

《大数据存储与管理》课程建设与教学改革研究

刘超, 董理君*, 邓泽, 王媛妮, 陈云亮

中国地质大学(武汉)计算机学院, 湖北 武汉

收稿日期: 2023年6月5日; 录用日期: 2023年7月17日; 发布日期: 2023年7月26日

摘要

高等学校课程建设与教学改革是提高本科教育人才培养质量的重要途径。大数据存储与管理课程是数据科学与大数据技术专业的核心必修课程, 理论和实践教学并重。作为新兴专业的新设课程, 从国内外课程内容设置的现状分析为抓手, 以培养大学生动手创新能力为目标, 提出了系统化的课程教学内容与实践教学建设体系, 强化了课程的实验设计的完整性、实验内容的关联性, 激发了学生的学习兴趣。经过三年教学实践与改革, 根据教学效果评价反馈分析, 验证了课程建设与教学改革取得良好的教学成效。

关键词

大数据存储, 大数据管理, 教学改革, 实践教学, NoSQL数据库

Research on Curriculum Construction and Teaching Reform of "Big Data Storage and Management"

Chao Liu, Lijun Dong*, Ze Deng, Yuanni Wang, Yunliang Chen

School of Computer Science, China University of Geosciences, Wuhan Hubei

Received: Jun. 5th, 2023; accepted: Jul. 17th, 2023; published: Jul. 26th, 2023

Abstract

The curriculum construction and reform of college teaching is an important way to improve the quality of undergraduate education talents. "Big Data Storage and Management" course is the core

*通讯作者。

文章引用: 刘超, 董理君, 邓泽, 王媛妮, 陈云亮. 《大数据存储与管理》课程建设与教学改革研究[J]. 创新教育研究, 2023, 11(7): 1881-1886. DOI: 10.12677/ces.2023.117279

compulsory course of data science and big data technology major, while the theoretical and practical teaching is equal to the focus. As an emerging professional new course, the related work of the content of the curriculum at home and abroad has been analyzed. With the goal of cultivating the hands-on innovation ability of college students, a systematic curriculum teaching content and practical teaching system have been proposed. The integrity of the experimental design and the correlation between the experimental content have been strengthened, by inspiring the learning interest of the college students. Through three years of teaching practice and reform, the excellent results of the curriculum construction and teaching reform have been verified based on the feedback of the teaching effect evaluation.

Keywords

Big Data Storage, Big Data Management, Teaching Reform, Practical Teaching, NoSQL Database

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着社会对大数据领域人才的需求越来越大,大数据专业方向的人才培养受到了高等学校的广泛关注。中国地质大学(武汉)在2018年初获批准新增“数据科学与大数据技术”本科专业(简称:大数据专业),计算机学院在2020年春季学期首次开设了《大数据存储与管理》课程。由于我国高等院校在《大数据存储与管理》的教学方面尚处于摸索阶段,缺乏成熟的、系统的和规范的大数据存储管理的教学体系和现成教材,因此,我校计算机学院的《大数据存储与管理》课程组面临教学内容组织和实践实施的巨大挑战。

《大数据存储与管理》课程的教学内容具有覆盖面广的特点,既有传统关系型数据管理的知识,也包含最新的NoSQL数据库内容。该课程的实验环节复杂,涉及了依托大数据平台的各类NoSQL数据库,如:键值数据库、列存储数据库、文档数据库、图数据库等。《大数据存储与管理》作为大数据专业的核心必修课程,在课程建设初期遇到诸多困难,没有教学大纲和内容设计,缺乏现成的教学课件。因此,迫切需要对教学内容和实验环节进行课程建设与教学改革。课程组通过充分调研讨论,申请了以创新能力培养为目标,研究系统化课程内容教学与实践教学体系改革的校级教学项目立项。通过三年的教学实践,迭代完善了课程教学内容设计,加强了《大数据存储与管理》的实验设计创新性、实验内容的关联性,激发了本专业学生的学习兴趣,培养了学生的创新实践能力。

2. 大数据存储与管理课程现状分析

据调查发现,国内的重点大学目前没有开设与《大数据存储与管理》直接对应的同名课程,相关理论教学内容分散在不同的课程中。例如:北京邮电大学开设了NoSQL数据库技术课程,提出以图、文档、键值、列数据库为核心的教学内容[1]。武汉大学把百万级国产数据库的内容引入数据库原理课程[2]。多门课程的内容分散,使得学生缺乏对大数据管理教学内容的整体认识,容易陷入片面化,局限性。

实验教学是培养大数据专业学生创新思维能力的重要教学手段,是学生实践能力和创新能力培养的必需环节。有一些高校开展了大数据管理技术相关的实践教学研究。例如,叶枫等人基于双层学习模式开展了NoSQL数据库实践教学的探讨[3]。党德鹏等人在北京师范大学从大数据技术角度提出了数据库

实验教学内容的改革创新实践[4]。西安交通大学的赵玺等人基于大数据管理与应用课程提出了实验教学平台的建设原则[5]。厦门大学大数据技术课程的实验环节依托 Hadoop 分布式平台, 设计了 HDFS 分布式文件系统, HBase 列数据库实验[6]。美国伯克利大学、华盛顿大学等国外知名高校开设有数据库系统、Web 数据管理等课程, 实验环节和大数据存储与管理课程的内容部分契合, 但是网络公开的英文资料难以直接借鉴使用。

总体而言, 传统的数据库课程的实验内容创新性不够, 现有的新型数据库管理课程缺少复杂的综合性实验, 不能充分培养学生的创新能力。本文提出的课程建设与教学改革研究是从本校大数据专业的《大数据存储与管理》课程的教学实际出发, 致力于研究一种系统化的教学内容建设和实践教学模式改革可行性方案。

3. 大数据存储与管理课程建设

本课程建设的目的是让本专业的学生了解大数据管理系统的发展历史, 掌握大数据管理的数据特征、系统特征和应用特征, 掌握大数据管理的数据模型与查询语言, 理解大数据管理的分层架构体系, 通过实验掌握大数据管理的关键技术。课程设置参考了多本已出版的大数据管理和 NoSQL 数据库系统相关的教材, 讲授内容既保留了传统 SQL 数据库的基本理论和原理, 也针对大数据行业需求, 更侧重非关系型数据库系统的技术和方法。课内实验采用云平台编程环境, 侧重学生的动手能力培养。

3.1. 课堂教学内容建设

围绕大数据存储的数据库原理, 非结构化数据管理的高效率、高可用基础理论, 以建设课程的大纲、教案、考核评价等内容为目标, 《大数据存储与管理》课堂教学共计安排 48 学时, 主要涉及十个部分的教学内容。① 大数据管理系统概述, 包括大数据管理的发展历史、大数据的特征和组成, 2 学时; ② 关系数据库模型与 SQL, 包括关系数据结构与定义、关系操作与关系代数、关系数据库与 SQL 语言、关系的完整性约束, 8 学时; ③ 键值对模型, 包括键值对模型的数据类型、操作类型、操作方法, 基于键值对模型的主流分布式系统 Dynamo, BigTable, 6 学时; ④ 文档模型与查询语言, 包括文档结构, 查询语言 XQuery、JSON, 文档数据库实例, 6 学时; ⑤ 图模型与类 SQL 查询语言, 包括图数据结构与定义, 图数据的操作匹配、查询, 查询语言 SPARQL, 典型的图系统 Neo4J, 6 学时; ⑥ 大数据管理系统的体系架构, 包括集中式体系架构、客户-服务器体系架构、并行体系架构和分布式数据库体系架构, 4 学时; ⑦ 大数据存储与组织, 包括硬件存储、索引技术、以及键值存储、列存储等常见的数据组织方法, 4 学时; ⑧ 分布式查询处理, 主要包括分布式查询处理的层次、数据分布策略、分布式查询处理执行方法, 4 学时; ⑨ 分布式事务, 包括分布式系统、分布式事务、图/键值/文档模型事务处理技术, 4 学时; ⑩ 分布式故障恢复, 包括故障种类、恢复协议, PAXOS 技术与 RAFT 技术, Hadoop 容错备份技术, 4 学时。

3.2. 课程实习内容建设

为了让本科生更好地掌握和运用课程理论知识解决实际问题, 实习课是必不可少的动手实践教学环节。《大数据存储与管理》实习课程共计 16 学时, 分四次安排实验课。主要目标是让同学们熟悉 MySQL, HBase, MongoDB, Neo4j 等四大数据库的安装与使用。为后续的大数据综合实习等实践课程打下坚实的存储管理应用的基础。其中, 四次实验课的基本操作安排如下。① MySQL 数据库实验: 基本表的定义与修改, 单表查询, 连接查询, 嵌套查询, 数据更新, 存储过程建立与调用等操作; ② HBase 数据库实验: 建立单表, 添加数据, 查询表内数据, 整表删除与按照行删除, 按照行删除, 搜索等操作;

③ MongoDB 数据库实验：建表，插入数据，索引，树查询，投影分页和排序，简单查询与复杂查询等操作；④ Neo4j 数据库实验：创建表节点与关系，删除指定关系，查询与更新，筛选与过滤，约束与索引操作等操作。

3.3. 课程建设成效分析

本课程组的老师们通过充分调研，反复研读参考教材，并结合自己的教学和科研实践，对《大数据存储与管理》课程的教学内容进一步调整并完善。根据最终编写的教学大纲，制定详细的教学方案和教学安排，并制作每一个讲的教案。在备课环节，我们以小组团队合作完成的教学课件为基础，着重梳理本课程的重要知识点，进一步做好每节课讲授内容的优化设计。与此同时，依托计算机学院的大数据专业实验室，搭建《大数据存储与管理》课程的教学实验平台，并编写详细的实验教学大纲，以方便学生在动手实践中掌握大数据存储与管理的核心技术和实验方法。课程组通过多轮的教学研讨打磨，总结本课程的经验教训，不断丰富和完善教学环节，正在积极申报《大数据存储与管理》的校级精品课程。

4. 系统化课程教学改革

4.1. 课程教学改革思路

以 2018 级开展第一轮《大数据存储与管理》授课的教学内容为基础，为了进一步提高本科生以大数据思维解决存储问题的创新能力，我们开展了课程改革研究。本课程改革的主要思路包括教学主体的两个方面。一是通过“探索 - 反馈 - 改进”，提高课程组老师们的课堂执教能力以及实践教学水平，二是完成课程教学以及实验环节的内容建设，培养本专业学生的大数据系统创新实践能力。通过课程改革，搭建统一的实验平台，设计创新型实验，提出合适的评价手段和考核方式。在本学院 2018 级大数据专业本科生中进行改革试点，培养学生的大数据管理专业技能以及实践动手能力，切实改善教学效果，形成符合大数据专业特点的新型教学评价体系。总结之后，在 2019 级和 2020 级本专业学生课程教学实践中进行推广。

4.2. 课程教学改革实践

我们围绕调研探索、做好前期开课准备。第一轮上课实践与学生深入交流，通过“调研实施、收集反馈、迭代改进”的三步走路线展开课程教学的改革实践。

首先，从 2019 年立项的 6 月到 9 月，收集并调研了相关资料，比较各个高校大数据数据管理相关课程的教学大纲、教学内容、教材选择等。走访学生、设计教学调查问卷，收集同学们的反馈建议。并调研国内重点大学大数据专业的实验环境，实验内容设置，教学效果等。课程组成员也参加全国大数据管理的教学培训课程。

其次，2019 年 10 月到 2020 年 2 月，调整和完善了教学大纲、编写教学方案设计、收集网络课程资源、设计课程教案。针对实验环节，采用统一的大数据实验技术、搭建实验环境、配套完善创新实验的设计，为 2020 年正式开课确定了有效教学内容，教学方法，考核体系等。

再次，2020 年 3 月到 2020 年 9 月，针对 2018 级大数据专业学生，开展第一轮上课。课程组成员充分研讨，相互听课提建议，完善教学内容的不足。正式开展系统化教学，以及创新实验模式验证。并根据探索的经验，撰写《大数据存储与管理》课程改革的教学反思总结文档。

最后，2020 年 10 月至今，在计算机学院继续开展《大数据存储与管理》的教学模式验证与优化，进一步改进教学内容、教学模式和考核评价方法，并检验实验教学效果。目前已经在 2018 级，2019 级，2020 级等本科生中开展了三轮迭代优化教学改革实践。

4.3. 课程教学改革效果

本课程教学改革实施过程中,取得了一系列的教学改革成果。① 课堂教学内容详实具体,引导学生自主、协作式学习,通过提问和讨论,改善传统数据库课程教学内容陈旧、缺乏系统性、教学效果差的现状。② 《大数据存储与管理》课程的实验设计考虑了大数据采集、大数据分析等关联课程的相互衔接。实验方式有连续性和系统性,加深学生对实验的原理理解,培养实验整体设计能力。③ 设计《大数据存储与管理》新型评价体系,全方位指导教学管理过程,强化学生的创新思维与创新实践能力,通过反馈改善课程的教学质量。④ 提出一种系统化的《大数据存储与管理》课程建设思路,教师依托课堂教学与研讨,学生参与课堂讨论学习,线下自主学习网络教学资源并参与交流互动,拓展大数据专业学生的专业视野和创新思维能力。⑤ 采用统一的实验技术,设计《大数据存储与管理》实验的创新性和综合型教学内容,强化实验平台的使用,案例结合行业大数据应用,激发学生的学习兴趣,提高学生的创新实践能力。⑥ 研究了一种适合大数据专业《大数据存储与管理》课程的教学考核体系,提高学生课堂听讲的效率和学习积极性,提高实验教学的满意度。

我们设计了一份《大数据存储与管理》课程教学改革的评价表,发给近三年来参与本课程教学的2018级、2019级、2020年级学生匿名填写,并进行统计分析。调查表一共设计了四个方面的小项,每个小项采用等级评分制(1表示效果很差,2、3、4表示不同程度的满意,5表示非常满意。3分是满意的分界线,低于3分表明学生持否定态度,高于3分表明学生持肯定态度),重点调查学生对课程内容的前沿性,实验设计的个人兴趣和难度分层,实验内容对创新能力培养是否有帮助,整个课程的学习收获满意度等方面进行评价。问卷调查统计结果如表1所示。

Table 1. The survey resulting data of teaching reform
表 1. 教学改革调查结果数据

教学班级	内容前沿性	实验兴趣难度	创新能力培养	课程满意度
2018 级学生	3.94	3.91	3.88	3.92
2019 级学生	4.08	4.02	3.97	4.01
2020 级学生	4.11	4.05	4.13	4.08

调查结果表明:本课程的理论教学和实验教学内容在历年迭代更新中,保持了课程内容跟进大数据存储研究的前沿技术;学生的实验兴趣也逐年明显进步,能够主动去解决编程问题;在本课程培养创新能力上多数学生较为满意,对课程的综合评价也是逐年提高。

5. 结束语

《大数据存储与管理》课程建设与教学改革研究,提出了系统化的课程教学体系,体现了课程内容的联系紧密、整体统一的实验设计。我们完成并优化了《大数据存储与管理》课程的教学大纲、教学课件、统一实验平台和创新实践设计等内容,全面提高了课程组的任课老师的教学水平,改善教学效果,同时也提高了大数据专业学生的学习兴趣 and 积极性,培养了大学生的动手创新实践能力。

基金项目

中国地质大学(武汉) A 类教学研究项目(2019A41, 2019A43)。

参考文献

- [1] 袁燕妮, 吴斌, 王柏. 新工科背景下 NoSQL 数据库技术课程建设探索与实践[J]. 软件导刊, 2022, 21(7): 21-26.

- [2] 刘斌, 彭煜玮, 余琨, 等. 国产化时代的数据库原理教学与实践改革[J]. 软件导刊, 2022, 21(11): 172-176.
- [3] 叶枫, 孙骏, 黄倩, 等. NoSQL 数据库实践教学: 结合 Benchmarking 的方法[J]. 软件导刊, 2022, 21(11): 162-165.
- [4] 党德鹏, 高虎, 杨竞帆, 等. 新时代数据库系统原理教材建设实践[J]. 计算机教育, 2022, 335(11): 200-204.
- [5] 赵玺, 冯耕中, 李轩涯, 等. 基于大数据管理与应用课程的虚拟仿真教学实验平台建设与应用[J]. 计算机教育, 2022, 336(12): 194-197.
- [6] 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(6): 29-37.