

# 计算机类大数据专业课程体系构建的研究

——以武昌首义学院为例

◆程海英

(武昌首义学院 湖北武汉 430064)

**摘要:** 大数据时代,许多行业企业越来越重视大数据的影响和作用。面对社会对大数据人才的实际需求,结合大数据专业的专业定位和培养目标,通过对人才培养方案的深入梳理,从课程体系结构、课程体系构建重组等方面进行分析和研究,目的是培养适应现代大数据产业发展需求的高素质应用型人才。

**关键词:** 大数据专业;课程体系;课程建设;培养目标

近年来,国务院相继推出了《关于运用大数据加强对市场主体服务和监管的若干意见》、《促进大数据发展行动纲要》、《大数据产业发展规划(2016-2020年)》,标志着我国大数据战略部署和顶层设计正式确立。教育部近5年获批的大数据专业近500个,涉及500多所学校,开设大数据专业,对师资和课程设置有较高要求,其课程体系涉及数学、统计、信息管理以及计算机科学等多个领域。通过整合优质资源,培养符合市场需求的人才专业建设之本。

本文以武昌首义学院计算机类大数据课程体系为研究对象,探索大数据应用型人才培养的模式,构建理论和实践课程体系,深化校企合作,提升合作交流层次和水平,借助学校转型发展的契机,培养有特色的大数据人才。

## 1 大数据行业发展对软件人才的影响

大数据已经上升为国家基础性战略资源,是新世纪的数字钻石矿。十三五时期是我国新旧产业和发展驱动转换接续的关键时期,国内市场应用需求处于爆发期,我国大数据产业发展面临重要的发展机遇。培养出足够的、合格的大数据人才,对我国在未来掌握大数据的核心价值起着至关重要的作用。

IDC预测,到2020年,企业基于大数据计算分析平台的支出将突破5000亿美元,未来5年的复合增长率达到34.1%;而未来3-5年,中国需要180万数据人才,但目前只有约30万人。与此同时,我国高校中云计算、数据科学等专业仍处于初级阶段,每年培养的人才远远不能满足行业的需要。因此,开设大数据专业,加速人才培养是当务之急。

鉴于大数据身的技术前沿性,大数据与各行业结合的复杂性,当前的大数据专业人才培养仍面临诸多问题和困难:就业岗位和能力素质需求不明确;人才培养体系和课程体系有待优化;实践教学软硬件环境和案例缺失;师资匮乏等。与业界大数据研发和应用经验丰富、用人需求旺盛的企业进行专业共建联合培养,是高效率、高质量培养专业人才的有效途径。

目前,针对大数据人才需求的岗位及对应的能力要求,如表1所示。

表1 大数据岗位与对应能力要求

技术岗位	业务能力要求
大数据应用开发工程师	根据业务需求分析、设计并实现用于数据管理、处理、分析和展现的应用系统
	熟悉 Java, Python, Scala 等开发语言
	熟练掌握基于 Hadoop 的大数据平台开发技术,包括 Storm, Spark 等
	熟练掌握典型的关系型和非关系型数据库系统的使用和开发
	熟悉 Linux 系统及 Shell 编程
大数据分析师	扎实的计算机网络及软件工程基础
	根据业务需求对数据进行建模及分析,通过一定的算法,定性、定量分析结果形成分析报告熟悉 SQL、Python、R 等开发语言
	熟悉 SQL、Python、R 等开发语言
	熟练掌握 Excel、Tableau 等数据统计分析工具
	熟练使用 Python 或 R 进行数据分析
	扎实的数学、算法

大数据运维工程师	负责大数据硬件、网络、软件、平台及应用系统的搭建、管理、维护、监控等
	熟悉 Linux 系统管理和维护,熟悉 Shell
	熟悉关系型和非关系型数据库系统的配置、管理、维护及优化
	熟悉 Hadoop 等大数据平台的规划、安装、配置、管理、监控、维护等工作
数据清洗员	熟悉虚拟化系统的规划、配置和管理
	根据业务需求编写数据清洗脚本;按照模板输出清洗干净的数据;校验数据
	熟悉 Java、R 或 Python
	熟悉 MySQL、Oracle 或 SQL Server 关系型数据库;熟悉某种非结构化数据库
	熟练使用 Excel 等数据统计分析工具

## 2 明确大数据人才的培养目标和课程体系

### 2.1 培养目标

大数据专业涉及应用数学、统计学、计算机科学等多领域的交叉学科。大学本科四年期间,学生主要学习应用数学、统计学、计算机科学的基本理论和基本知识,打好坚实的数学基础,受到系统而扎实的计算机编程训练,具备较强的数据分析和信息处理能力,能在大数据科学与工程领域从事数据分析管理、系统设计开发、大数据处理应用、科学研究等方面的工作,具备综合运用所学知识分析和解决实际问题的能力。

大数据专业强调培养具有多学科交叉能力的大数据人才。该专业重点培养具有以下三方面素质的人才:一是理论性的,主要是对数据科学中模型的理解和运用;二是实践性的,主要是处理实际数据的能力。三是应用性的,主要是利用大数据的方法解决具体行业应用问题的能力。

### 2.2 培养要求及课程体系

大数据专业培养掌握大数据科学与技术相关的基本理论和基本知识,系统地掌握数据科学与工程专业知识,具备大数据应用系统设计及开发的能力,以及一定的科研工作能力,达到知识、能力与素质的协调发展。毕业生在知识、能力和素质等方面应达到如下具体要求:

(1) 知识要求:具有良好的自然科学基础,扎实的信息科学基础;具有较好的人文社会科学、管理科学知识;熟练掌握大数据科学与技术核心专业知识和应用技术,主要包括信息论与编码、计算思维和数据科学、数据采集技术、云计算与数据中心、机器学习与模式识别、数据挖掘、大型数据库技术、数据可视化技术、图像视频与非结构化数据、分布式海量存储系统、大数据与领域建模、数据安全等。

(2) 能力要求:具备从事大数据应用系统设计与实现的能力,特别在数据分析、数据管理、数据存储等方面,受到较系统的工程训练,能发现、分析和解决实际工程技术问题。具备良好的工程项目交流、表达、组织、管理、协调与沟通的能力。了解信息学科、计算机学科、数据科学的发展动态,并掌握相关文献检索方法,具有基本的专业资料分析与综合的能力,良好的文档与科学论文撰写能力。具有较强的创新意识,一定的创新创业能力。

具体应具备以下的专业能力:

① 计算思维与系统能力:计算思维主要包括形式化、模型化描述和抽象思维与逻辑思维能力。大数据项目无时无刻不在体现“计算”与“分布式系统”的重要性,应具备运用计算方法和模型去求解问题、设计系统等能力。

② 程序设计与实现能力:掌握结构化程序设计和面向对象程序设计的基本思想、方法和技巧,具备高级语言编程解决行业应用实际问题的能力。大数据项目作为信息科学技术发展的新阶段、新成就、新内容,对于程序设计与实现能力的要求已不仅仅局限于软件实现,而是数据采集和数据呈现的无缝对接、相互支

撑和相互补充。

③大数据技术体系:了解大数据的行业发展及应用,大数据核心技术体系以及 Hadoop 生态系统,包括核心子系统的学习: HDFS、MapReduce、HBase 等基本子系统。

④数据预处理、分析与应用能力:掌握数据预处理、数据清洗、融合、数据分析等关键技术。具备利用各种大数据行业工具,对行业海量数据和信息进行分析并处理,实现智能化的决策和控制的能力;具备运用运筹学、机器学习、数据挖掘、专家系统等技术,为大数据行业应用提供智能支撑平台的能力。

⑤数据挖掘及应用开发能力:大数据项目专业最大的特点是“面向具体项目应用领域”,掌握从数据输入,传输,存储、分析、挖掘、分析处理和数据呈现等完整的系统逻辑,从系统级的角度去理解与认识算法与程序,需掌握常用的大数据挖掘和统计分析的语言: Python、R 语言等。

(3)素质要求:毕业生身心健康;有良好的道德修养,尊重生命、遵纪守法、诚信友善、乐于奉献;有高尚的民族精神,积极弘扬传统文化,热爱祖国,崇尚集体主义精神;有坚定的理想信念,拥护中国特色社会主义,贯彻科学发展观、和谐社会理论和“四个全面”思想。大数据专业核心及特色课程如表 2 所示。

表 2 大数据核心及特色课程

核心课程	课程核心知识点
数据科学概论	作为专业概论课程,全面介绍云计算、大数据相关的概念、技术及应用,使学生建立对本专业相关知识、技术及发展前景的初步认识,作为后续深化课程的引导。
Python 程序设计	程序设计。介绍 Python 语法和语言特性;常用操作库;以 Flask 为核心的 Python 的数据库访问技术和 Web 开发技术等。
Python 数据处理技术	介绍文本格式、数据类型、字符编码等基础知识,以及如何提取和清洗关系型数据库、网页文件、文档中的数据;使用 Python 进行数据分析,包括 matplotlib、pandas、数据加工、数据处理、数据可视化、信号处理、数据库、文本分析、互操作性和性能优化等内容,使学生掌握数据处理和分析的基本方法、工具和思路。
数据仓库	介绍数据仓库的基本概念、原理,以及建立数据仓库的方法和过程。主要内容包括:关系模型对象(主题、实体、元素、属性、关联、范式等)、业务模型和主题域、建立和维护键、日期建模、层次建模、事务建模、模型维护等;还包括如何使用 ETL 进行数据集成。
非结构化数据库系统	介绍 NoSQL 数据库的基本概念、安装配置、管理维护、数据访问和开发等内容。课程重点讲述 HBase、MongoDB、Redis 等 NoSQL 数据库,并介绍它们在分布式环境下的使用和开发。使学生建立对分布式数据库的基本认识,掌握初级的分布式数据库应用系统设计和开发方法,为后续深度课程奠定理论和实践基础。
分布式计算框架基础	介绍 Hadoop 分布式计算系统的基本概念、安装配置、分布式编程模型(Map/Reduce),分布式文件系统(HDFS),以及相关的调度、监控和维护工具,使学生建立对分布式计算系统的基本认识,掌握初级的分布式应用设计和实现方法,为后续深度课程奠定理论和实践基础。
数据可视化	介绍各种数据可视化的理论基础,以及平台和开发工具的设计和使用,包括 Excel、报表服务、Chart.js、D3.js、Tableau 等。使学生能够将大数据处理的结果以高效、灵活和友好的方式呈现。
内存计算与实时计算	介绍高性能的分布式计算框架,包括 Storm 和 Spark,以作为 Hadoop 框架之外的更为强大的选择方案。
分布式计算框架组件技术	分布式计算框架组件技术。介绍 Hadoop 分布式计算平台上的主流组件,包括 Hive、Pig、Sqoop、Flume、Kafka、Zookeeper 等。使学生具备完整的基于 Hadoop 生态系统的大数据应用设计和实现能力。
云计算技术	介绍云计算基本概念,虚拟化原理; Hyperv-V、VMWare、Docker、等虚拟化系统的安装、配置、使用、监控,以及应用程序在这些系统上的部署和管

理; OpenStack 部署结构、计算服务 Nova、网络服务 Neutron、存储服务、计量服务、身份认证服务、镜像服务、仪表盘服务、编排服务等,使学生能够顺利构建私有云系统。

高并发系统设计与实现 用于解决大型网站性能问题,能够承受大数据、高并发。主要涉及技术有: nginx、tomcat、memcached、redis 缓存、负载均衡等高级开发技术。

### 3 与企业协同合作构建大数据实践教学环节

教育部为了推动产学合作育人,实现高校人才培养与企业发展的合作共赢,组织大型技术先进企业与高校共同实施 2017 年大数据方向教学内容和课程体系改革产学合作项目。为了促进民办高校的转型发展,武昌首义学院与中软国际教育集团签署了专业共建合作协议,共同打造“中软国际产学研基地”,探索一条集“学、研、训、产”四位一体的创新型软件人才培养模式。学校与合作企业共同建设大数据实践教学环节,建设目标全面落实“产、学、研、用”一体化的思想和模式,从教学、实践、科研和使用多方面注重专业人才培养。具体目标如下:

(1)搭建大数据运行的基础环境,提供大数据专业课件,满足高校教师大数据的学习环境;

(2)满足高校大数据相关课程的实验的要求,学生可以通过大数据实验室结合理论教学进行大数据相关实验;

(3)搭建与企业大数据的实验与科研环境,将理论课程中学到的数据挖掘算法运用到实际的数据分析过程中,提升学生的动手操作和项目实践能力,使得学生所学与企业项目人才需求,满足学生大数据实训、利用大数据实验环境进行创新创业和科研需要,为学生走向社会奠定扎实基础;

(4)建设大数据实验室,真正在产业、学校、科研及实际项目中相互配合,发挥优势,形成生产、学习、科学研究、实践运用的系统运作模式;教师可以在开放的平台环境下开展大数据科研工作,提升教师的科研创新能力,充分提高“研”的成效。

校企合作方中软国际凭借自身在云计算和大数据领域的多年实践经验,联合华为等顶级软硬件供应商,推出包括虚拟化桌面教学系统、云计算虚拟化实验系统、大数据和商务智能实验系统在内的高端实验室建设解决方案。在项目实训环节中的可用案例有:烟草生产经营数据 BI 分析及预测;金审工程联网审计数据统计分析;民航旅客大数据分析系统(个性化旅行产品推荐;社会化媒体营销;智慧旅游)等。

### 4 结语

大数据环境下,为了培养具有良好科学素养,具备一定的数学、软件、计算机等方面知识的基础上,较全面地掌握大数据方向的基本理论和技术,能够从事数据采集、预处理、数据挖掘、大数据应用分析及开发、数据可视化等大数据相关工作的应用型大数据人才。大数据专业需要将培养目标和要求落到实处,系统性地开展大数据课程建设,结合行业和时代的需求,优化课程结构,突出课程的时代性和实用性,才能使培养的学生成为大数据的复合型人才。

### 参考文献:

- [1]郭伟光.全面深化教育改革背景下的民办高校全面风险管控策略研究[J].求知导刊,2016(1):70-71.
- [2]汪祖柱.大数据背景下的信息管理与信息系统专业课程建设[J].宿州学院学报,2015(2):119-121.
- [3]周煜.大数据时代出版行业发展趋势分析[J].中国出版,2014.4(上):18-20.
- [4]熊伟,洪玫.大学本科软件工程专业建设的探索与实践[J].理工高教研究,2010,29(1):59-61.
- [5]张敬伟,古天龙,杨青.“新工科”建设与大数据人才需求双驱动的数据库实践课程教学探索[J].工业和信息化教育,2018,(4):61-64,71.
- [6]李学国.基于成果导向培养大数据技术创新型人才研究[J].电脑迷,2018,(1):68-69.

教研项目:2017年湖北省教学研究项目“基于协同视域下的计算机类大数据课程体系与专业建设研究”(编号:2017500)。

作者简介:程海英,女,湖北武汉,副教授,硕士,研究方向为计算机算法、软件工程、大数据技术方向。工作单位:武昌首义学院信息科学与工程学院。