



山西职业技术学院

SHANXI POLYTECHNIC COLLEGE

计算机工程系
大数据技术与应用专业
人才培养方案

(2020 级)

二〇二〇年六月

目 录

一、专业名称及代码	1
二、招生对象	1
三、修业年限	1
四、培养目标与培养规格	1
(一) 培养目标	1
(二) 培养规格	3
五、课程设置	4
六、学时分配	6
七、教学进程总体安排	8
八、毕业标准	10
九、实施保障	11
附件 1 计算机专业群人才需求调研报告	21
附件 2 对应“1+X”项目职业技能等级证书标准	47
附件 3 大数据技术与应用专业核心课程标准	67
《Python 程序设计》课程标准	67
《Hadoop 系统及应用》课程标准	73
《HBase 数据库技术》课程标准	78
《Hive 编程技术与应用》课程标准	82
《Spark 技术与应用》课程标准	87
《大数据可视化技术》课程标准	91

一、专业名称及代码

专业名称：大数据技术与应用

专业代码：610215

二、招生对象

高中阶段教育毕业生或具有同等学力者。

三、修业年限

高等职业学校学历教育修业年限为3年。

四、培养目标与培养规格

（一）培养目标

1. 总体目标

培养思想政治坚定、德技并修、全面发展，适应山西区域经济发展需要，具有良好的职业素养，掌握大数据基本理论、方法与技能，面向大数据应用企事业单位，能够从事使用大数据技术，从事大数据系统搭建与运维、大数据获取与存储、大数据处理与分析等工作的高素质技术技能人才。

2. 职业知识目标

- （1）掌握必备的体育健身基础知识和相关心理健康知识；
- （2）掌握必备的计算机应用、英语、数学的基本知识；
- （3）掌握公共安全、自身安全防范的基本知识；
- （4）了解相关法律法规及国际通用惯例；
- （5）掌握计算机软硬件基础知识；
- （6）掌握程序设计、软件需求分析基础知识；
- （7）掌握数据库原理、操作系统原理基础知识；
- （8）具备网站设计、计算机网络基础知识；
- （9）了解软件部署、软件系统安装运维知识；
- （10）掌握大数据系统工作原理及其搭建维护知识；
- （11）了解云计算平台工作原理及基本操作知识；
- （12）掌握数据获取和预处理知识；
- （13）掌握 Map Reduce 编程知识；
- （14）掌握数据仓库知识；

- (15) 掌握数据处理基本知识;
- (16) 掌握数据分析基本知识;
- (17) 掌握数据挖掘基本知识;
- (18) 了解 IT 领域前沿动态, 掌握部分相关新知识;
- (19) 具备一定的经营管理知识、经济效益意识及成本核算知识。

3. 职业素质目标

(1) 职业素质

具有热爱祖国、遵纪守法, 科学的世界观、人生观和价值观, 具有良好的思想品德、社会公德和坚定的政治素养; 具有一定的科学素养和文学、艺术修养; 具有积极健康、乐观向上的身心素质; 具有爱岗、敬业、奉献、协作等职业素养; 具有诚信品格、服务意识、质量意识; 具有良好的职业道德与职业操守, 具备较强的组织观念和集体意识; 具有一定的数理与逻辑思维; 具有一定阅读并正确理解分析报告和项目建设方案的能力; 具有阅读本专业相关中英文技术文献、分析与处理, 进行文档管理的能力; 具有一定的工程意识和效益意识; 具有创新精神和开拓能力。

(2) 劳动教育课

劳动教育是青年学生形成正确世界观、人生观、价值观的基础。根据中共中央、国务院印发的《关于全面加强新时代大中小学劳动教育的意见》要求, 我们要以建构新时代劳动教育体系为经, 以提升劳动教育支撑保障能力为纬, 注重围绕创新创业, 结合学科和专业积极开展实习实训、专业服务、社会实践、勤工助学等, 重视新知识、新技术、新工艺、新方法应用, 创造性地解决实际问题, 使学生增强诚实劳动意识, 积累职业经验, 提升就业创业能力, 树立正确择业观, 具有到艰苦地区和行业工作的奋斗精神, 懂得空谈误国、实干兴邦的深刻道理; 注重培育公共服务意识, 使学生具有面对重大疫情、灾害等危机主动作为的奉献精神, 构建中国特色劳动教育模式的四梁八柱, 为职业院校劳动教育的加强提供了基本遵循。

4. 职业能力目标

- (1) 具有运用辩证唯物主义的基本观点及方法认识、分析、解决问题的能力;
- (2) 具有数据库系统的基本操作能力;
- (3) 具有基本的程序设计能力;
- (4) 具有阅读本专业相关中英文技术文献、资料的能力;
- (5) 具有熟练查阅各种资料, 并加以整理、分析与处理, 进行文档管理的能力;

(6) 具有通过系统帮助、网络搜索、专业书籍等途径获取专业技术帮助的能力；

(7) 具有解决实际问题的综合技术能力。

(8) 具有各种大数据系统安装、维护与维修能力，包括 Hadoop、HBase、Hive、Spark 等系统；

(9) 具有使用各种大数据系统解决实际问题的能力，包括 Hadoop、HBase、Hive、Spark 等系统；

(10) 具有获取数据和对数据进行预处理的能力；

(11) 具有一定的大数据应用开发能力，包含批处理、实时流式处理、随机查询、数据挖掘分析等模式；

(12) 具备基本的大数据分析和数据挖掘能力。

(二) 培养规格

1. 专业群与产业链的对应性

根据《山西省“十三五”战略性新兴产业发展规划》，对接新一代信息技术产业，组建以大数据技术与应用专业为核心、以计算机应用技术、信息安全与管理专业为骨干，以计算机信息管理、移动应用开发专业为支撑的计算机专业群。

随着云计算、大数据及人工智能技术应用的快速落地，“云、数、智”技术融合发展为专业群转型升级带来机遇。2018年10月，国家统计局公布了《战略性新兴产业分类（2018）》，新一代信息技术产业新增新兴软件和新型信息技术服务，互联网与云计算、大数据服务，人工智能等产业分类。根据学院“大数据”贯穿、“智能”主线的专业集群新生态建设布局，专业群增设云计算技术与应用、人工智能技术服务专业，面向新一代信息技术产业“云、数、智”架构，重构计算机专业群。

大数据产业链是新一代信息技术产业中与大数据相关的若干产业的组合，产业链包括大数据的产生与集聚、组织与管理、分析与发现、应用与服务等层级，每一层都包含相应的IT基础设施、软件和信息技术服务。

群内专业对接大数据产业链的“云、数、智”三大前沿产业，面向云计算系统部署与运维、云计算应用开发与服务、大数据系统运维、大数据应用开发、AI应用开发、AI产品服务、信息系统实施与运维、软件开发与测试、Web前端开发、信息系统安全与管理等岗位，从事数据采集与存储、数据组织与管理、数据分析与发现、数据应用与展示、系统安全与运维等技术工作。

2. 专业群人才培养对应岗位

专业群名称	专业名称	主要职业类别	对应岗位名称	职业资格证书或技能等级证书	对应“1+X”项目证书
计算机专业群	大数据技术与应用	软件和信息技术服务人员	大数据系统运维 大数据分析	大数据运维工程师 大数据分析工程师	大数据平台运维 大数据分析与应用

备注：对应“1+X”项目证书填准确证书名称和等级，若无对应证书填写“无”；无法对应专业群的专业单独编写。

3. 本专业职业岗位与核心能力

职业岗位	主岗位	拓展岗位
	大数据系统运维	大数据分析与应用
主要工作任务	大数据平台安装配置 大数据组件安装配置、维护及使用 大数据平台维护及监控 大数据平台优化	部门及岗位，主要完成 数据 分析 处理 、 数据挖掘预测、数据治理、数据可视化、 数 据大屏制作、 数据分析报告撰写
岗位核心能力	Hadoop 集群、安装配置及运行核心组件 执行客户大数据平台实施方案 监控大数据平台运行状态 查看管理大数据平台日志信息	数据抽取、清洗、转化等数据处理程序开发
对应核心课程	Hadoop 系统及应用 HBase 数据库技术	Hive 编程技术与应用 Spark 技术与应用 大数据可视化技术
对应“1+X”项目证书	大数据平台运维（初级、中级）	
“1+X”证书考核要点	大数据平台安装、配置、维护 大数据平台组件安装、配置、维护 大数据平台实施、监控、优化	

五、课程设置

（一）课程体系的构建理念

大数据技术与应用专业课程体系的构建过程遵循职业教育国家教学标准、教育部关于职业院校专业人才培养方案制定与实施工作的指导意见等指导性文件，在专业群整体设计的基础上，结合本专业人才培养目标、国家教学标准、党和国家在课程设置、教学内容等方面的基本要求、学校自身资源优势等方面制定专业课程体系。制定课程体系过程中遵循职业教育规律，遵循技术技能人才成长和高职学生身心发展规律，课

程设置突出大数据专业的应用性，按照计算机专业群各专业之间“基础共享，核心分立，拓展互选”的原则，构建科学合理的大数据技术与应用专业课程体系。

（二）课程体系的开发程序

根据企业人才需求及计算机专业群建设与改革需要，组织行业企业专家、各专业带头人、骨干教师进行专业群课程体系开发。在人才需求调研、毕业生跟踪调查的基础上，通过岗位需求分析，提炼岗位典型工作任务及岗位知识技能要求，基于典型工作任务和工作过程，结合职业教育规律，将专业知识、职业技能合理构建课程并序化，构建大数据技术与应用专业课程体系。

（三）课程体系的结构

底层共享课程（18 门）		核心分立课程（9 门）	拓展互选课程（5 门）
公共基础课（14 门）	专业基础课（4 门）		
国防教育与军事训练、入学教育	Java 程序设计	Python 程序设计★	计算机组装与维护
思想道德修养与法律基础	HTML 5 网页设计	Hadoop 系统及应用★	计算机网络基础
毛泽东思想和中国特色社会主义理论体系概论	Linux 操作系统	HBase 数据库技术★	大数据导论
大学语文	MySQL 数据库	Hive 编程技术与应用★	Shell 编程
应用数学		Spark 技术与应用★	专业英语
基础英语		大数据可视化技术	
体育		毕业设计（论文）	
形势与政策		跟岗实习	
心理健康		顶岗实习	
安全教育			
信息素养概论			
大学生职业发展与就业指导			
创新创业教育			
劳动教养			

备注：标注“★”的课程为专业核心课程

（四）核心分立课程简介

课程名称	Python 程序设计	开设学期	第三学期		
课程代码	0917006	参考学时	72	学分	4

本课程主要学习 Python 数据类型，控制结构，正则表达式等基本知识以及大数据处理模块等内容，学习 Python 环境搭建及 Python 数据类型，Python 基础及函数，Python I/O 及文件读写，Python 面向对象编程，Python 库介绍（Numpy，Pandas），Python 数据爬取，Python 数据采集，整理及清洗等方面的应用，使学生通过 Python 编程语言的网络爬虫抓取工具，实现对网络数据抓取，以及使用 Python 的 Scrapy 框架来进行爬虫的抓取。通过 Python API 接口，能对计算分析结果导出到 HDFS 中，并提供报表，日志分析等结果；通过 Python 实现数据的展现，使学生具备使用 Python 对大数据进行操作及综合分析的能力

课程名称	Hadoop 系统及应用	开设学期	第三学期		
课程代码	0917005	参考学时	108	学分	6

本课程主要学习 Hadoop 的体系架构，原理及搭建方法，HDFS 文件系统架构，原理及使用方法，MapReduce 计算框架的工作流程，基于 MapReduce 的大数据应用开发，使学生理解 Hadoop 在大数据技术中的地位，理解 Hadoop 的体系结构和工作原理，理解 HDFS 文件系统结构和工作原理，理解 MapReduce 计算框架的工作流程，具有搭建和配置 Hadoop 系统的能力，具有基于 MapReduce 编程的能力及使用 HDFS 文件系统的能力

课程名称	HBase 数据库技术	开设学期	第四学期		
课程代码	0917007	参考学时	68	学分	4

本课程主要学习 HBase 的体系架构及实现方法，逻辑模型及物理模型，HBase 的搭建方法和使用方法，HBase 表的设计，创建和使用，使学生了解 HBase 在大数据技术中的地位，理解 HBase 和关系型数据库的本质差别，掌握 HBase 的体系结构和工作原理，具备搭建和配置 HBase 系统的能力，具备设计，创建和使用 HBase 表的能力

课程名称	Hive 编程技术与应用	开设学期	第四学期		
课程代码	0917008	参考学时	68	学分	4

本课程主要学习数据仓库的基本概念和作用，学习 Hive 的体系架构及搭建方法以及 Hive DDL，DML 和 SQL 操作方法，HiveQL 使用实例，使用 Hive 建立数据仓库的过程等，使学生理解 Hive 在大数据技术中的地位，理解数据仓库与数据库的本质区别，掌握 Hive 的工作原理，使学生具备定义，创建和使用数据仓库的能力，搭建和配置 Hive 系统的能力，存储 Hive 元数据的能力

课程名称	Spark 技术与应用	开设学期	第四学期		
课程代码	0917009	参考学时	102	学分	6

本课程主要学习 spark 的基本概念和作用，spark 使用 Scala 语言进行实现，它是一种面向对象，函数式编程语言，能够像操作本地集合对象一样轻松的操作分布式数据集。本课程学习 Scala 基础，Scala 高级使用，Apache Spark 基础及架构，Apache Spark 分布式计算原理，Spark SQL 精华与 Hive 的集成，基于 Spark GraphX 的图形数据分析等内容。让学生对大数据拥有数据清理，分析的能力。

课程名称	大可视化技术	开设学期	第四学期		
课程代码	0917010	参考学时	36	学分	2

通过本课程的学习培养学生信息数据可视化处理能力。使学生掌握大数据可视化的基础理论和概念，可视化的类型与模型、数据可视化的过程及数据可视化的过程以及数据可视化的常用方法。熟练掌握当下流行的大数据可视化工具，如 Echarts、Tableau 等的使用，为可视化应用开发和维护奠定基础。

六、学时分配

表 6-1 教学活动按周分配表

学期	入学教育及军训	课堂教学	集中实训	教学周合计	机动	考试周	学期小计	假期	总计
1	2	14		16	1	1	18	6	24
2		18		18	1	1	20	6	26
3		18		18	1	1	20	6	26
4		17	1	18	1	1	20	6	26
5		0	18	18	1	1	20	6	26
6		0	20	20	0	0	20		20
总计	2	67	39	108	5	5	118	30	148

表 6-2 学期教学任务书

学期	课程代码	课程名称	课程类型	教学周数	建议周学时	学时数 (理论+实践)
第一学期	2100001	国防教育与军事训练、入学教育	C	2	√	0+48
	1200009	思想道德修养与法律基础	A	14	2	20+7
	1200026	形势与政策	A	√	√	8+0
	1200012	心理健康	A	14	1	14+0
	2100003	安全教育	A	√	√	4+0
	1200030	大学生职业发展与就业指导	A	14	1	16+5
	1110046	大学语文	A	14	2	30+0
	1110044	应用数学	A	14	4	48+8
	1110049	基础英语	A	14	2	28+0
	1400007	体育	B	14	2	4+24
	0911001	信息素养概论	B	14	4	12+44
	0511001	Java 程序设计	B	14	4	28+28
	0911009	计算机组装与维护	B	14	2	18+10
	0917004	HTML5 网页设计	B	14	2	15+15
	合计学时					245+189=434
第二学期	1200010	思想道德修养与法律基础	A	18	2	20+7
	1200027	形势与政策	A	√	√	8+0
	2100004	安全教育	A	√	√	4+0
	1110058	大学语文	A	18	2	30+0
	1110045	应用数学	A	18	2	24+4
	1110050	基础英语	A	18	4	56+0
	1400008	体育	B	18	2	4+24
	2100009	创新创业教育	B	18	2	12+20
	0911008	计算机网络基础	B	18	2	18+18
	0917002	Java 程序设计	B	18	4	32+40
	0917003	大数据导论	B	18	2	20+10

	0921107	Linux 操作系统	B	18	2	18+18
	合计学时					241+146=387
第三学期	1200037	毛泽东思想和中国特色社会主义理论体系概论	A	18	2	28+8
	2100005	安全教育	A	√	√	4+0
	1400009	体育	B	18	2	4+24
	1200028	形势与政策	A	√	√	8+0
	0914019	MySQL 数据库	B	18	4	28+44
	0917012	Shell 编程	A	18	2	15+15
	0917006	Python 程序设计	B	18	4	30+42
	0917005	Hadoop 系统及应用	B	18	6	44+64
		合计学时				
第四学期	1200029	形势与政策	A	√	√	8+0
	2100006	安全教育	A	√	√	4+0
	1200038	毛泽东思想和中国特色社会主义理论体系概论	A	18	2	28+8
	1200034	大学生职业发展与就业指导	A	18	1	16+5
	0917007	HBase 数据库技术	B	18	4	32+40
	0917008	Hive 编程技术与应用	B	18	4	32+40
	0917009	Spark 技术与应用	B	18	6	54+54
	0917010	大据可视化技术	B	18	2	18+18
	0911007	专业英语	B	18	2	30+0
		劳动教育课	A	√	√	16+0
		合计学时				
第五学期	2100007	安全教育	A	√	√	4+0
	0411164	跟岗实习	C	18	15w	0+390
	0411137	毕业设计（论文）	C	18	3w	0+78
		合计学时				
第六学期	2100008	安全教育	A	√	√	4+0
	0411142	顶岗实习	C	20	20w	0+520
		合计学时				
合计	实践学时数		1721		总学时	2622
	实践学时所占比例		65.6%			
说明： 1.课程类型：A 类（理论课） B 类（理论+实践课） C 类（实践课） 2.课程代码为教务管理系统中的课程代码，同一课程在不周学期开设使用不同代码。						

七、教学进程总体安排

表 7-1 教学进程安排表

课	序	课程名称	学时	考核	学时分配	学分
---	---	------	----	----	------	----

程 结 构	号					方式	第一学年		第二学年		第三学年		
			总学时	理论	实践		第一 学期 16 周	第二 学期 18 周	第三 学期 18 周	第四 学期 18 周	第五 学期 20 周	第六 学期 20 周	
公共 基 础 课 程	1	国防教育与军事训练、入学教育	48		48	综合评价	2w						2.5
	2	思想道德修养与法律基础	54	40	14	过程考核+测试	2	2					3
	3	毛泽东思想和中国特色社会主义理论体系概论	72	56	16	过程考核+测试			2	2			4
	4	形势与政策	32	32		综合评价	√	√	√	√			2
	5	心理健康	14	14		综合评价	1						1
	6	安全教育	24	24		综合评价	√	√	√	√	√	√	2
	7	体育	84	12	72	过程考核+测试	2	2	2				4.5
	8	大学语文	60	60		过程考核+测试	2	2					3
	9	应用数学	84	72	12	过程考核+测试	4	2					4.5
	10	基础英语	84	84		过程考核+测试	2	4					4.5
	11	信息素养概论	56	12	44	过程考核+测试	4						3
	12	大学生职业发展与就业指导	42	32	10	过程考核+测试	1			1			2.5
	13	创新创业教育	32	12	20	综合评价			2				2
	14	劳动教育课	16	16						√			1
小计			702	466	236		18	14	4	3			39.5
专 业 课 程	1	Java 程序设计	128	60	68	过程考核+技能测试	4	4					3+4
	2	HTML 5 网页设计	28	16	12	过程考核+测试	2						2
	3	Linux 操作系统	36	18	18	过程考核+技能测试		2					2
	4	MySQL 数据库	72	28	44	过程考核+测试			4				4
	5	Python 程序设计	72	30	42	过程考核+测试			4				4
	6	Hadoop 系统及应用	108	44	64	过程考核+测试			6				6
	7	HBase 数据库技术	72	32	40	过程考核+测试				4			4
	8	Hive 编程技术与应用	72	32	40	过程考核+测试				4			4
	9	Spark 技术与应用	108	54	54	过程考核+测试				6			6
	10	大数据可视化技术	36	18	18	过程考核+测试				2			2

	11	毕业设计（论文）	130	0	130	过程考核+测试					5w		5
	12	跟岗实习	390	0	390	过程考核+测试					15w		15
	13	顶岗实习	520	0	520	过程考核+测试					20w		20
小计			1772	332	1440		6	6	14	16			81
专业拓展课程	1	计算机组装与维护	28	18	10	过程考核+测试	2						1.5
	2	计算机网络基础	30	20	10	过程考核+测试		2					1.5
	3	大数据导论	30	20	10	过程考核+测试		2					1.5
	4	Shell 编程	30	15	15	过程考核+测试			2				1.5
	5	专业英语	30	30	0	过程考核+测试				2			1.5
小计			148	103	45		2	4	2	2			7.5
选修课程	1	公共选修课 1											
	2	公共选修课 2											
	3	公共选修课 3											
	4	公共选修课 4											
小计			100										4
合计			2622	901	1721		26	24	20	21			132

说明：

- 1.校内外集中实训、毕业设计、顶岗实习周学时按 26 学时计算；
- 2.标示“√”课程不占用正常教学时间，以讲座形式开展；
- 3.公共选修课学时不计入总学时，只计学分。

表 7-2 可开设的非限定性专业选修课一览表

课程名称	课程类型	教学周数	建议周学时	学时数	学分
计算思维和数据科学	A	15	2	30	1
Java Web 应用开发	A	15	2	30	1
数据挖掘基础	B	15	2	30	1
人工智能导论	A	15	2	30	1

表 7-3 实践教学项目一览表

实践教学项目	开设学期	开设地点	教学周数	总学时数
劳动教育课	3	实训室	16	16
跟岗实习	5		15	390
毕业设计	5		5	104
顶岗实习	6		20	520

八、毕业标准

（一）学分要求

学生须修完本专业培养方案中公共学习领域课（38.5 学分）、专业课程（81 学分）、专业拓展课程（7.5 学分）、公共选修课（4 学分），总学分达到 132 学分。

必修课、公共选修课（其中面授选修课 1 门，网络选修课三门或 60 课时）成绩合格。

（二）素质要求

三年修业期间，素质拓展达到合格标准，取得学院颁发的素质评定证书。

（三）职业资格证书要求

建议学生毕业前取得以下职业资格证书或技能等级证书其中之一。

类别	资格证（技能证）名称	考核等级	考核学期	要求	职业编码
通用资格	大数据分析员	初级 / 中级	6	合格	
	其它	初级 / 中级	6	合格	
职业资格	大数据运维工程师	初级 / 中级	6	合格	
	大数据工程师	初级 / 中级	6	合格	
	阿里云大数据专业认证	初级 / 中级	6	合格	
	其它	初级 / 中级	6	合格	
“1+X”项目证书	大数据平台运维职业技能	初级 / 中级	6	合格	
	数据采集职业技能	初级 / 中级	6	合格	
	大数据分析职业技能	初级 / 中级	6	合格	
	其它	初级 / 中级	6	合格	

九、实施保障

（一）师资队伍

本方案实施需要建立由专业带头人、骨干教师、“双师素质”教师、企业技术专家或技术能手共同组成的教学团队，生师比建议不高于 18:1；具有硕士学位教师占专任教师的比例达 35%以上；具有高级职务教师占专任教师的比例达 30%以上；专业基础课和专业课中双师素质教师比例达 70%以上；兼职教师承担实践课学时数比例达 50%。

1.专业带头人

校企各配置 1 名专业带头人。校内专业带头人应具有副高级以上技术职称，要有一定的云计算大数据专业知识，熟悉行业企业最新技术动态，能够站在云计算大数据专业领域发展前沿，把握专业技术改革方向，对本专业的前沿动态、行业发展、岗位需求等有较深入的了解，能够准确把握该专业建设与教学改革方向，具有对本专业发

展的规划能力。

2. 骨干教师

专业教学团队应配置骨干教师 4 名以上。骨干教师应具有中级及以上职称，要能够根据行业企业岗位群的需要开发课程、编写教材，及时更新教学内容；承担 2 门以上专业课，具有课程开发及教学设计的能力，能够合理利用各种教学条件，采用不同教学方法和手段组织教学；能够开发校本教材、实训指导书，制作多媒体教学课件，建设精品网络资源共享课；到校企合作企业挂职锻炼，熟悉毕业生所从事工作岗位的要求。骨干教师必须是“双师素质”教师。

3. “双师素质”教师

“双师素质”教师应具有高等学校助理讲师（或以上）教师技术职务，年度考核合格，又具备下列条件之一：近五年有六个月（可累计）以上企业工作经历；近五年主持（或主要参与）2 项应用技术研究，成果已被企业使用，效益良好；近五年主持（或主要参与）两项校内实践教学设施建设或提升技术水平的设计安装工作，使用效果好，在省内同类院校中居先进水平；具有中级（或以上）工程系列专业技术职称或国家注册执业资格证书、职业资格证书者。其他情况可由学院教学指导委员会认定。

4. 兼职教师

企业兼职教师应具有熟练的大数据相关岗位技术能力和一定的教学水平，从事相关岗位工作 3 年以上；具有中级以上专业技术职务或高级工以上职业资格或在本行业享有较高声誉、具有丰富实践经验和特殊技能的“能工巧匠”；企业兼职教师上课或担任学生实践指导任务前，需经过教育教学培训；企业兼职教师承担专业实践课及顶岗实习学时数达 50%以上，形成稳定的企业兼职骨干教师队伍。

5. 本专业教师知识、能力与素质要求

知识要求：

- (1) 掌握计算机的基本结构、工作原理，计算机网络的基本概念和技术；
- (2) 掌握软件工程的基本知识，熟悉软件工程每个阶段的任务和工具；
- (3) 掌握操作系统的特点及功能，熟悉存储系统、网络系统的结构和原理；
- (4) 掌握数据库系统的特点及功能，熟悉数据库表的设计和操作系统；
- (5) 熟悉面向对象的程序设计方法，包含 C#语言、JAVA 或.NET 语言、HTML 语言、PHP 语言、Python 语言；
- (7) 熟悉分布式系统的特点和核心技术，熟悉 HTTP、REST、SOAP 协议；
- (8) 熟悉大数据相关技术，包括数据获取和预处理、数据存储、数据应用开发、

数据挖掘和分析等；

(9) 熟悉大数据相关系统的技术架构、工作原理和使用技术，包括 Hadoop、HBase、Hive、Spark、Mahout 等；

(10) 熟悉虚拟化、云计算相关的技术；

(11) 熟悉信息安全相关标准及法规。

能力要求：

(1) 具有设计、开发、测试和部署 Web 应用的能力；

(2) 具有为软件系统开发 REST 接口的能力；

(3) 具有使用脚本搭建大数据技术相关系统的能力，包括 Hadoop、HBase、Hive、park、Mahout 等；

(4) 具有使用相关工具对数据进行预处理的能力；

(5) 具有设计、建立和使用数据仓库的能力。

素质要求：

(1) 拥护党的领导，拥护社会主义，热爱祖国，热爱人民；热爱教育事业，具有良好的师德风范；

(2) 掌握教育学理论，具备在教学中实施行动导向教学法的能力，灵活运用案例及项目教学法和任务驱动等方法实施课程教学；

(3) 具有教学设计能力、课堂教学能力、指导实训项目的能力等较高的教学技能；

(4) 具备一定的科研素养；

(5) 具备提高自身专业素质的能力，适应大数据技术的快速发展；

(6) 具有较强的敬业精神，具有强烈的职业光荣感、历史使命感和社会责任感，爱岗敬业，忠于职守，乐于奉献。

(二) 教学设施

1. 校内实践教学条件

校内实训室（基地）一览表

实训室名称	主要设备	实训功能
大数据实训室	计算节点服务器	对学生端提供服务支持
	Training 实训模块	操作系统级的真机模拟训练
	Reporter 报表模块	大数据分析报表实训
	VmWeb 在线虚拟化桌面模块	虚拟机实训
	Dashboard 虚拟机控制台模块	虚拟机管理

实训室名称	主要设备	实训功能
	Compute 计算模块	虚拟机配置
	Networking 网络模块	配置内部拓扑，网络划分实训
	ShiroX 权限控制模块	用户身份验证和权限控制管理
	Swift 分布式存储模块	静态数据的分布式存储
	BlockStorage 块存储模块	数据库和文件系统
	Images 镜像管理模块	VM 镜像
	Hadoop 2.7.1	Hadoop 实训
	Python2.7.12	Python 相关的大数据实训
	Pandas 0.9.12	数据分析实训
	计算机	学生操作用机实训
网络管理实训室	服务器	文件共享
	二层交换机	交换机配置实训
	堆叠模块	交换机堆叠实验
	模块化路由器	路由器配置实训
	路由器线缆	路由器实验
	三层交换机	交换机配置实训
	投影仪	多媒体教学
	计算机	学生操作用机
	网络机柜	机房组网实验
	交换机	机房组网、交换实验
	信息安全攻防平台	网络安全实训
	防火墙及配件	防火墙实训
路由器及配件	路由实验	
网站开发实训室	服务器	ERP 平台服务器
	计算机	学生操作用机实训
	二层交换机	机房组网实训
	三层交换机	机房组网实训
	投影仪	多媒体教学
组装维护实训室	计算机	计算机组装、维护实训
	组装工具包	
应用软件开发实训室	服务器	文件共享
	计算机	学生操作用机
	电子白板	教师教学、演示
	打印机	教师教学、实训

实训室名称	主要设备	实训功能
	投影机	多媒体教学
	平板电脑（苹果及安卓）	学生项目实训、演示
	笔记本电脑	用于苹果 IOS 系统软件开发
基础实训室（6~8个）	计算机	软件教学

2. 校外实习基地及要求

校外实训基地的主要功能：有利于学生掌握岗位技能、提高实践能力；满足学生半年以上顶岗实习的需要，从而实现学生在基地的顶岗后就业。校外实训基地有健全的规章制度及基于职业标准的员工日常行为规范，有利于学生在实训期间便养成遵纪守法的习惯，能真正地领悟到团队合作精神，同时能培养学生解决实际问题的能力。校外实习基地业有利于学校及时了解社会对人才培养的要求，及时发现问题，有针对性地开展教育教学改革。

本专业应通过政府、行业协会等平台或积极主动联系等方式，紧密联系行业企业，多渠道筹措资金，多形式开展合作。在校外实训基地建设中，积极寻求与国内外、区域内大型知名企业开展深层次、紧密型合作，建立 3~5 个相对稳定的校外实训基地，充分满足大数据技术与应用专业学生的综合实践能力及半年以上的顶岗实习需要。发挥企业在人才培养中的作用，由企业提供场地、办公设备、项目和技术指导人员，企业技术人员与教师共同组织和带领学生完成真实项目设计、施工、调试与维护，使学生真正进入企业项目实战，形成校企共建、共管的格局。

在学生顶岗实习过程中，每个实习基地聘用一名企业兼职联络员，负责所及提供企业生产信息，根据企业生产内容和专业教学进度协调实习实训相关事宜。专业应与各合作企业共同动态修订人才培养目标，共同制定实习方案，按照人才培养方案，聘请企业专家担任兼职教师，在专兼职教师的共同指导下，以实际工作项目为主要实习任务。学生通过在企业真实环境中的实践，积累工作经验，具备职业素质综合能力，达到“准职业人”的标准，从而完成从学校到企业的过渡，使学生达到企业用人标准。

（三）教学资源

课程资源建设：

课程资源建设应该从专业核心课程开始，逐步带动其他课程的教学资源建设，并在教学实践中不断完善。每一门课程应主要建设以下资源：

1. 课程标准，应说明课程定位、教学学时，对于前序课程的要求，以及后续课程的衔接，课程的教学目标、课程设计思路、课程内容及要求、教学手段方法及课程考

核方式建议等内容；

2. 课程的教学课件，包含 PPT、例题程序代码；
3. 理实一体化课程教案，详细的教学步骤设计和教学内容；
4. 实验题目、实验内容、实验环境、实验说明书和程序代码；
5. 练习题及其参考答案；
6. 实训任务清单（实训环境、实训说明、实训要求、实训内容、实训考核方法）、实训项目代码、实训指导书；
7. 考试标准、试题库和参考答案。

教材建设：

教材是教学内容的载体，可以呈现教学标准的内容，也可以体现教学方法。内容适度、结构合理的教材是教学质量保证的重要因素，建议从以下几方面加强教材建设。

1. 校企合作共建“理实一体化”教材

专业组教师要联合企业一线技术专家，紧贴大数据当前实际，合作完成教材编写。

教材要将真实项目引入教材，实现理论知识学习和实际应用一体化；教材要面向教学过程、结合学生实际合理设置理论教学和技能训练环节，实现“教、学、做”甚至是“教、学、做、考”合一。

教材以项目为核心，每一教学单元建议采用教学导航、课堂知识讲解、课堂实践、课外拓展的环节开展教学。教学单元结束后，通过“单元实践”进一步提升技能；相关课程结束后，通过“综合实训”提升学生的综合能力。

2. 选用高职高专优质或国家规划教材

网络资源建设：

以信息技术为手段，以网络为平台，构建体系完善、资源丰富开放式的专业教学资源。同时要善于整合、消化、吸收企业优秀教学资源，使其实行共享。

网络资源需从以下几方面进行建设：

（1）专业建设方案

专业建设方案包括：专业简介、专业人才培养方案、课程标准、教学文件等。

（2）职业技能标准

（3）精品资源共享课资源

①基本资源。基本资源应包含课程简介、课程标准、教学大纲、授课计划、教案、多媒体课件、学习指南、习题、实验实训项目、电子教材、试题库等。

②拓展资源。拓展资源是在基本资源基础上，面向学生和社会学习者扩展的自学、

培训、进修、检索、科普、交流等内容，体现课程技术特点并向产业领域扩展。拓展资源包括素材库、培训包、工种包、企业案例、参考网站等。

③课程视频。课程视频包括课程整体设计介绍、课程单元设计说明等课程设计指导，课堂授课、现场教学、实训实习等教学场景，原理结构、工作过程、业务流程、操作步骤、技术细节、安全禁忌等内容。

（4）人文素养教学资源

①品德德育教学资源库。包含思想道德修养与法律基础、毛泽东思想和中国特色社会主义理论体系概论、职业生涯规划与创业就业指导等课程的课程资源，思政网站等。

②基础文化课教学资源库。包含本专业开设的高职语文、高职数学、高职英语、计算机应用基础、体育等文化基础课程的课程标准、教材、课件、案例库、习题库、视频资料等教学资源。

③职业拓展教学资源库。包含本专业开设的心理健康、形式与政策、拓展学习领域课程及公选课等课程课程资源。

（四）教学方法

课堂教学的组织管理。系（部）聘任有相应学识水平、有责任心、有教学经验的专任或兼职教师任课。组织任课教师认真研究课程标准，组织编写或选用与标准相适应的教材和教学参考资料；要求教师认真履行教师岗位职责，按教学规律讲好每一节课；组织教师开展教学方法的讨论和研究，合理使用现代化教学手段，充分利用教学资源，保证课堂教学质量。理实一体化课程及实践性教学内容要严格按人才培养方案和课程标准的要求进行教学，充分发挥校内外实训基地的教学资源，任课教师要设计好每一节或每个项目的教学做环节，训练学生的专业基本技能和综合职业能力。

对于公共基础课，建议采用启发式授课方式，一讲授为主，配合简单实验，多采用案例法、推理法等，深入浅出地讲解理论知识，注重学生参与，可制作图表和动画，易于学生理解。

对于基本技能课和岗位能力课，建议采用训练考核的教学方法，在讲清原理的基础上以实践技能培养为目标，保证训练强度达到训练标准，实践能力达到技术标准。可采用情景化教学、互动教学，教师可进行演示，为学生提供较为详尽的训练指导、动画视频等演示资料。

对于综合能力课，可采用项目教学法、案例教学法、任务驱动教学法、模块化教学法等，按照项目获任务的实施流程展开教学，让学生间接学习工程项目经验。项目

教学法尽量配合小组教学法，可将学生分组教学，并在分组中分担不同的职能，培养学生的团队合作能力，积极指导学生学以致用，组织学生自主完成工作任务。

（五）教学评价

凡是培养方案规定开设的课程都要对学生进行考核评价。根据课程特点和性质采用多样化的考核方式和方法，考核重点放在学生的综合素质及能力的评价方面，注重过程性评价。

课程考核与评价建议采用多方面结合的形式，应采用过程考核与结果考核相结合，理论与实践相结合，笔试、口试、操作相结合，离线作业、在线作业、终结性考核相结合，校内教师评价与企业教师评价相结合，他评、自评、互评相结合，根据课程不同特点，各考核项分配不同的比例进行成绩评定。

（六）质量管理

（1）教学管理

①日常教学管理。为保证人才培养方案的有效实施，按照教务处统一的教学运行文件，教务处及系（部），对学院教学运行进行日常检查、抽查、和学期检查。一般采取听课、检查任课教师的教学文件、召开学生座谈会、对学生进行问卷调查等形式，对出现的问题及时纠正改进，以确保方案的正常运行。

②建立教学工作例会制度。根据学院教学工作需要，由教务处协助主管教学副院长定期和不定期召开教学工作会议，全体系（部）主任及相关部门人员参加。通过教学工作例会，传达并学习最新职教发展动态和教学改革理念，布置学院教学发展改革任务，了解系（部）日常教学及专业、课程建设工作进展情况，研究和处理人才培养方案执行中出现的各种问题等。

③系（部）教学管理。系（部）定期召开专业主任会议和任课教师会议，及时掌握教学过程情况，总结教学工作和教学管理工作经验，及时研究解决教学过程中出现的问题。各专业要在每学期初制定出工作计划，组织集体备课、观摩教学、开展教学研究，了解教师教学进展情况，按学院要求进行教学检查。

（2）教学质量监控体系

①教学督导委员会组织机构

建立院系两级教学督导委员会，分级管理，分工负责，协同监控。

院级教学督导委员会由学院党委书记任主任，分管教学工作和学生工作的两位副

院长任副主任，同时聘请具有丰富教学经验的在职或离退休教师、具有丰富管理经验的教学管理人员组成山西职业技术学院教学督导委员会。院级教学督导委员会由督导中心牵头，以教学目标和主要教学环节的宏观监控为主，在院领导的直接领导下，负责全校教学质量监控工作的总体协调，确保教学质量的稳步提高。主要工作职责：一是对专业设置的论证、专业人才培养方案及相关教学文件的审核；二是通过深入课堂、实验室、实习基地，客观掌握教学运行的全过程，提出督导建议，为学院有关教学决策提供参考依据。

系级教学督导委员会由系主任负责，成立由校企合作工作委员会和专家、优秀毕业生代表组成的人才培养质量监控小组。系级教学督导委员会的主要职责：以教学过程自我监控为主，在主要负责人的领导下，负责对本单位的整体教学工作、教师的教学情况、学生的学习情况进行监控。负责组织各专业的听课、试卷命题、阅卷、试卷质量分析、毕业论文质量分析等工作，并通过学院、系部、专业教研室组织的各类检查评估（教案、作业布置与批改、教学进度计划、学生评教、教师评学、教研活动的开展等），严把各个教学环节的质量。

②日常教学督导

听课制度：院级领导每月听课次数不少于 1 次；值班中层干部每周听课不少于 1 次；系（部）主任、副主任及系（部）书记每月听课不少于 2 次。学院和系（部）各级党政干部深入教学第一线，及时了解教学情况，倾听师生意见，发现并解决教学中存在的问题，避免教学一线与管理层的脱节，保证教学管理工作的针对性和有效性。

学生教学信息员制度：以专业班级为单位，确定思想品德优良，有参与教学管理的积极性，善于联系老师和同学，能客观反映广大学生的意见学生代表和学生干部，举行学期座谈会，填写任课教师评分表，给学生以畅通的渠道反映本系、本专业的教学管理、办学条件和教学质量中存在的问题并对教学提出意见和建议，使系部的管理和教学更加贴近学生、贴近实际。

教学检查与管理制度：从学期初到学期末，院、系两级安排不少于 2 次的集中教学检查，采取听（听课、召开座谈会听取师生的反映）、看（查看教学条件和管理软件）、查（抽查教案、学生作业、实验报告、实习报告、课程设计、毕业设计等）、评（对教学条件、状态、效果进行评价）。教学情况的检查工作贯穿始终，发现问题及时反馈并解决落实。

为保证人才培养方案的执行，在教学运行中严格执行学院制定的教学工作规范、教学计划、课程标准和教学进程，严格教学事故的认定与处理，严格执行教学评价制

度，严格执行课堂教学和实践教学过程检查制度，严格教学文件的规范管理，保证人才培养方案的顺利实施、教学秩序的稳定和教学质量的提高。

附件 1 计算机专业群人才需求调研报告

一、调研目的与对象

（一）调研目的

为适应山西省产业转型升级、高质量发展需要，掌握行业、企业现有计算机类人才现状，了解区域及全国未来几年对计算机相关专业人才需求及培养要求，从而为确定我院计算机专业群各专业培养目标与专业设置提供基本依据，我系组织人员对高职计算机专业群人才需求情况进行了专题调研。

（二）调研对象

国家和山西省政府官网，全国行业协会、行业龙头企业，山西省范围内行业大中型企业，山西省综改示范区入区企业，主流招聘网站等。

二、调研方法与内容

（一）调研方法

1. 检索、查阅国家和山西省新一代信息技术相关产业政策

通过政府官网等官方平台检索、查阅国家和山西省关于新一代信息技术产业政策，如《国务院关于印发“十三五”国家战略性新兴产业发展规划的通知》、《工业和信息化部关于印发软件和信息技术服务业发展规划（2016-2020年）的通知》、《国家统计局战略性新兴产业分类（2018）》、《山西省国民经济和社会发展第十三个五年（2016-2020）规划纲要》、《山西省“十三五”战略性新兴产业发展规划》、《山西省“十三五”信息产业发展规划》、《山西省软件和信息技术服务业 2020 年行动计划》等。了解国家和山西省相关政策。

2. 行业企业人士访谈

通过事先拟定的访谈提纲，与非 IT 生产型企业、非 IT 服务型企业、政府机构、公共服务型企业和 IT 生产型企业、IT 服务型企业管理层以及一线员工进行座谈、访谈，就软件与信息服务、互联网与云计算、大数据服务、人工智能等产业发展的空间和趋势、人才现状和培养需求等进行调研和咨询。

3. 问卷调查

采用了问卷、资料收集等多种方法，抽取省内多家非 IT 生产型企业、非 IT 服务型企业、政府机构、公共服务型企业和 IT 生产型企业、IT 服务型企业，了解专业群关联

产业从业人员的具体岗位及所需的知识、能力、素质要求等信息。

4. 文献查阅

在相关行业协会网站和主流招聘网站进行数据收集和整理, 查阅第三方数据公司的调查报告与行业领军企业发布的权威报告, 搜集专业群相关岗位(群)人才需求的资料和数据。

(二) 调研内容

1. 新一代信息技术相关产业背景;
2. 职业岗位要求;
3. 就业人群分析;
4. 人才需求情况

三、调研分析

进入 21 世纪以来, 学科交叉融合加速, 新兴学科不断涌现, 前沿领域不断延伸。云计算、大数据、物联网、移动互联网、人工智能等新一代信息技术的发展, 正加速推进全球产业分工深化和经济结构调整, 重塑全球经济竞争格局, 数字经济正成为驱动我国经济发展的重要力量。新一代信息技术创新异常活跃, 技术融合不断加深, 催生出一系列新产品、新应用和新模式, 极大地推动了新兴产业的发展壮大, 加快了产业结构调整步伐, 促进了产业转型升级, 改变了传统经济发展方式。

2010 年 10 月 10 日国务院印发《关于加快培育和发展战略性新兴产业的决定》(国发〔2010〕32 号), 列出了七大国家战略性新兴产业, 其中包括新一代信息技术产业。2011 年公布的《中华人民共和国国民经济和社会发展第十二个五年(2011—2015 年)规划纲要》明确了战略新兴产业是国家未来重点扶持的对象, 其中信息技术被确立为七大战略性新兴产业之一, 将被重点推进。之后, 《工业和信息化部关于印发软件和信息技术服务业发展规划(2016—2020 年)的通知》(工信部规〔2016〕425 号)、《国务院关于印发“十三五”国家战略性新兴产业发展规划的通知》(国发〔2016〕67 号)等一系列政策文件相继出台, 加速推动新一代信息技术产业发展。

山西省继 2016 年发布《山西省国民经济和社会发展第十三个五年(2016—2020)规划纲要》后, 陆续出台一系列相关产业政策, 支持新一代信息技术与大数据产业加快高质量发展, 主要政策有《山西省“十三五”战略性新兴产业发展规划》、《山西省“十三五”信息产业发展规划》、《山西省信息化促进条例》、《关于山西省大数据发展规

划（2017-2020年）的通知》、《山西省促进大数据发展应用2017年行动计划》、《山西省电子信息产业2018年行动计划》、《山西省软件和信息技术服务业2020年行动计划》等。近年来，山西省委、省政府将实施大数据战略、发展数字经济、建设智慧山西作为高质量转型发展的重要引擎，打造大数据、云计算、人工智能、物联网等产业集群，鼓励高校、职业院校与企业合作培养大数据专业型、复合型与跨界复合型人才，为大数据战略实施提供人才支撑。

近几年，随着我国人工智能、物联网、大数据和云计算的广泛运用，与此相关的高新技术产业成为我国经济新的增长点，对从业人员的需求大幅增长，形成了相对稳定的从业人群。在这一背景下，以较高的专业技术知识和能力为支撑的新一代信息技术几类新职业——大数据、人工智能、云计算和新兴软件新型信息技术服务人员应运而生。

根据新职业的定义，大数据工程技术人员指从事大数据采集、清洗、分析、治理、挖掘等技术研究，并加以利用、管理、维护和服务的工程技术人员；人工智能工程技术人员指从事与人工智能相关算法、深度学习等多种技术的分析、研究、开发，并对人工智能系统进行设计、优化、运维、管理和应用的工程技术人员；云计算工程技术人员从事云计算技术研究，云系统构建、部署、运维，云资源管理、应用和服务的工程技术人员。新兴软件新型信息技术服务人员从事新兴软件开发、网络与信息安全软件开发、互联网安全服务、新型信息技术服务等工作。

2018年10月，国家统计局公布了《战略性新兴产业分类（2018）》，新一代信息技术产业新增新兴软件和新型信息技术服务，互联网与云计算、大数据服务，人工智能等产业分类。

（一）大数据产业与行业调研分析

1. 产业背景

大数据产业指以数据生产、采集、存储、加工、分析、服务为主的相关经济活动，包括数据资源建设，大数据软硬件产品的开发、销售和租赁活动，以及相关信息技术服务。当前，智慧医疗、智慧城市、精准扶贫以及其他相关高新技术产业都离不开大数据的支撑，大数据技术在我国得到了较为广泛的应用。

（1）国家实施大数据战略，构建数字中国

大数据被认为是“未来的新石油”，也被比喻为21世纪的“钻石矿”，在社会生产、流通、分配、消费活动以及经济运行机制等方面发挥着重要的作用。2014年大数

据首次写入政府工作报告；2015年8月国务院颁布《促进大数据发展行动纲要》，大数据正式上升为国家发展战略。随后国家出台了一系列大数据政策，覆盖生态环境大数据、农业大数据、水利大数据、城市大数据、医疗大数据、交通旅游服务大数据等多层次下游应用市场，加快实施国家大数据战略。

同时，伴随大数据政策出台，各地政府相继成立了大数据管理机构，促进大数据产业发展，全国22个省区，200多个地市相继成立大数据管理部门，如图1所示。

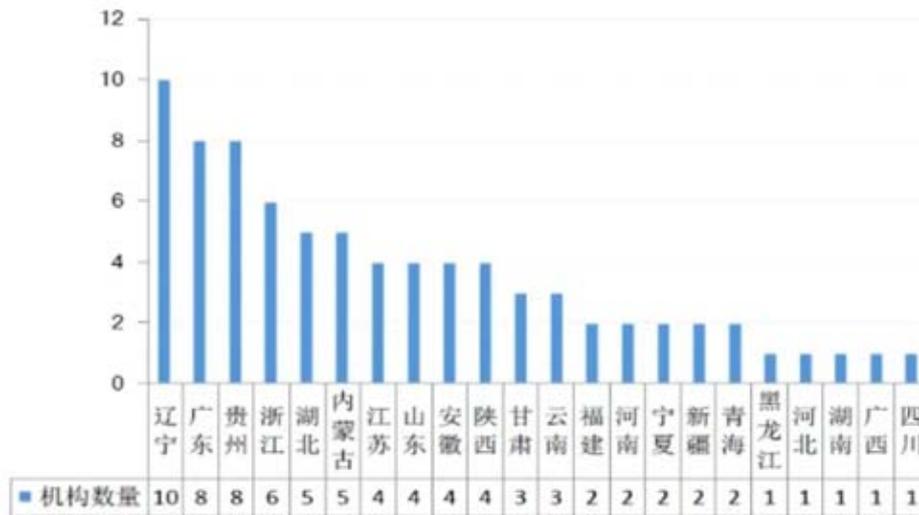


图1 各省大数据管理机构设置数量（单位：个）

(2) 大数据行业发展迅猛，产业规模巨大

2016年，工信部印发了《大数据产业发展规划（2016-2020年）》，全国大数据产业建设掀起热潮，目前已形成八大大数据综合试验区，建成100多个大数据产业园。伴随新一代信息技术、智慧城市、数字中国等发展战略逐步推动社会经济数字化转型，大数据的产业支撑得到强化，应用范围加速拓展，产业规模实现快速增长。

通过对1572家企业的调查结果显示，企业对数据分析的重视程度进一步提高，65.2%的企业已成立数据分析部门，24.4%的企业正在计划成立相关数据部门。

近四成的企业已经应用了大数据。在接受调查的企业中，已经应用大数据的企业有623家，占比为39.6%，垂直行业中如金融等领域大数据应用增加趋势较为明显。此外，24.3%的企业表示未来一年内将应用大数据。

对数据分析方式选择情况的调查显示，40.3%的企业采取实时处理动态数据并提供分析结果，占比最高；其次是分析历史数据和通过机器学习进行辅助决策，占比分别为32.3%和25.5%。不久的将来，随着人工智能技术的发展和普及，选择机器学习进行辅助决策的企业占比有望进一步提升。

2019年5月6日中国信息通信研究院发布《中国大数据与实体经济融合发展白皮书（2019年）》，书中综合国内外环境、新兴技术发展等多种因素，测算2018年我国大数据产业增速约为15%，产值达到5405亿元。另据赛迪数据显示，2018年中国大数据产业规模为4384.5亿元，同比增长23.5%；到2021年，中国大数据产业规模将超过8000亿元，如图2所示。



图2 2016-2021年中国大数据产业规模（单位：亿元）

从企业业务布局来看，大数据产业主要集中在华北、华东及中南地区，如表所示。

表1 2018年中国大数据产业区域分布情况

序号	地区	统计包含省份	占比(单位:%)
1	华北	北京市、天津市、河北省、山西省、 内蒙古自治区	27.5
2	华中 华南	河南省、湖北省、湖南省、广东省、 广西壮族自治区、海南省	26.0
3	华东	上海市、江苏省、浙江省、安徽省、 江西省、福建省、山东省	20.1
4	西南	重庆市、四川省、云南省、 西藏自治区、贵州省	14.2
5	东北	辽宁省、吉林省、黑龙江省	6.1
6	西北	陕西省、甘肃省、青海省、 宁夏回族自治区、新疆维吾尔自治区	6.1

(3) 数据资源资产化步伐稳步推进

2015年8月，国务院印发《促进大数据发展行动纲要》，明确“加快政府数据开放

共享,推动资源整合”。社会各界通过对数据资源的整合、利用,加速了数据流通共享以及数据资源化进程。2018年10月,《数据管理能力成熟度评估模型》发布实施,规范了各组织、机构数据管理和应用工作,提升国内数据管理和应用能力。2019年10月,在中国共产党第十九届中央委员会第四次全体会议上,中央首次公开指出“健全劳动、资本、土地、知识、技术、管理和数据等生产要素按贡献参与分配的机制。”这是中央首次在公开场合提出数据可作为生产要素按贡献参与分配,反映了随着经济活动数字化转型加快,数据对提高生产效率的乘数作用凸显,成为最具时代特征新生产要素的重要变化。

(4) 技术融合成为大数据发展主流

当前,大数据相关技术已基本成熟,逐步成为支撑型的基础设施,其发展方向也开始向提升效率转变,向个性化的上层应用聚焦。随着5G通信标准的落地,物联网、移动互联网、大数据、传统行业将深度融合,算力、流批、TA、模块、云数、数智等技术融合的趋势愈发明显,大量既懂大数据技术又懂其他相关行业技术的人才在大数据应用领域发挥着越来越多的作用。

(5) 数据安全受到业界普遍关注

近年来,大数据业界不断有安全事件曝出。2019年9月6日,位于杭州的大数据风控平台杭州魔蝎数据科技有限公司被警方控制,高管被带走,相关服务暂时瘫痪。同日,另一家提供大数据风控服务的新颜科技人工智能科技有限公司高管被带走协助调查。大数据安全合规的问题,特别是对于个人信息保护的问题,当前已成为整个社会 and 行业关注热点。

在全球不断收紧数据合规政策的大环境下,我国在数据法律监管方面也日趋严格规范。2019年以来,数据安全方面的立法进程明显加快。中央网信办针对网络安全审查、数据安全管理办法、儿童个人信息网络保护、个人信息出境安全评估等四项关于数据安全的管理办法相继发布征求意见稿。这些我国数据安全法律法规重点关注个人信息的保护,大数据行业整体合规也必然将以此作为核心。

2. 职业定义和工作任务

近年来,随着经济社会发展、科学技术进步和产业结构调整,新产业、新业态、新模式滋生孕育出许多新职业。大数据技术应用在各行各业的全面展开,我国社会需要越来越多的大数据工程技术人员。其职业定义和工作任务如下:

大数据工程技术人员职业定义：从事大数据采集、清洗、分析、治理、挖掘等技术研究，并加以利用、管理、维护和服务的工程技术人员。大数据工程技术人员主要工作任务：

- (1) 研究和开发大数据采集、清洗、存储及管理、分析及挖掘、展现及应用等有关技术；
- (2) 研究、应用大数据平台体系架构、技术和标准；
- (3) 设计、开发、集成、测试大数据软硬件系统；
- (4) 大数据采集、清洗、建模与分析；
- (5) 管理、维护并保障大数据系统稳定运行；
- (6) 监控、管理和保障大数据安全；
- (7) 提供大数据的技术咨询和技术服务。

3. 当前就业人群分析

(1) 学历层次

大数据人才的学历层次分为 4 个大类，分别是硕士及以上、本科、专科、专科以下，如图 3 所示。



图 3 大数据人才学历结构（单位：人）

可以看出，本科占比最高，其次是硕士及以上，专科占比只有 12.22%。大数据行业是新兴行业，目前学历要求比较高。

(2) 专业来源

专业来源分为 4 个大类，分别是数理类、经济管理类、计算机类及其他专业。计算机类占比最高，其次是数理类。项目组调研企业大数据人才的专业人数和占比见图 4。



图4 大数据人才专业来源（单位：人）

(3) 薪资水平分布

当前，大数据人才的薪资处于相对较高水平。薪资在1万元以下，占总人数的34.6%；1万元-2万元占比为35.64%；2万元以上占比为29.77%，如图5所示。



图5 大数据人才薪资水平分布（单位：人）

(4) 岗位类型及数量

目前企业提供的大数据岗位按照工作内容要求，可以分为以下几类：

- ①初级分析类，包括业务数据分析师、商务数据分析师等。
- ②挖掘算法类，包括数据挖掘工程师、机器学习工程师、深度学习工程师、算法工程师、AI工程师、数据科学家等。
- ③开发运维类，包括大数据开发工程师、大数据架构工程师、大数据运维工程师、

数据可视化工程师、数据采集工程师、数据库管理员等。

④产品运营类，包括数据运营经理、数据产品经理、数据项目经理、大数据销售等。

四类岗位的数量和占比见图 6。



图 6 大数据岗位类型结构（单位：人）

4. 行业人才需求情况

（1）整体需求

当前信息化对人类经济活动产生深刻影响，正渗透到生产生活方方面面，数据已经成为新的生产要素，大数据行业已成为人们按需使用信息处理、信息存储、信息交互资源的重要模式，也是进行大数据处理和深度挖掘的重要平台，大数据工程技术人员在我国现阶段及未来发挥的作用将日益凸显。

《大数据产业发展规划（2016-2020 年）》指出，目前大数据人才队伍建设亟需加强，大数据基础研究、产品研发和业务应用等各类人才短缺，难以满足发展需要。要建设多层次人才队伍，建立适应大数据发展需求的人才培养和评价机制。加强大数据人才培养，整合高校、企业、社会资源，推动建立创新人才培养模式，建立健全多层次、多类型的大数据人才培养体系。

根据天府大数据国际战略与技术研究院（简称“天府大数据研究院”）《2018 全球大数据发展分析报告》数据，2018 年我国大数据产业人才占整体就业人口规模的 0.23%，大约 179.4 万人。

猎聘《2019 年中国 AI&大数据人才大数据人才就业趋势报告》指出，2019 年中国大数据人才缺口高达 150 万。另据中国商业联合会数据分析专业委员会统计，未来中国

基础性数据分析人才缺口将达到 1400 万。

随着大数据、物联网、5G 等技术应用的不断发展，社会对该职业从业人员的需求日益增长。预计 2020 年中国大数据行业的人才需求规模将达到 210 万，2025 年前大数据人才需求仍将保持 30%-40% 的增速，需求总量在 2000 万人左右，大数据人才规模及增速如图 7 所示。

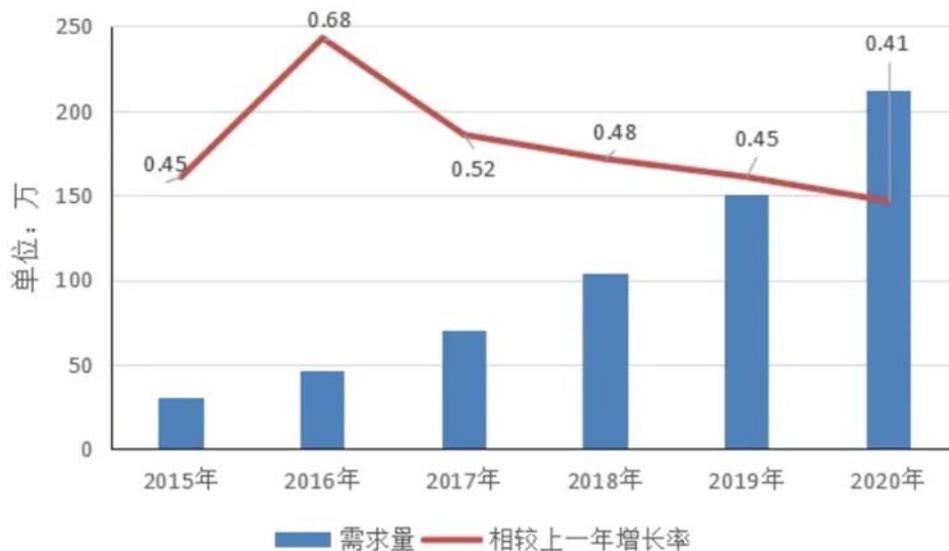


图 7 大数据人才规模及增速

(2) 行业发展对大数据相关岗位产生的影响

从业态变化的角度看，企业需要大量的复合型人才，即能够对数学、统计学、数据分析、机器学习和自然语言处理等多方面知识综合掌握的人才。从技术变化的角度看，深度神经网络等新兴技术的发展，弥补了传统分析挖掘技术在大数据时代的短板，这就需要大数据技能人才掌握深度学习方面的相关知识，适应大数据的分析挖掘需要。从运营方式的角度看，运营方式的变化要求运营人员提升运营前准备、运营中把握、运营后反馈、修正，提升预见能力和掌控能力。

目前企业对中高职层次的大数据人才相关岗位主要有：数据分析师、挖掘工程师、深度学习/算法/机器学习工程师、大数据开发工程师、大数据架构工程师、大数据运维工程师、数据可视化工程师、数据采集工程师、数据库管理员、数据运营经理、数据产品经理、数据项目经理、大数据销售工程师。可以看出，行业发展引发技术革命，相对应的岗位及要求也有所变化。

(1) 技术层面逐步由“万花筒”向“中国特色”、“中国制造”转变，中国标准逐渐成为业界标准，中国证书逐渐成为业界证书。

(2) 技术纵深发展和横向拓展, 引发企业对人才需求变化, 既有岗位重新细分的高精尖专才需求, 又有中等层次的广博复合型人才需求。(三) 岗位职责及技能要求根据调研情况整理, 大数据工程技术人员相关岗位的职责以及对大专以上学历人才的职业技能要求如表 2 所示。

表 2 大数据相关岗位的职责以及岗位技能要求

岗位	岗位职责	岗位技能
数据分析师	负责行业数据搜集、整理、分析, 并依据数据做出行业研究、评估和预测	数理统计基本知识, Excel, SQL, Python/R。
挖掘工程师	负责行业数据整理、挖掘, 并依据数据做出行业研究、评估和预测	常用数据挖掘算法, SQL, Python/R/Java。
深度学习/算法/机器学习工程师	负责利用各种神经网络模型及其算法并处理具体事务。	各种神经网络模型, Python/C++/Java, TensorFlow、Caffe等深度学习系统。
大数据开发工程师	使用编程语言开发大数据相关软件和应用系统	Java/Python/C++/Scala, Linux/Unix系统。
大数据架构工程师	负责大数据架构的设计与实施。	分布式系统原理, Linux/Unix系统及其脚本shell等, Hadoop、Spark等大数据框架及其组件Yarn, HBase、Hive、Pig等。
大数据运维工程师	负责大数据系统的运行和维护	Linux/Unix系统及其脚本shell等, Java。
数据可视化工程师	负责大数据可视化应用开发, 对数据分析结果多维度生动地体现。	前端框架及工具如jQuery、Vue.js、Webpack等, Web前端相关技术包括HTML/CSS/Javascript, 数据可视化框架如Echarts、Highcharts、D3.js等。

数据采集工程师	负责数据采集、预处理、标注等。	Linux/Unix系统，数据库如Mysql, redis, mongdb等，爬虫框架如Scrapy等等，web基础知识如HTML/JavaScript/CSS/xpath/url/Ajax/xml等，解析工具如HttpClient、jsoup、WebDriver、phantomjs等。
数据库管理员	负责数据库的运行和维护。	Linux/Unix系统，MySQL、SQL等数据库的运行机制和体系架构。
数据运营经理	负责数据的运营。	数理统计基本知识，运营方法，SQL。
数据产品经理	负责数据产品的销售。	工具如: Axure;Visio,Mindmanager,Project,PPT等，BI，SQL，产品规划能力，撰写需求文档能力。
数据项目经理	负责数据项目。	项目管理工具，PMP证书，梳理流程能力。
大数据销售工程师	负责大数据业务销售。	沟通能力，业务谈判能力。

（二）云计算产业与行业调研分析

1. 产业背景

云计算（cloud computing）是分布式计算的一种，指的是通过网络“云”将巨大的数据计算处理程序分解成无数小程序，然后通过多部服务器组成的系统进行处理和分析这些小程序得到结果并返回给用户。随着与云技术相关技术的发展，云服务已经不仅仅是一种分布式计算，而是分布式计算、效用计算、负载均衡、并行计算、网络存储、热备份冗余和虚拟化等计算机技术混合演进并跃升的结果，而且逐渐地将大数据技术、人工智能技术等技术融入到了云服务之中，其功能越来越强大。

在技术和价格双效推动下，全球云计算市场持续增长。根据 Gartner 的数据，包括 IaaS、PaaS、SaaS、流程服务、广告营销在内的云计算市场在 2016 年为 2196 亿美元，到 2020 年预计整体规模将达到 4114 亿美元，2016 至 2020 年的复合增长率为 17%。云计算作为智能社会的基础设施，其在未来社会中占有重要的位置，正是由于云计算在未来社会中承担着非常重要的技术地位。

2. 职业定义

云计算工程技术人员是指从事云计算技术研究，云系统构建、部署、运维，云资源管理、应用和服务的工程技术人员。主要工作任务：

(1) 研究、开发虚拟化、云平台、云资源管理和分发等云计算技术，以及大规模数据管理、分布式数据存储等相关技术；

(2) 研究、应用云计算技术、体系架构、协议和标准；

(3) 规划、设计、开发、集成、部署云计算系统；

(4) 管理、维护并保障云计算系统的稳定运行；

(5) 监控、保障云计算系统安全；

(6) 提供云计算系统的技术咨询和技术服务。

3. 当前就业人群分析

(1) 云计算人才区域分布

当前云计算行业技能型人才的需求保持持续增长，一线城市云计算人才需求最为明显。北京云计算人才缺口将近 12 万人，其次为上海、深圳和广州，分别突破 9 万人、7 万人和 6 万人，如图 8 所示。

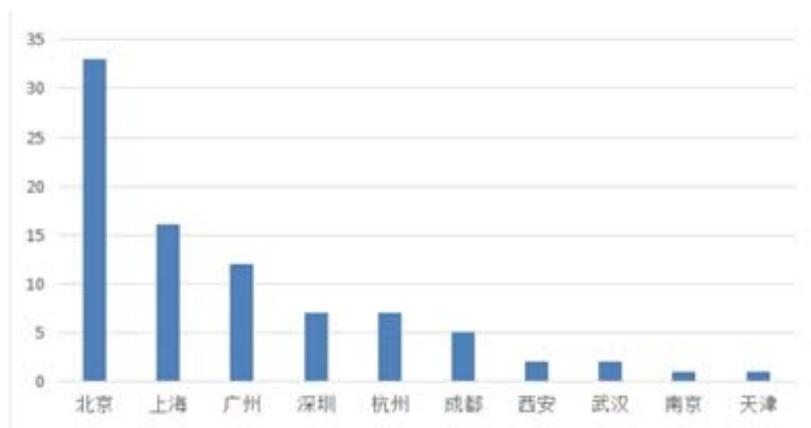


图 8 云计算技术人员地域分布

(2) 云计算人才岗位分布

云计算技术技能型大专人才的分布在设计研发等技术要求较强的岗位比例偏低，以技术服务、基础实施维护型技术岗位及销售工程师岗位为主；在基础硬件综合服务型岗位，应用研发综合服务企业以及云计算服务提供商对人才设置比较丰富，人才需求基数普遍较大；而在集成服务企业和云计算服务岗位，虽然企业规模偏小，需求较少，但是企业数量规模比例较大，主要以销售工程师、交付工程师、运维工程师、系统管理员为主，岗位技能要求相对偏低，是大专人才比较聚集的地方，如图 9 所示。

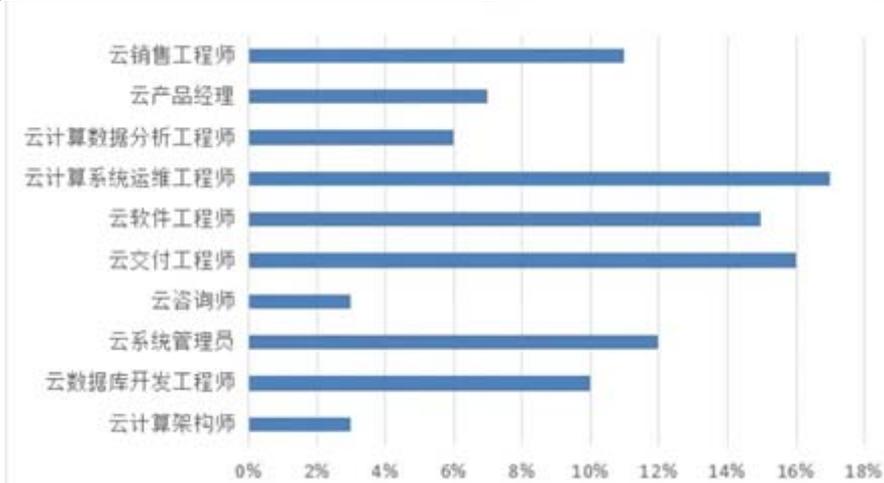


图9 云计算技术岗位分布

(3) 云计算人才薪酬待遇

在云计算领域细分岗位月均薪酬分布中，10000元以上成为基本标配。云计算领域人才月均薪酬在10000元以上的占比高达93%，30000元以上占比34%，反映出市场对于云技术专业技术人才的刚需。

(4) 云计算人才学历分布

云计算基础硬件综合服务型、应用研发综合服务类企业，人才学历层次结构呈倒金字塔型，本科占比最高；而集成服务、云计算服务提供商类企业，人才结构呈橄榄型，大专层次占比最高，如表3所示。

表3 云计算技术人才学历分布

企业类型	研究生及以上学历	本科	大专	中职及职业培训
基础硬件综合服务型厂家	6%	56%	35%	3%
应用研发综合服务企业	7%	59%	32%	2%
集成服务企业	2%	38%	54%	6%
云计算服务提供商	2%	36%	58%	8%

(5) 所在企业规模分析

云计算人才的岗位需求主要集中在100-499人规模的企业，占总需求量的45%，500-999规模企业占17%，1000以上规模企业占23%，50人以下企业占15%，如图10所示。

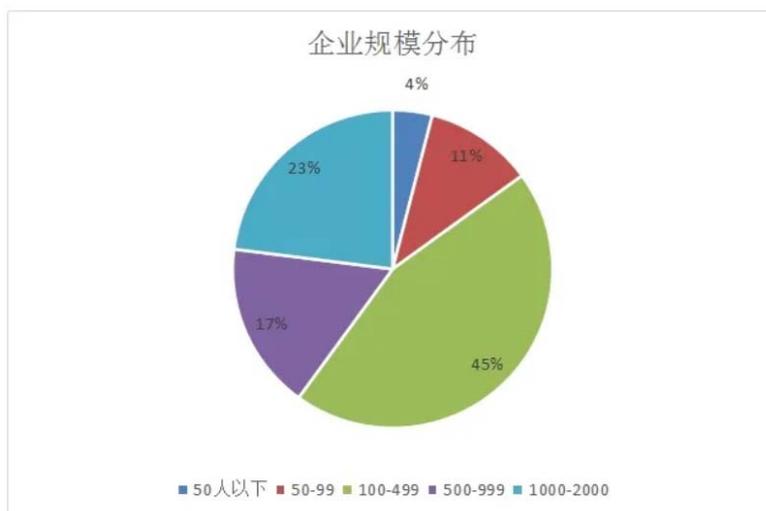


图 10 云计算技术人才所在企业规模分布

4. 职业发展通道

云计算主要岗位有云平台规划、部署、开发、服务和运维等岗位。云计算为人工智能提供发展所需的算力支撑，并通过辅助数据计算和存储为人工智能的发展提供支持，同时云计算也为海量的数据提供存储平台，使得数据能够有效被提取、处理和利用，所以云计算工程技术人员也可以向大数据或人工智能相关岗位发展。

（三）人工智能产业与行业调研分析

1. 产生背景

目前，人工智能已成为国家重要战略，也是我国供给侧改革的创新引擎。党的十九大报告提出要“加快建设制造强国，加快发展先进制造业，推动互联网、大数据、人工智能和实体经济深度融合”。人工智能已连续三年被写入政府工作报告。加快人工智能深度应用，培育壮大人工智能产业和人才供给，满足全球新一轮科技革命和产业变革趋势下人工智能人才需求，进而服务于科教兴国、创新驱动和人才强国等国家战略，已成为我国经济发展的重要支撑。

近三年来，国务院、国家发展改革委、工业和信息化部等多次颁布《新一代人工智能发展规划》《促进新一代人工智能产业发展三年行动计划(2018-2020)》等战略性和指导性文件共同推动人工智能的发展。《三年行动计划》提出，五个保障措施之一就是要加快人才培养，即要“吸引和培养人工智能高端人才和创新创业人才，支持一批领军人才和青年拔尖人才成长，支持加强人工智能相关学科专业建设，引导培养产业发展急需的技能型人才。”

由此可见，我国政府高度重视人工智能发展，将新一代人工智能技术的产业化和集

成应用作为发展重点。同时，也强调培养人工智能技术技能人才的重要性。

2. 职业定义

人工智能工程技术人员定义为从事与人工智能相关算法、深度学习等多种技术的分析、研究、开发，并对人工智能系统进行设计、优化、运维、管理和应用的工程技术人员。

人工智能工程技术人员主要工作任务：

- (1) 分析、研究人工智能算法、深度学习等技术并加以应用；
- (2) 研究、开发、应用人工智能指令、算法；
- (3) 规划、设计、开发基于人工智能算法的芯片；
- (4) 研发、应用、优化语言识别、语义识别、图像识别、生物特征识别等人工智能技术；
- (5) 设计、集成、管理、部署人工智能软硬件系统；
- (6) 设计、开发人工智能系统解决方案。

3. 当前就业人群分析

(1) 人工智能企业总量与分布状况

人工智能企业可划分为基础层、技术层和应用层。基础层以 AI 芯片、计算机语言、算法架构等研发为主；技术层以计算机视觉、智能语言、自然语言处理等应用算法研发为主；应用层以 AI 技术集成与应用开发为主。

据艾瑞咨询发布资料显示，2018 年我国人工智能相关公司总数达到 2167 家，其中应用层占比达到 77.7%，技术层和基础层企业占比相对较小，两者之和仅占到 22.3%；从技术类型分布来看，涉及机器学习的公司最多，占比 25.3%，其次大数据、云计算、机器人技术和计算机视觉的公司紧跟其后，整体分布相对均匀。具体分布如图 11 所示。

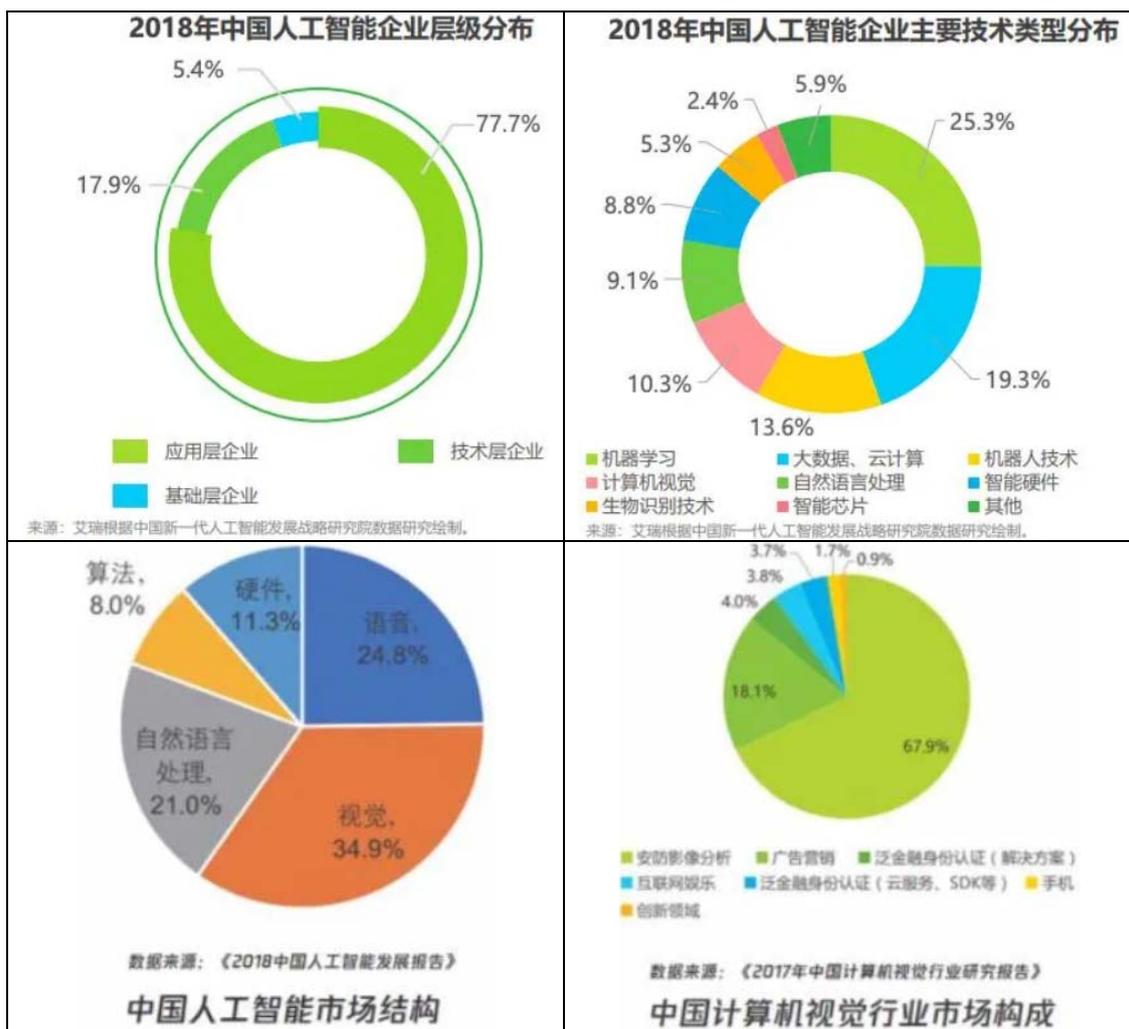


图 11 人工智能企业总量与分布状况

(2) 人工智能产业市场规模

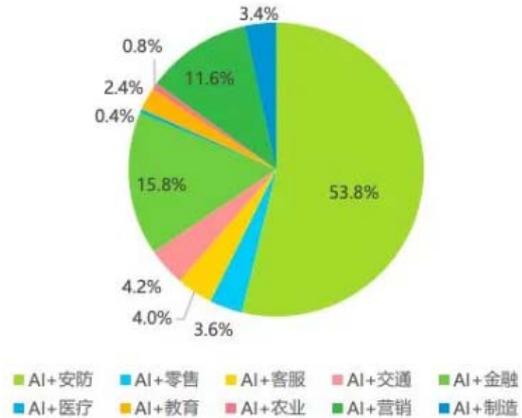
近几年，人工智能技术在实体经济中寻找落地应用场景成为核心要义，人工智能技术与传统行业经营模式及业务流程产生实质性融合，智能经济时代的全新产业版图初步显现，2019年人工智能核心产业规模预计突破570亿元，目前，安防和金融领域市场份额最大，工业、医疗、教育等领域具有爆发潜力，如图12所示。

2018-2022年中国人工智能赋能实体经济市场规模



来源：艾瑞根据专家访谈自主研究绘制。

2018年中国人工智能赋能实体经济各产业份额



来源：艾瑞根据不同行业市场规模的研究，进行统计建模所得。图表中相关统计口径，见下文细分领域内容。

图 12 人工智能产业市场规模

(3) 人工智能产业人才供需现状

随着人工智能概念的持续火爆，大批求职者主动向人工智能相关岗位靠近。根据《2017年全球人工智能人才白皮书》，过去几年中，我国期望在AI领域工作的求职者正以每年翻倍的速度迅猛增长，特别是偏基础层面的AI职位，如算法工程师，供应增幅达到150%以上。

为了对比国内AI人才供需情况，《白皮书》引入供需指数，该指数根据在特定时间段内的行业整体招聘需求量、活跃求职者存量以及招聘求职活跃度四个指标建模得出。从结果上看，目前国内AI人才供需指数逐年走高。2017年，国内AI人才供需较2015年提升11个百分点，表面上看人工智能人才供需已基本平衡，然而相关人才质量参差不齐。在对人才各项参数进行详细分析后得出，近三成期望在人工智能领域大展身手的求职者与AI雇主所要求的各项指标相距甚远，这部分人或为低学历求职者，或为刚初出茅庐，仅对基础编程略知，缺乏实际AI技能的初级程序员。说明我国AI人才不但严重紧缺，且这种趋势正由于人工智能企业增多而变得愈发严重，部分核心类岗位，如语音识别、图像识别工程师等，人才供需缺口更大。而且，由于合格AI人才培养所需时间远高于一般IT人才，人才缺口很难在短期内得到有效填补。

(4) 人工智能工程技术人员薪资水平现状

根据各大招聘网站的数据来看，人工智能行业的高薪主要分布在京津、长三角、珠三角及部分内陆省会城市。北京、上海、深圳及杭州的薪水位列第一方阵，月薪在1.8万左右；苏州、南京、广州及厦门位列第二方阵，月薪在1.4万左右；其他沿海及内陆

省会城市，如成都、重庆、长沙及济南等位于第三方阵，月薪在 1.3 万左右。其中，TOP 热门职位：深度学习算法工程师月薪可以达到 2.2 万；职位量方面，算法工程师需求遥遥领先。

根据测算，我国人工智能人才目前缺口超过 500 万，国内的供求比例为 1: 10，供需比例严重失衡。不断加强人才培养，补齐人才短板，是当务之急。

4. 职业发展通道

人工智能工程技术人员在企业中的最终角色是 CTO，其职业通道大致可分为初级工程技术人员、中级工程技术人员、高级工程技术人员。

初级工程技术人员在企业扮演的角色为：负责功能的实现方案设计、编码实现、疑难 BUG 分析诊断、攻关解决。

中级工程技术人员在企业扮演的角色为：开发工作量评估、开发任务分配；代码审核、开发风险识别/报告/协调解决；代码模板研发与推广、最佳实践规范总结与推广、自动化研发生产工具研发与推广。

高级工程技术人员在企业扮演的角色为：组建平台研发部，搭建公共技术平台，方便上面各条产品线开发；通过技术平台、通过高一层的职权，管理和协调各个产品线组。现在每个产品线都应该有合格的研发 Leader 和高级程序员了。

CTO 在企业扮演的角色为：业绩达成, 洞察客户需求，捕捉商业机会，规划技术产品，通过技术产品领导业务增长，有清晰的战略规划、主攻方向，带领团队实现组织目标。前沿与平台：到这个研发规模级别了，一定要有专门的团队做技术应用创新探索和前沿技术预研，而且要和技术平台团队、应用研发团队形成很好的联动作用，让创新原型试点能够很平滑地融入商业平台，再让应用研发线规模化地使用起来。研发过程管理：站在全局立场来端到端改进业务流程，为业务增长提供方便。组织与人才建设：公司文化和价值观的传承；研发专业族团队梯队建制建设、研发管理族团队梯队建制建设；创建创新激发机制，激发研发人创新向前发展，激发黑马人脱颖而出。

5. 未来市场需求

IDC 和 Forrester 发布了 2020 年及以后的人工智能（AI）预测。Forrester 表示，虽然外部“市场”可能会让企业对人工智能持谨慎的态度，但那些“勇敢”的企业将继续投资并扩大 AI 的布局。以下是 Forrester 的调查：53% 的全球决策者表示，他们已经实施、正在实施、或正在扩大人工智能的布局。29% 的全球开发人员在过去一年中从事过 AI/机器学习软件工作。

在全球实施边缘计算的公司中，54%的决策人员表示，边缘计算为他们处理当前和未来的 AI 需求提供了很大的灵活性。16%的全球 B2C 营销决策者计划今年将数据和分析技术（包括人工智能）的支出增加 10%及以上。

IDC 预测，到 2022 年，75%的企业将把智能自动化嵌入到技术和流程开发中，使用基于人工智能的软件来指导创新。到 2024 年，人工智能将整合到企业的每一个部分，在“结果即服务”（outcomes-a-service）的人工智能解决方案上，25%的总投资将用于推动规模创新和卓越的业务价值。人工智能将成为新的用户界面，并且重新定义用户体验。在未来几年，我们将看到人工智能和计算机视觉、自然语言处理和手势等新兴用户界面嵌入到每一种产品和设备中。

（四）软件和信息技术服务产业与行业调研分析

1. 产业背景

（1）产业整体形势

从整体来看，软件业保持平稳发展态势，进入结构优化、快速迭代的关键期。2019 年，在我国经济转型进入新常态，经济下行压力加大的背景下，软件和信息技术服务业作为数字经济之擎，产业规模保持较快增长，整体发展持续稳中向好。2019 年前三季度，软件业务收入 5.19 万亿元，同比增长 15.2%，增速同比提高 0.2 个百分点；利润总额增速小幅回升，全行业实现利润总额 6518 亿元，同比增长 10.8%，增速同比提高 0.9 个百分点。

2020 年，随着软件和信息技术服务业、大数据产业、“十三五”发展规划以及《促进新一代人工智能产业发展三年行动计划（2018—2020 年）》等国家政策贯彻落实进入加速冲刺的最后一年，云计算、大数据、人工智能等新兴技术应用将持续深化，软件产业的产品形态、服务模式、竞争格局将不断演进，软件将向云化、服务化、平台化、融合化、生态化趋势发展。随着软件和信息技术服务业在经济社会中的渗透力不断增强，软件服务将围绕主流软件平台体系构造产业生态，产业纵向、横向整合步伐加快，产业结构调整优化，产业生态环境不断优化。预计 2020 年，我国国民经济各个领域对软件和信息技术服务产业的需求更加旺盛，产业发展将以协同发展、融合创新、快速迭代为主要特征。新产业政策的预研、制定和发布将为产业发展和生态构建带来新的政策红利。

（2）产业细分形势

从细分形势来看，新兴技术加快创新应用，融合创新向深层次拓展。2019 年，对

基础软件、面向重点行业的高端软件的需求处于爆发增长的起点。工业软件前三季度产品实现收入 1277 亿元，同比增长 19.8%，高于软件和信息服务业平均增速。大数据、云计算保持快速发展态势，随着“百万企业上云”和制造业“双创”平台培育行动计划的推进，工业大数据应用价值持续深化。人工智能产业进入起步加速阶段，向全面商业化发展，其应用场景面向工业、安防、家居、医疗、物流、交通等行业迅速扩张。区块链产业形态不断成熟，包括人民银行、四大国有商业银行在内的 34 家银行正在进行区块链应用探索。阿里、腾讯、华为等均加大了在区块链领域的布局，金融、供应链、溯源、硬件、公益慈善 4 个领域区块链应用占比超过 60%。

2020 年，我国基础软件、高端软件发展进入攻坚克难的关键时期，具有自主知识产权的国产软件市场空间广阔。伴随着制造业等工业企业转型升级需求不断增加和工业互联网建设推进，依托专项支持和重点企业研发，工业研发设计软件将被优先重点发展，工业生产控制软件持续创新和迭代优化。云计算和大数据等与实体经济融合将更加深入，企业上云、设备上云进程将进一步加速。人工智能与经济社会发展深度融合将成为主线，在制造业转型升级中的作用更加突出，对智能制造的赋能效应将进一步释放。区块链将成为核心技术自主创新重要突破口，与实体经济实现深度融合，支撑经济高质量发展。预计 2020 年，软件与工业、金融、医疗、交通、智慧城市等各行业领域的融合将引发多领域、多维度、深层次变革，与制造业融合将不断催生新模式、新业态，智慧城市融合也将加速产业智能化升级。

2. 行业人才需求分析

我国软件与信息技术服务业市场发展迅猛。工信部指出，2019 年全国软件与信息技术服务业增速较快，实现规模和数量的双增长，规模以上企业超 4 万家，2019 年累计完成软件业务收入 71768 亿元，同比增长 15.4%，实现利润总额 9362 亿元，同比增长 9.9%。软件从业人数逐年增加，截至 2019 年，已达到 673 万人，比上年末增加 28 万人，同比增长 4.7%。2019 年人均实现业务收入 106.6 万元，同比增长 8.7%。从业人员工资总额 9086 亿元，同比增长 11.8%，人均工资增长 6.8%。由此可见，软件与信息技术服务业发展势头迅猛，是国民经济发展的重要领域，为劳动力市场提供更多的就业机会。

人才需求激增，人才市场供不应求。行业的发展必然引起人才和技术的流动，云计算、大数据以及网络传输等信息技术逐渐步入建设高峰期，对人才的需求更加迫切。工信部指出，我国部分新工科的人才缺口高达 750 万人。行业人均薪酬逐年递增，根据国

家统计局数据，信息传输、软件与信息技术服务业年平均工资 14.8 万元，是全国平均工资水平的 1.79 倍，侧面凸显人才市场供不应求，因此企业期望通过高薪引进人才。

人才分布不平衡。东部区域经济发展较快，具备产业聚集优势，人才吸引力度较大。西部地区多面临企业小、培训力度和发展空间不足、薪资较低等问题，导致人才“出走”，难以形成人才梯队，自身培育“造血”周期较长，而软件与信息技术服务行业产品技术更新速度较快。循环往复，导致西部与东部地区差距越来越大。

人才结构失衡的影响日益凸显。伴随着融合创新发展，软件人才结构失衡日益成为我国软件产业高质量发展面临的痛点。一方面，新兴产业的发展对于相关领域软件人才的需求迅速增长，大数据、区块链、工业互联网等新兴产业人才需求缺口逐步扩大，领军型人才、复合型人才缺乏，基础编程人才紧缺。另一方面，软件基础教育起点晚、人才培养机制滞后、“产学研用”融合体制低效以及高校新一代信息技术课程设置单一、陈旧导致我国软件产业人才供给不足，难以满足软件新技术和产业发展的实际需求。

四、结论与建议

（一）调研结论

1. 专业群定位

根据《山西省“十三五”战略性新兴产业发展规划》，对接新一代信息技术产业，组建以大数据技术与应用专业为核心、以计算机应用技术、信息安全与管理专业为骨干，以计算机信息管理、移动应用开发专业为支撑的计算机专业群。

随着云计算、大数据及人工智能技术应用的快速落地，“云、数、智”技术融合发展为专业群转型升级带来机遇。根据学院“大数据”贯穿、“智能”主线的专业集群新生态建设布局，专业群增设云计算技术与应用、人工智能技术服务专业，面向新一代信息技术产业“云、数、智”架构，重构计算机专业群。

新一代信息技术产业规模大、涉及面广，结构错综复杂。大数据产业链是新一代信息技术产业中与大数据相关的若干产业的组合，产业链整体布局完整，包括大数据的产生与集聚、组织与管理、分析与发现、应用与服务等层级，每一层都包含相应的 IT 基础设施、软件和信息技术服务。

群内专业对接大数据产业链的“云、数、智”三大前沿产业，面向云计算系统部署与运维、云计算应用开发与服务、大数据应用开发、大数据系统运维、AI 应用开发、AI 产品服务、信息系统实施与运维、软件开发与测试、Web 前端开发、信息系统安全与

管理等岗位，从事数据采集与存储、数据组织与管理、数据分析与发现、数据应用与展示、系统安全与运维等技术工作，专业群与产业链的对应关系如图 13 所示。



图 13 计算机专业群与产业链对应关系图

2. 专业群内各专业的群组关系

专业群以学院大数据技术与应用专业为核心，融合云计算和人工智能技术发展，按照数据采集、数据传输、数据存储、数据处理、数据展示、数据安全的产业逻辑主线，聚合计算机应用技术、信息安全与管理、计算机信息管理、移动应用开发等专业，面向新一代信息技术产业“云、数、智”三大领域，按照“专业基础相通、技术领域相近、工作岗位相关、教学资源共享”原则组建专业群。实现群内各专业的管理集约、资源集成共享和协同发展，形成专业集群优势。

3. 人才培养目标和就业岗位

专业群以立德树人为根本，培养理想信念坚定，德、智、体、美、劳全面发展，具有一定的科学文化水平，良好的职业道德、工匠精神和创新精神，具有较强就业能力、一定的创业能力和支撑终身发展的能力，掌握专业群知识和技术技能，面向软件和信息

技术服务等行业，能够从事专业群岗位工作的高素质技术技能人才。其中，大数据技术与应用专业主要面向大数据处理与分析岗位；云计算技术与应用专业主要面向云计算系统部署与运维、云计算应用开发与服务岗位；人工智能技术服务专业主要面向 AI 应用开发、AI 系统运维等岗位；计算机应用技术专业主要面向 Web 前端开发、数据恢复、系统运维等岗位；计算机信息管理专业主要面向（云）数据库开发、软件测试等岗位；移动应用开发专业主要面向移动 UI 设计、移动软件开发等岗位；信息安全与管理专业主要面向信息安全系统集成、网络安全运维、Web 安全管理等岗位。

4. 人才培养规格

（1）知识要求

①通用知识：体育健康基础知识；心理健康知识；计算机软硬件基础知识；计算机基本操作；英语、数学基本知识；IT 领域新技术基本知识；公共安全、安全防范基本知识；经营管理基本知识；项目管理基本知识；成本核算基本知识；法律法规及国际通用惯例基本知识。

②专用知识：各职业岗位专业知识。

（2）能力要求

①职业核心能力：交流表达能力；数字运算能力；革新创新能力；自我提高能力；与人合作能力；解决问题能力；信息处理能力；外语应用能力。

②行业通用能力：基本的数学思维能力；基本的编程能力；基本的英语阅读能力；基本的文档编写能力；基本的信息系统集成与维护能力；新媒体应用能力

③岗位专用能力：各职业岗位专门能力。

（3）素质要求

①爱党祖国、遵纪守法，树立科学的世界观、人生观和价值观，树立中国特色社会主义共同理想。

②具有诚实品质与劳动意识，不断积累职业经验，提升就业创业能力；

③树立正确择业观，具有到艰苦地区和行业工作的奋斗精神；

④具有积极健康、乐观向上的身心素质和良好的人文素养；

⑤具有爱岗、敬业、奉献、协作等职业素养；

⑥具有诚信品格、公共服务意识、精益求精的工匠精神；

⑦具有良好的职业道德与职业操守，具备较强的组织观念和集体意识；

⑧具有较强的团队合作精神，能够进行有效的人际沟通和协作，有社会、自然和谐相处；

⑨具有良好的生活习惯、行为习惯、自我管理意识和安全防范意识；

⑩时间意识、效率意识、成本意识与节约品质。

（二）实施建议

1. 搭建技术技能创新平台

与行业领军企业共建产业学院，为师资建设、人才培养和技术创新与服务提供快捷有效的技术技能创新服务平台。建立科技创新中心、认证中心，构建集人才培养、团队建设、技术服务于一体，资源共享、机制灵活、产出高效的人才培养与技术创新平台。建立技术服务中心，联合行业、企业、职业院校等成立大数据产教联盟，加强与地方政府、产业、行业深度合作，建设兼具科技攻关、英才培养、创新创业功能的产教融合平台。建设大师工作室和工匠工坊，打造高水平技术技能平台，兼具产品研发、工艺开发、技术推广、大师培育功能，助力区域产业转型发展。

2. 推进专业群人才培养模式改革

以立德树人为根本，满足不同特质的学生个性成长需要，依托技术技能创新平台，创新“校企协同，书证融通，分层分向，能力递进”人才培养模式。依据产业链岗位标准，将素质教育融入教育教学全过程，构建“底层共享、中层分立、高层互选”的课程体系。积极推动1+X证书试点建设与企业认证，对接国家学分银行。建立技能大赛体系，构建学生实践能力培养体系。落实“三全育人”，培养学生综合素质。

3. 打造高水平教师教学创新团队

加强师德师风建设，引进高层次人才、能工巧匠，加强专业（群）带头人和骨干教师培养和管理，提升专业（群）带头人的行业影响力，打造一支数量充足、专兼结合、结构合理，且具有国际视野的“双师型”教师教学创新团队。

4. 不断丰富教学资源，深入推进“三教”改革

构建大数据专业群教学资源平台，基于“互联网+”建设课程资源、培训资源和案例资源三个平台。积极开发线上精品教学资源，推进课程思政改革及“专创融合”课程开发。以专业核心课程为重点，建设适应理实一体化教学、适应模块化教学、信息化资源丰富的新型立体化教材，形成专业群系列化教材体系。实施“互联网+”教学方法改革，建立以“项目教学、教学做一体”为核心的教学模式，以培养学生能力为根本出发

点，全面推进线上线下混合式教学，全面推进案例教学、项目化教学、任务驱动等教学方法改革。探索实施教师分工协作的模块化教学模式。探索混合教学模式，打造优质课堂，推动课堂革命。

5. 建设产教融合实践教学基地

依托产业学院，统筹多方资源，打造立体化多功能产教融合实训基地，优化校外实践教学基地，创新校内外实践教学基地管理模式。满足实践教学、技能训练、技术创新及社会服务等需求，支撑专业群“分层分向”技术技能人才培养，服务企业技术改造与产品研发，助推山西区域产业转型升级。

6. 积极开展社会服务与国际交流

整合专业群优质资源，建立包含职教帮扶、社会培训和技术服务的全方位社会培训、服务体系。开拓国际交流新局面，助力“一带一路”建设，开展国际交流合作，开展教师境外学习、师资互访交流，引进高水平教育资源；开展海外人才培养培训，吸引“一带一路”沿线国家学生与企业员工接受教育，培养具有跨文化视野的国际化人才，推动中国培训标准和培训方案走向世界。为师生的国际化素质的培养以及跨国升学就业等提供条件，提升学生的国际竞争力和就业能力。

附件 2 对应“1+X”项目职业技能等级证书标准

大数据平台运
维 职业技能等
级标准

(2020年1.0版)

新华三技术有限公司 制定

2020年3月发布

前 言

本标准按照 GB/T 1.1-2009 给出的规则起草。

本标准起草单位：新华三技术有限公司、工业和信息化部教育与考试中心、无锡职业技术学院、浙江机电职业技术学院、南京信息职业技术学院、长沙民政职业技术学院、重庆电子工程职业学院、贵州轻工职业技术学院。

本标准主要起草人：刘小兵、姚明、肖李晨、于鹏、陈喆、蔡建军、卢涤非、聂明、邓文达、卢建云、汪洪、陈穆衍、白杨、陈永波。

声明：本标准的知识产权归属于新华三技术有限公司，未经新华三技术有限公司同意，不得印刷、销售。

1.范围

本标准规定了大数据平台运维职业技能等级对应的工作领域、工作任务及职业技能要求。

本标准适用于大数据平台运维职业技能培训、考核与评价，相关用人单位的人员聘用、培训与考核可参照使用。

2.规范性引用文件

下列文件对于本标准的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本标准。凡是不注日期的引用文件，其最新版本适用于本标准。

GB/T 35295-2017 信息技术 大数据 术语

GB/T 5271.1-2000 信息技术 词汇 第 1 部分：基本术语 中等职业学校专业教学标准（2018 年）

本科专业类教学质量国家标准

2019 年全国职业院校技能大赛 GZ-2019032 大数据技术与应用赛项规程 大数据安全标准化白皮书（2018 版）

大数据标准化白皮书（2018 版）

GB/T 37973-2019 《信息安全技术 大数据安全管理指南》

GB/T 37722-2019 《信息技术 大数据存储与处理系统功能要求》

GB/T 37721-2019 《信息技术 大数据分析系统功能要求》

ISO/IEC 20547-4《信息技术 大数据参考架构 第 4 部分：安全与隐私保护》ITU-T Y.3600《大数据 基于云计算的要求和能力》

3. 术语和定义

GB/T 35295-2017、国家、行业标准界定的以及下列术语和定义适用于本标准。

3.1 大数据 Big Data

具有体量巨大、来源多样、生成极快、且多变等特征并且难以用传统数据体系结构有效处理的包含大量数据集的数据。

注：国际上，大数据的 4 个特征普遍不加修饰地直接用 Volume、Variety、Velocity 和 Variability 予以表述，并分别赋予了它们在大数据语境下的定义： 体量 Volume：构成大数据的数据集的规模。

多样性 Variety：数据可能来自多个数据仓库、数据领域或多种数据类型。速度 Velocity：单位时间的数据流量。

多变性 Variability：大数据其他特征，即体量、速度和多样性等特征都处于多变状

态。

3.2 大数据系统 Big Data System

实现大数据参考体系结构的全部或部分功能的系统。

3.3 大数据服务 Big Data Service

基于大数据参考体系结构提供的数据服务。

3.4 集群 Cluster

集群就是一组计算机，它们作为一个整体向用户提供一组网络资源和服务，这些单个的计算机系统就是集群的节点（node）。集群具有可扩展性、高可用性、负载均衡及错误恢复的关键特性。

3.5 虚拟 Virtual

用来修饰一种功能单元，它看起来是实际的，但其功能是通过其他手段得以实现的。

3.6 虚拟机 Virtual Machine, VM（缩写词）

一种虚拟的数据处理系统，它看起来是在某个特定用户的独占使用下，但其功能是通过共享真实数据处理系统的各种资源得以实现的。

3.7 网络功能虚拟化 Network Function Virtualization

对路由器/路由选择、周界防护、远程访问鉴别以及网络流量/载荷监控等网络功能的虚拟应用实现。

注：网络功能虚拟化支持信息系统的高弹性、容错和资源管理，是应对大数据巨大数据体量下用户数据连接的峰、谷起伏问题的至关重要的应用。

3.8 本地虚拟化 Native Virtualization

大数据环境下的一种虚拟化基本形式，按此种形式，在本地裸机上运行管理程序，该程序管理由操作系统和应用组成的多个虚拟机。

3.9 主机虚拟化 Hosted Virtualization

数据环境下的一种虚拟化基本形式，按此种形式，在本地裸机上运行操作系统，在驻留客户操作系统和应用的顶层运行管理程序。

3.10 数据治理 Data Governance

对数据进行处置、格式化和规范化的过程。

注 1：数据治理是数据和数据系统管理的基本要素。

注 2：数据治理及数据全生命周期管理，无论数据是处于静态、动态、未完成状态还是交易状态。

3.11 链接数据 Linked Data 连接其他数据的数据。

3.12 分析 Analytics

根据信息合成知识的过程。

3.13 资源协商 Resource Negotiation

一种支持多租户以及要求高可用性和低延迟的环境的资源访问模式。

注：按此模式，资源管理器是若干节点管理器的集线器；各个客户（或用户）依次请求节点管理器中的应用管理器，紧接前一个请求者的后一个请求者分配到同一个或不同的节点管理器的应用管理器。根据中央处理器（CPU）和存储器可用情况为所请求的任务确定先后次序并在节点提供适当的处理资源。

3.14 集群管理 Cluster Management

在以非关系模型方式驻留数据的集群资源之间提供通信的一种机制。

3.15 数据处理 Data Processing

数据操作的系统执行。

注：术语“数据处理”不能用作“信息处理”的同义词。

3.16 数据管理 Data Management

在数据处理系统中，提供对数据的访问，执行或监视数据的存储，以及控制输入输出操作等功能。

3.17 数据挖掘 Data Mining

从大量的数据中通过算法搜索隐藏于其中信息的过程。

注：一般通过包括统计、在线分析处理、情报检索、机器学习、专家系统（依靠过去的经验法则）和模式识别等方法来实现。

3.18 数据中心 Data Center

由计算机场站（机房）、机房基础设施、信息系统硬件（物理和虚拟资源）、信息系统软件、信息资源（数据）和人员以及相应的规章制度组成的组织。

3.19 数据可视化 Data Visualization

借助图形化手段，清晰有效地传达与沟通信息，是关于数据视觉表现形式的科学技术研究。

3.20 数据平台框架 Data Platform Framework

用于指导实现结合相关应用编程接口（API）访问的逻辑数据组织和分发的集合。

注 1：此类框架一般还包含数据注册和连同语义数据描述（如格式化本体和分类）的元数据服务。逻辑数据组织的覆盖范围从简单限定的平面文件到完全分布式关

系数据存储或分栏数据存储。

注 2：这是大数据框架提供者可能提供的一种框架。

3.21 配置 Configuration

信息处理系统中的硬件和软件组织和互连起来的方式。

3.22 下载 To Download

将程序或数据从一个计算机传送到与之相连的资源较少的计算机上，通常是从主计算机传到个人计算机上。

3.23 上载 To Up Load

将程序或数据从一个与之相连的计算机传送到一个资源较多的计算机上，通常是从个人计算机传到主计算机上。

3.24 接口 Interface

两个功能单元共享的边界，它由各种特征（如，功能、物理连接、信号交换等）来定义。

3.25 软件工程 Software Engineering

将科技知识、方法和经验系统地应用到软件的设计、实现、测试和文档编制中，以优化软件的生产、技术支持和质量。

3.26 数据科学 Data Science

根据原始数据，经过整个数据生存周期过程凭借经验合成可用行动的知识的一种科学。

4. 适用院校专业

中等职业学校：计算机应用、计算机网络技术、网站建设与管理、软件与信息服务、电子与信息技术、网站安防系统安装与维护等专业。

高等职业学校：计算机应用技术、计算机网络技术、计算机信息管理、计算机系统与维护、软件技术、软件与信息服务、云计算技术与应用、大数据技术与应用、电子商务技术、人工智能技术服务等专业。

应用型本科学校：数据科学与大数据技术、计算机科学与技术、软件工程、网络工程、智能科学与技术、大数据管理与应用、信息管理与信息系统等专业。

5. 面向职业岗位（群）

主要面向从事大数据平台安装、配置、规划、部署、实施、优化升级以及大数据平台监控、管理、维护等相关工作的人员。从事大数据平台部署实施，监控、管理、运行维护等相关工作的人员。

6. 职业技能要求

6.1 职业技能等级划分

大数据平台运维职业技能等级分为三个等级：初级、中级、高级，三个级别依次递进，高级别涵盖低级别职业技能要求。

【大数据平台运维】（初级）：主要面向大数据平台安装配置、大数据组件安装配置、大数据平台基础实施、大数据平台简单维护及监控工作岗位。从事虚拟化软件安装与使用、基于 Linux 系统的常用服务安装配置、安装配置及运行 Hadoop 集群、安装配置及运行核心组件、执行客户大数据平台实施方案、监控大数据平台运行状态、查看管理大数据平台日志信息、监控大数据平台服务和资源状态等工作，掌握大数据平台安装和配置方法，理解 Hadoop 核心组件的功能及工作原理，掌握关键组件安装配置方法，理解大数据平台实施流程，熟悉常用集群监控工具的使用方法。

【大数据平台运维】（中级）：主要面向大数据平台高可用性部署实施、大数据组件维护及使用、大数据平台维护及大数据平台优化等工作岗位。从事 Hadoop 高可用集群部署及配置、Hadoop 组件维护及使用、Hadoop 集群节点管理及维护、大数据平台故障诊断及维护等工作，掌握高可用集群 Hadoop 配置方法，熟练使用 shell，掌握 Hadoop 集群节点及其服务的增删改等基本操作方法，熟悉常用系统性能诊断工具及集群监控管理工具，能独立排查和解决大数据平台常见问题，优化集群性能。

【大数据平台运维】（高级）：主要面向大数据平台规划、大数据平台安全管理、大数据平台资源管理、大数据平台优化及升级等工作岗位。从事大数据集群软硬件配置方案拟定、Hadoop 架构方案设计、Hadoop 组件部署方案规划、Hadoop 安全机制规划与实现、大数据平台资源配置及管理、大数据平台优化拓展与升级等工作，熟练运用 shell 进行平台运维，熟练掌握 Hadoop 生态圈组件的工作原理和使用方法，掌握

Hadoop 集群的安全管理机制和方法，熟悉 Hadoop 资源配置和管方法，掌握大数据平台优化策略和方法，熟练 Hadoop 集群软硬件升级方法和操作。

6.2 职业技能等级要求描述

表 1 大数据平台运维职业技能等级要求（初级）

工作领域	工作任务	职业技能要求
1.大数据平台安装	1.1 虚拟化软件使用	能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立创建虚拟机
		能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立删除虚拟机
		能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能

		<p>独立修改虚拟机</p> <p>能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立访问虚拟机</p>
	1.2 Linux 操作系统 安装和 使用	<p>能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立 Linux 操作系统安装</p> <p>能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立区分 Linux 系统种类</p> <p>能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立设置 root 密码</p> <p>能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立下载 Linux SSH 服务</p> <p>能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立安装 Linux SSH 服务</p> <p>能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立安装 Linux JDK</p> <p>能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立配置 JDK 环境变量</p>
	1.3Hadoop 平台安装	<p>能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立下载 Hadoop 安装包</p> <p>能根据大数据平台安装工作任务要求，按照大数据平台安装指导书，能独立解压 Hadoop 安装包</p> <p>能根据大数据平台安装工作任务要求，按照大数据平台安装指导书，能独立配置 Hadoop 环境变量</p> <p>能根据大数据平台安装工作任务要求，按照大数据平台安装指导书，能独立单机 Hadoop 安装</p> <p>能根据大数据平台安装工作任务要求，按照大数据平台安装指导书，能独立伪分布 Hadoop 安装</p> <p>能根据大数据平台安装工作任务要求，按照大数据平台安装指导书，能独立完全分布 Hadoop 安装</p>
2.大数据 平台 配置	2.1 主机 的基础环 境配置	<p>能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置主机网络 IP</p> <p>能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置主机 DNS</p> <p>能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置主机名</p> <p>能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能配置集群局域网络连通</p> <p>能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立生成 SSH 密钥对</p> <p>能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置 SSH 认证文件</p>
	2.2Hadoop	能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独

	<p>文件参数配置</p>	<p>立配置 <code>hdfs.site.xml</code> 文件参数 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置 <code>core.site.xml</code> 文件参数 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置 <code>mapred.site.xml</code> 文件参数 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置 <code>yarn.site.xml</code> 文件参数</p>
	<p>2.3 Hadoop 集群运行</p>	<p>能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置 Hadoop 格式化 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立启动 Hadoop 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立查看 Java 进程 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立查看 HDFS 的报告 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立使用浏览器浏览 Master 节点，查看节点状态 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立停止 Hadoop</p>
<p>3.大数据平台组件安装配置</p>	<p>3.1 HBase 组件安装配置</p>	<p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能了解 HBase 基本知识 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 HBase 基础环境 能根据大数据平台组件安装配置工作任务要求，按照大数据平台配置指导书，能独立安装 HBase 组件 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 HBase 组件 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立测试 HBase 组件功能</p>
	<p>3.2 Hive 组件安装配置</p>	<p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能了解 Hive 基本知识 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Hive 基础环境 能根据大数据平台组件安装配置工作任务要求，按照大数据平台配置指导书，能独立安装 Hive 组件 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Hive 组件 3.2.5 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立测试 Hive 组件功能</p>
	<p>3.3 Zookeeper 组件安装配置</p>	<p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能了解 Zookeeper 基本知识 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Zookeeper 基础环境</p>

		<p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台配置指导书，能独立安装 Zookeeper 组件</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Zookeeper 组件</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立测试 Zookeeper 组件功能</p>
	<p>3.4 Sqoop 组件 安装 配置</p>	<p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能了解 Sqoop 基本知识</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Sqoop 基础环境</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台配置指导书，能独立安装 Sqoop 组件</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Sqoop 组件能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立测试 Sqoop 组件功能</p>
	<p>3.5 Flume 组 件 安 装 配 置</p>	<p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能了解 Flume 基本知识</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Flume 基础环境</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台配置指导书，能独立安装 Flume 组件</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Flume 组件</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立测试 Flume 组件功能</p>
<p>4.大数据 平台 实施</p>	<p>4.1 大数 据平台 实施方 案的制 定</p>	<p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能正确理解客户需求</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能编写客户大数据平台实施方案</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能按要求正确执行客户大数据平台实施方案</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能验证客户大数据平台实施方案</p>
	<p>4.2 客户 培训方 案制 定</p>	<p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能独立使用文档制作工具</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能制定客户培训方案</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能独立演示操作客户大数据平台</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能培训客户使用大数据平台</p>

	<p>4.3 大数据平台出现的基本问题处理</p>	<p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能解决平台基础环境的基本问题</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能解决组件配置的基本问题</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能解决平台启动的基本问题</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能解决平台运行的基本问题</p>
<p>5.大数据平台监控</p>	<p>5.1 大数据平台运行状态监控</p>	<p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过图形监控运行状态</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过命令监控运行状态</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立制作大数据平台运行状态报表</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立查找运行状态异常基本问题</p>
	<p>5.2 大数据平台资源状态监控</p>	<p>根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过图形监控资源状态</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过命令监控资源状态</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立制作大数据平台资源状态报表</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立查找资源状态异常基本问题</p>
	<p>5.3 大数据平台告警信息监控</p>	<p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过图形监控告警信息</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过命令监控告警信息</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立制作大数据平台告警信息报表</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立查找告警信息异常基本问题</p>
	<p>5.4 大数据平台服务状态监控</p>	<p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过图形监控服务状态</p> <p>根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过命令监控服务状态</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立制作大数据平台服务状态报表</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立查找服务状态异常基本问题</p>
	<p>5.5 大数据平台日志信息</p>	<p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过图形监控日志信息</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指</p>

	监控	<p>导书，能通过命令监控日志信息</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立制作大数据平台日志信息报表</p> <p>能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能分析基本日志信息</p>
--	----	--

表 2 大数据平台运维职业技能等级要求（中级）

工作领域	工作任务	职业技能要求
1. 大数据平台系统架构	1.1 大数据特点及行用场景	<p>能根据大数据平台系统架构工作任务要求，按照大数据平台系统架构指导书，掌握大数据平台系统架构特点</p> <p>能根据大数据平台系统架构工作任务要求，按照大数据平台系统架构指导书，掌握大数据平台系统架构原理</p> <p>能根据大数据平台系统架构工作任务要求，按照大数据平台系统架构指导书，掌握大数据平台系统架构典型企业应用场景</p> <p>能根据大数据平台系统架构工作任务要求，按照大数据平台系统架构指导书，掌握大数据平台系统架构典型行业应用场景</p>
	1.2 大数据实施和运维流程	<p>能根据大数据平台系统架构工作任务要求，按照大数据平台系统架构指导书，掌握大数据实施和运维的工作职责</p> <p>能根据大数据平台系统架构工作任务要求，按照大数据平台系统架构指导书，掌握大数据实施和运维的工作能力素养</p> <p>能根据大数据平台系统架构工作任务要求，按照大数据平台系统架构指导书，掌握大数据项目实施工作流程</p> <p>能根据大数据平台系统架构工作任务要求，按照大数据平台系统架构指导书，掌握大数据运维的工作流程</p>
	1.3 大数据发展趋势	<p>能根据大数据平台系统架构工作任务要求，按照大数据平台系统架构指导书，掌握大数据发展历程</p> <p>能根据大数据平台系统架构工作任务要求，按照大数据平台系统架构指导书，掌握大数据未来发展趋势</p> <p>能根据大数据平台系统架构工作任务要求，按照大数据平台系统架构指导书，掌握 Hadoop 生态圈中企业应用典型实战案例</p> <p>能根据大数据平台系统架构工作任务要求，按照大数据平台系统架构指导书，掌握 Hadoop 生态圈中行业应用典型实战案例</p>
2. 大数据平台高可用部署	2.1 Hadoop 集群基础环境配置	<p>能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立配置集群主机之间时钟同步</p> <p>能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立关闭防火墙</p> <p>能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立配置 selinux</p> <p>能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立配置 IP 与主机名映射关系</p> <p>能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立配置 IP 与主机名映射关系</p> <p>能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立配置 IP 与主机名映射关系</p>

		部署指导书，能独立配置环境变量
	2.2 HadoopHA 集群配置	能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立配置高可用 Zookeeper 集群大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立配置高可用集群 Hadoop 文件参数能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立配置初始化 JournalNode服务 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立启动 JournalNode 服务
	2.3 Hadoop HA 集群 启动	能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立初始化 namenode 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立复制 namenode 元数据到其它节点 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立格式化 zkfs 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立启动 Hadoop 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立启动 Yarn 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立启动 MapReduce 任务历史服务器
3. 大数据 组件维护	3.1 Hbase 组件维护	能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立掌握 HBase 组件原理 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立卸载 HBase 组件 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立创建 HBase 库和表 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立在 HBase 库和表中查询、增加、切换、删除操作 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立在 HBase 库和表中插入数据 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立退出 HBase 库
	3.2 Hive 组件维护	能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立掌握 Hive 组件原理 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立卸载 Hive 组件 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立创建 Hive 库和表 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立在 Hive 库和表中查询、增加、切换、删除操作 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立在 Hive 库和表中插入数据 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护

		指导书, 能独立退出 Hive 库
	3.3 Zookeeper 组件维护	<p>能根据大数据平台组件维护工作任务要求, 按照大数据平台组件维护指导书, 能独立卸载 Zookeeper 组件</p> <p>能根据大数据平台组件维护工作任务要求, 按照大数据平台组件维护指导书, 能独立修改 Zookeeper 组件</p> <p>能根据大数据平台组件维护工作任务要求, 按照大数据平台组件维护指导书, 能独立配置 Zookeeper 决策选举</p>
	3.4 ETL 组件维护	<p>能根据大数据平台组件维护工作任务要求, 按照大数据平台组件维护指导书, 能独立卸载 Sqoop 组件</p> <p>能根据大数据平台组件维护工作任务要求, 按照大数据平台组件维护指导书, 能独立修改 Sqoop 组件</p> <p>能根据大数据平台组件维护工作任务要求, 按照大数据平台组件维护指导书, 能独立使用 Sqoop 进行数据发送和接收</p> <p>能根据大数据平台组件维护工作任务要求, 按照大数据平台组件维护指导书, 能独立下载 Flume 组件</p> <p>能根据大数据平台组件维护工作任务要求, 按照大数据平台组件维护指导书, 能独立修改 Flume 组件</p>
	3.5 Spark 组件维护	<p>能根据大数据平台组件维护工作任务要求, 按照大数据平台组件维护指导书, 能独立安装配置 Scala</p> <p>能根据大数据平台组件维护工作任务要求, 按照大数据平台组件维护指导书, 能独立安装 Spark</p> <p>能根据大数据平台组件维护工作任务要求, 按照大数据平台组件维护指导书, 能独立修改 Spark 参数</p> <p>能根据大数据平台组件维护工作任务要求, 按照大数据平台组件维护指导书, 能独立启动 Spark</p> <p>能根据大数据平台组件维护工作任务要求, 按照大数据平台组件维护指导书, 能独立卸载 Spark</p>
4. 大数据平台优化	4.1 Linux 系统优化	<p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 Linux 系统的内存</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 Linux 系统网络</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 Linux 系统磁盘</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 Linux 文件系统</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 Linux 系统缓冲区</p>
	4.2 HDFS 配置优化	<p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 dfs.replication 文件副本数</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立设置 dfs.block.size 数据块大小</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书,</p>

		<p>能独立设置 <code>mapred.local.dir</code> 磁盘目录</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立设置 <code>dfs.data.dir</code> 优化 IO 读写能力</p>
	<p>4.3 MapReduce 配置优化</p>	<p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 Reduce 节点个数</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 Reduce I/O 相关参数</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 reuduce shuffle 阶段并行传输数据的数量</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 tasktracker 并发执行的 reduce 数</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化可并发处理来自 tasktracker 的 RPC请求数</p>
	<p>4.4 Spark 配置优化</p>	<p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 Spark Streaming 配置</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 Spark 读取 Kafka</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化读取 Flume</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导, 能独立优化 Spark 写入HDFS</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导, 能独立优化 Spark Scala 代码</p>
<p>5. 大数据平台诊断与处理</p>	<p>5.1 Hadoop 及生态圈组件负载均衡诊断与处理</p>	<p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立对 Hadoop 集群及生态圈组件负载均衡进行诊断</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立对高可用 Hadoop 集群及生态圈组件负载均衡进行诊断</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立制定 Hadoop 及生态圈组件负载均衡问题解决方案</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立处理 Hadoop 及生态圈组件负载均衡问题</p>
	<p>5.2 集群节点故障诊断与处理</p>	<p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立使用集群日志对节点故障进行诊断</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立使用集群告警信息诊断节点故障</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立制定集群节点故障解决方案</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立处理集群节点故障</p>
		<p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书,</p>

	5.3 集群组件服务故障诊断处理	<p>能独立使用集群日志诊断组件服务故障问题</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立使用集群告警信息诊断组件服务故障问题</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立制定集群告警信息诊断组件服务故障问题解决方案</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立处理集群告警信息诊断组件服务故障问题</p>
--	------------------	---

表 3 大数据平台运维职业技能等级要求 (高级)

工作领域	工作任务	职业技能要求
1. 大数据平台规划	1.1 Hadoop 集群选型	<p>能根据大数据平台规划工作任务要求, 按照大数据平台规划指导书, 能正确选择处理器、内存、存储、网络</p> <p>能根据大数据平台规划工作任务要求, 按照大数据平台规划指导书, 能合理制定集群硬件配置方案</p> <p>能根据大数据平台规划工作任务要求, 按照大数据平台规划指导书, 能独立设计大数据集群网络方案</p> <p>能根据大数据平台规划工作任务要求, 按照大数据平台规划指导书, 能独立设计大数据集群网络高可用方案</p>
	1.2 Hadoop 平台架构设计	<p>能根据大数据平台规划工作任务要求, 按照大数据平台规划指导书, 能掌握规划 Hadoop 集群节点高可用方案</p> <p>能根据大数据平台规划工作任务要求, 按照大数据平台规划指导书, 能掌握规划 Hadoop 集群容量方案</p> <p>能根据大数据平台规划工作任务要求, 按照大数据平台规划指导书, 能掌握规划 Hadoop 行业方案</p> <p>能根据大数据平台规划工作任务要求, 按照大数据平台规划指导书, 能掌握规划 Hadoop 企业方案</p>
	1.3 Hadoop 组件部署	<p>能根据大数据平台规划工作任务要求, 按照大数据平台规划指导书, 能合理选择 Hadoop 集群组件</p> <p>能根据大数据平台规划工作任务要求, 按照大数据平台规划指导书, 能合理选择 Hadoop 集群组件版本</p> <p>能根据大数据平台规划工作任务要求, 按照大数据平台规划指导书, 能掌握 Hadoop 生态组件的工作原理</p> <p>能根据大数据平台规划工作任务要求, 按照大数据平台规划指导书, 能了解 Hadoop 生态圈的发展趋势</p>
2. 大数据平台安全管理	2.1 hadoop 安全机制实现	<p>能根据大数据平台安全管理工作任务要求, 按照大数据平台安全管理指导书, 能独立使用 HDFS 的各类安全机制</p> <p>能根据大数据平台安全管理工作任务要求, 按照大数据平台安全管理指导书, 能独立使用 HDFS 安全组件管理 HDFS 安全</p> <p>能根据大数据平台安全管理工作任务要求, 按照大数据平台安全管理指导书, 能独立使用 MapReduce 的各类安全机制</p> <p>能根据大数据平台安全管理工作任务要求, 按照大数据平台安全管理指导书, 能独立使用 MapReduce 安全组件管理 MapReduce 安全</p> <p>能根据大数据平台安全管理工作任务要求, 按照大数据平台安全管理指导书, 能独立使用 MapReduce 安全组件管理 MapReduce 安全</p>

		<p>据平台安全管理指导书，能独立使用 Yarn 的各类安全机制 能根据大数据平台安全管理工作任务要求，按照大数据平台安全管理指导书，能独立使用Yarn 安全组件管理Yarn安全</p>
	<p>2.2 大数据平台安全体系实现</p>	<p>能根据大数据平台安全管理工作任务要求，按照大数据平台安全管理指导书，能独立配置大数据平台安全 能根据大数据平台安全管理工作任务要求，按照大数据平台安全管理指导书，能独立配置大数据平台访问安全 能根据大数据平台安全管理工作任务要求，按照大数据平台安全管理指导书，能独立配置大数据平台存储安全 能台安全管理工作任务要求，按照大数据平台安全管理指导书，能独立添加 Hadoop 组件的控制权限 能根据大数据平台安全管理工作任务要求，按照大数据平台安全管理指导书，能独立修改 Hadoop 组件的控制权限</p>
<p>3. 大数据平台资源治理</p>	<p>3.3 大数据平台资源池治理</p>	<p>能根据大数据平台资源治理工作任务要求，按照大数据平台资源治理指导书，能独立配置静态资源池 能根据大数据平台资源治理工作任务要求，按照大数据平台资源治理指导书，能独立查看静态资源池状态 能根据大数据平台资源治理工作任务要求，按照大数据平台资源治理指导书，能独立配置动态资源池 能根据大数据平台资源治理工作任务要求，按照大数据平台资源治理指导书，能独立使用动态资源池管理资源</p>
	<p>3.1 大数据平台治理体系</p>	<p>能根据大数据平台资源治理工作任务要求，按照大数据平台资源治理指导书，掌握大数据资产地位确定 能根据大数据平台资源治理工作任务要求，按照大数据平台资源治理指导书，掌握大数据管理体系机制 能根据大数据平台资源治理工作任务要求，按照大数据平台资源治理指导书，掌握大数据的数据共享与开放 能根据大数据平台资源治理工作任务要求，按照大数据平台资源治理指导书，掌握大数据安全与隐私保护</p>
	<p>3.2 大数据平台数据治理</p>	<p>能根据大数据平台资源治理工作任务要求，按照大数据平台资源治理指导书，能独立对元数据治理 能根据大数据平台资源治理工作任务要求，按照大数据平台资源治理指导书，能独立对资源池中敏感数据进行标记 能根据大数据平台资源治理工作任务要求，按照大数据平台资源治理指导书，能独立对数据源与库中的关键字对接 能根据大数据平台资源治理工作任务要求，按照大数据平台资源治理指导书，能独立对数据源与库中的关键字整合</p>

4. 大数据平台优化	4.1 Linux 系统优化	<p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立避免使用 swap 分区优化</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立调整内存分配策略</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立修改 socket 监听参数</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化同时打开文件描述符上限</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立修改 Transparent Huge Pages</p>
	4.2 Hadoop 应用程序优化	<p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立减少大量小文件输入</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立合理使用分布式缓存</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立合理重用写数据类型</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立对 JVM 缓存调优</p>
	4.3 Hadoop 组件性能优化	<p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立配置 HDFS 集中缓存管理</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 MapReduce 调度参数</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 Yarn 内存配置</p> <p>能根据大数据平台优化工作任务要求, 按照大数据平台优化指导书, 能独立优化 Spark 配置</p>
5. 大数据平台升级	5.1 大数据平台升级	<p>能根据大数据平台升级工作任务要求, 按照大数据平台升级指导书, 能独立升级大数据平台处理器</p> <p>能根据大数据平台升级工作任务要求, 按照大数据平台升级指导书, 能独立升级大数据平台内存</p> <p>能根据大数据平台升级工作任务要求, 按照大数据平台升级指导书, 能独立升级大数据平台存储</p> <p>能根据大数据平台升级工作任务要求, 按照大数据平台升级指导书, 能独立升级大数据平台网络</p>
	5.2 大数据平台版本升级	<p>能根据大数据平台升级工作任务要求, 按照大数据平台升级指导书, 能独立升级 Hadoop 集群中的 HDFS 配置至高版本</p> <p>能根据大数据平台升级工作任务要求, 按照大数据平台升级指导书, 能独立升级 Hadoop 集群中的 MapReduce 配置至高版本</p> <p>能根据大数据平台升级工作任务要求, 按照大数据平台升级指导书, 能独立验证升级后的 Hadoop 集群运行正常</p> <p>能根据大数据平台升级工作任务要求, 按照大数据平台升级指导书, 能独立验证升级后的 Hadoop HA 集群运行正常</p>

	<p>5.3 Hadoop 组件版本 升级</p>	<p>能根据大数据平台升级工作任务要求，按照大数据平台升级指导书，能独立升级 Hadoop 组件至高版本 能根据大数据平台升级工作任务要求，按照大数据平台升级指导书，能独立验证升级后的 Hadoop 组件运行兼容性 能根据大数据平台升级工作任务要求，按照大数据平台升级指导书，能独立验证升级后的 Hadoop 组件集群运行兼容性 能根据大数据平台升级工作任务要求，按照大数据平台升级指导书，能独立验证升级后的 Hadoop 组件之间运行兼容性</p>
--	---------------------------------------	---

参考文献

- [1]国务院关于促进云计算创新发展培育信息产业新业态的意见
- [2]中等职业学校专业目录（征求意见稿）
- [3]普通高等学校高等职业教育（专科）专业目录及专业简介（截至 2018 年）
- [4]普通高等学校本科专业目录（2012 年）
- [5]中等职业学校专业教学标准（试行）
- [6]高等职业学校专业教学标准（2018 年）
- [7]本科专业类教学质量国家标准
- [8]2019 年全国职业院校技能大赛 GZ-2019032 大数据技术与应用赛项规程
- [9]国家职业技能标准编制技术规程（2018 年版）
- [10]中华人民共和国职业分类大典
- [11]战略性新兴产业分类（2018）
- [12]GB/T 4754-2017 国民经济行业分类
- [13]GB/T 5271.1-2000 《信息技术 词汇 第 1 部分：基本术语》
- [14]GB/T 1.1-2009 标准化工作导则
- [15]GB/T 35295-2017 《信息技术 大数据 术语》
- [16]GB/T 35274-2017 《信息安全技术大数据服务安全能力要求》
- [17]GB/T 35589-2017 《信息技术大数据技术参考模型》
- [18]GB/T 36073-2018 《数据管理能力成熟度评估模型》
- [19]大数据安全标准化白皮书（2018 版）
- [20]大数据标准化白皮书（2018 版）
- [21] GB/T 37973-2019 《信息安全技术大数据安全管理指南》
- [22] GB/T 37722-2019 《信息技术大数据存储与处理系统功能要求》
- [23] GB/T 37721-2019 《信息技术大数据分析系统功能要求》
- [24] ISO/IEC 20547-4 《信息技术 大数据参考架构 第 4 部分：安全与隐私保护》
- [25] ITU-T Y.3600 《大数据 基于云计算的要求和能力》

附件3 大数据技术与应用专业核心课程标准

《Python 程序设计》课程标准

一、课程基本信息

课程名称	Python 程序设计				
课程代码	0917006	学时	72	学分	4
授课时间	第三学期	适用专业	大数据技术与应用专业		
课程性质	岗位能力课程				
先修课程	Java 程序设计	后续课程	Hadoop 系统及应用		

二、课程定位

本课程是大数据技术与应用专业的岗位能力课程，目的是通过课堂教学和实践教学相结合，使学生能够深入理解 Python 及面向对象概念，清楚的了解 Python 软件开发工作流程，建立起应用程序的概念，最终能够掌握 Python 软件开发的基本方法、基本技能，培养利用 Pygame 等库进行 Python 应用软件产品的分析、设计、编码、测试的综合应用能力，培养科学的思维方法，灵活运用知识的能力，实验操作能力，使学生具有较强的发现问题、分析问题、解决问题的能力，具有毕业后的直接上岗能力（或经短期培训后上岗），并且毕业时已具备一定的 Python 软件开发经验。

三、课程设计思路

本课程主要以教材为参考，融入经典的项目案例，同时借鉴了一些企业级项目模块的雏形，有利于学生的理解与学习。不失传统性，任何一门语言课程都要了解该语言的书写风格与运行的平台特点，但是在学习过程中，又主要通过项目实例让学生进一步学习基础知识同时也见到了项目雏形。通过前期近一百多课时的学习，总体上掌握该门课程的特点与学习思路，然后利用较综合性项目培养学生对软件组织结构、思想、过程的设计。新课程理念要求充分的调动学生思考和动手能力，先学后教，精讲多练。

四、课程目标

（一）职业知识目标

1. 掌握 Python 语言基础；
2. 熟练掌握对象和类；
3. 掌握继承与多态；

4. 熟练掌握数组和字符串;
5. 掌握 Python 的异常处理;
6. 掌握 Python 的输入/输出;
7. 了解 GUI 程序设计;
8. 熟练掌握 Python 的集合;
9. 了解面向对象的常用设计模式;
10. 掌握 Python 程序设计的思想和方法。

(二) 职业能力目标

1. 培养学生规范编码和良好的程序设计风格;
2. 培养学生面向对象编程的思维和提高逻辑思维能力;
3. 培养学生发现问题、分析问题和解决问题的能力。

(三) 职业素质目标

1. 培养按时完成项目的良好习惯, 增强集体和社会责任感;
2. 培养团队协作和良好的沟通能力;
3. 培养学生务实、创新、诚实、守信的性格。

五、课程内容及要求

序号	教学内容	能力目标	知识目标	教学方法及手段	学时
1	计算机和程序	了解计算机软硬件及 Python 程序内容	<ol style="list-style-type: none"> 1. 了解计算系统中硬件和软件各自的作用。 2. 学习计算机科学家研究的领域和他们使用的技术。 3. 了解现代计算机的基本设计。 3. 了解计算机编程语言的形式和功能。 4. 开始使用 Python 编程语言。 	由教师示范、指导、讲解	4
2	编写简单程序	能够熟练应用输入输出赋值及循环语句完成简单程序编写。	<ol style="list-style-type: none"> 1. 知道有序的软件开发过程的步骤。 2. 了解遵循输入、处理、输出 (IPO) 模式的程序, 并能够以简单的方式修改它们。 3. 了解构成有效 Python 标识符和表达式的规则。 4. 能够理解和编写 Python 语句, 将信息输出到屏幕, 为变量赋值, 获取通过键盘输入的信息, 并执行计数循环。 	由教师示范、指导, 学生小组协作	4
3	数字计算	掌握数字计算	<ol style="list-style-type: none"> 1. 理解数据类型的概念。 2. 熟悉 Python 中的基本数值数据类型。 3. 理解数字在计算机上如何表示的基本原理。 4. 能够使用 Python 的 math 库。 5. 理解累积器程序模式。 6. 能够阅读和编写处理数值数据的程序。 	由教师示范、指导、学生小组协作完成	4
4	对象和图	掌握图形及 graphics 模块	<ol style="list-style-type: none"> 1. 理解对象的概念以及如何用它们来简化编程。 2. 熟悉 graphics 库中可用的各种对象。 3. 能够在程序中创建对象并调用适当的方法来 	由教师示范、指导、学生独立编程实现点、	4

			进行图形计算。 4. 了解计算机图形学的基本概念，特别是坐标系统和坐标变换的作用。 5. 了解如何在图形编程语境中使用基于鼠标和基于文本的输入。 6. 能够使用 graphics 库编写简单的交互式图形程序。	圆、圆柱类的设计与实现	
5	序列、字符串、列表和文	掌握序列、字符串、列表和文件的操作	1. 了解字符串数据类型以及如何在计算机中表示字符串。 2. 熟悉通过内置函数和字符串方法对字符串执行的各种操作。 3. 理解序列和索引的基本概念，因为它们适用于 Python 的字符串和列表。 4. 能够用字符串格式化来产生有吸引力的、富含信息的程序输出。 5. 了解在 Python 中读取和写入文本文件的基本文件处理概念和技术。 6. 了解加密的基本概念。 7. 理解和编写处理文本信息的程序。	由教师示范、指导、学生独立编程实现图形类的设计与实现	4
6	定义函数	掌握函数的编写	1. 了解程序员为什么将程序分成多组合作的函数。 2. 能够在 Python 中定义新的函数。 3. 理解 Python 中函数调用和参数传递的细节。 4. 利用函数来编程，减少代码重复并增加程序的模块性。	由教师示范、指导、学生小组协作完成图形类的设计与实现	4
7	判断结构	掌握判断语句	1. 利用 Python 的 if 语句来理解简单的判断编程模式及其实现。 2. 利用 Python 的 if-else 语句来理解两路判断编程模式及其实现。 3. 利用 Python 的 if-elif-else 语句来理解多路判断编程模式及其实现。 4. 理解异常处理的思想，并能够编写简单异常处理代码，捕捉标准的 Python 运行时错误。 5. 理解布尔表达式和布尔数据类型的概念。 6. 能够阅读、编写和实现使用判断结构的算法，包括使用系列判断和嵌套判断结构的算法。	由教师示范、指导、学生自主学习	4
8	循环结构和布尔	掌握循环结构和布尔值	1. 理解确定和不定循环的概念，以及它们用 Python 的 for 和 while 语句的实现。 2. 理解交互式循环和哨兵循环的编程模式，以及它们用 Python 的 while 语句的实现。 3. 理解文件结束循环的编程模式，以及在 Python 中实现这种循环的方法。 4. 能为涉及循环模式（包括嵌套循环结构）的问题设计和实现解决方案。 5. 理解布尔代数的基本思想，并能分析和编写涉及布尔运算符的布尔表达式。	由教师示范、指导、学生自主学习	4
9	模拟与设计	理解并掌握模拟与设计	1. 理解模拟的应用可能是解决现实问题的一种方式。 2. 理解伪随机数及其在蒙特卡罗模拟中的应用。 3. 理解并能应用自顶向下和螺旋式设计技术来编写复杂的程序。 4. 理解单元测试，并能将这种技术应用于复杂程序的实现和调试。	由教师示范、指导、学生自主学习认识常用组件	4
10	定义类	掌握类、对象、封装和控	1. 领会定义新类如何能为复杂程序提供结构。 2. 能够阅读并编写 Python 类定义。	由教师示范、指导、	4

		件	3. 理解封装的概念，以及它如何有助于构建模块化的、可维护的程序。 4. 能够编写包含简单类定义的程序。 5. 能够编写包含创新（程序员设计的）控件的交互式图形程序。	学生小组学习认容器与菜单	
11	数据集合	掌握数组的使用	1. 了解使用列表（数组）来表示相关数据的集合。 2. 熟悉用于操作 Python 列表的函数和方法。 3. 能够编程用列表管理信息集合。 4. 能够编程利用列表和类来构造复杂数据。 5. 了解用 Python 字典存储无顺序集合。	由教师示范、指导学生小组学习	4
12	面向对象设计	掌握面向对象思想及操作	1. 理解面向对象设计的过程。 2. 能够阅读和理解面向对象的程序。 3. 理解封装、多态和继承的概念，因为它们从属于面向对象的设计和编程。 4. 能够利用面向对象设计来设计中等复杂程度的软件。	由教师示范、指导学生小组学习	4
13	算法设计与递归	掌握查找、排序及递归算法	1. 理解分析算法效率的基本技巧。 2. 知道查找是什么，并且理解线性和二分查找的算法。 3. 理解递归定义和函数的基本原理，并能够编写简单的递归函数。 4. 深入理解排序，并理解选择排序和归并排序的算法。	由教师示范、指导学生小组学习完成画布上画线实训	4
14	Gui 的实现—Tkinter	掌握 Tkinter	1. 利用 kinter 实现 GUI	由教师示范、指导学生自主学习	8
15	综合实例：Pygame 游戏开发	掌握 Pygame	2. 利用游戏复习巩固前面的内容	由教师示范、指导学生自主学习	12

六、课程实施建议

（一）教学建议

1. 教学团队基本要求

担任本课程的教学团队由四到五名教师组成，所有主讲教师需要具有丰富的一线教学经验，能进行一体化教学，能按照本课程标准制定详细可行的授课计划，精心设计每一次课的教学过程。教学团队所有教师需具备一定的课堂控制能力和应变能力，做到因材施教。熟悉 python 语音的知识体系，具备 python 语言开发技能，具有资源开发能力。

2. 教学条件

实训室应具备相应的软硬件实训环境，学习版本应在 python3.7 及以上。

3. 教学方法与手段

本课程的实施，不仅应充分利用已开发的课程资源以及校外的课程资源，如相关企业的岗位实习和实践，做到校企结合、工学结合，开展理实一体化教学。

充分利用网络资源，了解最新的技术策略和实施方法，培养学生自主学习的能力。

4. 课程资源的开发与利用

(1) 学习包和教材，都是素材性的课程资源，但教材是知识的载体，而学习包是引导学生学习的载体。因此，要开发学习包，突出以学生为中心的学习过程，将本课程的教学内容与地区经济的发展密切结合，并体现学生职业生涯发展的需要。

(2) 教学中学生的问题、困惑、见解、情感和体验等都是动态生成的课程资源，老师应重视这些教学过程中动态生成的课程资源，并主动性和创造性地运用，使以项目为主的教学充满生机与活力。

(4) 积极开发教学资源：课程标准、实训指导书、授课计划等教学文件，以及课件、习题、案例库、网络方案、工具软件和网络资源等。

5. 教材选用

选用的教材必须符合本专业人才培养目标及课程教学的要求，适应高职高专学生使用。教材选用要优先选用教育部的国家级高职高专规划教材，提高国家规划教材的使用率，其次选择高等教育出版社及相关专业行业出版社的高职教材。

推荐参考资料：

- (1) 《Python 程序设计》（第 2 版） 清华大学出版社
- (2) 《从零基础入门学习 Python》 清华大学出版社
- (3) 《Python 程序设计基础》 人民邮电出版社
- (4) 与 Python 程序设计相关的其它书籍、网络资源

(二) 考核建议

考核主要由平时成绩、理论考试以及实训项目成绩组成。

1. 平时成绩

平时成绩强调平时的出勤、课堂纪律、课堂表现和作业，教师通过千分制的实施可以在整个授课过程中监控学生的学习效果以及学习态度等等。平时成绩占全部考试的 30%。

2. 理论考试

理论考试采用传统的笔试方式，为了更好的达到考核的目的，我们采用第三方出题的形式来组织理论考试。理论考试占全部考试的 40%。

3. 项目评审

实施项目内部考核及教师考核相结合的原则，在项目实训末期，要进行项目评审，根据项目的完成情况给学生打分。项目评审成绩占全部成绩的 30%（包括学生自我评价

及对同伴的评价、完成项目效果评价以及教师评价，其中自我评价(占 30%)、项目效果评价(占 30%)、教师评价(占 40%)。)

《Hadoop 系统及应用》课程标准

一、课程基本信息

课程名称	Hadoop 系统及应用				
课程代码		学时	108	学分	6
授课时间	第三学期	适用专业	大数据技术与应用		
课程性质	岗位能力课程				
先修课程	Java 程序设计, MySQL 数据库	后续课程	Spark 技术与应用, 大数据可视化技术		

二、课程定位

本课程是大数据技术与应用专业的岗位能力课程，目的是通过课堂教学和实践教学相结合，使学生能够深入理解以 Hadoop 为核心的生态圈，阐述了基本的大数据平台运维技术。在 Hadoop 生态圈中，从基础的虚拟机搭建、分布式及高可用集群的部署、各个核心组件的部署及操作步骤；让我们深入了解大数据处理平台的运行与维护，结合实践，具有毕业后的直接上岗能力，并且毕业时已具备初级大数据平台部署实施、监控、管理、运行维护的经验。

三、课程设计思路

本课程主要以教材为参考，融入经典的项目案例，同时借鉴了一些企业级项目模块的雏形，有利于学生的理解与学习。在学习过程中，主要通过项目实例让学生进一步学习基础知识同时也见到了项目雏形。通过前期近一百多课时的学习，总体上掌握该门课程的特点与学习思路，然后利用较综合性项目对 Hadoop 平台的结构、部署过程、各组件的工作原理有更深刻的认识，为以后的数据分析打下良好的基础。

四、课程目标

(一) 职业知识目标

1. 了解大数据基本概念；
2. 了解 Hadoop；
3. 了解 Hive 和 HBase；
4. 了解 HDFS 运行原理。
5. 了解 HDFS 高级知识；
6. 了解 MapReduce 编程模型；
7. 了解 YARN；

8. 了解 HDFS 新特性;
9. 了解 YARN 新特性;
10. 掌握安装 Hadoop 平台;
11. 掌握 HDFS 基本操作;
12. 掌握 MapReduce 应用开发。

(二) 职业能力目标

1. 能从事虚拟化软件安装与使用;
2. 能安装配置及运行 Hadoop 集群;
3. 能正确配置、运行、维护核心组件;

(三) 职业素质目标

1. 培养按时完成项目的良好习惯, 增强团队责任感;
2. 培养团队协作操作能力和良好的沟通能力;
3. 培养学生大数据的敏感度和对数据库的理解。

五、课程内容及要求

序号	教学内容	能力目标	知识目标	教学方法及手段	学时
1	初始大数据: 大数据基本概念, 大数据带来的挑战	掌握大数据的基本概念和基本特性, 以及大数据带给社会, 企业的挑战有哪些	1. 了解大数据的基本概念和基本特性。 2. 了解大数据带来的挑战, 大数据对现有数据库的挑战, 实时性的技术挑战, 对数据中心, 运维团队的挑战	由教师示范、指导学生小组学习	2
2	初始 Hadoop: Hadoop 概述, Hadoop 生态圈, Hadoop 应用案例	掌握 Hadoop 的基本概念, 掌握 Hadoop 的核心构成, 了解 Hadoop 生态圈中各个组件的功能	1. 了解 Hadoop 的概念以及 Hadoop 发展史 2. 掌握 Hadoop 框架构成, 分布式文件系统, 分布式计算系统, 资源管理系统 3. 掌握 Hadoop 生态圈	由教师示范、指导学生小组学习	8
3	安装 Hadoop 平台: 安装虚拟机, 安装 Linux 系统, 配置虚拟机运行环境	掌握虚拟机, CentOS 7, 主机的网络属性配置 集群的网络连通配置 集群主机之间 SSH 免密登录	1. 了解虚拟机概述 2. 掌握虚拟机的安装方式, 掌握 Linux 系统的安装方式 3. 能熟练配置主机网络环境 4. 能配置集群局域网络连通	由教师示范、指导学生小组学习 安装虚拟机, 熟悉网络环境的配置	8
4	Hadoop 文件参数配置 Hadoop 集群启停	Hadoop 的安装 安装 Hadoop 完全分布式环境	1. 能正确配置 Hadoop 各项文件参数 2. 能正确启停集群, 查看集群进程	由教师示范、指导、学生独立编程实现 Hadoop 的安装及使用	8
5	大数据平台高可用部署	能配置集群主机之间时钟同步 能配置高可用 Zookeeper 集群 能正确配置高可用集群 Hadoop 文件参数 能启动并初始化	1. 群主机之间时钟同步配置 2. 高可用 Zookeeper 集群配置 3. 高可用集群 Hadoop 文件参数配置 4. JournalNode 服务启动初始化 5. 高可用集群启动	由教师示范、指导、学生独立搭建部署 Hadoop 的高可用集群	8

		JournalNode 服务能启动并验证高可用集群运行正常	6.高可用集群 HDFS 自动切换模式配置		
6	HDFS 运行原理：HDFS 读写流程，HDFS 副本机制，HDFS 负载均衡，HDFS 机架感知	掌握 HDFS 文件读写流程 了解 HDFS 的副本机制 了解 HDFS 文件数据的负载均衡和机架感知	1.掌握 HDFS 文件的读写流程 2.了解 HDFS 的副本机制以及容错机制。 3.了解 HDFS 负载均衡以及机架感知	由教师示范、指导、学生小组协作完成分布式文件系统 HDFS 的读写流程	8
7	认识 MapReduce 编程模型：MapReduce 基础，MapReduce 编程模型，MapReduce 词频统计编程实例	掌握 MapReduce 的设计目标， 了解 MapReduce 的特点与 MapReduce 编程模型。	1.了解序列化与反序列的概念 2.掌握 Hadoop 的序列化 3.掌握 SequenceFile 概述与特点以及读操作 MapFile 写操作	由教师示范、指导、学生自主学习 MapReduce 的编制与运行	8
8	认识 HBase 列式数据库，熟悉配置部署及访问方式。	熟悉 HBase 的分布式搭建 掌握 HBase 的 shell 操作	1.熟悉 HBase 结构体系。 2.HBase 的 Shell 使用及命令操作。	由教师示范、指导、学生自主学习，并完成 HBase 部署及 Shell 命令使用	12
9	认识 Hive 编程技术与应用，熟悉配置部署及访问方式。	掌握 Hive 的三种安装模式 熟悉 Hive 的 SQL 语句数据操作	1.熟悉 Hive 结构体系及数据模型。 2.Hive 的 SQL 命令使用。	由教师示范、指导、学生自主学习 Hive 的安装模式，HQL 的命令及操作。	12
10	MapReduce 高级应用：使用 MapReduce 实现 join 操作，使用 MapReduce 实现排序，使用 MapReduce 实现二次排序，使用 MapReduce 合并小文件	掌握 MapReduce 实现 join 操作。 掌握了 MapReduce 排序的实现原理。	1.熟练掌握 MapReduce 实现 join 操作。 2.实现 MapReduce 排序，理解原理。	由教师示范、指导、学生自主学习 MapReduce 实现 join 操作，多加练习 MapReduce 排序，理解其原理。	8
11	初始 YARN：YARN 产生背景，YARN 简介，YARN 架构设计	了解 YARN；背景，简介，架构设计	了解 YARN；背景，简介，架构设计	由教师示范、指导学生小组学习了解 YARN 是什么东西，理解其含义，以及应用场景	8
12	了解 HDFS 新特性：HDFS NameNode 高可用机制，HDFS NameNode Federation，HDFS REST API，DistCp 工具	了解 HDFS 新特性、HDFS NameNode 高可用机制。熟练掌握 HDFS NameNode Federation，HDFS REST API，DistCp 工具	1.了解 HDFS 新特性、HDFS NameNode 高可用机制。 2.熟练掌握 HDFS NameNode Federation，HDFS REST API，DistCp 工具	由教师示范、指导学生小组学习明白 HDFS 新特性、HDFS NameNode 高可用机制。熟练掌握课上所讲的一些工具。	10
13	了解 YARN 新特性：ResourceManager 自动重启，ResourceManager 高可用机制	了解 YARN 新特性，ResourceManager 自动重启概念、ResourceManager 高可用机制	1.了解 YARN 新特性， 2.ResourceManager 自动重启概念、ResourceManager 高可用机制	由教师示范、指导学生小组学习 YARN 新特性。 ResourceManager 自动重启概念、明白高可用机制。	8

六、课程实施建议

（一）教学建议

1. 教学团队基本要求

课程教学团队专兼职师资团队成员不少于 5 人。其中，行业企业授课教师不低于 40%，“双师型”授课教师中 60%应具有中级或高级职称，技术服务团队成员不少于 1 人。

2. 教学条件

硬件要求：微机、运行内存、网络。

软件要求：JDK、Hadoop、Linux、虚拟机等软件。

3. 教学方法与手段

本课程利用授课平台，进行理实一体化教学。

4. 课程资源的开发与利用

教学资料开发：

（1）结合实际软件开发工程案例，进行课件制作、教学环境模拟。

（2）与企业专家合作编写符合技能培养的实验实训教材，是学生学有所用。

（3）提供网络资源学习平台，使学生不拘泥于固定的时间、地点学习，提高学生学习的灵活性。

教学资源使用：

（1）学习包和教材，都是素材性的课程资源，但教材是知识的载体，而学习包是引导学生学习的载体。因此，要开发学习包，突出以学生为中心的学习过程，将本课程的教学内容与地区经济的发展密切结合，并体现学生职业生涯发展的需要。

（2）教学中学生的问题、困惑、见解、情感和体验等都是动态生成的课程资源，老师应重视这些教学过程中动态生成的课程资源，并主动性和创造性地运用，使以项目为主的教学充满生机与活力。

（3）本课程的实施，不仅应充分开发校内的课程资源，还应利用校外的课程资源，如相关企业的岗位实习和实践，尽量做到校企结合、工学结合。

（4）积极开发教学资源：课程标准、实训指导书、授课计划等教学文件，以及课件、习题、案例库、网络方案、工具软件和网络资源等。

充分利用网络资源，了解最新的技术策略和实施办法，培养学生自主学习的能力。

5. 教材选用

选用的教材必须符合本专业人才培养目标及课程教学的要求，适应高职高专学生使用。教材选用要优先选用教育部的国家级高职高专规划教材，提高国家规划教材的

使用率，其次选择高等教育出版社及相关专业行业出版社的高职教材。

建议选用教材：《Hadoop 应用与开发》 人民邮电出版社

推荐参考资料：

(1) 《Hadoop 实战 2》作者：陆嘉恒

(2) 《Hadoop 权威指南》

(3) 《HBase 入门与实践》

(4) 《Hive 编程指南》

(二) 考核建议

考核主要由平时成绩、理论考试以及实训项目成绩组成。

1. 平时成绩

平时成绩强调平时的出勤、课堂纪律、课堂表现和作业，教师通过千分制的实施可以在整个授课过程中监控学生的学习效果以及学习态度等等。平时成绩占全部考试的 30%。

2. 理论考试

理论考试采用传统的笔试方式，为了更好的达到考核的目的，我们采用第三方出题的形式来组织理论考试。理论考试占全部考试的 40%。

3. 项目评审

实施项目内部考核及教师考核相结合的原则，在项目实训末期，要进行项目评审，根据项目的完成情况给学生打分。项目评审成绩占全部成绩的 30%（包括学生自我评价及对同伴的评价、完成项目效果评价以及教师评价，其中自我评价（占 30%）、项目效果评价（占 30%）、教师评价（占 40%）。）

《HBase 数据库技术》课程标准

一、课程基本信息

课程名称	HBase 数据库技术				
课程代码	0917007	学时	72	学分	4
授课时间	第四学期	适用专业	大数据技术与应用		
课程性质	岗位能力课程				
先修课程	Java 程序设计, Hadoop 系统及应用	后续课程	Spark 技术与应用		

二、课程定位

HBase 是一个分布式的、面向列的开源数据库，不同于一般的关系数据库，它是一个适合于非结构化数据存储的数据库。本课程主要讲述了 HBase 详细的架构原理及特点、HBase 内部各个角色的详细介绍、安装配置、HBase 的 Shell 操作、新旧版本的读写数据详细流程、HBase 的 API 操作。本课程是计算机应用技术专业的岗位能力课程，目的是通过课堂教学和实践教学相结合，使学生能够深入理解 HBase 的概念、原理、服务架构，最终能够掌握使用 MapReduce 以及 Hive 对 HBase 数据分析、Rowkey 设计、预分区设计、调优策略安装部署、集群配置及 web 环境搭建及综合应用的方法。

三、课程设计思路

本课程主要以选的教材内容为主，融入经典的项目案例，同时借鉴了一些企业级项目模块的雏形，有利于学生的理解与学习。不失传统性，任何一门语言课程都要了解该语言的书写风格与运行的平台特点，但是在学习过程中，又主要通过项目实例让学生进一步学习基础知识同时也见到了项目雏形。本课程分为三个大的部分对 HBase 进行系统化的介绍，由浅入深，最后以一个实践的项目为例说明怎么使用 HBase 以及怎样的设计表能够提高查询性能，总体上掌握该门课程的特点与学习思路。

四、课程目标

（一）职业知识目标

1. 掌握 HBase 应用场景、环境搭建；
2. 了解与 RDBMS 比较的优缺点；
3. 掌握表的模型；
4. 熟练 HBase 架构；
5. 了解数据存储模型；

6. 掌握 HBase Java API 使用；
7. 掌握 HBase 与 MapReduce 集成；
8. 熟练掌握 Hbase 部署环境准备 ；
9. 掌握 Hbase 伪分布式的启动及 hbase 命令的使用；
10. 掌握 Hbase shell 中表的 DDL 操作；
11. 掌握 Hbase 的读写流程；
12. 掌握 Hbase 表的设计。

(二) 职业能力目标

1. 具备规范设计表、建立表的预分区的能力；
2. 具备调优策略、安装部署、集群配置及 web 环境搭建的能力。

(三) 职业素质目标

1. 培养大数据的分布式存储思维；
2. 培养团队协作和良好的沟通能力；
3. 培养学生知识创新、不断学习的精神。

五、课程内容及要求

序号	教学内容	能力目标	知识目标	教学方法及手段	学时
1	HBase 基本使用及存储设计	掌握 HBase 基本概念、掌握 HBase 环境搭建	1. HBase 的介绍及发展 2. HBase 与 MYSQL 的存储比较 3. Hbase 部署环境准备	由教师示范、指导学生小组学习	8
2	Hbase 伪分布式配置	了解所有进程运行在同一个节点上,不同进程运行在不同的 JVM 当中	1. Hbase 伪分布式配置文件的修改 2. Hbase 伪分布式的启动及 hbase 命令的使用	由教师示范、指导学生小组学习	8
3	Hbase shell 命令	掌握 Shell 基本命令,了解 HBase shell 中的帮助命令	1.namespace 的常用操作 2.表的 DDL 操作	由教师示范、指导学生小组学习	8
4	Hbase 的模型	掌握 HBase 物理模型、HBase 体系架构	1. Hbase 的物理模型 2. Hbase 的存储模型 3.HBase 的数据存储	由教师示范、指导学生小组学习	8
5	Hbase 的 namespace 表、meta 表及读写流程	掌握创建 namespace 删除 namespace 查看 namespace 列出全部 namespace	1. Hbase 的读写流程 2. Hbase 中各组件的应用	由教师示范、指导学生小组学习	8
6	Hbase Java API-环境配置	掌握配置 Hbase Java API 的基本步骤	1. Java API-通过 get 读取 2. Hbase 表中数据 3. Java API-put、delete 4. Java API-Scan 和过滤器	由教师示范、指导学生小组学习	8
7	Hbase 与 MapReduce 集	掌握 HBase 与 MapReduce 的集成整合	1. 环境变量的配置 2. rowcounter 测试	由教师示范、指导、学生自主学习	10

	成	与常用操作	3 代码实现自定义需求	MapReduce 的编制与运行	
8	案例：Hbase 的设计及企业优化	掌握基于具体的业务场景的实际动手操作	1. Hbase 表的设计 2. Hbase 性能优化-配置 snappy 压缩 3. PHoenix 的编译及安装部署	由教师示范、指导学生小组进行企业及项目的设计	14

六、课程实施建议

(一) 教学建议

1. 教学团队基本要求

课程教学团队专兼职师资团队成员 4 至 5 人。其中，行业企业授课教师不低于 30%， “双师型” 授课教师中 60%应具有中级或高级职称， 技术服务团队成员不少于 1 人。

2. 教学条件

硬件要求：微机、网络。

软件要求： JDK、Hadoop、Linux、虚拟机等软件。

3. 教学方法与手段

本课程利用教学平台进行理实一体化化教学。

4. 课程资源的开发与利用

课程资料开发：

- (1) 结合实际软件开发工程案例，进行课件制作、教学环境模拟。
- (2) 与企业专家合作编写符合技能培养的实验实训教材，是学生学有所用。

(3) 提供网络资源学习平台，使学生不拘泥于固定的时间、地点学习，提高学生学习的灵活性。

课程资源使用：

(1) 学习包和教材，都是素材性的课程资源，但教材是知识的载体，而学习包是引导学生学习的载体。因此，要开发学习包，突出以学生为中心的学习过程，将本课程的教学内容与地区经济的发展密切结合，并体现学生职业生涯发展的需要。

(2) 教学中学生的问题、困惑、见解、情感和体验等都是动态生成的课程资源，老师应重视这些教学过程中动态生成的课程资源，并主动性和创造性地运用，使以项目为主的教学充满生机与活力。

(3) 本课程的实施，不仅应充分开发校内的课程资源，还应利用校外的课程资源，如相关企业的岗位实习和实践，尽量做到校企结合、工学结合。

(4) 积极开发教学资源：课程标准、实训指导书、授课计划等教学文件，以及课

件、习题、案例库、网络方案、工具软件和网络资源等。

充分利用网络资源，了解最新的技术策略和实施方法，培养学生自主学习的能力。

5. 教材选用

选用的教材必须符合本专业人才培养目标及课程教学的要求，适应高职高专学生使用。教材选用要优先选用教育部的国家级高职高专规划教材，提高国家规划教材的使用率，其次选择高等教育出版社及相关专业行业出版社的高职教材。

建议选用教材：《HBase 不睡觉书》 ISBN:9787302490555 清华大学出版社

推荐参考资料：

(1) 《HBase 权威指南》 人民邮电出版社（“十二五”国家重点图书出版规划项目）

(2) 《HBase 实战》 人民邮电出版社

(3) 《HBase 入门与实践》 人民邮电出版社

(4) 《HBase 分布式存储系统应用》 水利水电出版社

(二) 考核建议

考核主要由平时成绩、理论考试以及实训项目成绩组成。

1. 平时成绩

平时成绩强调平时的出勤、课堂纪律、课堂表现和作业，教师通过千分制的实施可以在整个授课过程中监控学生的学习效果以及学习态度等等。平时成绩占全部考试的 30%。

2. 理论考试

理论考试采用传统的笔试方式，为了更好的达到考核的目的，我们采用第三方出题的形式来组织理论考试。理论考试占全部考试的 40%。

3. 项目评审

实施项目内部考核及教师考核相结合的原则，在项目实训末期，要进行项目评审，根据项目的完成情况给学生打分。项目评审成绩占全部成绩的 30%（包括学生自我评价及对同伴的评价、完成项目效果评价以及教师评价，其中自我评价（占 30%）、项目效果评价（占 30%）、教师评价（占 40%）。）

《Hive 编程技术与应用》课程标准

一、课程基本信息

课程名称	Hive 编程技术与应用				
课程代码	0917008	学时	72	学分	4
授课时间	第四学期	适用专业	大数据技术与技术专业		
课程性质	岗位能力课程				
先修课程	Hadoop 系统及应用	后续课程	Spark 技术与应用		

二、课程定位

本课程是计算机应用技术专业的岗位能力课程，目的是通过课堂教学和实践教学相结合，使学生能够深入理解 Hive，Hive 是一个构建在 Hadoop 上的数据仓库框架，是应 Facebook 每天生产的海量新兴社会网络数据进行管理和学习的需求而产生和发展的。解决 MapReduce 学习成本高、项目周期短、MapReduce 实现复杂查询逻辑开发难度大的问题。通过本课程学生学习，学生能够掌握如何完成大数据集的批处理作业。

三、课程设计思路

本课程旨在介绍如何使用 Hive 的 SQL 方法 HiveQL 来汇总、查询和分析存储在 Hadoop 分布式文件系统上的大数据集合。课程通过大量的实例，首先介绍如何在用户环境下安装和配置 Hive，并对 Hadoop 和 MapReduce 进行详尽阐述，使学生掌握 Hive 如何在 Hadoop 生态系统中进行工作。

四、课程目标

（一）职业知识目标

1. 掌握 Hive 的安装及配置
2. 熟练 Hive 命令
3. 掌握数据类型和文件格式
4. 熟练掌握 Hive 的数据定义
5. 掌握 Hive 的数据操作
6. 掌握 Hive 查询
7. 掌握 Hive 视图
8. 熟练 Hive 索引

9. 了解模式设计
10. 掌握 Hive 调优方法
11. 了解其他文件格式和压缩方法
12. 掌握 Hive 函数。

(二) 职业能力目标

1. 培养学生规范编码和良好的程序设计风格；
2. 培养学生对大数据集分析处理的能力；
3. 培养学生发现问题、分析问题和解决问题的能力。

(三) 职业素质目标

1. 培养按时完成项目的良好习惯，增强集体和社会责任感；
2. 培养团队协作和良好的沟通能力；
3. 培养学生务实、创新、诚实、守信的性格。

五、课程内容及要求

序号	教学内容	能力目标	知识目标	教学方法及手段	学时
1	基础知识： Hadoop 和 MapReduce 综述	熟悉 Hadoop 和 MapReduce 计算 机模型。 理解 MapReduce 的基本原理。	1. 了解 haddop 和 MapReduce 综述 2. 掌握 MapReduce 的基本原理	引导启发学生自动 去学习查找了解 Hadoop 的起源及前 期。	4
2	Hive 基本 操作	能够掌握如何 安装 Hiew 的 Hadoop 集群。	1. 掌握安装 Hive 的 Hadoop 集群 2. 掌握 Hive 的启动方式方 3. 掌握 Hive 命令	由教师示范、指导， 学生小组协作	6
3	数据类 型和文 件格 式	掌握基本数据 类型	1. 掌握 Hive 中的基本数据类型 2. 集 合数据类型 3. 掌握文本文件数据编码 4. 了解读时模式	由教师示范、指导、 学生小组协作完 成本章的设计与实 现	6
4	HiveQL: 数 据定义	掌握 HiveQL 中 数据库和表的 操作	1. 掌握数据库的创建和修改 2. 掌握表：管理表、外部表、分区表 等、删除表、修改表	由教师示范、指导、 学生独立完成数据 库及表的操作。	8
5	HiveQL: 数 据操作	掌握数据的操 作	1. 掌握向管理表中装载数据。 2. 通过查询语句向表中插入数据 3. 单个查询语句中创建表并加载数 据 4. 导入导出数据	由教师示范、指导、 学生完成数据的操 作	6
6	HiveQL: 查 询	1. 掌握 HiveQL 查询	1. 掌握 SELECT...FROM...语句 2. 掌握 WHERE 语句 3. 掌握 GROUP BY 语句 4. 掌握 JOIN 语句等	由教师示范、指导、 学生小组协作完 成 HiveQL 查询功 能	6
7	HiveQL: 视 图	掌握 HiveQL 视 图。	1. 能够使用视图来降低查询复杂度 2. 能够使用视图限制基于条件过滤 数据 3. 能够通过动态分区中的视图和 map 类型	由教师示范、指导、 学生自主学习	4

8	HiveQL: 索引	掌握 HiveQL 索引	掌握 Hive 所有的创建、修改、显示及删除	由教师示范、指导、学生自主学习 Hive 索引。	6
9	模式设计	了解 Hive 模式设计	了解 Hive 设计的设计模式	由教师示范、指导, 学生自主学习认识常用设计模式。	6
10	调优	掌握 Hive 性能调优的方法。	了解并掌握 Hive 性能调优的方法	由教师示范、指导, 学生小组学习并练习 Hive 调优的方法。	4
11	其他文件格式和压缩方法	了解文件格式和压缩方法	掌握其他人间格式和压缩方法	由教师示范、指导学生小组学习	6
12	开发	了解 Hive 自身源代码	了解 Hive 自身源代码	由教师示范、指导学生小组学习	4
13	函数	掌握 Hive 中函数的用法	掌握 Hive 中函数的用法	由教师示范、指导学生小组学习完成函数练习。	6
14	Streaming	了解 Streaming 处理数据的方式。	了解 Streaming 处理数据的方式	由教师示范、指导学生自主学习	6
15	自定义 Hive 文件和记录格式	掌握自定义文件和记录格式	掌握文件和记录格式	由教师示范、指导、学生自主学习	6
16	Hive 的 Thrift 服务	了解跨语言的服务开发软件框架: Thrift	了解跨语言的服务开发软件框架: Thrift	由教师示范、指导学生自主学习	6
17	Hive 和亚马逊网络服务系统 (AWS)	掌握程序中各事件的综合协调处理	掌握较复杂业务的处理	学生自主学习、小组协作	10

六、课程实施建议

(一) 教学建议

1. 教学团队基本要求

课程教学团队专兼职师资团队成员 4 至 5 人。其中, 行业企业授课教师不低于 30%, “双师型”授课教师中 60%应具有中级或高级职称, 技术服务团队成员不少于 1 人。

2. 教学条件

硬件要求: 微机、网络。

软件要求: JDK、Hadoop、Hive、Eclipse、VMware、XmanagerEnterprise 等软件。

3. 教学方法与手段

本课程利用教学平台进行理实一体化化教学。

4. 课程资源的开发与利用

课程资源开发:

- (1) 结合实际软件开发工程案例，进行课件制作、教学环境模拟。
- (2) 与企业专家合作编写符合技能培养的实验实训教材，是学生学有所用。
- (3) 提供网络资源学习平台，使学生不拘泥于固定的时间、地点学习，提高学生学习的灵活性。

课程资源利用：

(1) 学习包和教材，都是素材性的课程资源，但教材是知识的载体，而学习包是引导学生学习的载体。因此，要开发学习包，突出以学生为中心的学习过程，将本课程的教学内容与地区经济的发展密切结合，并体现学生职业生涯发展的需要。

(2) 教学中学生的问题、困惑、见解、情感和体验等都是动态生成的课程资源，老师应重视这些教学过程中动态生成的课程资源，并主动性和创造性地运用，使以项目为主的教学充满生机与活力。

(3) 本课程的实施，不仅应充分开发校内的课程资源，还应利用校外的课程资源，如相关企业的岗位实习和实践，尽量做到校企结合、工学结合。

(4) 积极开发教学资源：课程标准、实训指导书、授课计划等教学文件，以及课件、习题、案例库、网络方案、工具软件和网络资源等。

充分利用网络资源，了解最新的技术策略和实施方式，培养学生自主学习的能力。

5. 教材选用

选用的教材必须符合本专业人才培养目标及课程教学的要求，适应高职高专学生使用。教材选用要优先选用教育部的国家级高职高专规划教材，提高国家规划教材的使用率，其次选择高等教育出版社及相关专业行业出版社的高职教材。

推荐使用选用教材：《Hive 编程技术与应用》 ISBN: 9787517069140
中国水利水电出版社

推荐参考资料：

- (1) 《Hadoop 权威指南》 Tom White 清华大学出版社
- (2) 《基于 Hadoop 与 Spark 的大数据开发实战》 人民邮电出版社

(二) 考核建议

考核主要由平时成绩、理论考试以及实训项目成绩组成。

1. 平时成绩

平时成绩强调平时的出勤、课堂纪律、课堂表现和作业，教师通过千分制的实施可以在整个授课过程中监控学生的学习效果以及学习态度等等。平时成绩占全部考试的 30%。

2. 理论考试

理论考试采用传统的笔试方式，为了更好的达到考核的目的，我们采用第三方出题的形式来组织理论考试。理论考试占全部考试的 40%。

3. 项目评审

实施项目内部考核及教师考核相结合的原则，在项目实训末期，要进行项目评审，根据项目的完成情况给学生打分。项目评审成绩占全部成绩的 30%（包括学生自我评价及对同伴的评价、完成项目效果评价以及教师评价，其中自我评价(占 30%)、项目效果评价(占 30%)、教师评价(占 40%)。）

《Spark 技术与应用》课程标准

一、课程基本信息

课程名称	Spark 技术与应用				
课程代码	0917009	学时	108	学分	6
授课时间	第四学期	适用专业	大数据应用与技术专业		
课程性质	岗位能力课程				
先修课程	Hadoop 系统及应用	后续课程	顶岗实习		

二、课程定位

在 Spark 作为当代大数据处理引擎，在非常短的时间里崭露头角，并且以燎原之势席卷业界。Spark 对曾经引爆大数据产业革命的 Hadoop、MapReduce 的进行改进，通过学习本课程，学生能够利用 Spark 去解决一些可能会遇到但是没有办法解决的问题。能够掌握何搭建一个 Spark 集群，学会使用 Spark shell，以及编写 Spark 应用程序来解决需要并行处理的问题。

三、课程设计思路

本课程前两章讲述搭建基础的 Spark，了解 Spark 基本的概念。重点介绍 Spark shell。Spark shell 是开发 Spark 应用原型时非常有用的工具。后续几章则会详细介绍 Spark API、如何将 Spark 应用运行在集群上，以及 Spark 所提供的更高层的程序库支持，例如 SQL（数据库支持）和 MLlib（机器学习库）。

四、课程目标

（一）职业知识目标

1. 了解 Spark 生态圈
2. 熟练 Spark shell
3. 熟练掌握 Spark API
4. 掌握如何将 Spark 运行在集群上
5. 掌握 Scala 基本编程方法
6. 掌握 DataFrame 编程方法
7. 掌握 Spark Streaming 对 Socket、HDFS 数据进行流式处理的方法
8. 了解 Spark Streaming 与 Flume、Kafka 的整合

（二）职业能力目标

1. 能使用 Spark 进行离线数据处理;
2. 能基于 Spark 技术进行海量数据的处理、分析、统计和挖掘;
3. 能基于 Spark 框架进行数据仓库的设计、开发和维护。

(三) 职业素质目标

1. 培养按时完成项目的良好习惯, 增强集体和社会责任感;
2. 培养团队协作和良好的沟通能力;
3. 培养学生务实、创新、诚实、守信的性格;
4. 培养 Spark 的分布式计算思维。

五、课程内容及要求

序号	教学内容	能力目标	知识目标	教学方法及手段	学时
1	Spark 数据分析导论	能从宏观角度理解 Spark 到底是什么	1. 了解 Spark 是什么 2. 理解 Spark 软件栈	引导启发学生独立完成设置工作	9
2	Spark 下载与入门	能够下载 Spark 并在本地模式下单机运行它	1. 掌握下载 Spark 2. 掌握 Spark 中 Python 和 Scala 的 shell 3. 理解 Spark 核心概念简介 4. 掌握 Spark 的独立应用	由教师示范、指导, 学生小组协作	9
3	RDD 编程	能创建 PairRDD 及其转换操作	1. 掌握 Spark 对数据的核心抽象 RDD 2. 掌握数据分区原理	由教师示范、指导、学生小组协作完成本章练习	12
4	键值对操作	掌握如何操作键值对 RDD	理解类的继承与多态性的概念及实现	由教师示范、指导、学生小组协作完成本章练习	9
5	数据读取与保持	掌握共享变量	1. 了解累加器、广播变量及基于分区进行的操作	由教师示范、指导、学生小组协作完成本章练习	9
6	Spark 编程进阶	1. 会处理常见异常 2. 会用户自定义异常	1. 理解异常的概念 2. 了解异常类的层次结构	由教师示范、指导、学生小组协作完成本章练习	10
7	在集群上运行 Spark	掌握分布式 Spark 应用环境架构及配置	1. 理解 Spark 运行时架构 2. 掌握 spark-submit 部署应用 3. 掌握 Spark 打包代码与依赖	由教师示范、指导、学生自主学习	12
8	Spark 调优与调试	会配置 Spark 应用, 了解在生产环境中的 Spark 工作负载如何调优和调试	1. 掌握如何配置 Spark 应用 2. 了解如何调优和调试生产环境中的 Spark 工作负载	由教师示范、指导、学生自主学习	9
9	Spark SQL	掌握 SparkSQL	1. 掌握如何在应用中使用 SparkSQL 2. 掌握用户自定义函数	由教师示范、指导、学生自主学习	10

10	Spark Streaming	会创建常用容器、菜单的对象、设置属性的及调用方法	1. 使用 Java 和 Scala 练习 SparkStreaming 模型	由教师示范、指导、学生自主学习	10
11	基于 MLib 的机器学习	会使用 MLib 函数库	1. 了解机器学习基础 2. 了解 MLib 中主要的算法, 以及它们的输入和输出类型	由教师示范、指导、学生自主学习	12

六、课程实施建议

(一) 教学建议

1. 教学团队基本要求

课程教学团队专兼职师资团队成员 4 至 5 人。其中,行业企业授课教师不低于 30%,“双师型”授课教师中 60%应具有中级或高级职称,技术服务团队成员不少于 1 人。

2. 教学条件

硬件要求: 微机、网络。

软件要求: JDK、Hadoop、VMware、Eclipse、spark 等软件。

3. 教学方法与手段

本课程利用教学平台进行理实一体化化教学。

4. 课程资源的开发与利用

课程资料开发:

- (1) 结合实际软件开发工程案例,进行课件制作、教学环境模拟。
- (2) 与企业专家合作编写符合技能培养的实验实训教材,是学生学有所用。
- (3) 提供网络资源学习平台,使学生不拘泥于固定的时间、地点学习,提高学生学习的灵活性。

课程资源使用:

(1) 学习包和教材,都是素材性的课程资源,但教材是知识的载体,而学习包是引导学生学习的载体。因此,要开发学习包,突出以学生为中心的学习过程,将本课程的教学内容与地区经济的发展密切结合,并体现学生职业生涯发展的需要。

(2) 教学中学生的问题、困惑、见解、情感和体验等都是动态生成的课程资源,老师应重视这些教学过程中动态生成的课程资源,并主动性和创造性地运用,使以项目为主的教学充满生机与活力。

(3) 本课程的实施,不仅应充分开发校内的课程资源,还应利用校外的课程资源,如相关企业的岗位实习和实践,尽量做到校企结合、工学结合。

(4) 积极开发教学资源: 课程标准、实训指导书、授课计划等教学文件,以及课件、习题、案例库、网络方案、工具软件和网络资源等。

充分利用网络资源，了解最新的技术策略和实施方式，培养学生自主学习的能力。

5. 教材选用

选用的教材必须符合本专业人才培养目标及课程教学的要求，适应高职高专学生使用。教材选用要优先选用教育部的国家级高职高专规划教材，提高国家规划教材的使用率，其次选择高等教育出版社及相关专业行业出版社的高职教材。

推荐选用教材：《Spark 快速大数据分析》 王道远 译 人民邮电出版社

推荐参考资料

(1) 《Hadoop 权威指南》 清华大学出版社

(2) 《基于 Hadoop 与 Spark 的大数据开发实战》 人民邮电出版社

(3) 《Hadoop + Spark 大数据巨量分析与机器学习整合开发实战》 林大贵 清华大学出版社

(二) 考核建议

考核主要由平时成绩、理论考试以及实训项目成绩组成。

1. 平时成绩

平时成绩强调平时的出勤、课堂纪律、课堂表现和作业，教师通过千分制的实施可以在整个授课过程中监控学生的学习效果以及学习态度等等。平时成绩占全部考试的 30%。

2. 理论考试

理论考试采用传统的笔试方式，为了更好的达到考核的目的，我们采用第三方出题的形式来组织理论考试。理论考试占全部考试的 40%。

3. 项目评审

实施项目内部考核及教师考核相结合的原则，在项目实训末期，要进行项目评审，根据项目的完成情况给学生打分。项目评审成绩占全部成绩的 30%（包括学生自我评价及对同伴的评价、完成项目效果评价以及教师评价，其中自我评价（占 30%）、项目效果评价（占 30%）、教师评价（占 40%）。）

《大数据可视化技术》课程标准

一、课程基本信息

课程名称	大数据可视化技术				
课程代码	0917010	学时	36	学分	2
授课时间	第4学期	适用专业	大数据技术与应用专业		
课程性质	岗位能力课程				
先修课程	Web 前端设计、MySQL 数据库、Python 程序设计等	后续课程	毕业设计、顶岗实习		

二、课程定位

本课程是计算机大数据专业方向的核心课程，是培养学生数据分析与可视化处理能力的支撑课程，同时也是计算机专业群其他专业方向的必修课程。

本课程通过“教、学、做”理论与实践一体化教学，使学生掌握数据分析与可视化的基础理论和概念、可视化的类型与模型、数据可视化的过程以及数据可视化的常见方法，同时熟练掌握常见的大数据可视化工具，如 Echarts、Tableau 等的使用，为可视化应用开发和维护奠定基础。

先修课程：Python 程序设计、Web 前端设计、MySQL 数据库、Linux 操作系统

后续课程：大数据综合项目、毕业设计、顶岗实习

三、课程设计思路

在专业建设委员会的指导和帮助下，课程组教师通过对企业调研和深入地学习，经过深入探讨确定大数据可视化课程的知识目标和能力目标。

本课程以大数据专业学生的就业为导向，采用理实一体化教学方法，使学生全面掌握大数据可视化理论实践内容，保证学生毕业后能够胜任有关数据可视化技术应用岗位的需求。

课程主要包括九个单元教学任务，教学中对每个项目的的能力目标、知识目标、教学内容作了具体的设计，并配套设计了相应的实训项目。学生通过本课程的学习，能够熟练掌握数据分析基础知识、数据可视化概貌、数据可视化的类型与模型、数据可视化过程、数据可视化常用方法、数据可视化交互、Echarts 数据可视化方法、Tableau 数据可视化入门、Tableau 数据可视化设计等内容。通过整体学习，保证学生最终具备运用数据可视化技术进行综合应用开发的能力。

四、课程目标

通过本门课程的学习，使学生全面地掌握当前主流数据分析和可视化技术的知识与应用技能，能应用基础方法和工具对现有数据进行分析并得出结论，能进行数据库设计，能使用 SQL 语言进行数据定义、数据查询和数据操纵等，能熟练使用可视化工具 Echarts、Tableau 制作图表，借助图形化手段，清晰有效的传达数据信息。

（一）职业知识目标

1. 掌握模拟分析、规划求解、数据分析工具库等数据分析基础知识；
2. 了解大数据可视化的概念、作用及意义；
3. 了解常见的可视化的类型及模型；
4. 了解数据可视化的设计流程及相关组件；
5. 了解常见的数据可视化交互方法；能够描述常见的可视化交互模型；
6. 掌握常见的可视化方法，包括统计图表可视化方法、图可视化方法及可视化分析方法的算法等；
7. 掌握可视化工具 Echarts 的使用，包括 Echarts 常用图表的制作，如折线图、饼图、散点图等。以及掌握 Echarts 高级用法，如图表混搭、异步数据加载、事件与行为等；
8. 掌握 Tableau 的安装与应用；掌握 Tableau 中数据源的应用；
9. 掌握 Tableau 中的函数与计算，包括运算符的使用、数值计算、字符串计算、日期计算等；
10. 掌握 Tableau 中常用图表分析；掌握地图的创建与绘制；掌握仪表盘和故事的创建与设置；
11. 掌握大数据可视化相关行业案例的综合应用，掌握行业数据分析可视化方法。

（二）职业能力目标

1. 能够按照指定的任务描述，利用所学知识和技术，分析和处理数据；
2. 具备 Echarts 应用与开发的能力；
3. 具备 Tableau 图表设计的能力；
4. 具备 Tableau 安装与配置的能力；
5. 具数据可视化开发综合应用的能力。

（三）职业素质目标

1. 培养学生获取新知识的能力和搜索信息的能力；
2. 培养学生独立的决策能力；
3. 培养学生分析问题、解决问题的能力；

4. 培养学生具有创新意识、创新精神和良好的职业道德；
5. 培养学生良好的自我表现、与人沟通能力；
6. 培养学生的团队协作精神。

五、课程内容与要求

序号	教学内容	能力目标	知识目标	教学方法及手段	学时
1	单元 1: 数据分析基础 大数据思维与技术; 大数据的发展; 模拟分析; 规划求解; 数据分析工具库;	(1) 能掌握数据分析计算方法; (2) 能熟练使用数据分析工具。	了解大数据思维与技术; 了解大数据的发展; 掌握模拟分析概念、计算方法; 掌握规划求解概念、计算方法; 了解数据分析工具库。	理实一体	2
2	单元 2: 数据可视化概貌 数据可视化概念; 数据可视化的作用与意义; 数据可视化的经典案例。	能利用所学理论知识理解数据可视化的经典案例。	了解数据可视化的概念; 了解数据可视化的作用及意义; 培养学生实践动手、分析解决问题的能力。	理论讲授 示范教学	3
3	单元 3: 数据可视化的类型与模型 可视化类型; 可视化模型。	具备可视化分析与设计能力。	了解常见的可视化类型; 了解常见的可视化模型;	理实一体	2
4	单元 4: 数据可视化过程 可视化的基本特征; 可视化流程; 可视化设计组件; 可视化框架设计整体思路。	能够理解可视化的基础理论知识; 具备可视化框架整体设计能力。	了解可视化的基本特征; 了解数据可视化的设计流程及相关组件; 掌握可视化框架设计整体思路。	理实一体 教学 典型案例 实训	3
5	单元 5: 数据可视化常用方法 统计图表可视化方法; 图可视化方法; 可视化分析方法的常用算法; 可视化方法的选择。	能熟练运用常见的可视化方法; 具备可视化分析与设计能力。	掌握常见的图表可视化方法; 掌握常见的图可视化方法; 掌握可视化分析方法的常用算法; 了解可视化方法的选择。	理实一体	3
6	单元 6: 数据可视化交互	能够掌握常见的可视化交互方法;	了解常见的数据可视化交互方法;	理实一体	3

	可视化交互方法分类； 可视化交互空间； 可视化交互模型。	具备对常见的可视化交互模型的认知描述能力。	了解数据可视化交互空间； 能够描述常见的可视化交互模型。		
7	单元 7: Echarts 数据可视化方法 Echarts 基础架构； Echarts 常见图表（柱状图、饼图、折线图、散点图、地图等）； Echarts 高级（多图表联动、异步数据加载、事件与行为等）。	具备 Echarts 图表设计开发应用能力。	掌握 Echarts 基础架构； 掌握可视化工具 Echarts 的使用，包括 Echarts 常用图表的制作，如折线图、饼图、散点图等。 掌握 Echarts 高级用法，如图表混搭、异步数据加载、事件与行为等。	理实一体教学 典型案例实训	6
8	单元 8: Tableau 数据可视化入门 Tableau 概述； 软件安装与下载； 连接数据源； Tableau 的基础操作； Tableau 数据导出。	具备 Tableau 安装配置及应用能力。	掌握 Tableau 的安装与应用； 掌握 Tableau 中数据源的应用； 掌握 Tableau 中的函数与计算，包括运算符的使用、数值计算、字符串计算、日期计算等； 具备使用 Tableau 创建各种图表的能力。	理实一体教学 典型案例实训	6
9	单元 9: Tableau 数据可视化设计 创建图表 表计算 创建地图 仪表盘与故事	具备使用 Tableau 创建各种图表的能力； 能运用数据可视化技术进行综合应用开发。	掌握 Tableau 中常用图表分析； 掌握地图的创建与绘制； 掌握仪表盘和故事的创建与设置； 具备可视化技术综合应用开发能力。	理实一体教学 典型案例实训	8

六、课程实施建议

（一）教材建议

1. 教学团队基本要求

担任本课程的教学团队由四到五名专业教师组成，所有主讲教师需要具有丰富的大数据可视化应用开发及管理经验，熟知大数据可视化体系和理论，具备大数据可视化设计、应用与开发、管理与维护的能力。同时要充分发挥团队作用，能够进行一体化教学，能按照本课程标准制定详细可行的授课计划，精心设计每一次课的教学过程。教学团队所有教师需具备一定的课堂控制能力和应变能力，做到因材施教。

2. 教学条件

1. 计算机硬件要求：Windows2000 及以上，CPU 主频>1GHz，内存容量>=512MB；
2. 软件要求：Echarts, Tableau；
3. 其他要求：本课程教学要求在多媒体机房实施，以实现“教、学、做”三体合一；多媒体教学软件，便于下发教学任务和收集学生课堂实践任务。

3. 教学方法与手段

1. 课堂讲授采用“教、学、做一体化”教学方式，以教学单元为模块，以案例应用引导学生学习技能，并通过动手实践，让学生自己理解并归纳、掌握数据分析与可视化的知识和常识。整个教学过程突出精讲多练的模式，用大量的课内外练习来达到教学目的。

2. 在教学过程中建议将学生分为若干小组，给定工作任务，启发学生通过动手实践以及对实践结果进行思考获取知识。

3. 同时在讲授的过程中，教学项目设置要充分发挥学生学习的自主性，贴近学生，贴近生活，突出学生学习和实践的交往性和职业性。

4. 课程资源的开发与利用

本课程建议按照现有的模块和任务选取教材。并拓展课程教案、活页式讲义、微视频等课程资源。

教材应充分体现工学结合、任务驱动、项目导向课程的设计思想；教材中的教学项目设置要充分发挥学生学习的自主性，贴近学生，贴近生活，突出学生学习和实践的交往性和职业性；教材应注重一线材料的整理，引入企业的常用案例及规范要求和考核标准；教材应适合高等职业学校的学生；完善精确教学使用的多媒体教学素材、教案和讲义等；要充分利用行业资源，为学生提供阶段实训，让学生在真实的环境中磨练自己，提升其职业综合素质；要充分利用网络资源，搭建网络课程平台，开发微视频课程；积极利用数字图书馆、电子期刊、电子书籍，使教学内容更多元化，以此拓展学生的知识和能力；充分利用信息技术开放实训中心，将教学与实训合一，满足学生综合能力培养的要求。

5. 教材选用

(1) 教材选取原则

选用的教材必须符合本专业人才培养目标及课程教学的要求，适应高职高专学生使用。教材选用要优先选用教育部的国家级高职高专规划教材，提高国家规划教材的使用率，其次选择高等教育出版社及相关专业行业出版社的高职教材。

所选用的教材应充分体现培养学生的职业能力为中心思想，以“教、学、做”为一体组织课堂内容，注重理论与实践相结合、教材内容与行业标准要求相结合，强调理论在实践过程中的应用。

(2) 推荐教材与参考书籍

何光威编著. 大数据可视化 电子工业出版社. 2018.

王文, 周苏编著. 大数据可视化 机械工业出版社. 2018.

周苏, 王文. 大数据可视化 清华大学出版社. 2016.

黄源, 蒋文豪, 徐受蓉. 大数据分析 清华大学出版社. 2019.

(二) 考核建议

本课程考核采用形成性考核方式，总评成绩由形成性考核的各项成绩组成，全面考核学生的动手能力、基础理论和平时学习状况，分数比例为：

综合素质评价（平时表现+通用能力考核）	20%
过程性考核（课堂实践考核+课后实践考核+单元测试）	40%
终结性考核（期终考核）	40%