

大数据专业创新实践平台建设及应用研究

李杰

(内蒙古工业大学数据科学与应用学院, 内蒙古呼和浩特 010080)

摘要: 新工科背景下大数据产业迅猛发展,对数据科学与大数据技术专业的实践教学提出了更高要求。探讨大数据专业实践教学的现状与问题,并结合实际提出创新实践平台建设方案,以提升实践教学质量。通过对数年办学经验进行梳理,发现实践教学存在与行业发展脱节、体系不完整、软硬件资源不足等问题。为此,构建一个融合虚拟化技术与容器技术的创新实践平台,采用高性能计算资源和大容量存储系统支持虚拟实验环境、项目实训、创新创业和竞赛等多样化实践需求,显著提升了学生的实践能力和创新思维,拓展了其技术视野。实践表明,创新实践平台不仅优化了教学资源配置,提高了大数据专业的实践教学效果,而且为培养符合工程认证理念的高质量大数据人才提供了有力支撑,同时也为相关学科实践教学提供了有益参考。

关键词: 新工科;大数据;实践教学;创新实践平台

DOI: 10.11907/rjdk.241664

中图分类号: G642

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2025)001-0164-05



Research on the Construction and Application of Innovative Practice Platform for Big Data

LI Jie

(School of Data Science and Application, Inner Mongolia University of Technology, Hohhot 010080, China)

Abstract: Under the background of new engineering disciplines, the rapid development of the big data industry has put forward higher requirements for practical education in data science and big data technology majors. Explore the current situation and problems of practical teaching in big data majors, and propose innovative practical platform construction plans based on actual situations to improve the quality of practical teaching. Through years of experience in running schools, it has been found that practical teaching has problems such as disconnection from industry development, incomplete system, and insufficient software and hardware resources. To this end, an innovative practice platform integrating virtualization technology and container technology is constructed, which uses high-performance computing resources and large capacity storage systems to support diverse practical needs such as virtual experimental environments, project training, innovation and entrepreneurship, and competitions. This significantly enhances students' practical abilities and innovative thinking, and expands their technical horizons. Practice has shown that innovative practice platforms not only optimize the allocation of teaching resources and improve the practical teaching effectiveness of big data majors, but also provide strong support for cultivating high-quality big data talents that conform to the concept of engineering certification. At the same time, they also provide useful references for practical teaching in related disciplines.

Key Words: emerging engineering; big data; practical teaching; innovative practical platform

0 引言

2023年底召开的中央经济工作会议明确提出要以科技创新推动产业创新,特别是以颠覆性技术和前沿技术催生新产业、新模式、新动能,发展新质生产力^[1]。相较于传

统生产力而言,新质生产力是指大量运用大数据、人工智能、物联网、云计算等新技术与高素质劳动者、现代金融、数据信息等要素紧密结合而催生的新产业、新技术、新产品和新业态^[2]。2023年,我国组建国家数据局,将数据与土地、劳动力、资本和技术并列为生产要素,数据成为关键生产要素和重要战略资源。教育部于2016年增设数据科

收稿日期:2024-08-05

扫描二维码阅读全文:



基金项目: 教育部产学研合作协同育人项目(202101203025);内蒙古自治区直属高校基本科研业务费项目(JY20220283)

作者简介: 李杰(1990-),男,内蒙古工业大学数据科学与应用学院讲师,研究方向为区块链与访问控制。

学与大数据技术本科专业,截至2024年,已批准730余所高校设立大数据相关专业。这一举动符合国家长远发展战略规划,对新质生产力的发展具有支撑作用。

数据科学与大数据技术是以大数据分析为核心,以计算机科学与技术、统计学、数学等学科交叉融合,面向大数据时代人才需求的新兴工科专业,旨在培养大数据产业发展亟需的具备大数据理念、拥有大数据思维及分析应用技术的高素质应用型人才。在教学培养过程中仅靠理论和传统的实验教学很难进行专业知识与实践技能的融合与锻炼^[3]。大数据专业注重工程实践,对实践环境要求较高,必须对传统实践教学加以改革和创新,提供虚拟仿真的大数据分析挖掘场景和软硬件环境,使学生切实参与数据开发、数据分析挖掘、数据运维等工作,从根本上培养其工程实践能力和数据思维^[4]。

1 大数据专业实践教学现状

大数据是涉及计算机、统计学和数学等多学科的交叉学科,要求相关人才具备丰富的理论知识和较强的实践动手能力^[3]。大数据专业面向大数据行业培养具备大数据开发、数据分析挖掘、数据运维能力的创新型人才,因此实践能力培养是重中之重^[5-6]。然而,目前的实践教学体系仍存在许多问题。

1.1 实践教学与行业发展脱节

在大数据领域,技术和应用的发展速度极快。如果不能及时更新课程内容和教学方法,就会导致所教授的技术和工具与当前行业的技能需求不匹配^[4]。例如,学校仍在教授旧的数据处理技术,而市场上已经普遍使用新型大数据处理框架和工具,这种脱节导致学生毕业后需要花费额外的时间来适应工作要求,影响其就业竞争力^[7]。

1.2 实践教学体系不完整

一个完整的实践教学体系应包括从基础到高级的各项实践技能训练,涵盖统计学、数据采集、数据清洗、数据预处理、数据分析挖掘和可视化等全过程^[8]。受实践教学软硬件资源以及师资力量等限制,许多高校不能系统性地与实践能力的培养,学生难以形成系统性数据思维^[9]。

1.3 实践教学软硬件资源不足

大数据专业实验和实训通常需要高性能的计算资源和大容量的数据存储能力。在专业建设初期,相关配套资源有限,缺乏完善的软硬件环境,导致学生无法学习使用最新的数据处理工具和平台,实践教学效果大打折扣^[10]。对于分布式数据处理、大规模数据分析挖掘等场景,学校无法提供足够的计算存储资源和数据处理环境,无法满足实践教学需求^[11]。

内蒙古工业大学数据科学与应用学院于2018年开设数据科学与大数据技术专业,以地区经济发展和社会需求为导向,以培养大数据领域高素质工程型人才为目标,重视理论与实践相结合。经过数年建设,学院围绕实践教

育、创新创业、竞赛、实训等主题,打造了“实验教学—创新实践—竞赛—实训”一体化创新实践教学平台。

2 大数据专业创新实践平台建设

通过高年级学生的问卷调查结果和用人单位的反馈深入分析内蒙古工业大学大数据专业实践教学过程中存在的问题。问卷调查对象为数据科学与大数据技术专业大三和大四的学生,共发放问卷100份,回收有效问卷92份,回收率为92%。此外,向近3年毕业并从事大数据相关工作学生的所在单位进行了调查,共收到50份有效反馈。调查结果显示,实践教学主要存在以下两个方面问题:一是教学资源滞后。65%的学生认为现有教学资源更新不及时,无法跟上大数据技术的发展速度;70%的用人单位表示新入职员工对最新技术和工具的掌握程度不足;二是软硬件资源限制。72%的学生反映机房实验环境配置较低,难以满足复杂的大数据处理需求;60%的用人单位提到毕业生在实际工作中无真实分布式大数据处理经验,技能生疏,影响了工作效率。这是由于传统的计算机相关专业实践教学以单机为主,配合虚拟机软件进行课程实验。然而,单机普遍无高性能GPU,无法构建真实的分布式场景,提供分布式数据处理环境,亦无法满足人工智能模型训练要求^[12-13]。因此,内蒙古工业大学在专业实验室建设过程中进行了大量调研和论证,将业界流行的云计算与容器技术融合,作为平台的基础设施^[14]。

云计算具有可扩展性、灵活性和成本低等优点。对于高校来说,构建私有云平台不仅可以解决硬件资源受限的问题,而且可以通过虚拟化技术实现资源的动态分配与共享管理,从而满足不同课程、项目的个性化需求。容器是一种轻量化技术,能够将应用程序及其依赖项打包在一起,从而确保其在任何环境中都能一致运行^[15]。容器启动速度快、占用资源少,非常适合在有限的硬件资源下部署多样化实验环境。此外,容器还支持版本管理和镜像共享,有助于维护教学资源的一致性和更新效率。

大数据专业创新实践平台的建设紧密围绕实验教学、项目实训、创新创业、竞赛等内容开展,平台的设计理念如图1所示。根据大数据专业培养方案的要求,围绕核心课程开展实验教学,设计与现实世界数据分析处理相结合的实验项目,利用真实数据集开展数据分析挖掘实验。通过项目实训提升学生的项目管理能力和团队协作水平,通过全流程参与实际数据分析项目切实提高学生的数据分析挖掘实践技能。课下,创新实践平台可作为创新创业工作坊,为学生提供软硬件环境;还可利用该平台强大的计算存储资源支撑学生参与学科竞赛,提升其专业技能和国际视野。

创新实践平台架构如图2所示。平台以通用服务器、异构存储介质、高速网络设备和GPU计算单元作为底层硬件基础设施,采用虚拟化技术将计算资源、存储资源、网络

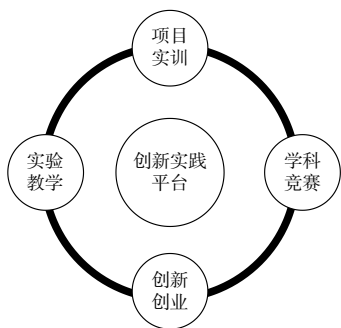


Fig. 1 Innovative practice platform concept
图1 创新实践平台理念

资源池化以形成计算资源池;然后采用开源的私有云操作系统 OpenStack 及 Docker 容器技术批量构建虚拟机及容器,统一接入实践教学管理系统对平台软硬件资源进行综合管理与调度,为大数据课程提供实验教学环境,为实训项目提供计算资源,为课程竞赛和创新创业项目提供工坊,全方位服务实践教学。

2.1 硬件基础设施

底层硬件基础设施是创新实践平台的基石,旨在为实践教学提供高效、稳定的环境。硬件采用通用的 x86 架构服务器,确保能够承载大数据处理所需的各类计算任务;提供充足的空间,确保学生能够存储大量数据集和实验结果,以支持长周期的数据处理任务;高性能网络交换设备确保数据传输的速度和稳定性;高性能 GPU 计算单元支持学生进行图像识别、深度学习领域的实验。单独设立机柜存放硬件基础设备,采用链路聚合+交换机堆叠、双电源+UPS 的方式保障硬件稳定运行,同时对机房动环信息进行实时监控。

2.2 融合架构与虚拟化

采用计算虚拟化、存储虚拟化和网络虚拟化及 GPU 计算单元集群技术对软硬件资源进行虚拟化,屏蔽底层硬件

异构架构,形成计算资源池,方便软件进行集中调度管理。资源池化降低了软硬件间的耦合性,利于后期进行硬件扩展及设备更新。创新实践平台能够突破单个硬件算力的限制,灵活地对计算资源进行集中分配与调度,确保每位学生都能获得所需的计算资源。

2.3 虚拟机与容器

采用开源的私有云操作系统 OpenStack 及时下流行的 Docker 容器技术构建虚拟实验实训环境。OpenStack 提供了强大的虚拟机管理功能,包括创建、启动、停止以及监控虚拟机等,师生可以利用 OpenStack 的 DashBoard 组件直观管理和使用虚拟机;其同时支持自定义系统镜像制作,教师能够预先上传操作系统镜像,方便配置实验和实训环境,提高了硬件设备的利用率。Linux 实训实验环境采用时下流行的轻量级 Docker 容器技术进行虚拟环境生成和资源隔离。Docker 容器可以快速部署并提供一致性运行环境,减少了因环境配置问题导致的教学延误。通过 Kubernetes (K8S) 进行管控和调度能够同时提供更多虚拟执行环境,方便学生进行环境配置和分布式数据处理系统的构建。人工智能领域的模型构建和训练采用 GPU 计算单元搭建的深度学习集群进行集中式资源分配,以满足高性能计算的需求。

2.4 实践教学管理

采用自主研发的实践教学管理平台对底层软硬件资源进行统一管理,从而满足实验教学、创新实践、竞赛、实训需求。在实验教学中细粒度地分配虚拟机和容器资源,统一分发镜像和计算资源,方便学生配置实验环境,多实例操作完成实验。在创新实践和竞赛模式下粗粒度地分配虚拟机和硬件资源,提供给学生更大的自主权,方便其自主配置和整合资源,扩展其创新思维。在实训过程中,以小组为单位划分计算资源,下放管理权限,由项目组



Fig. 2 Architecture of the innovative practical platform
图2 创新实践平台架构

自主申请使用资源。同时,创新实践平台提供了丰富的教学资源 and 实训案例,包含大数据专业相关课程的教学视频、虚拟实验环境、实训数据集和分析挖掘模型,且能够对实验室中的学生机进行操作系统镜像下发、广播控制、文件下发、作业收集、投票、点名等操作,满足日常实践教学。此外,创新实践平台具备良好的开放性和扩展性,支持硬件设备扩展和软件动态更新升级,以适应技术发展和实践教学体系的优化调整。

3 创新实践平台应用与实践效果

实践教学的主要目标为培养学生的实操能力和创新能力,使其能够掌握大数据处理的核心技术和工具;核心理念为以学生为中心,为其提供丰富的实践资源和灵活的教学模式,激发其学习兴趣和主动性。根据学生的学情和兴趣,基于实践平台提供个性化学习路径和学习资源,同时鼓励团队协作,培养合作沟通能力。课前布置实践任务,发布教学资料,分配软硬件资源,下发基础镜像,学生熟悉任务,预习相关知识;课中分配任务,统一管理调度资源,下发相关配置文件,学生快速配置实验环境,协作完成实践任务;课后在线收集实践成果,监控实践进度,并提供辅导和支持,同时通过平台展示成果,实现共享与反馈。典型课程内容,如大数据处理技术可利用平台快速搭建分布式计算集群,引用大型数据集,显著提升教学效率;实时流数据处理能够实时处理大量数据流,分析结果;深度学习模型训练利用 GPU 集群调整参数,缩短训练时间,提升运行效果。

创新实践平台建成后,经过 3 年教学实践,目前已有 3 届毕业生,日常为大数据专业 4 个年级共 280 名学生提供实验教学、课程设计和项目实训的软硬件环境。平台已建立实验项目 32 项,配置相关实验环境源镜像 8 个,基本可以满足大数据专业日常实验教学任务。平台通过采购实训项目资源以及与企业协同共建,现有 18 门大数据相关实训课程,55 个配套数据集,结合虚拟化技术,提供多种主流大数据分析处理组件,满足课程设计、项目实训教学需求。平台引入优质在线课程 32 门,同时整合疫情期间积累的在线教学资源,基本涵盖了大数据的各个方面,为学生课下自学提供了丰富的选择。同时,平台提供核心专业课程的在线练习与交流答疑,方便学生课下巩固学习相关课程;针对实验实训,提供代码助手辅助解决技术问题。课后,平台作为智慧数据工作坊,学生可以在校园网范围内自由访问获取线上教学资源,完成实验、实训、课程作业,参与学科竞赛。

本文采用定量数据分析和定性案例研究的方式考察创新实践平台的有效性。首先,通过问卷调查的方式收集内蒙古工业大学数据科学与大数据技术专业大二到大四学生对创新实践平台的反馈意见,问卷内容包括教学资源实用性、计算资源可用性、系统易用性等方面。共发放问

卷 145 份,回收有效问卷 133 份,回收率为 91.7%。其次,记录分析平台资源使用情况,包括计算资源、存储资源、GPU 集群等。再次,对学生实践任务完成比例进行分析,以评估平台对学生学习效果的影响。最后,与任课教师和学生进行座谈,了解其具体感受、遇到的问题以及改进建议。调查结果显示,有超过 85% 的学生认为实践教学平台显著提高了实验、实训学习效率,提供了充足的软硬件资源,是传统单机无法比拟的;72% 的学生表示,相较于传统实验环境,其更愿意在平台上进行实验。统计数据显示,平台投入使用后,每学期 CPU 平均使用率为 65%、内存使用率约为 70%、存储空间利用率也超过了 50%,软硬件资源得到了有效利用。引入实践平台后,学生的实验、实训完成率由 71% 上升至 89%,特别是大数据处理技术课程实验的成功率提高了近 20%。通过 Docker 容器技术,学生能够快速部署实验环境,极大缩短了准备时间,可以轻松构建多种类型的大数据实验环境,覆盖数据采集、清洗、存储、分析等多个环节,帮助学生全面掌握大数据处理流程。同时,计算资源的动态分配和热插拔技术能够确保每位同学都获得足够的计算资源,避免了因资源不足导致的学习中断。访谈结果显示,平台大大减少了教师的物理设备维护工作量,学生也能够快速搭建实验环境,从而更加专注于理解和解决问题。

近 3 年来,学院通过创新实践平台累计完成了 576 学时的实验教学和 45 门课程实训;学生获得“互联网+”比赛国家级铜奖 2 项、自治区级金奖 4 项,获得挑战杯、蓝桥杯等自治区级奖项多项。图 3 为大数据专业招生以来实践教学的加权综合达成度数据。可以看出,自 2020 年创新实践平台投入使用后,实践教学的综合达成度整体有所提升。

综上所述,创新实践平台有效解决了传统实践中软硬件资源受限的问题,提高了资源利用率,教学过程中实践任务的完成率显著提高,学生提交的代码质量大幅提升、错误率降低;教师和学生均对实践平台给予了积极反馈,认为其显著提升了教学和学习体验;学生在竞赛、创新创业活动中取得多项成果,实习和就业情况也随之改善。

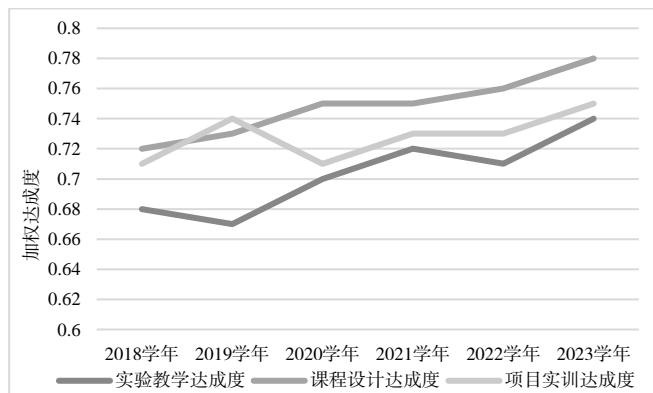


Fig. 3 Comprehensive achievement degree of practical teaching

图 3 实践教学综合达成度

4 结语

为提高大数据专业学生的实践能力和创新思维,本文利用虚拟化技术和容器技术构建了一个开放性创新实践平台。该平台通过在高性能服务器与大容量存储系统上构建计算资源池,显著提升了数据处理与分析能力,有效解决了传统单机或虚拟机环境在实践教学过程中面临的性能瓶颈。该平台采用的虚拟化技术和容器技术支持高度的系统隔离和快速的资源调配,使教学环境具备高效率和高可配置性的特点。容器化的应用部署进一步降低了系统维护的复杂度,同时具备了快速部署和升级的能力。基于该平台,学生能够在一个接近真实世界的技术环境中进行实验和实训,不仅提升了实践教学的实用性和技术的前瞻性,而且增强了处理复杂大数据任务的能力。此外,该平台的高可扩展性、灵活性使其可以满足不同场景下学生的创新实践需求,对于支撑应用型人才培养、探索自主学习、促进师生协同等具有一定价值。

尽管创新实践平台在实践教学方面有显著优势,但在实际应用中仍存在一些局限性。例如,OpenStack和Docker技术在业界应用广泛,但在教育领域普及程度较低,尤其是OpenStack私有云平台部署维护相对复杂,对运维人员要求较高,数据安全性和学生隐私保护成为新的挑战。此外,如何确保资源分配的公平性和透明性也值得深入探讨。未来可从以下几个方面进一步优化平台:①简化部署与维护流程,开发一套自动化运维部署工具,降低技术门槛,方便快捷搭建实践平台;②研究设计严格的数据加密和访问控制机制,确保数据和隐私安全;③优化调度算法,提高资源利用率。

参考文献:

- [1] WANG F. The coordinated advancement of new productive forces and traditional productive forces [N]. 21st Century Business Herald, 2024-03-06 (005):3.
王峰. 发展新质生产力要与提升传统生产力协调推进[N]. 21世纪经济报道, 2024-03-06 (005):3.
- [2] DU Z. The president's first mention of "new productive forces" and its connection to emerging industries and future industries [N]. China Economic Guide, 2023-09-19 (001):3.
杜壮. 总书记首提“新质生产力”与新兴产业、未来产业有何关联? [N]. 中国经济导报, 2023-09-19 (001):3.
- [3] SUN K W, DENG X, WANG J. Research on the practical teaching system of data science and big data technology major under the background of new engineering [J]. Higher Education Journal, 2023, 9 (14): 5-8.
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, 9 (14): 5-8.
- [4] ZHU L L, CAO H X, KAN X, et al. Research on the construction of data science and big data technology major in application-oriented undergraduate colleges under the background of new engineering [J]. Information System Engineering, 2024(3):166-169.
朱灵龙, 曹海啸, 阚希, 等. 新工科背景下应用型本科高校数据科学与大数据技术专业建设研究[J]. 信息系统工程, 2024(3):166-169.
- [5] ZHOU H, JIANG Z Y. Construction of big data technology and application major under the background of industry-education integration [J]. Computer Education, 2021(9):108-112.
周欢, 江兆银. 产教融合背景下的大数据技术与应用专业建设[J]. 计算机教育, 2021(9):108-112.
- [6] CAO F, ZHANG H, CAO F Y, et al. Construction and practice of data science and big data technology major based on OBE [J]. Software Guide, 2023, 22(6):162-166.
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.
- [7] LI J, QIAO J T, WANG J, et al. Analysis and enlightenment of big data major construction content [J]. Computer Education, 2024(1):157-161.
李杰, 乔金涛, 王捷, 等. 大数据专业建设内容分析及启示[J]. 计算机教育, 2024(1):157-161.
- [8] WANG Y, CONG Y Z, HUANG L, et al. Construction of cloud classroom for big data practical course system [J]. Computer Education, 2022 (4): 75-79.
王岩, 丛一之, 黄岚, 等. 大数据实践课程体系的云课堂建设[J]. 计算机教育, 2022(4):75-79.
- [9] GUO G W, CHEN D M, CAI Y. Research on talent training model of data science and big data technology major under the background of new engineering [J]. Journal of Henan Institute of Education (Natural Science Edition), 2023, 32(4):73-76.
郭改文, 陈丹敏, 蔡艳. 新工科背景下数据科学与大数据技术专业人才培养模式研究[J]. 河南教育学院学报(自然科学版), 2023, 32(4): 73-76.
- [10] YANG C Y, FANG J W, ZHOU X L, et al. Improvement of engineering practical education quality from the perspective of industry-education integration: a case study of data science and big data technology major in local applied undergraduate colleges [J]. Higher Education Journal, 2021, 7(31): 69-72, 77.
杨彩云, 方建文, 周小龙, 等. 产教融合视角下的工程实践教育质量提升——以地方应用型高校的数据科学与大数据技术专业为例[J]. 高教学刊, 2021, 7(31):69-72, 77.
- [11] WANG H Y, HU Y X, ZHANG W, et al. Exploration and practice of big data technology professional training teaching under the background of new engineering [J]. Education and Teaching Forum, 2024 (18): 93-96.
王海燕, 胡玉雪, 章文, 等. 新工科背景下大数据技术专业实训教学探索与实践[J]. 教育教学论坛, 2024(18): 93-96.
- [12] LIU S N. Research on practical teaching platform under the background of massive data [J]. Big Data Time, 2024(6):41-45.
刘胜楠. 海量数据背景下实践教学平台研究[J]. 大数据时代, 2024 (6):41-45.
- [13] ZUO D, LUO J, HUANG W. Teaching design and practice of big data technology based on engineering education certification [J]. Electronic Technology, 2024, 53(4):164-167.
左丹, 罗佳, 黄玮. 基于工程教育认证的大数据技术教学设计与实践[J]. 电子技术, 2024, 53(4):164-167.
- [14] YU Y F, ZHANG Y F, DENG J B, et al. Research and application of the practical teaching system of "cloud platform+service" [J]. Laboratory Research and Exploration, 2024, 43(1):209-213, 226.
余宇峰, 张云飞, 邓劲柏, 等. “云平台+服务”的实践教学体系研究与应用[J]. 实验室研究与探索, 2024, 43(1): 209-213, 226.
- [15] LI L. Application of Docker container technology in building computer experiment platforms [J]. Network Security and Informatization, 2024 (6):114-116.
李磊. Docker容器技术在构建计算机实验平台中的应用[J]. 网络安全和信息化, 2024(6):114-116.

(责任编辑:尹晨茹)