

职业本科 Hadoop 大数据技术课程建设研究

何永亚 苏秀芝

湖南软件职业技术大学 湖南 湘潭 411100

【摘要】：随着数字经济的发展，大数据技术已成为推进各行业转型升级的重要力量，大数据人才也在市场上变得非常炙手可热。职业本科学校在高等职业教育向本科层次延伸的过程中更是承担着培养高层次技术技能型人才的任務。Hadoop 是大数据技术框架的基础，在很大程度上决定着 Hadoop 课程的教学效果。但是，目前职业本科学校开设的 Hadoop 课程普遍面临课程教学内容与产业实践需求相脱节、实践环节少、师资“双师型”特征不明显等问题。基于此，本文以职业本科学校 Hadoop 课程建设为基础，从培养目标、课程体系、实践平台、教学模式四个维度出发，探索一条 Hadoop 大数据技术课程建设的新路径，为培养服务产业的技术人才提供一定依据。

【关键词】：hadoop 大数据技术；课程建设

DOI:10.12417/2705-1358.25.19.012

1 引言

随着大数据技术不断向垂直产业融合，已经渗透到了人们的日常生活中，在电商平台上的个性化推荐，在金融领域上的风险评估，在医疗行业的疾病预测以及在交通系统的智能调度上都有着很深入的应用。正是如此，大数据技术的发展空间将会更加广阔，在这种大环境下，职业本科院校的大数据技术与应用专业也有其优厚的发展机遇和严峻的挑战。

由于职业本科院校设立不久的缘故，其在专业的建设过程中面临较大的困难，首先便是投入的资金不足，在建设实训场地，设备等方面还无法满足大数据技术教学对硬件设备方面的要求。其次便是教师队伍的师资力量不够充足，能够讲授 Hadoop 等核心课程的教师少之又少。职业本科学生底子相对高职有很大提高，但是由理论到实践还是存在距离。而且对于职本来说，由于它要面对的就业市场上依然是主要是需要既懂理论又懂技术的复合型人才，除了能满足岗位技术条件要求外，还能帮助企业做好岗位相关问题的辅导答疑。

因此，作为大数据核心技术的 Hadoop 课程，在职业本科院校中如何开展教学才能更贴合培养需求？为帮助学生扎实掌握专业知识与技术，助力大数据专业课程教学质量提升及学校办学水平提高，可从两方面着手：一方面，强化实训与科研环境的建设及改造，为学生提供贴近实际的实践平台；另一方面，基于大数据工程技术专业的人才培养目标，结合 Hadoop “基础 + 应用”的课程设置思路，依据大数据行业发展需求，

系统设置 Hadoop 基础知识、HDFS 程序开发、MapReduce 开发、实战案例等课程内容，并根据专业教学实际情况动态调整教学内容。通过这样的方式，着力培养学生的动手能力、分析及解决问题的能力，使他们具备胜任大数据运维、大数据分析师等岗位的专业素养。

2 职业本科 Hadoop 课程的定位与培养目标

聚焦“应用导向、产教融合”的职业本科教育，相较于研究型本科 Hadoop 课程应当淡化理论深度、与高职结合 Hadoop 课程应减少技能广度，突出“技术应用+问题解决”的复合型能力培养。该模式适合职业本科教育自身特点的同时也符合了市场中大数据人才需求。

3 差异化定位分析

相较普通本科而言，职业本科 Hadoop 课程应该弱化分布式系统的底层原理数学推导，增加框架组件实战配置调优的内容；普通本科学校重在培养学生的理论研究能力，Hadoop 课程详细剖析了分布式系统的底层算法与数学模型，职业本科学校更加关注学生是否能够应用 Hadoop 框架实现本职工作、是否能够针对实际业务需要配置调整框架参数使得整个系统性能和稳定性更高。

相比于高职课程中侧重于数据操作而言，职业本科学习的 Hadoop 课程中应当增加对学生数据思维的培养，即学生可以从业务场景中找出真正的数据问题，并不是简单地去完成一些

工具的操作。高职学生在学习完 Hadoop 的各种工具后可以轻松完成具体的工作任务，但是对任务背后涉及的业务逻辑及数据的实际意义却不清楚。而职业本科的学生不仅要掌握如何运用好这些工具，而且还要了解每个业务的整个业务流程，能针对大量的数据，发现业务问题并用 Hadoop 技术解决问题。

4 三维能力培养目标

4.1 技术应用能力

Hadoop 生态中核心组件(HDFS、MapReduce、YARN、Hive、HBase 等)的部署、配置、维护及排障能力；具备搭建中小型数据处理集群的能力，掌握对于常见故障等处理手段。学生的职责是明白了解 Hadoop 生态各组件的安装部署、日常工作运维管理等等内容，在实践工作时能及时地发现问题，有效地解决问题。

4.2 业务解析能力

金融、电商等领域可以通过大数据将自己不同行业所具有的业务需求进行转化，即把它们转化成为大数据的处理流程，其中有一些业务需求可以分解成数据采集（Flume）、数据存储（HDFS）、数据清洗（MapReduce）、数据分析（Hive）、数据可视化（Davinci）这样的完整流程。各个行业的业务需求各不相同，学生们要了解不同行业的业务特点与数据特点，并能将不同行业的业务需求转换为具体的大数据处理步骤，并能根据不同的业务场景去选取不同的技术手段和方法。

4.3 持续学习能力

熟悉 Hadoop 和 Spark、Flink 等其他框架的技术差异性，能够依据情况选用合适的工具，了解开源社区资源查找及应用的方法。大数据技术高速发展，常常会有各种新技术出现，尽管 Hadoop 是大数据核心技术框架，但是总有一些情况下 Spark、Flink 等会更适合这个场景，这就需要我们去了解它们之间的差异。作为一名大学生必须要知道各种框架的不同之处以及其适合使用的场景，然后选取对应的工具进行应用。

5 “三阶六模块” 课程体系设计

按照“理论够用、实践突出”的原则，在职业本科学校的课程体系中，“学-练-用”是课程的基本规律。三阶梯度的教学板块由“基础层-应用层-项目层”构成，使教学活动呈环状结构。

5.1 基础层：构建知识框架

Hadoop 核心原理

基于“分布式系统入门、HDFS 架构、MapReduce 编程模型、YARN 资源调度”这一主线，辅以配套的单节点伪分布式

部署实验让学生自己动手做，达到让学生从中体会到分布式计算。

Linux 与 Shell 编程

突出大数据集群运维的常用命令；介绍 Linux 系统的基本原理以及主要的命令工具，在此基础上，要求学生掌握 Linux 系统常见的操作使用方法，学会管理用户和进程、文件系统等。让 Linux shell 编程可以帮助我们实现自动化运维，解放生产力，编写对应的集群分发和远程执行脚本等，代替人工去繁琐地配置工作，减少出错的概率。

SQL 与数据思维

强化 HiveQL 与传统 SQL 区别对比，教会了学生分组、聚合及窗口函数等操作以实现数据聚合、拆解维度的目的，而在案例过程中需根据业务要求运用 HiveQL 对相关的数据库信息进行提取。

5.2 应用层：强化组件协同

数据采集与处理

集合 Flume（日志采集）+Sqoop（数据迁移）工具完成“MySQL 订单数据 HDFS Hive 数据仓库”迁移工作，重点培养学生的数据格式转化、数据增量同步能力；而数据是通过不同的数据源产生并获取的，数据源有关系型数据库、日志文件、web 页面等等，在获取数据之后都要送到 Hadoop 集群上进行数据处理、数据分析等工作。对于关系型数据库而言可以使用 Flume 完成日志数据采集工作，而 sqoop 能完成关系型数据库到 Hadoop 集群的数据迁移工作，学生们应当学会熟练掌握这种工具，并能够根据实际情况设计出合理完善的数据迁移流程。

NoSQL 数据库应用

围绕 HBase 学习了 HBase 数据库基本知识，如比较关系型数据库，使用 Java API 对比学生上节学习的“用户标签实时存储”案例，运用行键设计、列簇划分；之后简要介绍了 HBase 的数据模型、存储模型、数据分布、扫描、Java API 的使用方法等。HBase 与传统的关系型数据库比较来说，在数据模型、存储方式、查询方式等方面都有较大不同。要求学生能有清晰的认识，并且可以针对实际业务来设计合理的 HBase 表结构，并且能够使用 Java API 来实现数据的增删改查。

5.3 项目层：深化综合实战

整合 Hive, mysql 和 Davinci，输出用户活跃度、消费偏好等指标看板，培养学生数据洞察能力；电商用户的画像对于电商企业开展精准营销有着非常重要的作用，通过对用户的在线行为数据分析，给每一个用户打上对应的标签让学生运用 hive

分析用户行为数据，把分析的结果保存到 mysql 中，再利用 davinci 进行可视化展示，形成一张直观的指标看板。

6 虚拟化集群部署方案

使用 VMware 搭建 3 节点集群，方便学生能在真实情况下完成大规模集群的部署与管理工作；使学生能够体验真实运维环境下的实验流程，例如，“NameNode 故障恢复”的实验，人为杀死主节点，并且观察 ZKFC 自动切换的过程，理解和学习 HDFS 的高可用机制。让学生在安全环境下做各种实验操作，培养学生对各种故障处理能力。

7 企业级案例库开发

问题故障库：整理了 30 种常见的故障，并编制了相应的解决方案库，在学生动手过程中遇到了相关问题时，可直接在故障库中查找其对应的解决方案来解决所面临的问题，从而加强对于 Hadoop 系统知识的学习与掌握。

各类资源库：提供集群部署脚本、组件和可视化模板等各种材料，能大大提升学生们对于实验环节的实践效率，避免同学们重复进行大量的繁琐工作，可以把时间更多的花在去真正解决一些具体问题上面。

8 “双师协同的教学模式创新

“双师型”教学团队组建

校内教师：有三年以上的数据相关项目经验和企业顶岗实践经验，能带领团队一起研究大数据相关的前沿技术并成功应用；在校教师需要明白教学规律及教学对象特点，将知识及技能传授给学生，并将企业顶岗实习的经验带回到教学过程中，让学生跟上时代的步伐，接受行业的新技术、新知识、新信息、新技能等。

企业导师：邀请大数据运维工程师和数据分析师进行授课，“直播运维实战”“项目评审”等方式来讲解本课程的知识以及更新行业相关的知识。企业的导师有着多年企业的工作经验，能给学生带来真实的企业案例，以及企业的技术要求，让学生的了解企业的一些工作情况和相关的工作模式等。

9 “情境化”教学方法实施

案例驱动教学：以“互联网金融项目离线分析”为主线，

在各章学习知识点的基础上，拆解成“需求分析-技术选型-方案实施-方案优化-迭代优化”项目的各个阶段，每节均以具体任务开展。采用案例驱动教学法，把一些比较抽象的知识点穿插到某个具体的案例当中去，使学生更好的理解和接受这些知识点；又通过做具体的任务来掌握一些相关的知识、技能，进而可以更好的解决一些现实的问题。

翻转课堂设计：翻转课堂打破了原有的教学逻辑顺序，借助在线课程让学生在课前预习的基础上掌握了本节课的基础知识，在课堂中进行案例研讨、互动交流等方式来进一步了解和掌握本节课的知识点以及重点和难点。

10 多元立体的教学评价体系

多维度评价指标：将学生的学习过程全面纳入评价范围，包括线上学习表现（视频学习完成率、线上讨论参与度、在线测试成绩）、线下实践操作（实践项目完成质量、操作熟练度、故障解决能力）、以及期末考试成绩等多个维度。各维度赋予不同权重。

过程性评价与终结性评价结合：在教学过程中，通过线上平台实时跟踪学生学习进度，定期进行阶段性评价。课程结束后，结合期末考试成绩进行终结性评价，综合考量学生整个学期的学习情况，全面、准确地评估学生对 Hadoop 大数据技术的掌握和应

11 结语

职业本科 Hadoop 大数据技术课程建设，需要紧扣“应用导向、产教融合”的核心，在定位上凸显“技术应用 + 问题解决”的复合型能力的培养，与普通本科课程形成差异化特色。“三阶六模块”课程体系、虚拟化集群部署、企业级案例库等建设，为学生搭建了从理论到实践的桥梁。“双师协同”的教学模式及情境化教学方法，进一步提升了教学质量与学生的岗位适配能力。

未来，随着大数据技术的持续发展，职业本科 Hadoop 课程需不断动态调整教学内容与模式，强化与产业的深度对接，持续优化课程体系与实践平台，以更好地培养出符合市场需求的高层次技术技能型大数据人才，为数字经济发展提供坚实的人才支撑能力。

参考文献：

- [1] 陈洁,张文翔.大数据视角下计算机科学与技术专业建设探究[J].软件导刊,2016(10).
- [2] 岳昆,陈红梅,王丽珍.大数据时代本科数据库课程体系改革设想[J].计算机教育,2015(11).
- [3] 周晴红.Hadoop 大数据开发技术课程实践教学[J].办公自动化,2021(20).