

大数据技术与应用专业建设规划

(2019-2022)

一、专业建设背景

2016年，习近平总书记指出，“大数据战略作为‘十三五’期间的十四大国家战略之一，是中国经济发展新的驱动力。”至此，大数据战略上升到国家战略。未来五年信息化建设将重点实施网络强国战略，并且明确提出实施网络强国战略，实施国家大数据战略。根据《国务院关于印发促进大数据发展行动纲要的通知》，主要分解了三大任务和十项工程。

1、三大任务

- 加快政府数据开放共享，推动资源整合，提升治理能力
- 推动产业创新发展，培育新兴业态，助力经济转型
- 强化安全保障，提高管理水平，促进健康发展

2、十项工程

- 政府数据资源共享开放工程、国家大数据资源统筹发展工程
- 政府治理大数据工程、公共服务大数据工程
- 工业和新兴产业大数据工程、现代农业大数据工程
- 万众创新大数据工程、大数据关键技术及产品研发与产业化工程
- 大数据产业支撑能力提升工程、网络和大数据安全保障工程

1. 大数据产业发展趋势

大数据在2011年由全球最著名的管理咨询公司麦肯锡首次提出并迅速在全球范围内引起共鸣。短短几年，大数据应用已经在各行各业形成不可阻挡产业技

术革新浪潮。当今社会已经进入了大数据的时代，大数据的应用正在逐步从互联网、金融扩展到教育、政务、交通物流、医疗健康等各大社会和经济领域，其必将深刻影响中国未来社会和经济的发展进程。大数据的产业规模也在急速增长，如下为部分相关数据：

- IDC 预计，全球大数据技术及服务市场 2016 年收入将达 238 亿美元
- 据麦肯锡预测，预计到 2020 年，美国的大数据可创造 3800 亿-6900 亿美元的价值
- 2014 年，我国大数据市场规模为 84 亿元，2015 年接近 116 亿元，增速为 38%
- 与 2014 年中国信息产业规模 15 万亿元相比，大数据市场还是一个非常小的市场，还处在非常初级的阶段

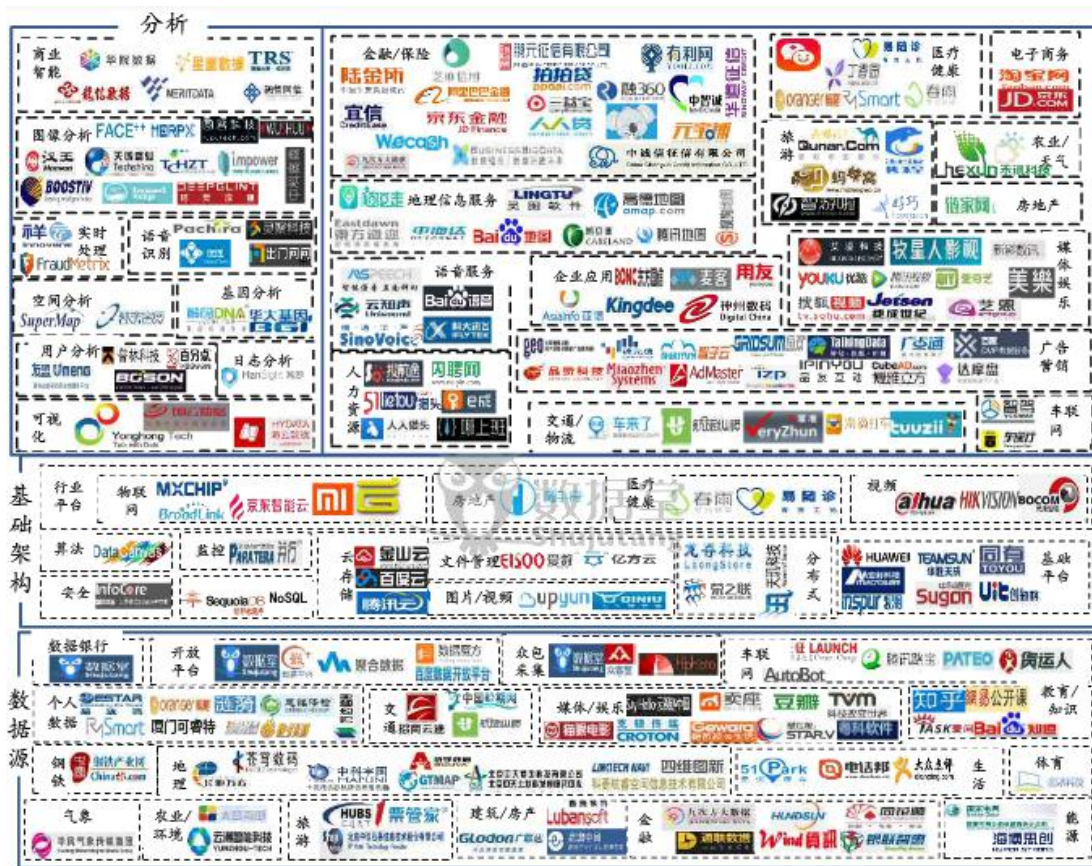
2. 大数据人才需求分析

根据领英最近发布的一份研究表明，过去四年从事数据分析的专业人士数量翻了一倍。相关工作的招聘量急剧上升，这一趋势反映了人们对高端数据分析人才的需求在不断增加，因为企业需要他们对数据进行各式处理分析，这必然会对 2016 年甚至以后的大数据局面产生影响。

另外，麦肯锡发布大数据人才需求报告并预测大数据人才短缺，到 2018 年，在“具有深入分析能力的人才”方面，美国可能面临着 14 万~19 万人的缺口，而“可以利用大数据分析来作出有效决策的经理和分析师”缺口则会达到 150 万。国内大数据人才有多大的缺口？目前尚无权威研究报告指出，不过今年 2 月万达集团的一份“求贤帖”颇能说明问题：万达开出 50 万~70 万元年薪吸纳大数据架构师、大数据研发工程师和算法工程师，应者寥寥。同时，从如下 Gartner 研究报告中也可以看出来大数据人才的缺口。

- Gartner 公司预测大数据将为全球带来 440 万个 IT 新岗位和上千万个非 IT 岗位
- Gartner 公司预测在未来 6 年，仅在美国本土就可能面临缺乏 14 万至 19 万具备深入分析数据能力人才的情况，同时具备通过分析大数据并为企业做出有效决策的数据的管理人员和分析师也有 150 万人的缺口。
- 涵盖了大数据的数据开发工程师、大数据分析师、数据架构师、大数据后台开发工程师、算法工程师等多个方向
- 数据岗位需求在上海是 5 万多，北京则 10 万多

大数据产品地图分布及人才就业去向图如下所示：



3. 大数据技术与应用专业设置

教育战略需要服务于国家战略，从国家大数据战略发展来看，随着大树产业的飞速发展，未来社会对大数据的人才需求将剧烈增长，为服务国家大数据战略，我们要提前做好大数据人才培养的准备，满足社会需求。2016 年 9 月，教育部研

究决定正式批准高职大数据技术与应用专业（专业代码：610215）（《普通高等学校高等职业教育（专科）专业目录 2016 年增补专业》）。

二、 专业人才培养目标

本专业人才培养主要技术岗位方向有大数据应用开发、大数据分析挖掘、大数据系统运维等方向，学生的知识体系符合大数据行业需求的复合型、创新型和实用型人才。本专业方向特别注重培养能够为企事业单位提供进行大数据系统的搭建、管理、和运维技术和能力的学生。学生在学习计算机基础课程和计算机系统管理的同时，接受大数据系统和应用知识的培养，并进行各种计算机系统，大数据平台系统，大数据应用系统搭建、配置、管理、及运维实训。在获得计算机系统管理技术的基础上，学生通过大量的案例与实践操作，可以熟练掌握大数据系统管理所需的各种专业知识和能力，具备一定的职业素养，为他们从事大数据这个尖端行业的系统管理工作奠定坚实的基础。本专业方向重点培养具有大数据应用、大数据分析以及大数据系统管理与运维方向的，应用型高技能人才。

三、 人才培养方向及职业能力分析

通过对主流招聘网站、主流互联网公司、大数据相关产业联盟对人才的需求和岗位能力要求分析，了解到人才培养目标中涉及的三个主要技术岗位需要学生掌握的能力有如下几个方面。

- 需要理解大数据系统的基本概念、掌握大数据应用的设计与开发方法、数据挖掘与大数据分析方法和、以及大数据系统搭建运维方法；
- 经过大数据应用开发工程实践的基本训练，具备从事大数据应用开发、大数据分析、或大数据系统运维的基本能力。

对上述能力目标进行细化与分解形成如下所示的岗位职业能力分析简表。

岗位方向	岗位职业能力
大数据应用工程师	<ol style="list-style-type: none"> 1、掌握大数据的基本概念和应用领域 2、掌握程序设计和实现方法 3、掌握使用 HBase、HDFS 存储和管理数据的方法 4、具有开发 Java Web 应用的能力 5、具有基于 MapReduce 开发大数据应用开发的能力 6、了解大数据系统的架构和设计方法 7、了解大数据系统的基本搭建和调优方法 8、了解大数据的发展趋势和应用领域 9、了解云计算的基本概念和作用
大数据分析工程师	<ol style="list-style-type: none"> 1、掌握大数据的基本概念和应用领域 2、掌握程序设计和实现方法 3、掌握使用 HBase、HDFS 存储和管理数据的方法 3、具有使用 Mahout 数据挖掘算法进行数据分析的能力 4、具有基于 MapReduce 开发大数据应用开发的能力 5、了解大数据系统的架构和设计方法 6、了解大数据系统的基本搭建和调优方法 7、了解大数据的发展趋势和应用领域 8、了解云计算的基本概念和作用
大数据系统运维工程师	<ol style="list-style-type: none"> 1、掌握大数据的基本概念和应用领域 2、掌握大数据系统的架构和设计方法 3、掌握计算机网络基本原理和基本知识 3、熟练掌握 Linux Shell 的使用 4、熟悉 Linux 操作系统 5、具有大数据系统的搭建、调优和运维能力 6、了解虚拟化技术 7、了解大数据的发展趋势和应用领域 8、了解云计算的基本概念和作用

四、专业课程体系设计

1. 专业课程设置思路

以学生实际动手能力为主导，项目式教学为驱动，理论、技术与实践课程一体化教学为主。

1.1 以理论课程为辅，使学生了解技术概况

- 理论是大数据认知的必经途径，也是被广泛认同和传播的基线
- 讲解大数据的概念、作用、挑战和各种技术概念

1.2 以技术实验课程过程为主导，使学生掌握使用技术的方法

- 技术是大数据价值体现的手段和前进的基石
- 讲解用于大数据应用开发过程中的各种系统和工具

1.3 通过项目实战课程使学生具有灵活使用技术解决实际问题的能力

- 培养学生的综合实践能力，以满足社会就业的需要

2. 专业课程知识体系

根据课程设置总体思路以及大数据专业岗位的职业能力分析，大数据专业需要掌握的知识体系如下表所示：

知识类别	知识名称	知识描述
理论知识	大数据概念	大数据概念出现的历史，大数据定义和主要特征，大数据与云计算、物联网、互联网的关系，大数据的重要意义、大数据的处理流程
	大数据应用领域	通过介绍典型的大数据应用实例的设计思想、主要方法和应用过程等讲解大数据应用的系统化的开发步骤和关键性问题的解决方法
	大数据关键技术与挑战	1、大数据集成：数据异构性和数据质量问题 2、大数据分析：数据形式多样化、数据处理的实时性、索引结构的复杂性等 3、大数据隐私问题：隐私保护和数据分析

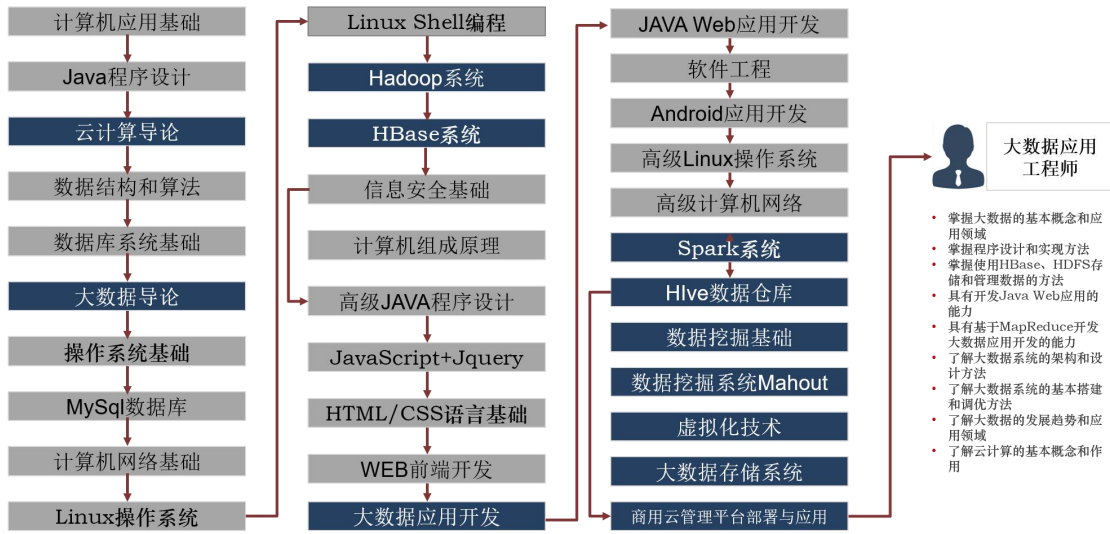
		<p>的矛盾</p> <p>4、大数据管理 易用性问题以及性能测试基准</p>
	大数据存储和管理技术	<p>1、大数据存储、管理 和调用</p> <p>2、分布式文件系统</p> <p>3、去冗余及高效低成本的大数据存储技术</p> <p>4、新型数据库技术（键值数据库、列存数据库、图存数据库以及文档数据库等）</p> <p>5、异构数据融合技术</p> <p>6、分布式非关系型大数据 管理与处理技术</p> <p>7、大数据索引技术和大数据移动、备份、复制</p>
	大数据分析及挖掘技术	<p>1、数据挖掘分为分类或预测模型发现、数据总结、聚类、关联规则 发现、序列模式发现、依赖关系或依赖模型发现、异常 和趋势发现等</p> <p>2、数据准备：从大数据中心 存储的数据中选取所需数据并整合成用于数据挖掘 的数据集</p> <p>3、规律寻找：用某种方法将数据集所含规律 找出来</p> <p>4、规律表示：以用户可理解的方式（如可视化）将找出的规律表示出来</p>
实验知识	新型数据库系统	<p>1、HBase 产生的背景和现状</p> <p>2、HBase 的功能、作用和特点</p> <p>3、HBase 的逻辑模型、物理模型、数据一致性模型</p> <p>4、HBase 的系统架构及实现方法</p> <p>5、HBase 的搭建方法和使用方法</p>
	分布式文件系统	<p>1、Hadoop 出现的背景、应用现状和发展趋势</p> <p>2、Hadoop 的优点、功能与作用</p> <p>3、Hadoop 的体系结构</p> <p>4、HDFS 文件系统</p> <p>5、Hadoop 与分布式开发</p> <p>6、Hadoop 应用案例</p> <p>7、Hadoop 系统搭建与调优</p>
	分布式处理模式	<p>1、MapReduce 分布式并行计算框架</p> <p>2、编程模型：Map 和 Reduce 函数</p> <p>3、MapReduce 工作流程</p> <p>4、分布式资源管理架构 YARN</p> <p>5、基于 MapReduce 的大数据应用开发</p>

	数据挖掘工具	<ol style="list-style-type: none"> 1、 Mahout 数据挖掘系统 2、 海量数据的数据挖掘实现 3、 Mahout 的搭建与部署 4、 Mahout 聚类、分类、关联、推荐算法的实现 5、 Mahout 各种挖掘算法的使用方法
	基于内存的分布式计算框架	<ol style="list-style-type: none"> 1、 Spark 的历史、现状和特点 2、 Spark 的计算模式和运行流程 3、 Spark 的使用方法 4、 基于 Spark 的大数据应用开发 5、 基于 Spark 的数据挖掘方法 6、 基于 Spark 的流式数据处理 7、 Spark 的搭建、部署与调优
项目实践	项目开发流程	<ol style="list-style-type: none"> 1、 选题、需求分析、数据收集及预处理 2、 探索性数据分析、模型选定分析 3、 模型验证、分析报告

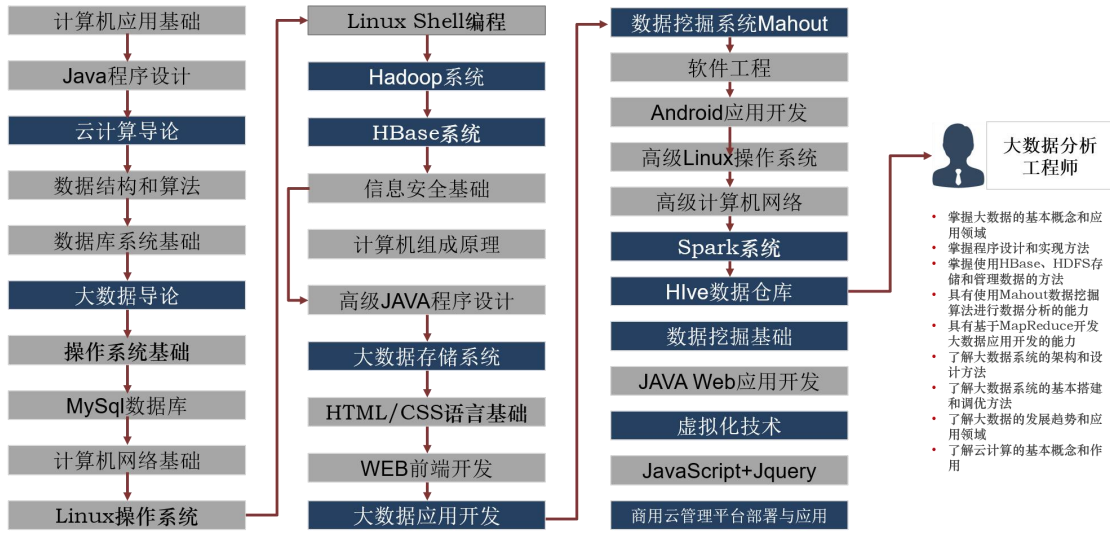
3. 岗位发展学习路径

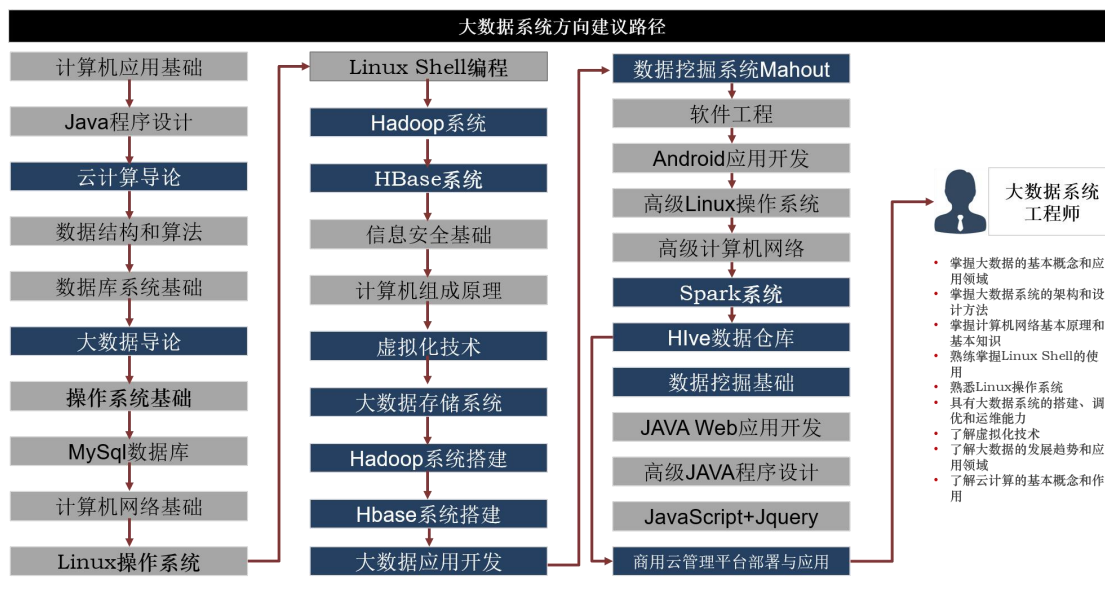
根据每个岗位需要掌握的知识和技能的不同，对上述知识体系进行分解与组合，形成如下图所示的大数据岗位知识学习路径图。

大数据应用方向建议路径



大数据分析方向建议路径





五、 教学计划设计

通过对上述岗位发展学习路径的分析，并结合学校、学生的学习规律，推荐的大数据技术与应用专业教学计划表如下所示：

序号	课程方向	课程名称	第一学期	第二学期	第三学期	第四学期	第五学期	第六学期	总计
1	专业基础课	静态网页制作	72						
2		C 语言程序设计	72						
3		数据库技术		68					
4		Java 语言编程基础		108					
5		HTML5/CSS3 技术		108					
6		JavaScript 技术			68				
7	专业核心课	Java 语言编程高级			108				
8		JSP/Servlet 动态网页			108				
9		Python 语言			108				
10		Hadoop 大数据				108			
11		HBase 分布式数据库				108			
12		SSM 框架				108			
13		网络爬虫					84		
14	专业拓展课	Redis 缓存				68			
15	综合实践课	HTML 静态网页综合实训					84		
16		行业数据分析综合实训					84		

17	大数据综合实训					84		
18	物流数据分析综合实训					84		

六、教学资源建设

经过多年的实践经验发现，教学资源开发必须通过校企合作共建的模式，才能得到适合学教学、满足企业用人需求的教材。通过校企合作共建教学资源，才能更好将学校的教学经验和企业的人才需求融合在一起。这样开发出来的教学资源不仅有助于课程内容与企业岗位需求的粘合度，同时也有助于根据市场的变化及时做出调整。如下课程为校企联合共建课程的部分样例介绍。

● Hadoop 基础

章节	理论课	实验课	合计
第一节课 Hadoop 知识背景	2 学时	0 学时	2 学时
第二节课 Hadoop 集群架构	2 学时	0 学时	2 学时
第三节课 Hadoop 文件系统 HDFS	2 学时	6 学时	8 学时
第四节课 Hadoop 命令行	1 学时	1 学时	2 学时
第五节课 Hadoop 计算模式简介	1 学时	1 学时	2 学时
第六节课 HDFS 基本概念	2 学时	2 学时	4 学时
第七节课 HDFS 运行机制	2 学时	2 学时	4 学时
第八节课 Block 的划分及存储方式	2 学时	2 学时	4 学时
第九节课 HDFS 的备份和文件管理机制	2 学时	2 学时	4 学时
第十节课 HDFS 文件的读写流程	2 学时	2 学时	4 学时
第十一节课 HDFS 的 JAVA API	2 学时	2 学时	4 学时
第十二节课 MapReduce 的基本概念	1 学时	1 学时	2 学时
第十三节课 MapReduce 相关类的功能及调用	2 学时	4 学时	6 学时
第十四节课 MapReduce 程序的运行机制	2 学时	2 学时	4 学时
第十五节课 Combiner 的概念和使用	2 学时	2 学时	4 学时
第十六节课 Partitioner 的概念和使用	2 学时	4 学时	6 学时
合计	29 学时	33 学时	62 学时

● Hbase 基础

章节	理论课	实验课	合计
第一节 HBase 概述	2 学时	2 学时	3 学时
第二节 HBase 表的设计	2 学时	6 学时	8 学时

第三节 PUT 方法：单行 put、KeyValue 类	2 学时	2 学时	4 学时
第四节 PUT 方法：写缓冲区、put 列表、原子操作	2 学时	2 学时	4 学时
第五节 GET 方法：单行 get、result 类	2 学时	0 学时	2 学时
第六节 GET 方法：get 列表、获取数据的方法	2 学时	2 学时	4 学时
第七节 DELETE 方法：单行 delete、列表 delete、原子性操作	2 学时	2 学时	4 学时
第八节 批量处理操作、行锁	2 学时	2 学时	4 学时
第九节 扫描	2 学时	2 学时	4 学时
第十节 过滤器简介	2 学时	0 学时	2 学时
第十一节 比较过滤器：行过滤器、列族过滤器、列名过滤器、值过滤器、参考列过滤器	2 学时	4 学时	6 学时
第十二节 专用过滤器：单列值过滤器、单列排除过滤器、前缀过滤器、分页过滤器	2 学时	2 学时	4 学时
第十三节 专用过滤器：行键过滤器、首次行键过滤器、包含结束的过滤器、时间戳过滤器、列计数过滤器、列分页过滤器、列前缀过滤器、随机行过滤器	2 学时	4 学时	6 学时
第十四节 附加过滤器	2 学时	2 学时	4 学时
第十五节 自定义过滤器	2 学时	2 学时	4 学时
第十六节 计数器：单计数器、多计数器	2 学时	2 学时	4 学时
第十七节 协处理器简介	2 学时	0 学时	2 学时
第十八节 总结	2 学时	0 学时	2 学时
总结	36 学时	36 学时	72 学时

七、 师资队伍建设

1. 专兼结合教学团队建设

通过校企合作，共建稳定的专兼结合双师队伍，采用“走出去、请进来”的方式培养骨干教师，提升教学团队职教能力、工程素养和社会服务能力，做实兼职教师队伍，形成相对稳定的兼职教师队伍和完善的管理办法。

专业遵循高职教育师资成长的客观规律，充分考虑校内教师实践性培养和企业兼职教师教学能力提升的实际需求，通过在企业设立联络工作站，派遣校内教

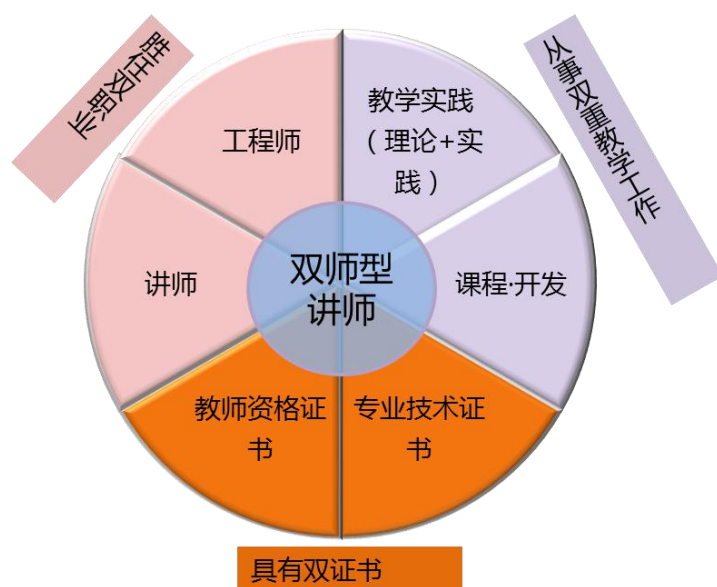
师驻场顶岗实习，提升网络工程项目实践能力，在学校设立企业导师工作室，让企业兼职教师来校开展专业讲座、指导生产性实训。

基于双方互派教师，定期轮岗机制，把与企业合作共同培养教师专业实践能力作为双师素质教师培训工作的重要途径，双方通过制定教师/工程师能力发展计划，有计划、分方向地安排专业教师到企业进行顶岗锻炼、网络工程师到学校参与实践教学。鼓励教师/工程师参加职业技能培训，考取高质量的职业资格证书。

定期让教师“上讲台、下企业、进项目、出成效”，有效培养教师的专业建设能力、技术服务能力和团队合作能力，逐步建立校企技术研讨和经验交流的合作机制，并扩大高技能兼职教师指导实践技能课程的比例。

2. 双师型教师培养

根据教育部大力培养双师型教师的指导性文件与评估标准，我们向合作院校提供阶段式的骨干教师培养体系，通过集中培训、顶岗实习、长期挂职锻炼三个阶段向讲师灌输网络项目实施与维护流程标准，以掌握当前热门应用及典型工程案例应用方案，并将之融入到教学过程中，从而提升教师企业项目指导水平，最终达到教育部双师型讲师实践能力标准。



■ 培养目标

通过集中培训或项目式培训,为高校教师全面而又深入的介绍 Spark、Hadoop 平台的构建流程,涉及 Spark、Hadoop 系统基础知识,概念及架构, Spark、Hadoop 实战技巧, Spark、Hadoop 经典案例等技能。

1. 帮助高校教师对 Spark、Hadoop 生态系统有一个清晰明了的认识;
2. 了解大数据中 Spark、Hadoop 系统适用的场景;
3. 掌握 Spark、Hadoop 等初中级应用开发技能;
4. 搭建稳定可靠的 Spark、Hadoop 集群,满足生产环境的标准;
5. 了解和清楚大数据应用的几个行业中的经典案例,包括阿里巴巴,腾讯等。

八、 时间教学环境建设

根据专业人才培养方案,在实训过程中需配备符合实际教学需求的专业实训室教学装备,以满足专业课程的一体化教学、生产性实习、综合实训、职业技能培训、技能鉴定等教学需要。实训项目与工作岗位任务对接,实训设备选型尽可能与实习实践及就业目标企业业务过程中常用设备保持一致。

大数据实验室采用以下 3 层架构作为建设理念。

硬件资源层:基于高性能计算与海量存储节点构建的运算资源池,作为云计算各项实验学习环境的主要承载平台。

业务平台层:面向教学活动中的实验课程与项目实训业务提供流程化支撑。

资源平台层:提供教学活动中必不可少的理论学习类资源、实验学习类资源、项目实战案例。