



晋城职业技术学院
JINCHENG INSTITUTE OF TECHNOLOGY

信息工程系

大数据技术专业 核心课程标准

佐证材料

2021年11月

《云计算》核心课程标准

课程编码	0408005	课程类别	专业核心课
计划学时	68	课程类型	B类
适用专业	大数据技术与应用专业	课程性质	必修
开课学期	第二学期	学分	4
先行课程	计算机应用基础，面向对象程序设计-Java	开课单位	信息工程系
平行课程	SQL-server 数据库应用	考试类型	考试
后继课程	Python 语言，Hadoop, R 语言，数据清洗，云计算平台运维		

一、课程性质与定位

本课程是面向信息工程系大数据技术与应用专业学生的核心课程，是云计算的基本概念、发展现状、主要平台的部署及关键技术、虚拟化与容器技术、云计算的实用化、国内外云计算服务与大规模应用、环境云(envicloud.cn)和万物云(wanwuyun.com)典型行业应用介绍与剖析等内容，为后续的大数据实训课程打好坚实的基础。

二、课程设计与理念

(一) 以“工种（岗位）技能标准”设计课程

本课程具有很强的实践性，目标是使学生通过本课程所规定的全部教学内容的学习，能够对云计算的由来、概念、原理和实现技术有个基本的认识，熟悉云计算的主要产品和工具以及掌握其技术原理和应用方法，了解云计算的主要研究热点与应用领域，认清云计算的发展趋势和前景。

（二）理论教学与实践教学相结合，以实践教学为中心重点培养学生的职业能力

本课程采用理论与实操一体化教学，理论与实操紧密联系，环环相扣，将理论与实操对应起来，使理论真正起到指导实操的作用。传统教学重理论轻实践实训，改革后的本课程侧重实训实操教学，强调学生职业能力与动手能力的培养。理论教学围绕实操转，教学以学生职业能力为根本，以学生职业能力的培养引领教学全过程。

（三）采用项目教学与任务驱动教学法相结合的方式进行教学

本课程系统介绍了云计算的理论知识、主流技术和实战应用，包括大数据与云计算、Google 云计算原理与应用、Amazon 云计算 AWS、微软云计算 Windows Azure、Hadoop2.0: 主流开源云架构、Hadoop2.0 大家族、虚拟化技术、OpenStack 开源虚拟化平台、云计算数据中心以及云计算核心算法等内容，并深度剖析了国内云计算技术发展和云计算在互联网领域的展望。期望学生对云计算技术有比较深入的理解，能够紧跟云计算的发展前沿，从具体应用场景出发，利用所学的云计算知识解决行业应用问题。

（四）坚持校企合作开发课程的理念

本课程在设计与开发过程中始终坚持校企合作的理念，经常与南京云创大数据公司保持合作与联系，还经常深入到

大数据培训公司及其相关企业进行调查研究，实时掌握企业对大数据人才的需求与任职要求，与企业一起研讨教学内容，探究教学方法，与企业合作开发设计课程。

三、课程目标

（一）总体目标

通过本课程的学习，学生能够能够对云计算的由来、概念、原理和实现技术有个基本的认识，熟悉云计算的主要产品和工具以及掌握其技术原理和应用方法，了解云计算的主要研究热点与应用领域，认清云计算的发展趋势和前景。

（二）技能与知识目标

培养德、智、体、美全面发展的，能够较快适应生产、建设、管理、服务等一线岗位需要的，面向电信、零售、银行、金融、政府等部门的云平台管理、应用与分析的相关工作岗位，具有云计算理论知识，具有较高综合素质与良好职业素养，能从事云计算运维与大数据技术应用相关基础工作的发展型、复合型、创新型技术技能人才。

（三）能力与素质目标

使学生通过本课程所规定的全部教学内容的学习，能够对云计算的由来、概念、原理和实现技术有个基本的认识，熟悉云计算的主要产品和工具以及掌握其技术原理和应用方法，了解云计算的主要研究热点与应用领域，认清云计算的发展趋势和前景。

四、课程教学内容及时分配

序号	项目名称	学习任务	学习内容	学时
1	大数据与云计算	熟悉云计算的概念与意义、云计算发展现状、云计算实现机制、云计算的成本优势以及云计算与大数据的关系等内容。	云计算的概念与意义、云计算发展现状、云计算实现机制、云计算的成本优势以及云计算与大数据的关系等内容。	4
2	Google 云计算原理与应用	熟练掌握 Google 云计算技术，主要包括：Google 文件系统 GFS、分布式计算编程模型 MapReduce、分布式锁服务 Chubby、分布式结构化数据表 Bigtable、分布式存储系统 Megastore、分布式监控系统 Dapper、海量数据的交互式分析工具 Dremel、内存大数据分析系统 PowerDrill 和 Google 应用程序引擎等内容。	Google 云计算技术 Google 文件系统 GFS 分布式计算编程模型 MapReduce 分布式锁服务 Chubby 分布式结构化数据表 Bigtable 分布式存储系统 Megastore 分布式监控系统 Dapper 海量数据的交互式分析工具 Dremel 内存大数据分析系统 PowerDrill Google 应用程序引擎等内容。	8
3	Amazon 云计算 AWS	熟悉 AWS 的基础存储架构 Dynamo、弹性计算云 EC2、简单存储服务 S3、非关系型数据库服务 SimpleDB 和 DynamoDB、关系数据库服务 RDS、简单队列服务 SQS、内容推送服务 CloudFront 等各项主要服务，重点掌握其中所涉及的核心概念、基本原理和重要技术。	Dynamo 架构的主要技术、弹性计算云服务 EC2、简单存储服务 S3、简单队列服务 SQS。 Dynamo 架构的主要技术、EC2 的关键技术、S3 的安全措施、SQS 的消息取样和可见性超时值	8
4	微软云计算 Windows Azure	熟悉微软云计算平台、微软云操作系统 Windows Azure、微软云关系数据库 SQL Azure 和基础架构服务等组件，理解其中所涉及的核心概念、基本原理和重要技术。	Windows Azure 的计算服务和存储服务、SQL Azure 关键技术、AppFabric 关键技术、Windows Azure 服务平台。 SQL Azure 关键技术、AppFabric 关键技术。	4
5	Hadoop2.0: 主流开源云	了解 Hadoop2.0 的由来、相关项目及应用，熟练完成	Hadoop 2.0 部署、分布式文件系统 HDFS、分布式操作系	4

	架构	Hadoop2.0 环境部署，熟练掌握 Hadoop 2.0 体系架构下分布式文件系统 HDFS、分布式操作系统 Yarn 和分布式数据处理 MapReduce 的具体使用。本章主要内容都需要在实际操作中学习，因此实践教学内容是本章的重要部分。	统 Yarn、Hadoop 2.0 访问接口及编程接口。 HDFS 体系架构、Yarn 体系架构、Hadoop 2.0 编程接口。	
6	Hadoop2.0 大家族	熟练掌握 ZooKeeper、Hbase、Pig、Hive、Oozie、Flume 及 Mahout 等分布式组件的使用方法。本章主要内容都需要在实际操作中学习，因此实践教学内容是本章的重要部分。	ZooKeeper、Hbase、Hive、Mahout。	4
7	虚拟化技术	熟练掌握服务器虚拟化、存储虚拟化、网络虚拟化和桌面虚拟化的原理与技术，从这四个方了解虚拟化技术在云计算中的地位和应用，并通过 VMware 公司的部分产品实例，理解虚拟化的一些实现方法。	服务器虚拟化、存储虚拟化、网络虚拟化。	4
8	OpenStack 开源虚拟化平台	了解 OpenStack 的定义与主要服务，着重掌握计算服务 Nova、对象存储服务 Swift、镜像服务 Glance 的基本原理和应用场景。	计算服务 Nova、对象存储服务 Swift、镜像服务 Glance。 RabbitMQ。	4
9	云计算数据中心	了解云数据中心的主要特征，熟悉数据中心网络部署方式，掌握自动化管理功能和容灾备份的系统原理，而对于绿色节能技术只需了解即可。	数据中心网络体系结构、容灾备份的关键技术。	4
10	云计算核心算法	Paxos 算法解决分布式系统中信息一致性问题，DHT 算法解决分布式网络的应用层选路问题，Gossip 协议解决分布式环境下信息高效分发问题，为便于读者更透彻地理解云计算技术，本章节将对这三种云计算核心	Paxos 算法、DHT 算法、Gossip 协议。Paxos 算法原理、DHT 算法原理与实现、Gossip 协议通信方式。	8

		算法进行详细介绍,要求学生熟练掌握其原理与实现方法。		
11	中国云计算技术	本章节通过梳理和分析国内云计算技术的发展概况,总结我国云计算发展形势,要求学生熟悉国产云存储技术、大数据库技术、云视频监控技术以及云服务等关键技术、产品与平台,从而能够更加直观地了解国内云计算发展态势。	淘宝分布式文件系统 TF、云创 cStor 分布式文件系统、阿里巴巴 OceanBase、云创数据立方 DataCube、cVideo 云视频监控系统、阿里云服务、万物云服务。	8
12	总结与展望	了解主流商业云计算解决方案的优劣对比,了解主流开源云计算系统的技术差异,了解云计算在互联网和信息社会发展中所处位置,能够方便学生更好地掌握本书的主体内容,理解云计算发展方向。	主流商业云计算解决方案比较、主流开源云计算系统比较。	4

五、考核评定办法

改革传统的学生评价手段和方法,采用阶段性评价、过程性评价与目标评价相结合、项目评价、理论与实践一体化评价模式。关注评价的多元性,将课堂提问、学生作业、平时测验、项目考核、技能目标考核作为平时成绩,占总成绩的 60%,期末作品汇报占总成绩的 40%。

评价中应注意学生动手实践中分析问题、解决问题能力的考核,对在学习和应用上有创新的学生应予特别鼓励,全面综合评价学生能力。

六、教学建议

（一）教学条件

本课程中实践内容居多，要求课堂教学应在实训室上课，教师边讲解，学生边实践。

（二）师资要求

本课程需要教师具有一定的基础理论知识、较广泛的专业知识以及相关学科的基本知识，以便解决教学、科研、实践工作中不断涌现出的新的问题；要求教师具备较高的教学能力，良好的知识传导能力和系统的教学设计能力；还要求教师具有同行业的实践经验或经历，与行业系统密切联系的沟通渠道，具备善于实践并指导学生实践的能力。

（三）教学方法

本课程遵循“学其所用，用其所学”的原则，注重培养学生的动手操作能力。采用以项目为导向，以工作任务为驱动，以典型案例为依托的方式，提高学生学习兴趣，激发学生的成就动机，积极引导提升职业素养，提高职业道德。课堂教学关键是上机实践操作，教学与实践操作相结合，教师操作示范和学生上机操作，学生提问与教师解答、指导有机结合，让学生在“教”与“学”过程中学有所获。

在教学中采用项目教学与任务驱动相结合的方式，学生既可以接触到真实的项目，又可以将项目分成小的学习任务应用在教学中；除此之外还采取理论与实践相结合的教学方

法，自主学习与小组合作学习相结合的教学方法。

（四）教学资源的开发与利用

本课程是我系与南京云创大数据科技股份有限公司合作开发的课程，所以在教学中可以将公司中项目分解为多个任务，选出典型案例作为教学的资源。

（五）评价标准

教学评价不再像以往一样采取简单的考试，而采用阶段性评价和总结性评价相结合的方式，学生每次实验的作品和最终的作品都将作为最终评价学生学习的依据；同时在教学评价中不仅仅是教师给予学生评价，也会聘请南京云创大数据科技股份有限公司的相关人员参与到评价中，从企业的角度给学生指导以及评价建议。

七、推荐选用教材

《云计算》 由刘鹏教授作为主编率领团队编写，电子工业出版社。

八、参考文献（含课程网站）

序号	书名	编者	出版社	版本
1	大数据	刘鹏	电子工业出版社	1
2	大数据库	刘鹏	电子工业出版社	1
3	大数据实验手册	刘鹏	电子工业出版社	1
4	实战 Hadoop2.0 从云计算到大数据	叶晓江、刘鹏	电子工业出版社	2

《大数据导论》核心课程标准

课程编码	0408006	课程类别	专业核心课
计划学时	68	课程类型	B类
适用专业	大数据技术与应用专业	课程性质	必修
开课学期	第一学期	学分	4
先行课程	计算机基础	开课单位	信息工程系
平行课程	云计算、网络基础	考核类型	考试
后继课程	Hadoop, 数据清洗		

一、课程性质与定位

本课程是面向信息工程系大数据技术与应用专业学生的核心课程，是了解大数据技术框架和生态系统，具备大数据相关编程技术框架基础知识、程序设计能力、了解非大数据数据库理论知识、多数据源整合、掌握大数据进行预处理、检验和清洗学习的前提基础理论课程。

二、课程设计与理念

《大数据导论》是了解和学习大数据的基础条件，通过课程了解大数据基本概念，大数据的架构，大数据的采集方式和预处理，常用的 ETL 工具，简单熟悉数据仓库的构建模式，大数据的存储，数据挖掘的方法，以及大数据的可视化技术，从而更好的将大数据技术应用在各行业领域，更深入地开展大数据技术的应用研究。从基础开始，通过理论与实际案例相结合，帮助学生由浅入深进行学习，逐步清理大数据的核心技术和发展趋势。

三、课程目标

(一) 总体目标

培养能够较快适应生产、建设、管理、服务等一线岗位需要的，面向电信、零售、银行、金融、政府等部门的大数据技术应用与分析的相关工作岗位，具有大数据技术应用与云计算理论基础知识，掌握大数据存储、清洗、管理、建模和分析的基本技能，了解大数据技术应用框架与其生态系统，具有较高综合素质与良好职业素养的发展型、复合型、创新型技术技能人才。

(二) 技能与知识目标

具备大数据应用理论基础知识，了解大数据技术框架和生态系统，具备大数据基础技术框架知识，了解熟悉大数据应用、大数据架构、大数据采集与预处理、大数据存储、大数据分析、大数据可视化等概念。

(三) 能力与素质目标

- 1.对大数据基础理论、架构有深刻理解；
- 2.熟悉大数据集群构建基础理论；
- 3.熟悉主流大数据应用的架构体系以及各种中间件技术。

四、课程教学内容及学时分配

序号	项目名称	学习任务	学习内容	学时
1	大数据的概念及其应用	学习掌握大数据的基本概念、来源、特征意义、表现形态与应用场景	大数据的基本概念 大数据来源 大数据的特征意义 大数据的表现形态 大数据的应用场景 企业中应用 物联网中应用 在线社交网络中的应用等	8
2	大数据的架构	学习掌握大数据架构基础知识,包括云计算基本概念特点与应用,大数据架构解决方案、Hadoop 体系架构	云计算的概念 云计算的特点 云计算的服务方式 云计算的应用 大数据的分类 数据类型 大数据解决方案 Hadoop 概述 Hadoop 核心组件 上机实训	8
3	大数据采集及预处理	学习掌握大数据采集、预处理与常用 ETL 工具	大数据采集的概念 采集工具 采集方法 数据清洗 数据集成 数据转换 数据归约 ETL 工具概念 常用 ETL 工具及比较	8
4	大数据的存储	学习掌握大数据存储相关问题、存储方式、数据仓库理论	大数据存储面临的挑战问题 大数据分布式存储系统 NoSQL 数据库 云存储 数据仓库的组成、构建步骤 数据集市	8
5	大数据的分析	学习掌握大数据分析的概念与分类、数据分析方法、数据挖掘技术	数据分析的概念和作用 数据分析的类型 数据分析的方法 数据来源问题 数据分析活动步骤 分析数据练习 数据挖掘概念 数据挖掘常用算法	10

			数据挖掘分类方法 数据挖掘聚类方法 数据挖掘关联规则 数据挖掘工具 数据挖掘算法应用	
6	大数据可视化	学习掌握大数据可视化基础、方法、软件与工具	数据可视化的特征 数据可视化的作用 数据可视化的流程 数据文本可视化 网络图可视化 多维数据可视化 数据可视化工具：Echarts 等	10
7	大数据的商业应用	学习了解国外大数据应用经典案例、国内大数据应用经典案例	国外大数据案例中数据资源重要性、相关性学习 数据商业价值 国外医疗行业大数据应用 国内大数据在智慧城市、保险、医疗、交通、环境、农业、零售、舆情、物流、房地产、地震、互联网等行业的应用	10

五、考核评定办法

本课程的考核评价手段和方法，采用阶段性、过程性项目评价、理论与实践一体化评价模式。关注评价的多元性，将课堂提问、学生作业、平时测验、项目考核、技能考核作为平时成绩，占总成绩的 60%，期末书面测试占总成绩的 40%。

六、教学建议

（一）教学条件

采用多媒体教学的方法，有较为丰富图片及影像学习资料供学生参考；图片来源主要是生产、生活中的应用实例，其结构合理、内容新颖、案例经典具有理论性、示范性、可读性、实用性相结合的特点。

（二）师资要求

本课程需要教师具有一定的基础理论知识、较广泛的专业知识以及相关学科的基本知识，以便解决教学、科研、实践工作中不断涌现出的新的问题；要求教师具备较高的教学能力，良好的知识传导能力和系统的教学设计能力；还要求教师具有同行业的实践经验或经历，与行业系统密切联系的沟通渠道，具备善于实践并指导学生实践的能力。

（三）教学方法

本课程注重培养学生的动手操作能力。通过通俗易懂、实用性的经典案例来分析讲解平面构成的概念、理论、类型、规律，采用基础理念教学与实践相互融合的方法，启发学生的设计思维的创造性，注重培养学生设计制作与表现的动手能力，使学生尽快熟悉平面构成设计制作与应用操作。

（四）教学资源的开发与利用

本课程是我系与南京云创大数据科技股份有限公司合作开发的课程，所以在教学中可以将公司中的经典案例及最新的设计思路用于实际的教学过程中，使学生学习理论知识的同时能受到最新、最美设计作品的熏陶，提高学生的审美能力。

（五）评价标准

教学评价不再像以往一样采取简单的考试，而采用阶段性评价和总结性评价相结合的方式，学生每次任务的作品和

最终的作品都将作为最终评价学生学习的依据。

七、推荐选用教材

《大数据导论》 刘鹏教授作为丛书总主编率领团队编写，清华大学出版社出版。

八、参考文献（含课程网站）

序号	书名	编者	出版社	版本
1	大数据	刘鹏	电子工业出版社	1
2	大数据库	刘鹏	电子工业出版社	1
3	大数据实验手册	刘鹏	电子工业出版社	1
4	实战 Hadoop2.0 从云计算到大数据	叶晓江、刘鹏	电子工业出版社	2

《Python 语言》核心课程标准

课程编码	0408008	课程类别	专业核心课
计划学时	68	课程类型	B 类
适用专业	大数据技术与应用专业	课程性质	必修
开课学期	第三学期	学分	4
先行课程	大数据导论 面向对象程序设计-Java Linux 操作系统 SQL-server 数据库应用	开课单位	信息工程系
平行课程	NoSQL 数据库 Hadoop		
后继课程	R 语言 数据清洗 云计算平台运维 大数据系统运维		

一、课程性质与定位

本课程是面向信息工程系大数据技术与应用专业学生的核心课程，通过 Python 语言编程教学，使学生学会该门编程语言，并能在大数据行业相关实践中提高编程技术素养及能力。

本课程以搭建 Python 开发基础平台为起点，重点阐述 Python 语言的基础知识和三个典型的项目实战案例。以理论引导、案例驱动、上机实战为理念打造 Python 语言学习的新模式。

二、课程设计与理念

本课程具体内容分为两大部分：第一部分以 Python 编程语言基础知识普及为主，分别介绍了 Python3 概述、基本语法、流程控制、Python 组合数据类型、字符串与正则式、函数、模块、类和对象、异常处理、文件操作；第二部分：以项目实战为核心，以学以致用为导向，以切近生活的案例为

依托，分别介绍 Python 爬虫项目实战、Python 数据可视化项目实战、Python 数据分析项目实战。

三、课程目标

（一）总体目标

培养德、智、体、美全面发展的，能够较快适应生产、建设、管理、服务等一线岗位需要的，面向电信、零售、银行、金融、政府等部门的掌握 Python 编程技术和使用，具有较高综合素质与良好职业素养，能从事云计算运维与大数据技术应用相关开发工作的发展型、复合型、创新型技术技能人才。

（二）知识目标

了解熟悉 Python 编程语言基础知识，掌握 Python 基本语法、流程控制、Python 组合数据类型、字符串与正则式、函数、模块、类和对象、异常处理、文件操作；学习了解 Python 项目实战，学习模仿练习 Python 爬虫、数据可视化、数据分析项目实战编程模式，掌握编程实战技能。

（三）能力与素质目标

- 1、熟悉 Python 基础知识。
- 2、熟练掌握 Python 语言基础编程。
- 3、熟练使用编程环境开发工具。
- 4、有良好的代码习惯，结构清晰，命名规范，逻辑性强。

四、课程教学内容及时分配

序号	项目名称	学习任务	学习内容	学时
1	Python 简介	学习掌握 Python 基本概述	Python 语言背景简介	1
2	Python 环境构建	构建 Python 编程环境	在不同操作系统中构建 Python 编程环境	1
3	Python 程序简析	基础入门编程练习	第一个程序 Hello World! 10	2
4	Python 简单函数	输入输出函数	print() 函数 input() 函数	2
5	Python 基本语法	变量, 表达式	用常量和变量 用运算符和表达式 type() 函数的使用 help() 函数的使用	6
6	Python 流程控制	条件流程控制 循环流程控制	使用条件语句 使用 for 语句 使用 while 语句 使用 break 语句 使用 continue 语句	8
7	Python 组合数据类型	组合数据类型	列表的使用 元组的使用 字典的使用 集合的使用	8
8	Python 字符串与正则表达式	字符串与正则表达式	使用字符串处理函数 正则表达式的使用 使用 re 模块	6
9	Python 函数	函数使用	声明和调用函数 在调试窗口中查看变量的值 使用函数参数和返回值 使用闭包和递归函数 使用 Python 的内置函数	6
10	Python 模块	模块命名、安装、调试	使用 datetime 模块 使用 sys 模块 使用与数学有关的模块 自定义和使用模块	6
11	Python 类和对象	类的定义封装	声明类 类的继承和多态 复制对象	6

12	Python 异常	异常处理	利用 try-except 处理除数为零的异常 自定义异常的使用 关键字的使用 内置异常处理语句的使用	4
13	Python 文件操作	文件操作	文件操作 目录操作 I/O 函数的使用	4
14	项目实战	实战练习	爬虫程序 数据可视化 数据分析	8

五、考核评定办法

本课程的考核形式分为讲授部分考核和实践环节考核，两者成绩之和为本课程最终成绩。讲授部分成绩包括课堂讨论、回答问题、考勤、结课考试占比例 60% 实践环节考核的内容以课程训练作业为主，重点考核学生对知识与技能的掌握情况和创新能力。实践环节成绩占本课程总成绩 40%。

六、教学建议

（一）教学条件

多媒体教室，有较为丰富图片及影像学习参考资料；有相对接的校外实训企业。

（二）师资要求

主讲教师应具备本科或硕士研究生学历，具有相关从业背景，有丰富的行业经验，了解前沿的技术发展趋势及理论知识，掌握一定的教学方法与教学艺术，能综合运用各种教学方法开发设计课程。

（三）教学方法

深入钻研教材，精心设计教案，善于从多渠道获取反馈信息。及时从多方面调控课堂教学；从反馈的信息中，发现教学现状与教学目标之间的差距，及时、全面、有效地调控；通过多种途径培养学生的自控能力。教学中，教师应引导学生不断解剖自己，及时找出知识与学习方法上的缺陷，加强学生的自我适应、自我调控能力。以学生为起点，以学生的发展为目标，即一切教学活动都是为了学生，教师对学生的起点了解得愈清楚，教学也就愈切合学生的需要。优化教学过程，按照教学的规律和原则要求来制定和选择一个最好的教学方案，用尽可能少的教学时间，获得最佳的教学效果。

学生须做到课前阅读教材，课后参阅具有针对性的设计作品，以加深对基本原理的理解，同时准备课堂提问。

通过对本课程的学习,让学生理解相关的编程算法设计,掌握编程语言在运用中的一些相关规律及技法,培养学生掌握科学的编程设计理论和观点,从而为今后的专业课程打下良好的理论及实践基础。

- 1.配以大量的实战项目详细讲解编程设计内容；
- 2.将所学的内容分成多个课题训练，这些课题从不同的角度训练学生编程设计的能力。

七、推荐选用教材

《Python 语言》由刘鹏教授作为丛书总主编率领团队编

写，清华大学出版社出版。

八、参考文献

序号	书名	编者	出版社	版本
1	大数据	刘鹏	电子工业出版社	1
2	大数据库	刘鹏	电子工业出版社	1
3	大数据实验手册	刘鹏	电子工业出版社	1
4	实战 Hadoop2.0 从云计算到大数据	叶晓江、刘鹏	电子工业出版社	2

《大数据系统运维》核心课程标准

课程编码	0408016	课程类别	专业核心课
计划学时	68	课程类型	B类
适用专业	大数据技术与应用	课程性质	必修
开课学期	第四学期	学分	4
先行课程	大数据导论 Python语言 云计算	开课单位	信息工程系
平行课程	数据清洗 数据标注工程	考试类型	考试
后继课程	顶岗实习		

一、课程性质与定位

本课程是面向信息工程系大数据技术与应用专业学生的核心课程，本课程针对大数据系统运维类岗位，以大数据相关系统运行维护过程中的各类运维技能为核心教学内容，实训为主，理实合一。

二、课程设计与理念

《大数据系统运维》将详细讲述大数据系统的运行维护过程中的各个主要阶段及其任务，包括配置管理、系统管理及日常巡检、故障管理、性能管理、安全管理、高可用性管理、应用变更管理以及升级管理，并介绍了相应的操作方法。通过以上内容的学习与上机实践，学生将了解大数据系统运维的要素，并掌握其操作要领。

三、课程目标

(一) 总体目标

培养德、智、体、美全面发展的，能够较快适应生产、建设、管理、服务等一线岗位需要的，面向电信、零售、银行、金融、政府等部门的大数据平台管理、掌握各类大数据平台运维基本技能，具有较高综合素质与良好职业素养，能从事云计算与大数据系统运维技术应用相关工作的应用型、复合型、创新型技术技能人才。

(二) 具体目标

1. 知识目标

学习并上机实践掌握包括大数据平台的配置管理、系统管理及日常巡检、故障管理、性能管理、安全管理、高可用性管理、应用变更管理以及升级管理能力。

2. 能力与素质目标

- 1、计算机网络管理能力。
- 2、Linux 操作系统管理能力。
- 3、虚拟化管理能力。
- 4、分布式存储管理能力。
- 5、云安全管理能力。

四、课程教学内容及学时分配

序号	项目名称	学习任务	学习内容	学时
1	配置管理	学习掌握大数据平台配置管理内容、方法、工具及上机练习	配置管理术语与定义 软件配置应用 硬件配置应用 配置流程 配置自动化 CMDB 数据库 自动配置工具 其他常用运维工具 上机练习	6
2	项目系统管理及日常巡检	学习掌握大数据项目系统建设、对象管理、内容管理、系统管理、管理规范、巡检管理及上机练习	系统建设技术方案、部署实施、测试验收 对象管理软件、硬件、数据及供应商管理 事件、问题、配置、变更、发布、知识、日志、备份管理 资产管理、监控管理、流程管理、外包管理 系统管理的标准、制度、规范 巡检内容、方式、流程 上机练习	8
3	故障管理	学习掌握集群结构、故障的发现、处理、后期管理与上机练习	集群结构 故障发现与影响分析 故障诊断处理 故障排除处理 故障知识库建立于更新 故障预防 上机练习	6
4	性能管理	学习掌握性能分析、性能监控工具与性能优化技术, 进行上机练习	性能分析因子 性能分析指标 GUI 监控工具 集群 CLI 工具 其他常用监控工具 Hadoop 集群配置 Hadoop 集群性能优化 作业优化	8

5	安全管理	学习掌握运维安全管理概念、资产安全管理、应用安全管理、安全威胁管理、安全措施等	运维安全概述 资产设施安全、设备安全技术安全 数据安全 人为失误管理 外部攻击管理 信息泄密管理 灾害管理 安全制度规范 安全防范措施	8
6	高可用性管理	学习掌握大数据系统运维中高可用性技术、业务连续性管理技术及上机练习	高可用性技术概念 高可用技术架构 容灾、监控与故障转移 灾备系统 应急预案设计 日常演练设计	2
7	应用变更管理	学习掌握变更管理基本理论、变更管理流程、变更配置管理与练习	变更目标管理 变更范围管理 变更种类 管理原则 变更组织架构、管理策略 变更流程控制管理 变更配置管理	8
8	升级管理	学习掌握 Hadoop、Spark、HiveSQL、ZooKeeper 升级管理及上机练习	Hadoop 升级风险管理 HDFS 数据与元数据升级 YARN 升级配置 Spark 特性、生态与升级管理 HiveSQL 体系结构、配置与升级 ZooKeeper 打击、集群模式配置	8
9	服务资源管理	学习掌握大数据系统平台业务能力设计管理、服务能力设计管理、服务资源整合管理	大数据平台业务需求评估 需求趋势预测 人员能力动态管理 服务成本动态管理 技术与工具管理 不同角色的责权划分 用户、供应商、厂商的典型协作方式	8

五、考核标准与成绩评定方法

改革传统的学生评价手段和方法，采用阶段性评价、过程性评价与目标评价相结合、项目评价、理论与实践一体化评价模式。关注评价的多元性，将考勤、作业占总成绩的10%，课堂汇报、小组合作、平时测验、作为平时成绩，占总成绩的60%，期末作品汇报占总成绩的30%。

评价中应注意学生动手实践中分析问题、解决问题能力的考核，对在学习和应用上有创新的学生应予特别鼓励，全面综合评价学生能力。

六、教学建议

（一）教学条件

本课程中实践内容居多，要求课堂教学应在实训室上课，可以小组汇报，也可以实训操作。

（二）师资要求

本课程需要教师具有一定的基础理论知识、较广泛的专业知识以及相关学科的基本知识，以便解决教学、科研、实践工作中不断涌现出的新的问题；要求教师具备较高的教学能力，良好的知识传导能力和系统的教学设计能力；还要求教师具有同行业的实践经验或经历，与行业系统密切联系的沟通渠道，具备善于实践并指导学生实践的能力。

（三）教学方法

本课程遵循“学其所用，用其所学”的原则，注重培养

学生的动手操作能力。采用以项目为导向，以工作任务为驱动，以典型案例为依托的方式，提高学生学习兴趣，激发学生的成就动机，积极引导提升职业素养，提高职业道德。课堂教学关键是上机实践操作，教学与实践操作相结合，教师操作示范和学生上机操作，学生提问与教师解答、指导有机结合，让学生在“教”与“学”过程中学有所获。

在教学中采用项目教学与任务驱动相结合的方式，学生既可以接触到真实的项目，又可以将项目分成小的学习任务应用在教学中；除此之外还采取理论与实践相结合的教学方法，自主学习与小组合作学习相结合的教学方法。

（四）评价标准

教学评价不再像以往一样采取简单的考试，而采用阶段性评价和总结性评价相结合的方式，每次课堂在网络教学平台的评价作为阶段性评价，学生最终作品由教师，学生，企业共同评价，作为总结性评价，在教学评价中不仅仅是教师给予学生评价，也会聘请新浪印业有限公司和景潮印业有限公司相关人员参与到评价中，从企业的角度给学生指导以及评价建议。

七、选用教材标及推荐教材和参考用书（或课程网站）

（一）推荐教材

《大数据系统运维》 刘鹏教授为丛书总主编率领团队编写，清华大学出版社出版。

(二) 参考文献 (含课程网站)

序号	书名	编者	出版社	版本
1	大数据	刘鹏	电子工业出版社	1
2	大数据库	刘鹏	电子工业出版社	1
3	大数据实验手册	刘鹏	电子工业出版社	1
4	实战 Hadoop2.0 从云计算到大数据	叶晓江、刘鹏	电子工业出版社	2

《数据清洗》核心课程标准

课程编码	0408011	课程类别	专业核心课
计划学时	68	课程类型	B类
适用专业	大数据技术与应用	课程性质	必修
开课学期	第四学期	学分	4
先行课程	大数据导论 Python 语言 云计算	开课单位	信息工程系
平行课程	大数据系统运维 数据标注工程	考试类型	考试
后继课程	顶岗实习		

一、课程性质与定位

本课程是面向信息工程系大数据技术与应用专业学生的核心课程，是培养掌握大数据清洗的一般理论与实际操作应用模式技术的专业实践性课程。

二、课程设计与理念

本课程系统地讲解了数据清洗理论和实际应用，共分为8章：第1章主要介绍数据清洗的概念、任务和流程，数据标准化概念及数据仓库技术等；第2章主要介绍 Windows 和类 UNIX 操作系统下的数据常规格式、数据编码及数据类型转换等；第3章介绍 ETL 概念、数据清洗的技术路线、ETL 工具及 ETL 子系统等；第4章介绍了 Excel、Kettle、OpenRefine、DataWrangler 和 Hawk 的安装及使用等；第5章介绍 Kettle 下文本文件抽取、Web 数据抽取、数据库数据抽取及增量数据抽取等；第6章介绍数据清洗步骤、数据检验和数据错误处理，数据质量评估及数据加载；第7章介绍网页结构，利

用网络爬虫技术进行数据采集，利用 JavaScript 技术进行行为日志数据采集等；第 8 章介绍 RDBMS 的数据清洗方法和数据脱敏处理技术等。

三、课程目标

（一）总体目标

培养德、智、体、美全面发展的，能够较快适应生产、建设、管理、服务等一线岗位需要的，面向电信、零售、银行、金融、政府等部门的掌握大数据存储、清洗、管理、建模和分析的基本技能，具有较高综合素质与良好职业素养的发展型、复合型、创新型技术技能人才。

（二）技能与知识目标

学习掌握数据清洗的概念、任务和流程，数据标准化概念及数据仓库技术；了解 Windows 和类 UNIX 操作系统下的数据常规格式、数据编码及数据类型转换等；掌握 ETL 概念、数据清洗的技术路线、ETL 工具及 ETL 子系统的使用，Excel、Kettle、OpenRefine、DataWrangler 和 Hawk 的安装及使用，Kettle 下文本文件抽取、Web 数据抽取、数据库数据抽取及增量数据抽取，数据清洗步骤、数据检验和数据错误处理，数据质量评估及数据加载，网页结构，利用网络爬虫技术进行数据采集，利用 JavaScript 技术进行行为日志数据采集，RDBMS 的数据清洗方法和数据脱敏处理技术等。

（三）能力与素质目标

具备大数据应用理论基础知识，了解大数据技术框架和生态系统，熟悉 Hadoop 技术框架操作，具备程序设计能力，掌握数据清洗操作，熟悉 ETL 工具的使用，掌握使用 ETL 工具进行定制化清洗技术，能够进行常规的统计分析、报表分析，能够提出问题，并能制定详细实施计划。

四、课程教学内容及学时分配

序号	项目名称	学习任务	学习内容	学时
1	数据清洗概述	数据清洗简介 数据标准化 数据仓库简介	数据科学过程、定义、任务、 流程、环境、实例说明 数据标准化概念、常用方法 数据仓库定义、组成要素、 分类、相关技术、常用工具 简介 上机练习	8
2	数据格式与编码	文件文本格式 数据编码 数据转换	常见文本格式 xls 及xlsx 文件格式 JSON 文本格式 HTML 和 XML 文本格式 数据类型、类型间转换 字符编码 空值和乱码 电子表格转换 RDBMS 数据转换	8
3	数据清洗基本技术方法	ETL 入门 技术路线 ETL 工具 ETL 子系统	ETL 解决方案、基本构成、 技术选型 文本清洗路线 RDBMS 清洗路线 Web 内容清洗路线 ETL 功能、开源 ETL 工具 抽取、清洗和更正数据、数 据发布、管理 ETL	10
4	数据清洗常用工具及基本操作	Microsoft Excel 数据 清洗基本操作 Kettle 简介及基本操 作 OpenRefine 简介及基	Excel 数据清洗概述 Kettle 软件概述、基本操 作、实例操作 OpenRefine 软件概述、基 本操作、数据清洗实例操作	8

		本操作 DataWrangler 简介及 基本操作 Hawk 简介及基本操作 上机练习与实训	DataWrangler 软件概述、 基本操作、数据清洗实例操 作 Hawk 软件概述、基本操作、 数据清洗实例操作	
5	数据抽取	文本文件抽取 Web 数据抽取 数据库数据抽取 上机练习与实训	制表符文本抽取 CSV 文件抽取 HTML 文件抽取 JSON 数据抽取 XML 数据抽取 数据导入导出 ETL 工具抽取 SQL 到 NoSQL 抽取	10
6	数据转换与加载	数据清洗转换 数据质量评估 数据加载 上机练习与实训	数据清洗 数据检验 错误处理 数据评估指标 审计数据 数据加载的概念 数据加载的方式 批量数据加载 数据加载异常处理	8
7	采集 Web 数据实 例	网页结构 网络爬虫 行为日志采集 上机练习与实训	DOM 模型 正则表达式 网络爬虫简介 网络爬虫异常处理 用户实时行为数据采集 用户实时行为数据分析	8
8	清洗 RDBMS 数据 实例	8.1 准备工作 数据库数据清洗 数据脱敏处理	准备待清洗的数据集 搭建操作环境 数据导入 MySQL 缺失值清洗 格式内容清洗 逻辑错误清洗 非需求数据清洗	8

五、考核评定办法

改革传统的学生评价手段和方法，采用阶段性评价、过

程性评价与目标评价相结合、项目评价、理论与实践一体化评价模式。关注评价的多元性，将课堂提问、学生作业、平时测验、项目考核、技能目标考核作为平时成绩，占总成绩的 60%，期末作品汇报占总成绩的 40%。

评价中应注意学生动手实践中分析问题、解决问题能力的考核，对在学习和应用上有创新的学生应予特别鼓励，全面综合评价学生能力。

六、教学建议

（一）教学条件

本课程实践内容居多，要求课堂教学应在实训室上课，教师边讲解，学生边实践。

（二）师资要求

本课程需要教师具有一定的基础理论知识、较广泛的专业知识以及相关学科的基本知识，以便解决教学、科研、实践工作中不断涌现出的新的问题；要求教师具备较高的教学能力，良好的知识传导能力和系统的教学设计能力；还要求教师具有同行业的实践经验或经历，与行业系统密切联系的沟通渠道，具备善于实践并指导学生实践的能力。

（三）教学方法

本课程遵循“学其所用，用其所学”的原则，注重培养学生的动手操作能力。采用以项目为导向，以工作任务为驱动，以典型案例为依托的方式，提高学生学习兴趣，激发学

生的成就动机，积极引导学生提升职业素养，提高职业道德。课堂教学关键是上机实践操作，教学与实践操作相结合，教师操作示范和学生上机操作，学生提问与教师解答、指导有机结合，让学生在“教”与“学”过程中学有所获。

在教学中采用项目教学与任务驱动相结合的方式，学生既可以接触到真实的项目，又可以将项目分成小的学习任务应用在教学中；除此之外还采取理论与实践相结合的教学方法，自主学习与小组合作学习相结合的教学方法。

（四）教学资源的开发与利用

本课程是我系与南京云创大数据科技股份有限公司合作开发的课程，所以在教学中可以将公司中项目分解为多个任务，选出典型案例作为教学的资源，按照公司对产品的质量要求制作。

（五）评价标准

教学评价不再像以往一样采取简单的考试，而采用阶段性评价和总结性评价相结合的方式，学生每次任务的作品和最终的作品都将作为最终评价学生学习的依据；同时在教学评价中不仅仅是教师给予学生评价，也会聘请南京云创大数据科技股份有限公司的相关人员参与到评价中，从企业的角度给学生指导以及评价建议。

七、推荐选用教材

《数据清洗》由刘鹏教授作为丛书总主编率领团队编写，

清华大学出版社出版。

八、参考文献（含课程网站）

序号	书名	编者	出版社	版本
1	大数据	刘鹏	电子工业出版社	1
2	大数据库	刘鹏	电子工业出版社	1
3	大数据实验手册	刘鹏	电子工业出版社	1
4	实战 Hadoop2.0 从云计算到大数据	叶晓江、刘鹏	电子工业出版社	2

《数据标注工程》核心课程标准

课程编码	0408012	课程类别	专业核心课
计划学时	68	课程类型	B类
适用专业	大数据技术与应用专业	课程性质	必修
开课学期	第三学期	学分	4
先行课程	大数据导论 Python语言 云计算	开课单位	信息工程系
平行课程	数据清洗 大数据系统运维	考试类型	考试
后继课程	顶岗实习		

一、课程性质与定位

本课程属于高职大数据技术与应用专业核心课程之一，其目的是培养学生掌握大数据标注工程的方法与技能，并能够灵活的运用这些方法和技能进行独立的大数据标注任务、管理、设计，不断提高实际操作能力，通过对教材的学习与实验训练，培养学生实际动手能力。

二、课程设计与理念

本课程配合教材，使用浅显易懂的语言，系统教授了数据标注的基本概念、分类、流程、质量检验、管理和应用等。通过理论与实战相结合的方式，帮助学生由浅入深进行学习，从而真正掌握数据标注的核心技术、实施和管理方法。

三、课程目标

（一）总体目标

培养德、智、体、美全面发展的，能够较快适应生产、建设、管理、服务等一线岗位需要的，面向电信、零售、银

行、金融、政府等部门的掌握大数据标注基本技能，了解大数据技术应用框架与其生态系统，具有较高综合素质与良好职业素养的发展型、复合型、创新型技术技能人才。

（二）知识与技能目标

掌握大数据标注工程的方法与技能，并能够灵活的运用这些方法和技能进行独立的大数据标注任务、管理、设计

（三）能力与素质目标

- 1、熟悉数据标注工程基础知识。
- 2、熟练掌握数据标注工程基本操作。
- 3、熟练使用数据标注任务工具。
- 4、有良好的操作规范习惯。
- 5、有设计管理数据标注任务的能力。

四、课程教学内容及学时分配

序号	项目名称	学习任务	学习内容	学时
1	数据标注概述	数据标注的起源与发展 数据标注的应用场景 数据标注工程的意义	什么是数据标注 数据标注分类概述 数据标注流程概述 出行行业数据标注 金融行业数据标注 医疗行业数据标注 家居行业数据标注 安防行业数据标注 公共服务数据标注 电子商务数据标注 有监督的机器学习	8
2	数据采集与清洗	标注对象 数据采集 数据清洗	主要的数据来源 常见的标注数据 数据采集方法 数据采集流程	10

			标注数据采集 数据清洗方法 数据清洗流程 MapReduce 数据去重	
3	数据标注分类	图像标注 语音标注 文本标注	什么是图像标注 图像标注应用领域 什么是语音标注 客服录音数据标注规范 什么是文本标注 文本标注应用领域	8
4	数据标注质量检验	数据质量影响算法效果 数据标注质量标准 数据标注质量检验方法	图像标注质量标准 语音标注质量标准 文本标注质量标准 实时检验 全样检验 抽样检验	8
5	数据标注管理	数据标注工厂设计 数据标注管理架构 数据安全管理与质量管理体系 数据标注项目评估 数据标注订单管理 数据标注客户关系管理	数据标注工厂设计 数据标注管理架构 数据存储安全管理要求 工厂人员行为管理 溯源体系建设 质量管理体系建设	10
6	数据标注应用	自动驾驶数据标注 智能安防数据标注 智能医疗数据标注	自动驾驶的发展 自动驾驶的 9 种数据标注 智能安防的发展分析 智能安防的 5 种数据标注 智能医疗的发展 智能医疗应用的 4 种数据标注	8
7	数据标注实战	实战环境搭建 医疗影像标注 遥感影像标注 车牌图像标注 人像数据标注	标注工具安装环境搭建 LabelImg 标框标注工具的使用方法 Labelme 工具的安装与使用方法 医疗影像标注 遥感影像标注 车牌图像标框标注 车牌图像分类标注 行人图像标注 人脸数据标注	16

五、考核评定办法

1. 平时成绩 30%
2. 模块训练成果考核 30%
3. 期末考评 40%

六、教学建议

（一）教学条件

在教学过程中，应充分利用校内实训基地的作用，使教学与实训紧密联系，保持学习内容与岗位工作内容的一致性，提高学生的职业能力。

（二）师资要求

主讲教师应具备本科或硕士研究生学历，具有相关从业背景，有丰富的行业经验，了解前沿技术发展趋势及理论知识，掌握一定的教学方法与教学艺术，能综合运用各种教法开发设计课程。

（三）教学方法

提倡了教学观念的转变，强调课程内容的新颖性和时代特征。打破了传统教学过多强调概念和灌注式的教学方式，将教材内容“问题化”，让学生自己学会提出问题与解决问题，结合本课程的特点加强学生设计思维的训练。

1. 实行任务导入、理论讲述、项目工作的新型教育模式，采用启发、实例制作为主体的教学方法。

2. 教师要充分利用教学资源，拓宽学生学习渠道，改进

学生学习方式，提高教学效果，增强教学的开发性和灵活性。利用多媒体教学手段，在教学过程的适当时间进行播放、讲解和分析，让学生在熟悉基本知识的同时调动学生参与理论与实践学习的积极性。

3.注重课程的实验性和操作性，激发学生主动学习的兴趣。在注重原理性教学基础上，拓宽视野，融会贯通，强调学生的理解力、分析力和创造力。

（四）教学资源的开发与利用

本课程是我系与南京云创大数据科技股份有限公司合作开发的课程，所以在教学中可以将公司中项目分解为多个任务，选出典型案例作为教学的资源，按照公司对产品的质量要求制作。

（五）评价标准

教学评价不再像以往一样采取简单的考试，而采用阶段性评价和总结性评价相结合的方式，学生每次任务的作品和最终的作品都将作为最终评价学生学习的依据；同时在教学评价中不仅仅是教师给予学生评价，也会聘请南京云创大数据科技股份有限公司的相关人员参与到评价中，从企业的角度给学生指导以及评价建议。

七、推荐选用教材

《数据标注工程》 刘鹏教授为丛书总主编率领团队编写，清华大学出版社出版。

八、参考文献（含课程网站）

序号	书名	编者	出版社	版本
1	大数据	刘鹏	电子工业出版社	1
2	大数据库	刘鹏	电子工业出版社	1
3	大数据实验手册	刘鹏	电子工业出版社	1