

渭南师范学院 大数据与智能信息处理技 术研究工程中心建设

工 作 方 案

计算机学院

2020 年 11 月

目 录

1 大数据专业建设国家政策	4
1.1 大数据的“国家战略”高度	4
1.2 教育部“产教融合，协同育人”指导方向	5
2 大数据行业发展	6
3. 高校大数据专业申报情况	8
4 大数据人才就业前景分析	9
4.1 岗位分类	9
4.2 人才缺口	9
4.3 招聘要求	10
4.4 薪资待遇	11
4.5 职业发展	12
5 高校大数据专业建设现状分析	13
5.1 课程体系内容缺乏	13
5.2 师资培养建设不足	13
5.3 大数据科研工作环境不足	14
5.4 大数据实践平台建设空白	14
5.5 学生就业培养不足	15
6 新开普大数据专业建设解决方案	15
6.1 教学体系建设	16
6.2 课程体系建设	17
6.3 师资培养建设	22

1 大数据专业建设国家政策

1.1 大数据的“国家战略”高度

党的十八届五中全会公报提出要实施“国家大数据战略”，这是大数据第一次写入党的全会决议，标志着大数据战略正式上升为国家战略。五中全会，开启了大数据建设的新篇章。

中共中央政治局 12 月 8 日下午就实施国家大数据战略进行第二次集体学习。中共中央总书记习近平在主持学习时强调，大数据发展日新月异，我们应该审时度势、精心谋划、超前布局、力争主动，深入了解大数据发展现状和趋势及其对经济社会发展的影响，分析我国大数据发展取得的成绩和存在的问题，推动实施国家大数据战略，加快完善数字基础设施，推进数据资源整合和开放共享，保障数据安全，加快建设数字中国，更好服务我国经济社会发展和人民生活改善。

习近平在主持学习时发表了讲话。他指出，大数据是信息化发展的新阶段。随着信息技术和人类生产生活交汇融合，互联网快速普及，全球数据呈现爆发增长、海量集聚的特点，对经济发展、社会治理、国家管理、人民生活都产生了重大影响。世界各国都把推进经济数字化作为实现创新发展的重要动能，在前沿技术研发、数据开放共享、隐私安全保护、人才培养等方面做了前瞻性布局。要推动大数据技术产业创新发展，构建以数据为关键要素的数字经济，运用大数据提升国家治理现代化水平，运用大数据促进保障和改善民生，要切实保障国家数据安全，善于获取数据、分析数据、运用数据，是领导干部做好工作的基本功。各级领导干部要加强学习，懂得大数据，用好大数据，增强利用数据推进各项工作

的本领，不断提高对大数据发展规律的把握能力，使大数据在各项工作中发挥更大作用。

1.2 教育部“产教融合，协同育人”指导方向

**中华人民共和国中央人民政府**
www.gov.cn

简 | 繁 | EN

国务院 总理 新闻 政策 互动 服务 数据 国情

首页 > 国务院公报 > 2018年第1号

【字体：大 中 小】  打印   +

国务院办公厅关于 深化产教融合的若干意见

国办发〔2017〕95号

各省、自治区、直辖市人民政府，国务院各部委、各直属机构：

进入新世纪以来，我国教育事业蓬勃发展，为社会主义现代化建设培养输送了大批高素质人才，为加快发展壮大现代产业体系作出了重大贡献。但同时，受体制机制等多种因素影响，人才培养供给侧和产业需求侧在结构、质量、水平上还不能完全适应，“两张皮”问题仍然存在。深化产教融合，促进教育链、人才链与产业链、创新链有机衔接，是当前推进人力资源供给侧结构性改革的迫切要求，对新形势下全面提高教育质量、扩大就业创业、推进经济转型升级、培育经济发展新动能具有重要意义。为贯彻落实党的十九大精神，深化产教融合，全面提升人力资源质量，经国务院同意，现提出以下意见。

一、总体要求

（一）指导思想。

全面贯彻党的十九大精神，坚持以习近平新时代中国特色社会主义思想为指导，紧紧围绕统筹推进“五位一体”总体布局和协调推进“四个全面”战略布局，坚持以人民为中心，坚持新发展理念，认真落实党中央、国务院关于教育综合改革的决策

习近平总书记在党的十九大报告中强调“着力加快建设实体经济、科技创新、现代金融、人力资源协同发展的产业体系”，并曾多次强调“人才是创新的根基”，“把人才作为支撑发展的第一资源”，“创新驱动实质上是人才驱动”。李克强总理也指出，人力资源培训和开发属于供给侧结构性改革重要内容。当前，我国人才的教育供给和产业需求在结构、质量、水平上还不能完全适应，特别是随着新增劳动年龄人口增速下降，人才供需的结构性矛盾凸显。可以说，深化产教融合，是推进人才和人力资源供给侧结构性改革一项非常迫切的任务。近年来，这一问题受到社会各方广泛关注，也是全国两会代表委员的热点议题。贯彻中央领

导同志重要指示精神，认真回应社会关切，要求我们必须加快破解制约产教融合的政策瓶颈，将新发展理念落到实处，以人才发展引领产业转型升级，推动产教融合成为转型升级的“助推器”、促进就业的“稳定器”、人才红利的“催化器”。

大数据专业相对于其他专业来说，技术复杂性，创新性显得尤为突出，技术的更新进步一日千里，与产业迅速地发展有着很强的关联性。在本专业上产教融合，校企合作，协同育人起着很重要的作用，同时也符合党的十八届五中全会《建议》“建设现代职业教育体系，推进产教融合、校企合作”的政策方针。

2 大数据行业发展

当前主要发达国家，互联网大国，运用大数据技术手段改善社会治理，推动经济发展，提升政府服务能力及监管能力已成为大势所趋。大数据潜在的巨大影响也已经作为发达国家的战略资源，纷纷给出了大数据开放共享的战略方案。

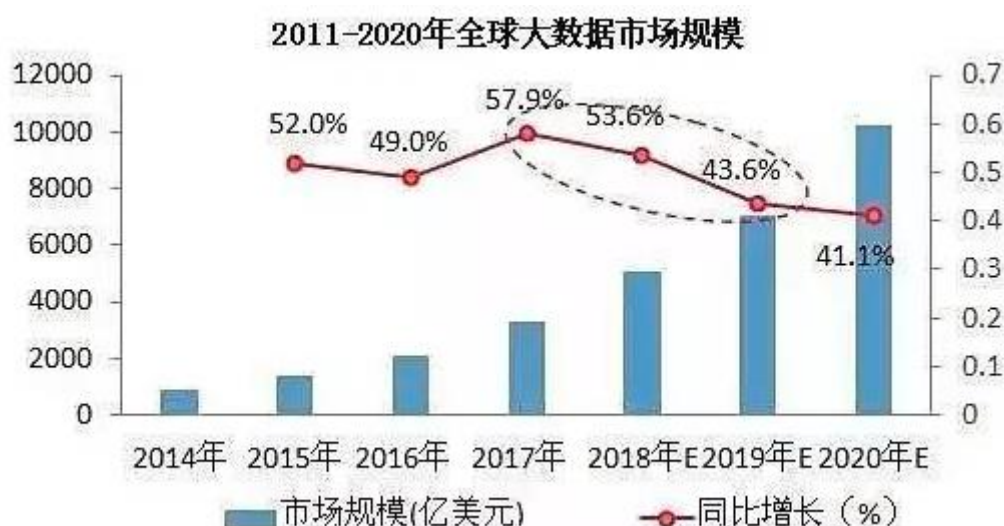
中国作为世界第二大经济体，同时作为互联网、移动互联网应用规模全球第一的大国，拥有丰富的数据资源和应用市场优势，大数据技术的发展日新月异，初具规模。大数据的关键技术研发取得了长足的进步，涌现出了一批创新企业和创新应用。一些地方政府已经启动了大数据相关工作。

目前，从发展阶段来看，我国大数据产业处于快速推进期，中国和美国几乎同一时期关注大数据产业，但与美国存在一定的差距，究其原因，美国是全球信息技术产业的领头羊，在硬件和软件领域都拥有超一流的实力，早在大数据概念火热起来之前，美国信息技术产业在大数据领域已经有了很多技术积累，这使得美国的大型信息技术企业可以迅速转型为大数据企业，从而推动整个大数据产业

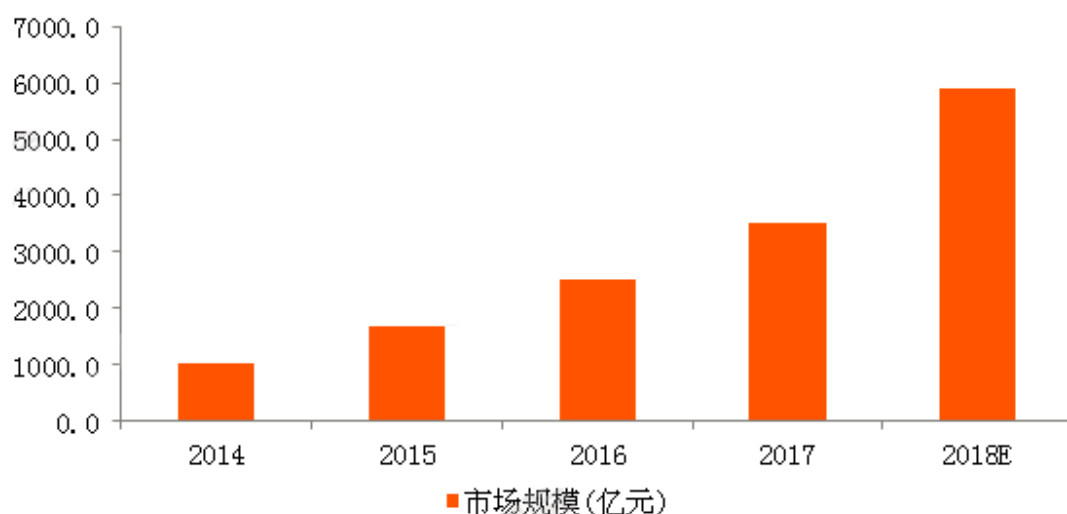
在美国的发展壮大。另外中国数据大多数都掌握在政府手里，数据源比美国相对封闭，数据分析受到局限，影响大数据的发展。



虽然目前中国在大数据领域稍滞后美国，但是从全球范围来看，大数据产业已经开始处于概念热潮的峰值滑落阶段，而我国大数据产业市场规模仍保持超高速增长。



2014-2018年中国大数据市场规模情况及预测



(数据来源：中商产业研究院)

3. 高校大数据专业申报情况

当习总书记将大数据作为国家战略提出后,我国教育部门相应国家战略方向,所有高校单位都开始纷纷积极申报大数据专业或学科。在教育部公布的高校新增专业名单中,第一批成功申请该专业的高校共有3所,为北京大学、对外经济贸易大学及中南大学,于2016年2月获得教育部批准。时隔一年,2017年3月17日,教育部公布第二批“数据科学与大数据技术”专业获批高校。此次获批高校共32所。加上 首批3所高校,我国已有35所高校获批该专业。2018年1月18日,又共有208所职业院校获批“大数据技术与应用”专业,255所本科院校获批“数据科学与大数据技术”专业。

站在互联网“风口”上的大数据,直接催热了大学里的大数据专业。



4 大数据人才就业前景分析

4.1 岗位分类

信息化行业中针对大数据就业一般分为三个方向：大数据系统研发类人才、大数据应用开发类人才和大数据分析类人才。就业岗位分为 10 类：

ETL 研发、Hadoop 开发、可视化（前端展现）工具开发、信息架构开发、数据仓库研究、OLAP 开发、数据科学研究、数据预测（数据挖掘）分析、企业数据管理、数据安全研究。

4.2 人才缺口

现阶段国内大数据人才缺口极大，全球最顶尖管理咨询公司麦肯锡 (McKinsey) 出具的一份详细分析报告显示，预计到 2018 年，大数据或者数据工

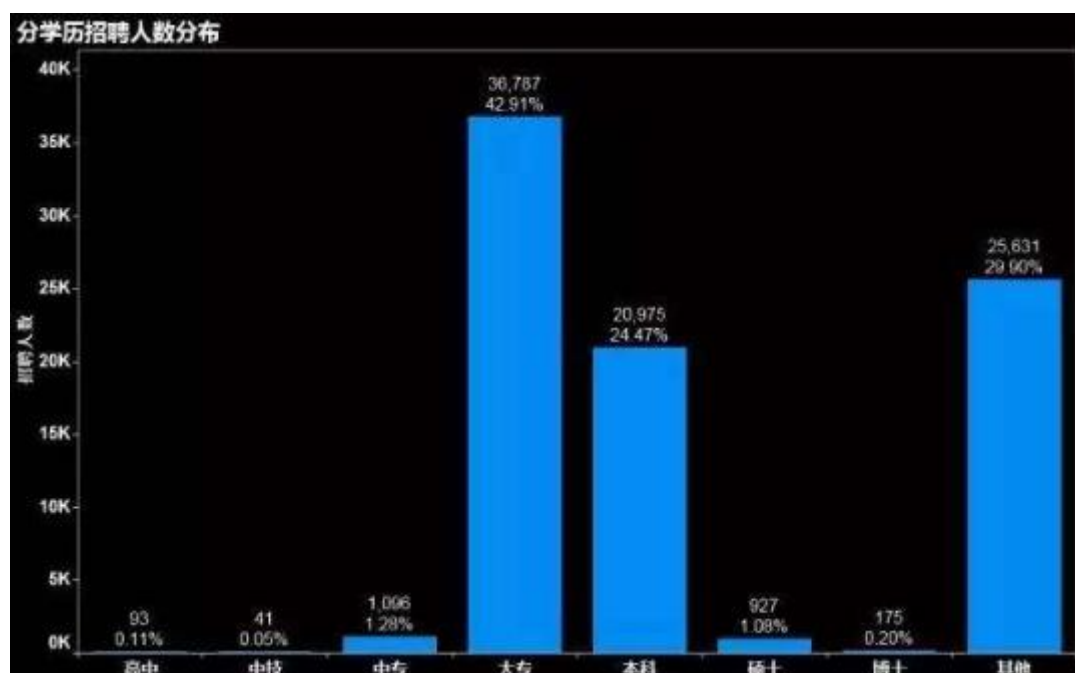
作者的岗位需求将激增，其中大数据科学家的缺口在 140000 到 190000 之间，对于懂得如何利用大数据做决策的分析师和经理的岗位缺口则将达到 150 万！

对数据的占有和控制，将成为国家间和企业间新的争夺焦点。大数据正在成为继云计算之后的新的热词，大数据时代已然来临，大数据背后，隐匿着巨大的商机。包括 IBM，微软，谷歌，亚马逊等一大批知名企业纷纷掘金这一市场。

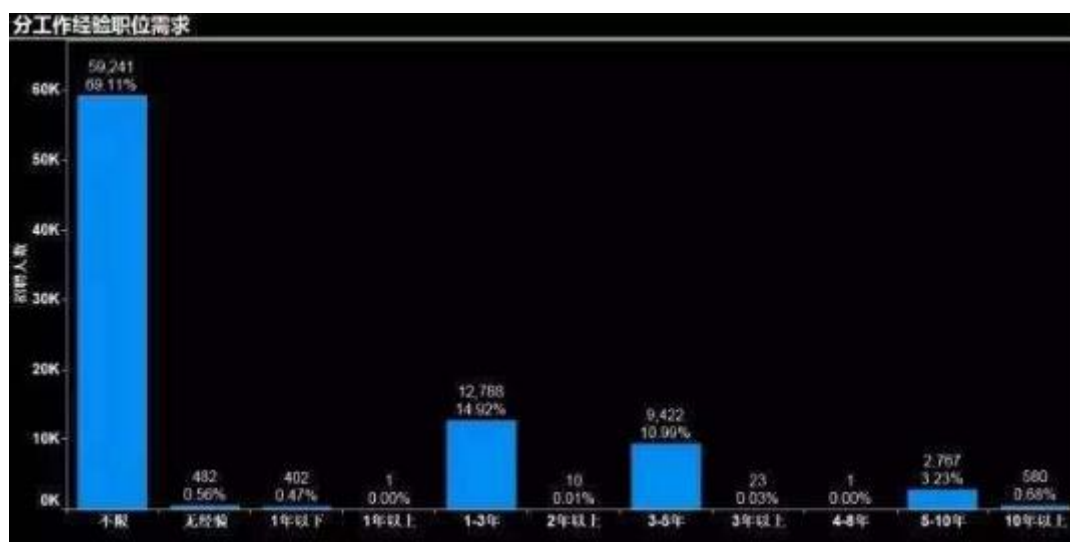
中国商业联合会数据分析专业委员会统计，未来中国基础性数据分析人才缺口将达到 1400 万，而在 BAT 企业招聘的职位里，60%以上都在招大数据人才。

4.3 招聘要求

从工作经验来看，69.1%的企业对求职者的要求是经验不限，这很大一部分说明的人才的稀缺，这对于应届大学生而言，无疑是一大机遇，因为企业对于工作经验要求的低，而应届大学生只需要通过培训就能掌握该岗位应有的技能，就能轻松上岗拿高薪，这也充分说明当下学习大数据绝对是千载难逢的黄金时间点。



(学历需求现状：大专最多，本科次之！ 来源：IDC 中国)



(工作经验需求现状：半数企业经验不限！ 来源：IDC 中国)

4.4 薪资待遇

人才紧缺带来的最直观的现象就是薪酬的提升。一个大数据工程师的月薪轻松过万，一个有几年工作经验的数据分析师的薪酬在 30 万~50 万元之间，而更顶尖的大数据技术人才则是年薪轻松超百万，成为各大互联网和 IT 公司争夺的对象。因而甚至有观点认为，大数据专业正在成为求职者进入大公司的捷径。

1. Hadoop 大数据开发方向

Hadoop 开发工程师入门薪资已经达到了 8K 以上，工作 1 年可达到 1.2W 以上，具有 2-3 年工作经验的 hadoop 开发工程师年薪可以达到 30 万—50 万，一般需要大数据开发的公司基本上都是大公司，Hadoop 开发工程师是目前 BAT 企业、大数据公司最急需的人才，所以学习大数据专业也是进大公司的捷径！

2. 数据挖掘、数据分析&机器学习方向

机器学习职位薪水范围：12k--30K，最大值 30k+若干股票，最小值 12k+若干期权，40%人员待遇在 25k，20%人员待遇在 18k，剩余 40%人员待遇在

13k；数据挖掘薪水范围：12k--30K，最大值 30k+若干股票，最小值 12k，40% 人员待遇在 23k，20%人员待遇在 18k，剩余 40%人员待遇在 14k。

4.5 职业发展

事实上，大到世界 500 强，BAT 这样的公司，小到创业公司，他们都需求数据人才。目前，大数据人才数量较少，因此大多数公司的数据部门一般都是扁平化的层级模式，大致分为数据分析师、资深研究员、部门总监 3 个级别。大公司可能按照应用领域的维度来划分不同团队，而在小公司则需要身兼数职。有些特别强调大数据战略的互联网公司则会另设最高职位——如阿里巴巴的首席数据官。这个职位的大部分人会往研究方向发展，成为重要数据战略人才。

另一方面，大数据工程师对商业和产品的理解，并不亚于业务部门员工，因此也可转向产品部或市场部，乃至上升为公司的高级管理层。



5 高校大数据专业建设现状分析

我国国内高校大数据学科专业的建设情况呈现雨后春笋，争相申报的火热态势。但是存在的矛盾和问题也很突出。问题总体存在以下 3 个方面：

5.1 课程体系内容缺乏

从 2016 年第一批高校申报大数据专业开始才过去 2 年时间，对于大数据专业教学时间最长的这三所院校，教学时间也才不足 2 年而已。大数据的教学过程中仅仅停留在理论阶段是不会有太大的教学成果体现的，必须贯穿大量的实验实践内容，学生理论与实践相结合，才能真正呈现出务实可观的教学成果。复旦大学赵卫东教授说：现阶段高校对大数据方向专业或学科教学普遍存在课程体系空白或者课程体系偏理论，少实践与实验。国内高校大数据分析教学大多偏重理论知识的传授，学生难以接触实际问题，应用能力还存在很多提高的空间，更多的则是刚刚申报成功的高校，根本不知道怎么进行课程开展和实施。教学资料的匮乏，课程内容的不成熟甚至是空白，让“讲什么”，“怎么讲”成为这些高校目前面临的尴尬问题。

5.2 师资培养建设不足

目前国内大数据产业的兴起处于蓬勃发展的初级阶段，大数据技术人才的紧缺直接导致了大数据专业教学的教师队伍人才匮乏，缺少高技术水平和具有国际视野的任课教师，也正是我国大数据产业的刚刚兴起，教师理论及技术水平参差不齐，导致学校专业的建设质量，学科的成长发展以及专业整体教学水平受到了

很大程度的制约。

从现阶段高水平大数据人才具备的必要素质角度出发,高校师资队伍应该具备较高的理论水平,同时也具备企业级大规模应用工作经验,不仅能够对理论研究做进一步延伸推论,还能够将理论与企业应用实践相结合进行教学实施。高校一不具备这样的教师人才储备,二找不到适合自己院校特色的大数据人才。

5.3 大数据科研工作环境不足

以互联网为代表的通信技术和以物联网为代表的传感技术的持续创新和广泛应用使人类的数据化能力和范围快速扩张,产生的能够被测量和记录的数据量越来越大,而我们对事物、现象等的测量、记录也更加频繁和细致。这种扩张让我们看到大量从宏观到微观、从自然到社会的观察,计算、传播等仪器设备和活动都在快速产生着海量、多样的数据。

海量数据的剧增,势必让量变引起质变,直接在科学研究领域,为科研模式带来极大的改变。继实验科学、理论科学、计算科学之后出现了第四种研究范式,即‘数据密集型科学’,成为大数据时代下的新科研模式。

然而现阶段大部分高校的科研模式依然在第四范式之外徘徊,原因在于学校不具备数据密集型科研模式所需要用的科研设备以及科研方法。

5.4 大数据实践平台建设空白

普通计算机专业学生一台机器即可完成学习练习任务,但是大数据课程的实验至少需要2~3台机器才能满足需要。高校资源不可能满足这样的硬件诉求。大数据课程前后课程关联紧密,要求对前课实验成果进行保留。所以实验机和学生

之间存在专属性！这点又呈几何倍数增加了硬件需要！

近两年，随着大数据产业技术的发展，企业实际应用中处理的数据量和技术要求都有了飞跃的提升。

基于以上两点，高校传统的实验室亦或是传统机房已经远远无法满足大数据方面新型技术的应用要求，紧随行业发展要求的现代化大数据实践环境不仅要求高配置的硬件配备，还需要满足高校日常教学的课程实施，并为之提供系统的教学实验环境。

5.5 学生就业培养不足

现阶段，大量社会企业很难找到合适的大数据人才，而且大数据岗位在任何企业公司都是关键咽喉岗位，企业纵然对此类人才求贤若渴，但依然秉持宁缺勿滥原则，导致企业人才招聘困难，成效低下。

而学校对于如何培养出符合企业需求的人才又深感困惑，在当前人才市场“供大于求”的严峻情况下，“毕业即失业”的高校毕业生越来越多。学生的实习实训，就业培养显得最为关键，现在企业宁缺勿滥的人才选择标准，直接将缺乏实际项目经验的毕业生过滤掉了。学生在学校没有机会接触实际项目，没有真实案例进行实战演练，没有行业技术大牛从旁进行宝贵经验的灌输，因此学生不能胜任企业的需求。

6 新开普大数据专业建设解决方案

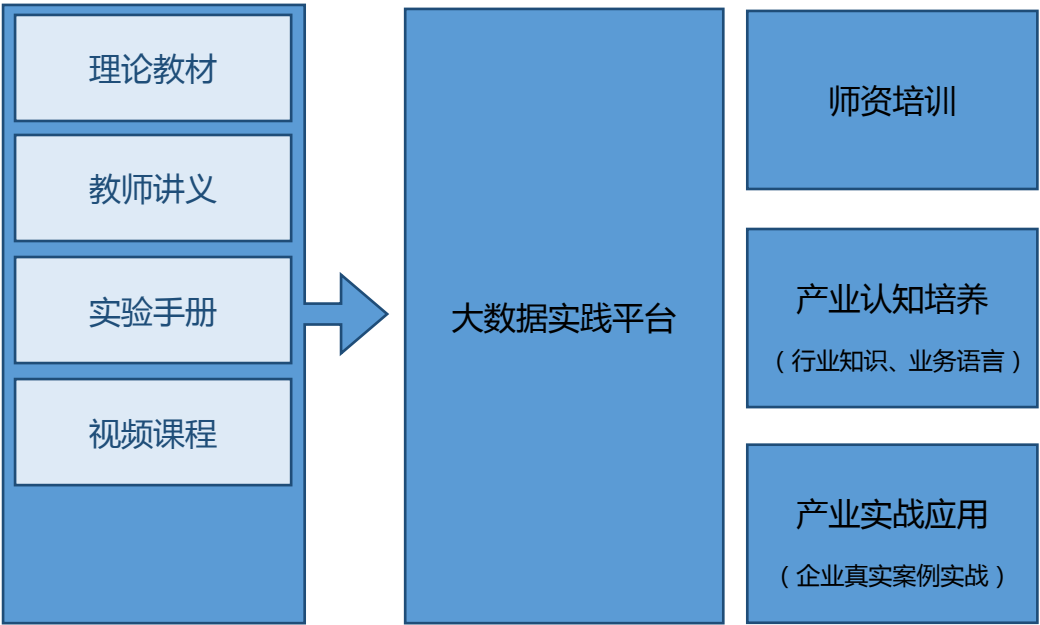
大数据实验室平台是为学校学生、老师、科研人员和管理员进行大数据专业教学上课、在线实验、在线考试以及进行相关大数据技术科研工作的教学科研平

台。大数据实验室平台通过私有云计算及虚拟化技术使得学生、老师和科研人员可以获取自己专属的虚拟机，从而进行在线教学实验和科研工作。平台还提供了一些教学辅助功能如：大数据课程管理系统，大数据考试系统，大数据远程协助系统等，以辅助教师和学生进行课堂教学工作，提高教学效率，优化用户体验。平台还内置配套的国内一流高校及行业专家倾力打造的大数据专业教学内容，包括：课程教材、实验手册、实验案例和海量行业数据。

大数据实验室平台涵盖了云计算虚拟机管理系统、大数据课程管理系统、大数据在线实验系统、大数据远程协助系统、大数据在线考试系统、大数据实验室管理系统、大数据科研系统。

6.1 教学体系建设

新开普大数据专业教学体系建设集学习实践平台、理论教材、实验手册、教师讲义、视频课程、实验数据、行业认知训练、产业实战应用课以及师资培训于一体，解决大数据专业课程“教什么”、“怎么教”、“由谁来教”、“教到什么程度”4个关键问题。教学体系内容涵盖面较为广泛。平台教学体系架构图如下：



6.2 课程体系建设

教育部在“教育部关于印发《教育信息化“十三五”规划》的通知》中指出，“当前，云计算、大数据、物联网、移动计算等新技术逐步广泛应用，经济社会各行业信息化步伐不断加快，社会整体信息化程度不断加深，信息技术对教育的革命性影响日趋明显。党的十八大以来，特别是中央网络安全和信息化领导小组成立后，党中央、国务院对网络安全和信息化工作的重视程度前所未有，“互联网+”行动计划、促进大数据发展行动纲要等有关政策密集出台，信息化已成为国家战略，教育信息化正迎来重大历史发展机遇。”。

“积极利用云计算、大数据等新技术，创新资源平台、管理平台的建设、应用模式。各地要根据信息化教学的实际需求，做好资源平台建设规划论证，充分利用现有通信基础设施，加快推进区域平台建设和与国家教育资源平台的协同服务。鼓励企业根据国家规定与学校需求建设资源平台，提供优质服务。”

我们依托国家“十三五”规划纲要及教育部“行动计划”，开展课程体系建设要以学科建设为指导，强调培养具有大数据实践能力的大数据人才。该专业重点培养具有以下两方面素质的人才：一是工具的掌握，掌握数据采集和数据分析的基本工具；二是数据分析能力，掌握实用数据分析和初步数据建模能力。推荐课程内容包括：《Linux 操作系统》、《数据库原理与应用》、《Java 程序设计》、《Java Web 应用》、《Python 程序设计》、《NoSQL 数据库应用》、《数据仓库与数据治理》、《数据分析与数据挖掘》、《R 语言统计分析》、《Hadoop

核心技术应用》、《Spark 大数据处理与应用》、《数据交换与数据清洗》、《数据可视化技术应用》、《Mahout 数据处理与算法分析》、《数据结构与算法》、《Zookeeper 核心技术应用》、《Kafka 核心技术应用》、《计算机网络》。

以上课程内容提供完整课程体系并可预置在平台中。

课程名称	课程描述
《Linux 操作系统》	《Linux 操作系统》是为高校计算机相关专业编写的 Linux 操作系统教材。全书共分 11 章, 内容包括 Linux 操作系统安装、Linux 桌面管理、Linux 常用命令、文本文件编辑、用户管理、设备管理、DHCP 服务器、Samba 服务器、FTP 服务器、DNS 服务器和 Apache 服务器。
《数据库原理与应用》	系统全面地阐述了数据库系统的基本概念、基本原理和应用技术, 主要内容包括数据库技术概述、关系数据库、关系数据库的标准语言 SQL、关系数据库设计、数据库保护、网络数据库、网络数据库管理系统 SQL Server 2000、分布式数据库系统、XML 数据库等。 本书概念清楚、重点突出、章节安排合理, 理论与实践结合紧密。每章配以丰富的习题、案例及实验, 有助于读者加深对内容的理解、掌握并巩固概念; 案例为读者提供了真实的数据库应用场景, 有助于读者从实际应用的角度出发, 联系所学理论, 掌握所学内容; 而实验为读者提供了将理论与实践相结合的具体上机操作途径, 最终巩固所学内容。习题、案例与实验的设计也是本书一个比较突出的特点。
《Java 程序设计》	《Java 程序设计》通过对 Java 编程语言的全面介绍, 引导快速地掌握 Java 编程语言的核心内容并学会灵活运用所学的语言知识及面向对象的编程思想。《Java 程序设计》共分 9 章, 内容包括 Java 语言概述、面向对象编程初步、Java 的基本语法、类库与数组、面向对象编程深入、Applet 程序、图形用户界面编程、异常处理和输入输出, 以及多线程编程。
《Java Web 应用》	通过 Java Web 开发常用技术及框架 (Spring、SpringMVC、Hibernate、Mybatis) 结合数据库知识, 实现 Java Web 开发内容学习。总体内容结构清晰、划分合理。将帮助学生更完整的了解 Java 研发体系。
《Python 程序设计》	本书是一本针对所有层次的 Python 读者而作的 Python 入门书。全书分两部分: 首部分介绍用 Python 编程所必须了解的基本概念, 包括 matplotlib、NumPy 和 Pygal 等强大的 Python 库和工

	具介绍，以及列表、字典、if 语句、类、文件与异常、代码测试等内容；第二部分将理论付诸实践，讲解如何开发三个项目，包括简单的 Python 2D 游戏开发，如何利用数据生成交互式的信息图，以及创建和定制简单的 Web 应用，并帮读者解决常见编程问题和困惑。
NoSQL 数据库应用	通过系统的学习 NoSQL 的来源及应用场景，结合 Redis、MongoDB、Neo4j、Hbase 完成对键值数据库、文档数据库、图形数据库、列族数据库的学习。实现 NoSQL 数据库的完全掌握及应用。
数据仓库与数据治理	通过元数据、主数据、数据质量、数据剖析、数据分析、数据集成、数据监控等实现数据仓库管理及数据治理。
数据分析与数据挖掘	通过讲解数据采集、数据存储、数据分析、数据挖掘实现数据借助数据分析及数据挖掘工具实现数据分析及数据挖掘方面的基础、进阶内容全覆盖。通过学习 EXCEL、SPSS、SAS、R 语言对数据进行深度剖析，实现从分析到挖掘的进阶。
R 语言统计分析	《R 语言与统计分析》以数据的常用统计分析方法为基础，在简明扼要地阐述统计学基本概念、基本思想与基本方法的基础上，讲述与之相对应的 R 函数的实现，并通过具体的例子说明统计问题求解的过程。《R 语言与统计分析》注重统计的思想性、方法的实用性和计算的可操作性。
《Hadoop 核心技术应用》	《Hadoop 核心技术》是一本技术深度与企业实践并重的著作。本书从使用、实现原理、运维和开发 4 个方面对 Hadoop 的核心技术进行了深入的讲解。详细讲解了 HDFS 存储系统、MapReduce 计算框架，以及 HDFS 的命令系统；结合源代码，深度分析了 MapReduce、HDFS、Streaming、Pipes、Hadoop 作业调度系统等重要技术和组件的架构设计、工作机制和实现原理；结合百度的实际生产环境，详细讲解了 Hadoop 集群的安装、配置、测试以及管理和运维；详细讲解了 HadoopStreaming、Pipes 的使用和开发实践，以及 MapReduce 的编程实践和常见问题。
《Spark 大数据处理与应用》	全面、深入讲解 Spark 大数据分析的各种技术和方法作为一个通用大数据计算框架，Spark 广泛应用于多种典型大数据分析场景。通过这本书，你将了解到：Spark 的机制与生态系统、Spark 算法分析与应用、典型的数据分析场景、Spark 在日志分析、情感分析、协同过滤、个性化推荐、社交网络分析、搜索、文本分析等场景的应用、Spark 与其他开源工具及系统的集成、本书抽象出典型数据分析应用场景，适当简化，只进行主干介绍，讲解 Spark 应用开发、数据分析算法，并结合其他开源系统与组件，

	勾勒出数据分析流水线全貌。
《Zookeeper 核心技术应用》	从分布式一致性的理论出发,向读者简要介绍几种典型的分布式一致性协议,以及解决分布式一致性问题的思路。同时,本书深入介绍了分布式一致性问题的工业解决方案——ZooKeeper,并着重向读者展示这一分布式协调框架的使用方法、内部实现及运维技巧,旨在帮助读者全面了解 ZooKeeper,并更好地使用和运维 ZooKeeper。
《Kafka 核心技术应用》	中的大量实例来源于作者在实际工作中的实践,具有现实指导意义。相信读者阅读完本书之后,能够全面掌握 Kafka 的基本实现原理及其基本操作,能够根据书中的案例举一反三,解决实际工作和学习中的问题。此外,在阅读本书时,读者可以根据本书对 Kafka 理论的分析,再结合 Kafka 源码进行定位学习,了解 Kafka 的设计和思想以及更多的编码技巧。
《计算机网络》	由于本书的重点立足于计算机网络的基本原理,同时兼顾了 Internet 体系结构与 TCP/IP 协议等内容,因此对于学习计算机网络课程的本科生和研究生,本书都是的教材或教学参考书。本书每章后面给出了大量练习题,有助于教师根据教学目的酌情安排课后练习。
数据交换与数据清洗	数据交换与数据清洗在当下大数据环境下显的犹为重要。通过原理、场景、方式方法的介绍再结合 ETL 工具的使用,实现数据交换与数据清洗的学习。
数据可视化技术应用	通过系统的学习数据可视化工具实现宏观态势可视化、设备仿真可视化、数据统计分析可视化、硬件数据可视化等方面的可视化技术及行业的学习。
Mahout 数据处理与算法分析	Mahout 是 Apache Software Foundation (ASF) 旗下的一个开源项目,提供一些可扩展的机器学习领域经典算法的实现,旨在帮助开发人员更加方便快捷地创建智能应用程序。Mahout 包含许多实现,包括聚类、分类、推荐过滤、频繁子项挖掘。此外,通过使用 Apache Hadoop 库, Mahout 可以有效地扩展到云中。
数据结构与算法	以基本数据结构和算法设计策略为知识单元,系统地介绍了数据结构的知识与应用、计算机算法的设计与分析方法,主要内容包括线性表、树、图和广义表、算法设计策略以及查找与排序算法等。《数据结构与算法》注重理论与实践相结合,内容深入浅出,可以作为高等院校计算机学科相关专业的教材或参考书,同时对计算机科技工作者也有参考价值。
《深度学习与人工智能》	深度学习正在引发一场深刻的技术革命,这是人类首次如此接近思维的本质。像手写体识别,脸像识别这类系统,过去的思路是

	从业务背景中线提取特征，然后产生若干辨识逻辑，再形成算法编程实现，但对于像 imagenet 那样要对上百万的图片进行上千个分类识别的问题，以往的技术就傻眼了，别说逻辑，连特征的提取都因为过于复杂而没办法进行。现在流行的深度学习网络的方法，把逻辑隐藏在成百上千万的神经网络权值里，让特征被自动识别与提取，却能得出让人吃惊的高准确率。给出通用的框架，通过大量学习数据训练出合适的权值，权值就是逻辑，这是未来的方向，那种先设计算法敲代码的日子该一去不复返了，以后甚至程序员的工作都由机器全部完成也不是没有可能。 本书旨在全面介绍深度学习与人工智能的各种算法模型与应用，指导读者全面掌握深度学习各个算法的形成与演变的基本脉络，帮助读者快速入门乃至精通深度学习与人工智能，并将各种卷积神经网络模型落地到实际应用。 本书一共分为三部分。第一部分（第一章）介绍神经网络与深度学习的发展演变过程，整体概述本书内容；第二部分（第二章到第七章）从浅到深，介绍了神经网络的发展历程中的各种网络模型与实际应用；第三部分（第八章到第十四章）介绍了深度学习发展历程中的各种经典卷积神经网络、残差神经网络与循环神经网络的网络结构与实际应用；第四部分（第十五章到第十八章）从深度学习任务角度出发，介绍了目标检测、语义分割模型的经典模型，并讲解了如何迁移学习将经典模型化为己用，最后介绍了 AlphaGO 的设计思路，可供读者借鉴使用。 本书可供对深度学习与人工智能感兴趣的读者作为参考教材使用，通过 Imagenet 比赛、图像分割、目标识别等经典案例，逐步深入卷积神经网络的核心，掌握深度学习的精髓，领略 Alpha GO 战胜人类棋手的奥秘。
《机器学习》	大数据时代，机器学习是人工智能领域中一个极其重要的研究方向。随着科学技术的发展，各个企业可能积累了大量的数据，如何从这些数据中挖掘出有商业价值的信息或模式，是目前企业所面临的大数据挑战之一。 本书内容涵盖了机器学习各个领域的算法，回归分析，有监督学习，半监督学习，无监督学习，自然语言处理等。内容结构上注重知识的实用性与可操作性，算法结合案例，提供完全的代码实现，帮助读者快速掌握知识并应用到实处。全书采用深入浅出、循序渐进的讲解方式，帮助初学者快速入门学习。 本书一共十九个章节，大致可以分为以下几部分。第一部分（第一章到第三章），介绍机器学习基础知识与数据的基本处理方法；第二部分（第四章到第十八章）介绍了机器学习中的常用算法（回

	归分析、决策树、随机森林、神经网络、kmeans 聚类等），并结合案例讲解具体算法的代码实现，第三部分（第十九章）介绍了概率图模型算法以及机器学习在自然语言处理中的应用案例。本课程的受众主要是没有经过专业训练的 IT 专业人员，他们可能是程序员，运维，IT 系统架构师等等，也适合没有经过科班训练的数据分析师。数据分析是一个业务+算法+IT 的交叉领域，同时熟悉这三方面知识的人，可以玩转大数据，产生无穷无尽的花样，产生巨大的价值。但很无奈的情况是这种人才太少，IT 人员即使熟悉本公司的业务，但同时又熟悉算法的人极少。一般只能做一些简单的维度统计，指标计算等等，如果说到开发更高智能的系统，知识上鞭长莫及。本课程的目标，正是要打破知识的鸿沟，向 IT 人员普及算法知识，并把这些知识用于实际项目，把中国的机器学习应用能力提升一个台阶。
--	---

6.3 师资培养建设

现在国内高校大数据专业面临的最重要最严峻的一个问题就是缺乏具备国际视野和高水平大数据技术的专业任课老师。为此新开普提供以“理论+实践”的培训形式，结合成熟的课程体系，为各高校客户任课老师进行系统综合的课程培训，同时提供给老师实际真实项目参与，帮助打通其理论知识与产业实践无法结合的关卡，对自身的教学素质做进一步提升，从而能够顺利胜任专业的授课工作。