《大数据开发技术应用实践 综合项目实践二》课程介绍

随着计算机存储能力的提升和复杂算法的发展，近年来的数据量成指数型增长，这些趋势使科学技术发展也日新月异;商业模式发生了颠覆式变化。大数据可以把人们从旧的价值观和发展观中解放出来，从全新的视角和角度理解世界的科技进步和复杂技术的涌现，变革人们工作、生活和思维的看法。大数据的应用十分广泛，通过对大规模数据的分析，利用数据整体性与涌现性、相关性与不确定性、多样性与非线性及并行性与实时性研究大数据在公共交通、公共安全、社会管理等领域的应用。大数据与云计算、物联网一起使得很多事情成为可能，必将会是新的经济增长点，大数据随着以数据科学为核心的计算机技术的迅猛发展 ，推动了社会科学与自然科学等跨科学研究的发展，因此全国各地的大数据研究具有深刻而广泛的意义。

2014年，大数据首次写入政府工作报告，大数据逐渐成为各级政府关注的热点；2015年9月，国务院发布《促进大数据发展的行动纲要》，大数据正式上升至国家战略层面，十九大报告提出要推动大数据与实体经济的深度融合；2019年10月，党的十九届四中全会首次将数据纳入生产要素范畴；2021年3月发布的“十四五”规划中，大数据标准体系的完善成为发展重点；2022年12月，《中共中央国务院关于构建数据基础制度更好发挥数据要素作用的意见》发布，以数据产权、流通交易、收益分配、安全治理为重点，系统搭建了数据基础制度体系的“四梁八柱”。现阶段，数据成为除土地、劳动力、资本、技术以外的又一项关键生产要素，数据的战略价值越来越重要。

随着互联网和物联网的发展，大数据处理和实时数据分析成为许多行业的关键需求，如金融、电子商务、医疗保健等。掌握这些技术能够帮助学生满足行业对大数据处理和实时分析的需求；在就业前景方面，大数据处理和实时数据分析技术已经成为许多企业招聘的重点。学生通过学习这门课程，将增强他们在大数据领域就业的竞争力；在技术发展方面：大数据领域的技术发展迅速，掌握这些技术将使学生在未来能够跟上潮流，为企业提供更好的数据处理和分析解决方案。在此背景下，我们开设了《大数据开发技术应用实践 综合项目实践二》这门课程，本课程是在学生具有linux基础的前提下学习，以项目制的教学方式，系统性的教授学生项目开发过程中的各个模块与功能，旨在锻炼学生的项目编写能力，业务能力。为大数据开发，数据分析工程师等职位打下坚实基础。

本系列课程的第一大特点是注重学生的实践能力培养，针对高校在实践教学中的痛点，首次提出“工单训练教学法”的概念，以企业真实需求为导向，使学生能紧紧围绕企业实际应用需求来学习技能，将学生需掌握的理论知识通过企业项目案例的形式与实际应用进行衔接，从而达到知行合一、以用促学的目的。本系列课程的第二大特点是以大数据技术应用为核心，紧紧围绕大数据技术应用闭环的流程进行教学。本系列课程涵盖企业云计算技术中的各个环节，符合企业大数据技术应用的真实场景，使学生从宏观上理解大数据技术在企业中的具体应用场景和应用方法。

本课程采用项目化教学的教学方式，”项目教学法”是指师生通过共同实施一个个完整的项目工作而进行的教学活动，其指导思想是将一个相对独立的项目任务交给学生完成，让学生在实施项目的过程中把握每一个环节的基本要求和重难点。

项目化教学极大地调动学生的积极性：在项目实施过程中，学生时常感受到成功的喜悦，这更能强化学生的学习积极性；培养学生的多种能力：项日教学要分小组完成，通过小组内及小组间的充分交流、讨论、决策等，提高学生的合作能力，强化学生的团队意识。提供了培养综合职业能力的机会，实现了与职业岗位的无缝对接。

在深化教师、教材、教法“三教”改革和“书证融通”“赛证融通”的人才培养实践过程中，本系列课程将根据师生的反馈意见和建议及时改进、完善，后期出版教材时，努力成为大数据时代的新型“研发、编写、使用、反馈”螺旋式上升的系列教材建设样板。

《大数据开发技术应用实践 综合项目实践二》是采用linux为基础，结合各种数据处理相关技术的项目制课程，共分为两个部分，指导篇和拓展篇，系统的讲解了数据分析与处理领域知识和技能，在基础理论上以精通，够用为原则，介绍与项目比较紧密的理论知识；在实践上，以新颖，实用为原则，介绍目前比较流行的新技术与新写法；在设计上结合目前比较典型的，应用比较广泛的数据算法与编程语法。本课程内容详实，资料丰富，配套有工单、PPT、视频和每个工单的输出产物，便于学生主动学习、理解和掌握；本课程具有较强的可用性和可操作性，从真实案例出发，进行一定的优化和完善，方便学生了解真实业务逻辑。

全书共2个项目，具体内容安排如下。

项目一和项目二分别为电商数据处理一期和二期，由老师带领学生从零到一逐步实现，巩固学生基础知识。

项目使用Linux、VMware和finalshell搭建基础项目环境。

项目一为电商数据处理一期购物网站数据是最原始，也是组成大数据重要的类型之一；购物网站上产生的商品数据揭示了用户的品名、价格和日期、商品参数，反馈商品特征等信息，对这些数据进行深度的挖掘和分析，有助于电商挖掘和预测用户行为，为商家决策提供数据支持。包括，环境搭建，数据抽取，数据清洗和数据分析。

项目二采用sqoop读取数据，再用pyspa抽取数据，使用脚本对数据进行清洗和分析。