OBE 教育理念^[11],以校企协同育人机制为支

撑,通过模块化课程链与项目式教学法的有机融合,系统设计出大数据专业的实践学习路径,即“大数据理论基础→大数据环境运维→大数据采集与清洗→大数据存储与计算→大数据分析可视化→数据挖掘→大数据应用→大数据创新”,实施“三阶八环”产教融合实践项目教学,见图2。该模式具体包含:

(1) 基础能力层:校企共建“企业数据治理”案例库,实施数据采集与清洗实训项目;

(2) 核心技术层:引入华为/广州泰迪等公司真实业务场景,开展分布式计算与可视化分析项目;

(3) 创新应用层:基于产业需求命题,组织“金融风险/智慧物流”等跨学科创新项目。

通过“项目贯穿-能力递进”的实施路径,确保每个教学模块均对应企业实际工程环节,实现从技术掌握到创新应用的能力跃迁。

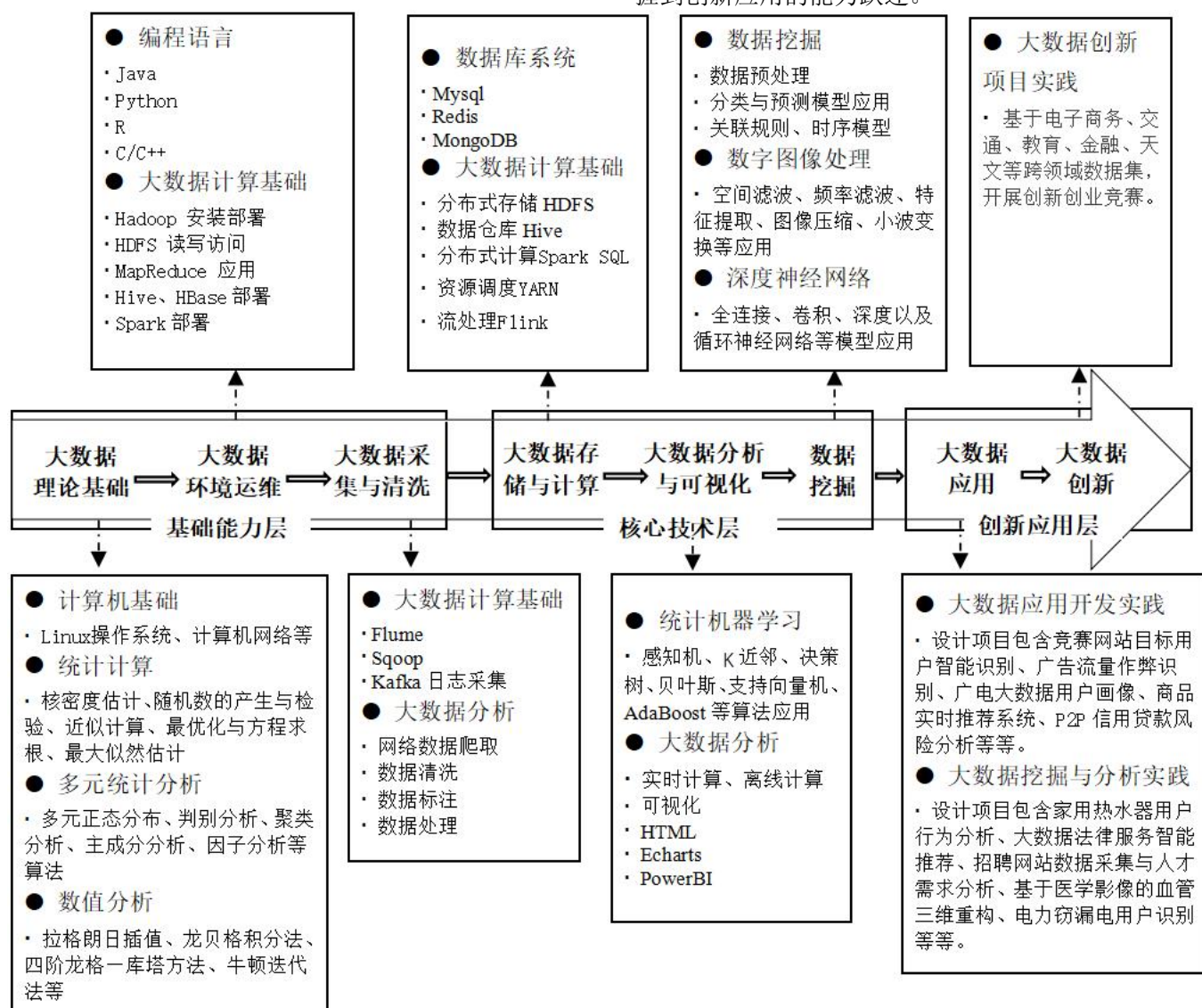


图2 大数据专业实施的“三阶八环”产教融合实践项目教学

以大数据创新项目实践课程为例,采用校企协同育人模式,与行业领军企业(如百度、华为等)建立深度合作。课程设计遵循“产业需求导向、真实项目驱动”的原则,由合作企业提供源自实际业务场景的项目课题与竞赛任务。以华为数据科学创新实践项目为例,其课程开发实施“三阶对接”机制:

(1) 知识体系对接:基于学生前期课程学习基础设计渐进式课题;

(2) 能力标准对接:参照企业岗位能力模型制定培养目标;

(3) 评价体系对接:引入企业技术专家参与全过程考核评估。

其中一个项目是基于企业实际业务场景,要求对日增量达万亿级的异构数据源(包括 MySQL、PostgreSQL、Hive、HBase、MongoDB、Redis 等 6 类数据库系统)进行跨平台对比分析,且限定在不迁移原始数据的约束条件下完成。采用项目驱动式教学模式,要求学生组建跨学科协作小组,在为期一周的密集实践中完成包括分布式环境搭建、多源异构数据预处理、跨平台查询优化及系统性能分析等系列工程任务。实践过程中实施全过程质量管理机制,学生需每日提交技术实践报告,并最终通过项目验收答辩。企业深度参与教学全过程:在项目启动阶段由企业专家解析任务需求与评估标准;实施阶段配备企业导师提供技术指导与过程督导;最终由企业技术团队基于行业标准对项目成果进行多维度评估,并提供专业优化建议。本研究通过为期三年的纵向追踪研究发现,大数据创新项目实践课程对提升学生工程实践能力具有显著效果,能快速提升学生的数据实战能力,深化对大数据技术应用的理 解,提升学生的职业认知。

3.3 整合优质实验资源,集成功能完备、操作便捷的智能化实践平台

本研究基于现有计算机实验室硬件基础设施,通过深度整合云计算资源与大数据技术生态,构建了面向新一代大数据人才培养的智能化实践教学平台。该平台采用云原生的服务管理架构,有机融合了 Hadoop、Spark 等开源技术栈与多元异构数据源,并引入企业级工程实践标准。通过校企协同共建,实验平台实现了功能模块化与操作智能化,完整覆盖数据采集、存储、处理、分析及应用的全流程实践场景。该平台建设围绕两个核心维度展开:

(1) 开源技术生态与企业实践标准的深度对接

平台采用模块化架构设计,完整集成了 Hadoop 3.3、Spark 3.2 等主流开源框架,并参照行业标准进行了教学化改造。同时平台引入金融、电商等领域的

真实数据工程规范,开发了包含数据质量检测、作业调度监控等企业级功能的实验模块,确保教学实践与产业需求的无缝衔接。

(2) 云计算弹性资源与传统实验室硬件的有机融合

平台创新性地采用混合云架构,将本地物理集群(20 节点,总内存 5TB)与云端虚拟资源池(支持动态扩展)进行统一纳管。通过 YARN 资源调度系统和容器化技术(Docker+K8s),实现计算资源的智能分配与弹性伸缩,既保障了基础实验的稳定性(服务协议(SLA)不低于 99%),又满足了峰值负载时的资源需求(可扩展至 100+计算节点)。平台提供资源利用率实时监控看板,使硬件使用效率提升至 85%以上。

故而该平台既有效提升了现有硬件资源的利用率(达 85%以上),又通过虚实结合的实验环境设计,为学生提供了兼具理论基础训练与真实工程实战的综合培养平台,显著提升了大数据专业人才的技术应用能力和工程素养。

大数据智能化实践平台构建了完整的硬件与软件基础设施体系,其核心组成包括基础计算设备集群、云计算服务平台、大数据开发工具集以及多元化数据资源。在硬件层面配置了高性能服务器、图形处理器、大容量存储设备等专业计算资源;软件层面整合了虚拟化云计算管理平台,实现计算资源的智能调度与弹性分配,有效支撑各类实验教学和实训需求。平台同时部署了主流编程语言开发环境和大数据技术生态体系,涵盖 Python、Java 等编程工具及 Hadoop 分布式框架等开源技术组件,并配备多源异构数据资源库,为大数据教学与研究提供全面的技术支持。平台内多元化数据来源于:

(1) 整合政府开放数据(如深圳市政府数据开放平台的民生数据集),为机器学习、数据挖掘等基础课程提供标准化实验素材;

(2) 采购气象、医疗、地理等领域的商业数据库,支撑分布式存储与并行计算等核心课程项目;

(3) 配置光学动作捕捉系统、七轴协作机器人等智能采集设备,生成多模态实验数据,用于开展轨迹还原、数据处理、身份检测等不同主题的创新实践项目,以培养学生创新实践动手能力。

通过多层次的数据获取渠道和多元化的技术支撑体系,该平台有效促进了理论教学与实践应用的深度融合,为培养具备扎实专业技能和创新实践能力的大数据人才奠定了坚实基础。

4 实验室运行管理方案

本校教学实验室遵循“分散规划、集中建设、统一管理、协同运行”理念。其中数据科学与大数据技术、经济管理与会计学以及计算机技术三个专业，在统一管理的构架下，与各学院形成有效互动，支持开放共享与创新，保障实验教学平台及设备资源的高效利用。

4.1 跨学科实验资源柔性调度机制

大数据实验室采用跨学科资源共享机制，面向数据科学与大数据技术、计算机科学、经济管理等多个专业开放实验空间，有效缓解实验课程排课压力，显著提高实验室利用率。实验教学团队实施柔性师资配置方案，大数据专业教师既承担本专业数学与计算机类实验课程，同时动态支援计算机专业数学方向实验教学。这种师资协同模式不仅有助于实现教学资源优化配置，还能促进跨专业教学经验交流，为实验课程体系的持续改进提供有力支撑。通过建立多专业协同的实验教学共同体，既完善了师资的动态互补机制，又提升了实验教学的整体质量。

4.2 多层次实验平台的信息化协同管理

大数据实验室构建了多层次实验教学平台，涵盖基础实验、综合实践和创新实践等教学环境，面向全校各专业开放大数据相关实验课程。实验室通过部署信息化管理系统，集成设备开放共享功能模块，实现实验设备全生命周期管理，包括设备参数、运行状态和使用情况的实时监控与动态调度。在实验室建设过程中，通过完善管理制度和优化资源配置，建立了科学规范的运行机制，显著提升了实验设备的使用效益和教学服务质量。

4.3 竞赛牵引的创新能力培养模式

大数据实验室在保障大数据专业本科实验教学核心任务的同时，充分发挥创新实践平台的优势资源，为数学建模、数据挖掘、人工智能等高水平学科竞赛提供全方位支持。实验室不仅配备 GPU 加速器、高性能服务器和专业数据集等硬件资源，更建立了完善的竞赛指导体系，包括系统化的培训课程和专业的导师团队，通过理论与实践相结合的培养模式，有效提升学生的创新能力和竞技水平，为应用型创新人才培养提供坚实的实践支撑。这种教学与竞赛相互促进的运行机制，既强化了学生的专业技能，又培养了解决复杂问题的综合素质，实现了人才培养质量的全面提升。

5 实验室建设成效

本校大数据实验中心构建了“硬件设施-管理体系-育人机制”协同发展的新型实验教学平台，形成了设备支撑、制度保障、人才培养的良性循环生态，为创新实践教育办学积累了丰富经验，初步培育了一批优

秀的高质量新工科人才。通过六年建设，在以下方面取得显著成效：

(1) 人才培养成效显著

近三年累计培养大数据专业毕业生 300 余人，就业于阿里、腾讯等头部企业的比例达 30%，人才输出规模与质量同步提升。指导学生参加如“数学建模”、“挑战杯”、“大数据挑战赛”等高水平学科竞赛，获国家级竞赛奖项 20 多项、省部级奖项 50 余项，省级大创项目立项 2 项，创新能力培养成果突出。本科生以第一作者发表 SCI/EI 论文 3 篇，获授权发明专利 5 项，科研育人初显成效。

(2) 教学科研协同发展

大数据教研室构建了“基础实验（30%）-专业实验（30%）-综合实践（20%）-创新实践（20%）”的阶梯式课程体系，开发《大数据计算基础》和《大数据实践》等 20 多门实验课程，编写累计 18 余万字实验指导书。实验教师参与国家级科研项目 2 项，省部级科研项目 5 项，参与发表高水平学术论文 2 篇，教研成果丰硕。

(3) 社会服务辐射广泛

年均开展“AI 开放日”等科普活动 10 余次，覆盖中小學生 500 余人次，科普教育呈常态化。另外也接待了 10 余个高校领导队伍的参观和访问，提供实验室建设思路和方案指导，实验室建设经验得到大力推广。

6 结束语

经过六年的实践探索，大数据实验室建设已取得显著成效。实践证明，该建设方案有效提升了大数据创新人才的培养质量，优化了实践教学体系，具有重要的实践意义和推广价值。在新工科建设的背景下，大数据专业实验室建设仍需进一步完善实验教学体系、丰富教学内容、整合优质资源，并构建创新生态系统。通过深化产教融合与科教融合，持续推进实验室建设，将为达成新工科人才培养的目标提供有力支撑。

参考文献

- [1] 陈洁菲, 蒋彭. 新工科背景下成人高校大数据人才培养改革探索[J]. 计算机教育, 2023 (3): 44-50.
- [2] 吕品, 于文兵. 基于三螺旋模型的应用型工业大数据人才培养策略[J]. 计算机技术与教育学报, 2023, 11 (3): 115-119.
- [3] 孙开伟, 邓欣, 王进. 新工科背景下数据科学与大数据技术专业实践教学体系研究[J]. 高教学刊, 2023 (14): 5-8.
- [4] 秦艳姣, 雷建军, 晏轲. 新工科背景下大数据专业实践教学改革[J]. 计算机教育, 2025 (1): 86-90.
- [5] 吴湘宁, 彭建怡, 罗勋鹤, 等. 高校大数据实验室及实验体系的规划与建设[J]. 计算机系统应用, 2020, 29 (11):

- 47-56.
- [6] [史梦安,任艳,叶倩,等. 应用型本科数据科学与大数据技术专业建设研究[J].软件导刊, 2021, 20(11): 235-239.
- [7] 石兵,熊盛武,饶文碧,等. 数据科学与大数据专业建设研究与实践[J].计算机教育, 2021(4): 88-92.
- [8] 何金凤,马天宏,杜明珠. 工程认证背景下的大数据创新实验室建设[J].软件导刊, 2022, 21(7): 12-15.
- [9] 崔晓龙,张敏,张磊,等. 新工科背景下应用型大数据人才培养课程群研究与建设[J].实验技术与管理, 2021, 38(2): 213-218.
- [10] 张明慧,郭欣. 大数据科学与技术专业课程群育人探索与实践[J].计算机技术与教育学报, 2022, 10(2): 66-69.
- [11] 曹峰,张虎,曹付元,等. 基于 OBE 的数据科学与大数据技术专业建设与实践[J].软件导刊, 2023, 22(6): 162-166.