

基于华为 MRS 平台的大数据实践课程探索

辛宇, 李国庆, 钱江波, 严迪群

宁波大学信息科学与工程学院, 浙江 宁波 315211

摘要

针对大数据实践教学中存在的大数据集群环境搭建、环境管理、组件兼容与硬件维护等教学难题, 以华为 MRS 大数据平台为实践教学平台进行课程探索。论述了 MRS 大数据平台教学的优势, 提出了“学、练、践、思”的改革思路, 即理论教学、实验教学、综合实践、课程评价, 并从知识体系、实验教学、综合实践任务、课程考核方式等方面介绍了教学改革的实施过程。该课程已在华为“智能基座”产教融合协同育人基地开展示范教学, 有效提高了学生在大数据技术方面的自主研发能力与创新思维能力。

关键词

华为 MRS 平台; 大数据教学; 课程改革

中图分类号: G642

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2025006

Big data practice course exploration based on Huawei MRS platform

XIN Yu, LI Guoqing, QIAN Jiangbo, YAN Diqun

Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China

Abstract

In response to the teaching challenges in big data practice education, such as building big data cluster environments, environment management, component compatibility, and hardware maintenance, Huawei MRS big data platform is used for course exploration as a practical teaching platform. The advantages of the Huawei MRS big data platform in teaching were discussed. The reform concept of “learning, teaching, practicing and thinking” was proposed, which included theoretical teaching, experimental teaching, comprehensive practice and curriculum evaluation. The paper discussed the implementation process of curriculum reform from the aspects of knowledge system, experimental teaching, comprehensive practice tasks and course assessment methods. This course has been demonstrated at Huawei “Intelligent Foundation” industrial-education integration collaborative education base, effectively improving students' independent research capabilities and innovation skills of big data technology.

Key words

Huawei MRS platform, big data teaching, curriculum reform

0 引言

随着信息技术的不断发展和数据分析的应用普及,大数据已成为当今时代的重要驱动力之一,对各行各业产生了深远的影响。在这个数据爆炸的时代,如何从海量的数据中提取有价值的信息,并运用于决策、创新和解决实际问题,成为大数据分析亟待解决的问题^[1]。2022年6月,国务院印发《关于加强数字政府建设的指导意见》,提出构建开放共享的数据资源体系。在这样的背景下,面向国产化大数据平台的大数据实践课程改革显得尤为重要和迫切^[2-3]。

大数据是物联网、云计算、人工智能课程的核心,也是计算机和软件工程类专业的必修课程,旨在培养学生运用大数据思维解决现实问题的能力^[4-5]。传统的大数据教学侧重于理论知识的传授,缺乏对实际应用能力的培养,与行业需求和发展趋势脱节,难以满足学生的实践学习需求和社会的实际需求,因此,基于大数据平台的大数据实践课程改革成为大数据教学的迫切问题^[6-7]。

华为“智能基座”产教融合协同育人基地由教育部、华为于2020年年底联合发起,首批布局72所高校,旨在深化信息技术领域人才培养模式改革和协同创新,着力构建以信息技术领域关键核心技术为基础的产业与人才生态^[8]。基于华为MRS (MapReduce service) 平台的大数据实践课程改革以华为弹性云服务器 (elastic compute service, ECS)、MRS平台为大数据课程的实践平台,开展多个面向应用的大数据处理项目;面向国产化技术对教学进行升级,旨在培养学生在国产大数据平台上的工程能力。

1 大数据实践教学存在的问题

大数据技术是现代化软件工程实践的主流技术,包含多种软件编程知识。因此,大数据实践教学涉及多种技术的综合性应用,应侧重于培养学生的实践能力,通过实践提升学生对理论知识体系的学习效果。然而,大数据实践教学容易出现实践能力培养不足、教学内容过于理论化、教学设备及实验平台难以满足综合性实践需求等问题。

(1) 实验教学方法缺乏工程能力的培养

随着大数据技术的迅速发展,新技术和工具不断涌现,如Hadoop、Spark等。大数据技术需要与实践应用深度绑定,动手实践有助于学生更深入地理解理论知识。为适应大数据的快速发展,应以实践应用为导向,教学内容与实际应用场景相结合,强调学生的技能培养和职业准备。

(2) 教学内容理论化

理论教学侧重于大数据的架构原理及大数据系统的设计原理,实践教学侧重于大数据软件的应用与程序实现,这导致理论教学与工程应用脱节,增加了学生理论理解的难度。此外,由于大数据处理技术是近年来的新型软件技术,掌握大数据技术需要长期参与相关项目,但目前一些教师主要从事理论研究,往往缺乏实际大数据项目经验,导致教学内容偏于理论,缺乏实际应用方面的教学指导和案例分享。

(3) 教学设备难以满足大数据实践要求

大数据教学通常以分布式集群为实验平台,需要使用高性能计算设备和大容量存储设备来处理和存储大规模数据集。目

前,大多数教学实验室难以为每一名学生提供所需的硬件与数据,因此大数据的实践教学需要搭建单机虚拟平台环境(如VirtualBox、VMware等)。虚拟平台环境的运行效率依赖于单机硬件资源,往往无法达到理想的大数据集群效果。

(4) 考核方法重理论而轻实际

目前全日制教育往往以传统笔试为考核制度,理论成绩占比大,对实践应用的考核不足。大数据课程通常注重实际操作和项目实践,不仅仅停留在理论层面,笔试难以全面考查学生的实际操作水平和应用能力。此外,大数据涉及多个领域的知识,包括数据处理、数据分析、可视化、机器学习等,这些内容更适合以实践操作的方式进行考核。

2 基于华为 MRS 平台的大数据教学改革思路

2.1 MRS 平台优势

基于MRS大数据云平台的实践教学应用具有资源弹性化、易于管理和易于全局监控等优势,能够为学生提供更加丰富的实践学习体验。

- 弹性化配置。华为MRS是华为云推出的一款托管式大数据服务平台,可随时进行集群扩容/缩容、升级Master节点规格、设置弹性伸缩策略,可根据教学需求随时调整资源规模,满足不同规模和复杂度的实践项目需求,使用成本比自建Hadoop集群的低。

- 易于平台运维。大数据云平台由专业的云服务提供商进行管理和维护,可以及时更新和部署最新的大数据技术和工具,保证教学内容与行业发展同步。此外,MRS提供统一的可视化大数据集群管理界

面,具有服务启停、配置修改、健康检查等服务,并提供集群管理监控的可视化、便捷告警功能,可有效提高学习实践的系统运维效率。

- 易于全局监控。华为MRS平台提供了专业的平台管理工具FusionInsight Manager,该工具提供集群、服务、角色实例的操作功能,可实时监测大数据平台中的大数据组件运行状态。教师可以通过大数据云平台对学生的实践活动进行管理和监控,了解学生的学习进度和表现,查看学生的实验进度并及时提供实践指导。

- 大数据组件兼容性好。大数据的生态组件较多,如Flume、Kafka、Flink、Spark等,在实践教学中,极易出现生态组件之间的版本以及Java版本不兼容的问题,在实践过程中学生会花费大量时间用于处理大数据组件间的版本适配问题。利用MRS平台创建的多种应用场景集群,比如Hadoop集群、HBase集群、Kafka集群等,完全与MRS平台相兼容,可有效提高实验效率。

2.2 教学改革方法

基于大数据云平台的教学以成果导向教育(outcome based education, OBE)理念为导向,从理论到实践,以具备利用大数据处理技术解决实际业务问题的能力为指引,设定课程目标,并在课程评价时,以大数据技术知识原理、技术掌握和解决问题的能力为具体要求,设计可量化的评价方法,及时发现问题并对教学过程进行干预,从而形成一个持续改进的闭环^[9]。基于OBE理念,本课程提出以下4个课程目标。

目标1:理解不同的大数据架构的工作原理,深入把握其存储和计算模式在应用领域中的适用性和局限性,能够评估、比

较和设计大数据应用领域复杂工程问题的解决方案。

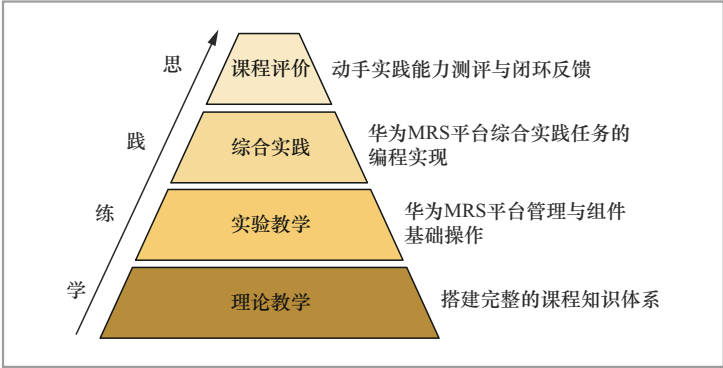
目标2：能够理解MRS及多种大数据平台的设计原则和实现细节，并对多种大数据技术进行对比分析，在实际应用中能够结合所学的专业知识，针对特定问题分析并设计大数据技术方案。

目标3：能够研究大数据应用系统的架构、数据处理方式等，分析结果，并改进系统的性能及可用性，掌握MRS等大数据平台生态环境各部分的功能、结构以及协作方式。

目标4：能够根据所学的大数据处理技术知识，主动了解计算机新技术的发展动态，能够运用所学知识自主解决生产实践中的各类复杂工程问题。

根据课程目标，本文制定了如图1所示的大数据教学整体改革思路，即理论教学、实验教学、综合实践、课程评价方面。

第一，搭建完整的课程知识体系，为学生梳理大数据技术知识脉络，对大数据课程基础知识（如Hadoop、Spark、NoSQL、ETL、分布式协调服务技术等理论知识）进行归纳。在此基础上，教师通过理论讲解与兴趣引领，指导学生对技术进行归类与思考，并以MRS的生态软件为实例平台，改善大数据教学内容理论化的问题，避免理论与实践脱节。



第二，在实验教学过程中，利用华为MRS的文档资源与操作手册，教师指导学生搭建自己的大数据MRS平台，并利用华为FusionInsight Manager工具进行MRS的管理，实现大数据组件安装、角色管理、安全管理等大数据运维实践，组织学生利用MRS进行Hadoop、Spark、NoSQL、ETL、分布式协调服务等基本技术的实践，提高实验教学对学生工程能力的培养。

第三，在综合实践过程中，教师指导学生学习华为MRS平台的大数据开发指导与技术指南，并布置基于Hadoop、Spark、HBase、Kafka、Flink等大数据组件的综合应用实践作业，指导学生参考组件SDK进行综合实践任务的编程，充分利用MRS软件与硬件平台的集成化优势，培养学生的实践应用能力与自主解决问题的能力。

第四，在课程评价方面，教师组织学生对所有完成的实践作品进行集体点评，学生分析他人实践作品的优劣之处；收集学生对实践创新性和个人收获的反馈，据此调整和改进课程安排，从而形成课程持续改进的闭环机制。

3 基于华为MRS平台的大数据教学实施过程

基于华为MRS平台的大数据教学改革需要在基础理论知识教学的基础上，提供真实的行业大数据场景，让学生更加充分地了解大数据行业的需求并积累实践经验。对此，本文在课程理论教学、实验教学、综合实践、课程评价4个环节开展课程改革，建立如图2所示的面向华为MRS平台的教学实践与评价体系，从而全方面提升学生的实践水平与思维能力。

3.1 知识体系

按照大数据领域的主要方向（如Hadoop技术、Spark技术、流处理技术、NoSQL技术等）整理相关知识，大数据主流技术见表1。同一技术方向可以采用层级结构从整体到细节逐步展开，这强调了同一技术方向中内部技术分类的逻辑关联和连贯性。构建如图3所示的大数据技术的课程知识图谱，该知识图谱展示了大数据课程中各个知识点之间的关系和结构。大数据课程知识图谱能够帮助学生理清课程结构，明确理论知识与大数据技术之间的关联。学生可以通过知识图谱快速定位某个知识点以及其在大数据生态系统中的位置，确保对各个知识点的学习能够循序渐进。根据知识图谱，可将华为MRS平台的FusionInsight Manager工具以及华为云数据治理中心DataArts Studio应用于大数据技术以及分析理论的教学，理论知识与实际应用相结合能够帮助学生快速构建大数据知识思维体系。

3.2 实验教学

在大数据实验教学中，学生直接参与

真实的数据处理、分析和解决问题的过程，通过亲身实践加深对相关概念和技术的理解。大数据实验教学包括平台搭建和数据管理、作业执行过程管理、批量离线数据处理和流式数据处理等。

针对平台搭建和数据管理，可以使用FusionInsight Manager、DataArts Studio等平台工具，帮助学生快速完成在线开发和调试MRS HQL/SparkSQL脚本、拖曳式地开发MRS作业，完成MRS与其他20多种异构数据源之间的数据迁移和数据集成。MRS强大的作业调度与可视化监控功能有助于培养学生自主搭建大数据平台环境与对大数据平台进行运维的能力。

针对MRS作业执行过程管理，可以使用MRS中的MapReduce任务或Hadoop生态系统中的其他组件（如MapReduce、Spark、HQL和SparkSQL等）进行测试，进行作业处理流程中的作业提交、作业调度、数据分布、任务执行和结果输出等操作，使学生快速掌握管理MRS作业执行过程的方法。

针对批量离线数据处理和流式数据处理，MRS中常用的批量离线数据处理框架包括MapReduce、Spark等，流式数据处理常用的框架包括Kafka、Spark Streaming、Flink等，可以根据实验需求选择合适的工具和框架，设计数据处理的逻辑和流程，帮助

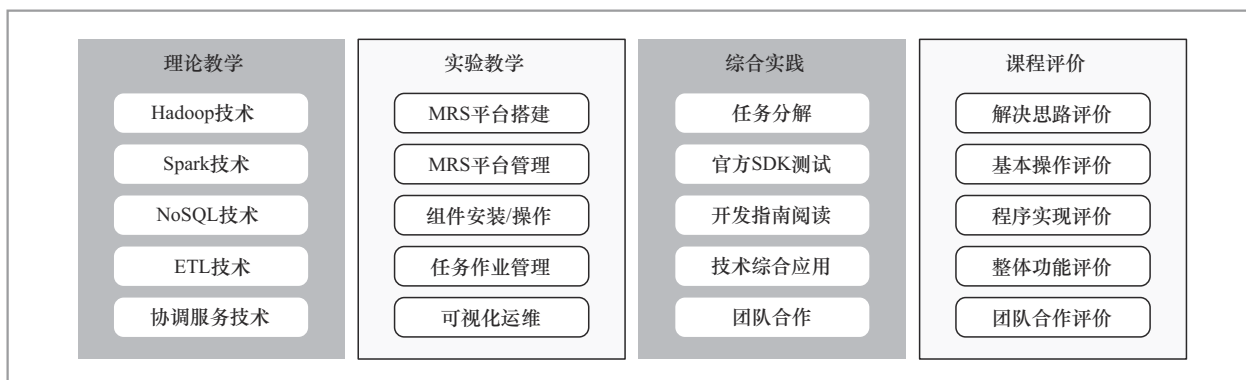


图2 面向华为MRS平台的教学实践与评价体系

表1 大数据主流技术

技术分类	技术实例	学习难点	MRS平台是否支持
Hadoop	Hadoop	配置和优化Hadoop集群,理解大数据存储和大数据处理的底层架构	支持
	HDFS	在分布式文件系统环境中管理数据,对数据复制、容错和NameNode性能进行调优	支持
	Yarn	资源分配和作业管理,理解Hadoop集群的任务处理及资源调度机制	支持
	MapReduce	理解MapReduce的分区、shuffle和编程方法,以及算法性能调优	支持
	Hive	对Hadoop、HBase中的大规模数据进行Hive查询和分析,掌握数据仓库的管理方法	支持
Spark	Spark RDD	理解弹性分布式数据集(Resilient Distributed Datasets,RDD),掌握多种RDD对对象的处理、转换和优化	支持
	Spark SQL	利用DataFrame API,在Spark应用程序中实现高效的SQL查询和数据操作	支持
流处理	Spark Streaming	掌握Dstream的编程方法以及有状态流的处理方法	支持
	Flink	基于Flink有状态处理、事件时间处理和容错处理3种处理技术的设计和实现	支持
NoSQL	HBase	理解HBase的Rowkey索引机制,并根据Rowkey索引机制实现HBase的读写性能优化	支持
	Redis	理解内存缓存和存储中的高可用性、数据持久性和高效数据结构,配置Redis集群	-
	MongoDB	面向文档学习数据库的语法和操作符,以及如何编写高效的查询语句	-
	InfluxDB	以高吞吐量和低延迟为目的,对时间序列数据存储和查询进行监控和分析用例	-
	Neo4j	学习Cypher查询语言、图数据建模以及图数据的检索	-
ETL	Flume	理解Flume的架构以及数据源(Source)、数据通道(Channel)、数据接收器(Sink)、拦截器(Interceptor)和转换器(converter)的使用方法	支持
	Kettle	理解转换(Transformation)、作业(Job)、步骤(Step)、连接(Connection)的配置方法,以及数据清洗方法	-
消息队列	RabbitMQ	学习RabbitMQ架构和组件交换机(Exchange)、队列(Queue)和绑定(Binding)的运行原理,以及多种消息路由及其路由策略	-
	Kafka	理解Kafka生产者、消费者、主题等组件的工作原理,以及分区、复制、故障容错等工作内容	支持
协调服务	ZooKeeper	理解ZooKeeper分布式锁和Watcher机制,实现ZooKeeper的配置优化	支持
	Dubbo	掌握Dubbo的微服务架构配置方法,Dubbo注册表、负载均衡和容错机制的实现方法	-

学生掌握批量离线数据处理的和流式数据处理的MRS技术架构及其实践应用。

3.3 综合实践任务

大数据课程的基础知识包含Hadoop、MapReduce、Spark、HBase、Hive、Flink、Kafka等理论知识与基础组件，通过综合

实践帮助学生了解在真实场景中如何使用MRS进行作业管理。典型的综合实践案例涉及大数据处理、日志分析、推荐系统等领域的的数据收集与清洗、数据存储与管理、批量离线处理和流式数据处理等任务，具体分类如下。

- 数据收集与清洗任务：从真实的数据源中获取数据，并进行清洗和预处理，

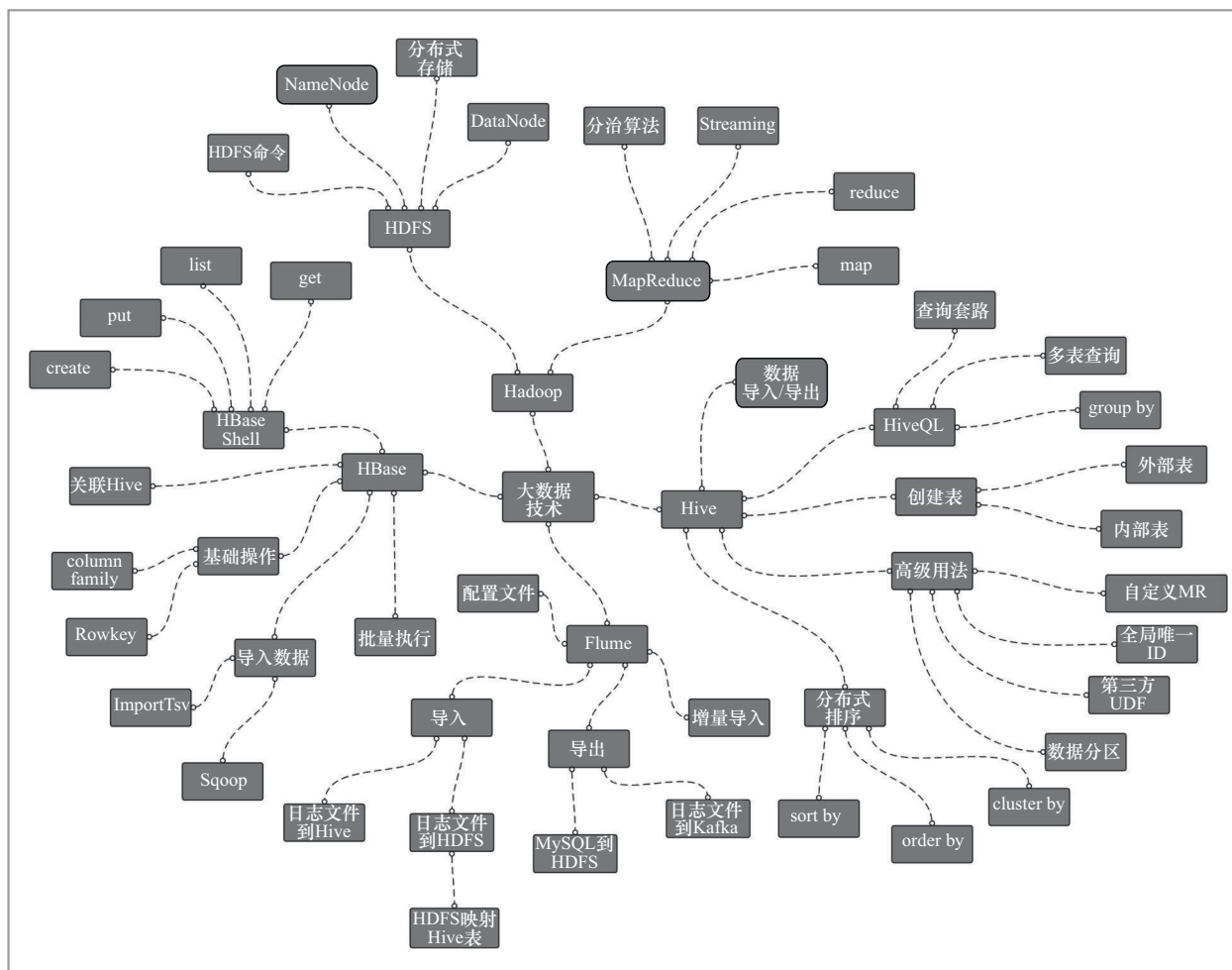


图3 大数据技术课程知识图谱

使用 Flume、Sqoop 等 ETL 工具进行数据清洗、去重、缺失值处理等操作，确保数据的质量和完整性。

- 数据存储与管理任务：选择合适的数据库或数据仓库，设计数据模型，进行数据分区和索引等。

- 批量离线处理任务：使用批量处理框架，如 Hadoop、MapReduce 或 Spark，对大规模数据进行批量处理，包括数据分析、特征提取、模型训练等。

- 流式数据处理任务：使用流式处理框架，如 Kafka 或 Flink，对实时产生的数据流进行处理和分析，包括流式计算、实

时监控、实时预测等。

综合实践任务鼓励学生组队进行项目实践，有助于培养学生的团队合作和沟通能力。此外，综合实践任务应及时跟进行业发展和技术更新，持续更新教学内容和案例，确保教学内容与实际需求相符。

3.4 课程考核方式

本课程的实践教学考核由平时表现、基础实验和综合实践3个部分构成，其中，基础实验考核是对学生大数据平台环境搭建、大数据组件基本操作及大数据技术应

用的考察，综合实践是对学生理论知识的应用能力，以及自主解决问题能力的考察。为了培养学生的动手实践能力及解决复杂工程问题的能力，综合实践任务的权重需要提高。大数据课程实践教学考核评价方式见表2。

4 基于华为 MRS 平台的大数据教学改革效果

本课程于2020年10月成为华为“智能基座”产教融合协同育人基地建设的课程，在计算机科学与技术专业本科生与研究生教学中开设了实验班，每学期学生人数为100人，目前已完成3届学生的授课。在理论知识学习方面，在华为 MRS 平台进行实验教学，学生可以快速掌握大数据组件之间的技术关联，加深对基础概念的理解。在实验及实践教学方面，学生通过对华为 MRS 云平台操作手册和编程指南的研究，学习基本的大数据平台管理及基本组件操作，并借助 MRS 平台设计交通堵塞实时预警、日志文件数据清洗、电子商务推荐系统等大数据技术的典型应用，这充分锻炼

了学生在大数据技术方面的自主研发能力与创新思维能力。此外，在传统大数据实验教学中，实验环境（包括搭建实验环境、安装生态软件与软硬件兼容性调优）约占1/3课时，且受实验硬件限制，实验教学大多采用单机模式或伪分布式模式进行大数据操作模拟。采用 MRS 作为实验平台后，学生的实践效率提高了1/3，学生的实践案例数量提高了1倍。团队合作的实践项目有助于培养学生的团队合作和沟通协调能力。基于华为 MRS 平台的大数据实践课程改革全方面地提高教学质量，为学生提供了更加丰富的学习体验。

目前，华为 MRS 平台应用于实践教学的缺点在于学生过分依赖平台化的大数据服务容易导致学生欠缺大数据生态软件底层运维知识。因此，在后续理论教学中，需要更加关注学生对大数据生态技术栈的基础知识点的掌握。未来本课程将重新归纳大数据技术知识图谱，帮助学生在 MRS 平台进行快速实践的同时，形成完整的大数据知识体系。

5 结束语

随着大数据技术的广泛应用，大数据

表2 大数据课程实践教学考核评价方式

考核环节	考核内容	考核方式	权重	课程目标
平时表现	考勤及问答	平时成绩	10%	目标1、目标4
基础实验	环境搭建	实验报告	10%	目标1~目标3
	MRS平台管理	实验报告	10%	
	大数据组件应用	实验报告	10%	
	MRS作业管理	实验报告	10%	
综合实践	任务思路及架构设计	系统展示	20%	目标1~目标4
	程序实现及功能展示	系统展示	20%	
	团队合作	系统展示	10%	

课程已成为计算机类专业本科生及研究生的核心课程。在实践教学过程中,以华为MRS平台为大数据应用实践平台,不仅解决了大数据集群环境搭建、环境管理、组件兼容与硬件维护等实践难题,还对学生进行了国产大数据平台应用的启蒙教育。此外,采用华为MRS平台避免了大数据软件不兼容的问题,学生可以将学习精力放在大数据项目的开发和操作中,这符合新工科教学改革及产学研融合的教育教学目标。

参考文献:

- [1] 夏大文,王林,张乾,等.大数据应用技术课程教学改革与实践[J].大数据,2020,6(4): 115-124.
- XIA D W, WANG L, ZHANG Q, et al. Teaching reform and practice of big data application technology course[J]. Big Data Research, 2020, 6(4): 115-124.
- [2] 郭玉臣,李育灵,王鹏凯.基于华为鲲鹏云平台的多维嵌入式实践教学改革[J].计算机教育,2023(5): 135-139.
- GUO Y C, LI Y L, WANG P K. Multidimensional embedded practice teaching reform based on Huawei Kunpeng cloud platform[J]. Computer Education, 2023(5): 135-139.
- [3] 杨阳,郝玉婷,陶丽,等.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].高等教育研究,2023(5): 54-59, 116.
- YANG Y, HAO Y T, TAO L, et al. Teaching reform and practice exploration of data analysis courses in the big data era[J]. Research in Higher Education of Engineering, 2023(5): 54-59, 116.
- [4] 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.
- CUI X L, ZHANG M, ZHANG L, et al.

Research and construction of curriculum group of applied big data talent training under background of new engineering[J]. Experimental Technology and Management, 2021, 38(2): 213-218.

- [5] 朝乐门.数据科学导论的课程设计及教学改革[J].计算机科学,2020,47(7): 1-7.
- CHAO L M. Course design and redesign for introduction to data science[J]. Computer Science, 2020, 47(7): 1-7.
- [6] 易艳红,张晶,张聪.应用型本科信息管理与信息系统专业大数据方向课程改革研究[J].图书馆学研究,2019(7): 15-20.
- YI Y H, ZHANG J, ZHANG C. Research on curriculum reform of big data orientation for applied undergraduates majored in information management and information system[J]. Research on Library Science, 2019(7): 15-20.
- [7] 张祖平.新工科背景下大数据专业导论课程的改革与探索[J].大数据,2018,4(6): 38-45.
- ZHANG Z P. Reform and exploration for introduction course of big data professional under the background of emerging engineering education[J]. Big Data Research, 2018, 4(6): 38-45.
- [8] 张伟文,林穗,刘冬宁.基于华为“智能基座”的操作系统课程建设[J].计算机教育,2024(3): 106-110.
- ZHANG W W, LIN S, LIU D N. Construction of operating system course based on Huawei's "smart base"[J]. Computer Education, 2024(3): 106-110.
- [9] 张晓明,李海生,陈明,等.基于OBE理念的大数据技术开发课程教学设计与实践[J].计算机教育,2021(8): 166-170.
- ZHANG X M, LI H S, CHEN M, et al. Teaching design and practice of big data technology development course based on OBE concept[J]. Computer Education, 2021(8): 166-170.

作者简介



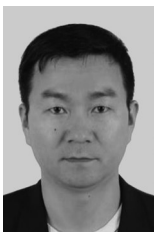
辛宇（1987–），男，博士，宁波大学信息科学与工程学院副教授，主要研究方向为人工智能、大数据计算等。



李国庆（1994–），男，博士，宁波大学信息科学与工程学院讲师，主要研究方向为人工智能、智能计算等。



钱江波（1974–），男，博士，宁波大学信息科学与工程学院教授，主要研究方向为人工智能大模型与云计算等。



严迪群（1979–），男，博士，宁波大学计算机系主任、副教授，主要研究方向为人工智能数据与模型安全等。

收稿日期: 2024-05-20
通信作者: 辛宇, xinyu@hrbeu.edu.cn
基金项目: 教育部产学合作协同育人项目(No.202102058024); 浙江省高等教育教学改革研究项目(No.JG20190128); 宁波大学教学改革研究项目(No.JYXM2025123)
Foundation Items: Ministry of Education Industry–University Cooperative Education Project (No.202102058024), Zhejiang Higher Education Teaching Reform Research Project (No. JG20190128), Ningbo University Teaching Reform Research Project(No. JYXM2025123)