云聚精英 创新创造

与聪明人一起做精彩的事！

南京云创大数据科技股份有限公司
Nanjing Innovative Data Technologies,Inc.
股票代码：835305　网　站：http://www.cstor.cn
电　话：4008855360 025-83700385
地　址：南京市白下高新技术产业园中国云计算创新基地A栋9层

---

云创大数据
www.cStor.cn

# 大数据实验室建设方案

## Solution for building a big data lab

教学体系+实验体系+实战应用+效果测评

南京云创大数据科技股份有限公司
Nanjing Innovative Data Technologies,Inc.

---

## 云创大数据简介

南京云创大数据科技股份有限公司是专业从事大数据、人工智能、云计算、云存储技术领域研发及产品销售的高新技术企业，被选为中国云计算创新基地理事长单位，被评为国家自主创新示范区瞪羚企业、国家高新技术企业、培育独角兽企业、江苏省规划布局内重点软件企业、江苏省高科技高成长30强、江苏省科技小巨人企业。公司于2015年12月28日在新三板挂牌（股票简称：云创数据，股票代码：835305）。

南京云创大数据科技股份有限公司
Nanjing Innovative Data Technologies,Inc.

在发展过程中，云创大数据始终高度重视大数据和人工智能的教育事业。响应教育部高等教育司号召，云创大数据参与产学合作协同育人项目，先后获批128个“教育部产学合作协同育人项目”。2019年，云创大数据将进一步构建需求导向的人才培养模式，为高校、企业培养大数据与人工智能应用人才提供教材、教学方案、课程体系、实验平台、应用场景等产品与服务，协助高校与企事业单位培养应用型稀缺人才，促进教育链、人才链、产业链与创新链有机衔接。

中国云计算创新基地
China Cloud Computing Innovation Base

## 目录

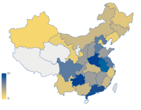
Contents

一、　迫切需求 …… 2
二、　云创优势 …… 3
三、　BDRack大数据实验平台（本科版） …… 6
四、　BDRack大数据实验平台（高职版） …… 12
五、　BDRack大数据实验平台（金融版） …… 15
六、　BDRack大数据实验平台（数学统计版） …… 17
七、　BDRack大数据实验平台（电子商务版） …… 20
八、　平台服务 …… 22
九、　师资培训 …… 24
十、　培养方案 …… 26
十一、认证证书 …… 26
十二、网站资源 …… 27
十三、专家团队 …… 27
十四、成功案例 …… 28

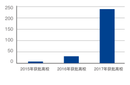
---

## 迫切需求

据《人民日报》报道，未来5-10年，我国大数据人才缺口将突破150万。高校纷纷启动大数据人才培养计划，283所高校获批“数据科学与大数据技术”专业，208所职业院校获批“大数据技术与应用”专业，但是专业建设困难重重，面临基础设施配置难、实验内容缺乏、配套教材缺失、专业师资紧缺、实训项目缺乏等内容。



全国新增“大数据技术与应用”专业学校省市分布情况



“数据科学与大数据技术”专业获批高校数增长图

- **1）基础设施**
学生实验需要大量的集群环境，一次性基础设施投入较高。
- **2）实验内容**
生态体系庞大复杂，技术发展迅速，实验设计花费时间。
- **3）教材体系**
高速发展的新领域，市场上高质量成体系的教材缺失
- **4）师资培训**
教师中精通大数据技术的人才较少
- **5）实训项目**
缺乏切合业界大数据实际应用的实训项目
- **6）考试系统**
实验考核、试题设计、考试组织管理、试卷批阅占用大量教师时间，考核成绩主观性强

## 云创优势

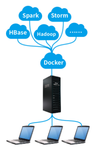
经过多年积累，云创大数据建立了大数据教学和实验的完整体系：

| | 本科 | 高职 |
|---|---|---|
| 大数据实验平台（标准版） | ✓ | ✓ |
| 大数据实验平台（金融版） | ✓ | ✓ |
| 大数据实验平台（电子商务版） | ✓ | ✓ |
| 大数据实验平台（数学统计版） | ✓ | ✓ |

与普通课程相比，云创BDRack大数据实验平台的优势如下：

| | 课程模式 | 教学体系 | 实验体系 | 实战应用 | 效果测评 |
|---|---|---|---|---|---|
| 大数据实验平台 | 边学边练 | 系统教材 | •实验手册 •实验代码<br>•实验数据 •教学视频 | 实战丰富<br>注重实操 | 线上测试 |
| 普通课程 | 理论学习 | 单一课程 | 资源单一，实验难跑通 | 缺乏实战<br>枯燥无趣 | 人工答卷 |

BDRack
大数据实验一体机

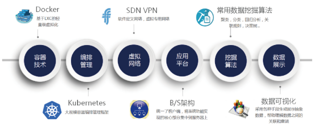

Spark
Storm
HBase
Hadoop
……
Docker

大数据
PPT
大数据实验手册
实验视频
实验数据

---

云创大数据实验室建设方案提供集硬件环境、软件环境、教学体系、课程体系、实验体系以及师资培训于一体的建设思路，提供了完整的大数据教学、实验体系及配套资源，包含实验环境、实验代码、实验手册、实验数据、系列教材、配套PPT、课程视频、师资培训等内容，“一站式”解决大数据教学实验难题。

云创BDRack大数据实验平台基于Docker容器技术，目前已经在数十所高校落地实施，受到一致好评。


Docker
基于LXC的轻量级虚拟化
SDN VPN
软件定义网络，虚拟专用网络
常用数据挖掘算法
聚类，分类，回归分析，关联规则，决策树...
容器技术
编排管理
虚拟网络
应用平台
挖掘算法
数据展示
Kubernetes
大规模容器编排管理框架
B/S架构
统一了客户端，将系统功能实现的核心部分集中到服务器上
数据可视化
采用各种手段生动展示抽象数据，帮助理解数据之间的关联和意涵

- 支持单点和集群的动手实验
- 内置数据挖掘等教学实验数据
- 实验集群相互隔离，互不干扰
- 即学即用，一键重启，用完即走

目前，BDRack大数据实验平台具有经济型、标准型与增强型三种规格，通过发挥实验设备、理论教材、实验手册等资源的合力，可满足数据存储、管理、计算、挖掘、可视化等多样化的教学科研需求。

| 配套/型号 | 经济型 | 标准型 | 增强型 |
|---|---|---|---|
| 管理节点 | 1台 | 3台 | 3台 |
| 处理节点 | 6台 | 8台 | 15台 |
| 上机人数 | 30人 | 60人 | 150人 |
| 理论教材 | 《大数据》50本 | 《大数据》80本 | 《大数据》180本 |
| 实验教材 | 《大数据实验手册》PDF版 | 《大数据实验手册》PDF版 | 《大数据实验手册》PDF版 |
| 配套PPT | 有 | 有 | 有 |
| 配套视频 | 有 | 有 | 有 |
| 免费培训 | 提供现场实施及3天技术培训服务 | 提供现场实施及5天技术培训服务 | 提供现场实施及7天技术培训服务 |

---

## BDRack大数据实验平台（本科版）

BDRack大数据实验平台（本科版）实验涵盖Hadoop生态、Python数据爬取分析挖掘生态、R语言数据分析挖掘生态、大数据实战、数据挖掘及可视化等原理验证、综合应用、自主设计及创新的多层次实验内容，项目实战均有真实的项目数据或实时数据予以支持，并集成考试系统、开发工具等应用，每个实验呈现详细的实验目的、实验内容、实验原理和实验流程指导，参照手册即可轻松完成每个实验。

实验三 HDFS实验：读写HDFS文件

### BDRack大数据实验平台（本科版）实验清单

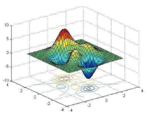
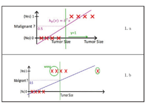
| 实验分类 | 序号 | 实验名称 | 实验手册 | 实验代码 | 实验数据 | 教学视频 |
|---|---|---|---|---|---|---|
| Hadoop生态圈集群实验 | 1 | 大数据基本操作 | ✓ | ✓ | ✓ | ✓ |
| | 2 | HDFS实验：部署HDFS集群 | ✓ | ✓ | ✓ | ✓ |
| | 3 | HDFS实验：读写HDFS文件 | ✓ | ✓ | ✓ | ✓ |
| | 4 | YARN实验：部署YARN集群 | ✓ | ✓ | ✓ | ✓ |
| | 5 | MapReduce实验：单词计数 | ✓ | ✓ | ✓ | ✓ |
| | 6 | MapReduce实验：二次排序 | ✓ | ✓ | ✓ | ✓ |
| | 7 | MapReduce实验：计数器 | ✓ | ✓ | ✓ | ✓ |
| | 8 | MapReduce实验：JOIN操作 | ✓ | ✓ | ✓ | ✓ |
| | 9 | MapReduce实验：分布式缓存 | ✓ | ✓ | ✓ | ✓ |
| | 10 | Hive实验：部署Hive | ✓ | ✓ | ✓ | ✓ |
| | 11 | Hive实验：新建Hive表 | ✓ | ✓ | ✓ | ✓ |
| | 12 | Hive实验：Hive分区 | ✓ | ✓ | ✓ | ✓ |
| | 13 | Spark实验：部署Spark集群 | ✓ | ✓ | ✓ | ✓ |
| | 14 | Spark实验：SparkWordCount | ✓ | ✓ | ✓ | ✓ |
| | 15 | Spark实验：RDD综合实验 | ✓ | ✓ | ✓ | ✓ |
| | 16 | Spark实验：Spark综例 | ✓ | ✓ | ✓ | ✓ |
| | 17 | Spark实验：Spark-sql | ✓ | ✓ | ✓ | ✓ |
| | 18 | Spark实验：Spark Streaming | ✓ | ✓ | ✓ | ✓ |

| 实验分类 | 序号 | 实验名称 | 实验手册 | 实验代码 | 实验数据 | 教学视频 |
|---|---|---|---|---|---|---|
| Hadoop生态圈集群实验 | 19 | Spark实验：GraphX | ✓ | ✓ | ✓ | ✓ |
| | 20 | Zookeeper实验：部署Zookeeper | ✓ | ✓ | ✓ | ✓ |
| | 21 | Zookeeper实验：进程协作 | ✓ | ✓ | ✓ | ✓ |
| | 22 | HBase实验：部署HBase | ✓ | ✓ | ✓ | ✓ |
| | 23 | HBase实验：新建HBase表 | ✓ | ✓ | ✓ | ✓ |
| | 24 | Storm实验：部署Storm | ✓ | ✓ | ✓ | ✓ |
| | 25 | Storm实验：实时WordCountTopology | ✓ | ✓ | ✓ | ✓ |
| | 26 | Flume实验：文件数据Flume至HDFS | ✓ | ✓ | ✓ | ✓ |
| | 27 | Kafka实验：Kafka订阅推送示例 | ✓ | ✓ | ✓ | ✓ |
| | 28 | Pig实验：Pig版WordCount | ✓ | ✓ | ✓ | ✓ |
| | 29 | Redis实验：Redis部署与简单使用 | ✓ | ✓ | ✓ | ✓ |
| | 30 | Redis实验：MapReduce与Spark读写Redis | ✓ | ✓ | ✓ | ✓ |
| | 31 | MongoDB实验：读写MongoDB | ✓ | ✓ | ✓ | ✓ |
| | 32 | LevelDB实验：读写LevelDB | ✓ | ✓ | ✓ | ✓ |
| | 33 | Mahout实验：K-Means | ✓ | ✓ | ✓ | ✓ |
| | 34 | 聚类算法实验：使用Spark实现K-Means | ✓ | ✓ | ✓ | ✓ |
| | 35 | 分类算法实验：使用Spark实现SVM | ✓ | ✓ | ✓ | ✓ |
| | 36 | 关联规则算法实验：使用Spark实现FP-Growth | ✓ | ✓ | ✓ | ✓ |
| Hadoop实战实验 | 37 | 综合实战：车牌识别 | ✓ | ✓ | ✓ | |
| | 38 | 综合实战：搜索引擎 | ✓ | ✓ | ✓ | |
| | 39 | 综合实战：推荐系统 | ✓ | ✓ | ✓ | |
| | 40 | 综合实战：环境大数据 | ✓ | ✓ | ✓ | |
| | 41 | 综合实战：智能硬件大数据托管 | ✓ | ✓ | ✓ | |
| | 42 | 综合实战：贷款风险评估 | ✓ | ✓ | ✓ | |
| Python编程基础实验 | 43 | Python基础：流程控制 | ✓ | ✓ | ✓ | |
| | 44 | Python基础：列表和元组 | ✓ | ✓ | ✓ | |
| | 45 | Python基础：字典 | ✓ | ✓ | ✓ | |
| | 46 | Python基础：文件操作 | ✓ | ✓ | ✓ | |
| Python Hadoop实战实验 | 47 | Python MR：Python统计全国各省城市数量分布 | ✓ | ✓ | ✓ | |
| | 48 | Python MR：Python统计上海2016年每月历史天气 | ✓ | ✓ | ✓ | |
| | 49 | Python MR：Python统计上海2016年每月空气质量 | ✓ | ✓ | ✓ | |
| | 50 | Python MR：Python统计北京和上海2016年月均气温对比 | ✓ | ✓ | ✓ | |
| | 51 | Python MR：Python统计北京和上海2016年空气质量对比 | ✓ | ✓ | ✓ | |
| Python数据挖掘算法实现实验 | 52 | Python算法：决策树分类 | ✓ | ✓ | ✓ | |
| | 53 | Python算法：随机森林分类 | ✓ | ✓ | ✓ | |
| | 54 | Python算法：朴素贝叶斯分类 | ✓ | ✓ | ✓ | |
| | 55 | Python算法：K最近邻分类 | ✓ | ✓ | ✓ | |

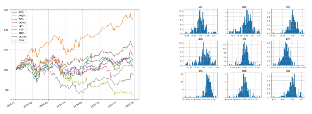
| 实验分类 | 序号 | 实验名称 | 实验手册 | 实验代码 | 实验数据 | 教学视频 |
|---|---|---|---|---|---|---|
| Python数据挖掘算法实现实验 | 56 | Python算法：支持向量机分类 | ✓ | ✓ | ✓ | |
| | 57 | Python算法：K-means聚类 | ✓ | ✓ | ✓ | |
| | 58 | Python算法：DBSCAN聚类 | ✓ | ✓ | ✓ | |
| | 59 | Python算法：回归分析 | ✓ | ✓ | ✓ | |
| | 60 | Python算法：Apriori关联规则 | ✓ | ✓ | ✓ | |
| Python数据挖掘实战实验 | 61 | Python实战：随机森林分类空气质量 | ✓ | ✓ | ✓ | |
| | 62 | Python实战：区域经纬度聚类 | ✓ | ✓ | ✓ | |
| | 63 | Python实战：回归预测空气指数 | ✓ | ✓ | ✓ | |
| R语言编程基础实验 | 64 | R语言基础：流程控制 | ✓ | ✓ | ✓ | |
| | 65 | R语言基础：文件操作 | ✓ | ✓ | ✓ | |
| | 66 | R语言基础：数据帧 | ✓ | ✓ | ✓ | |
| | 67 | R语言基础：因子操作 | ✓ | ✓ | ✓ | |
| R语言数据挖掘算法实现实验 | 68 | R语言算法：决策树分类 | ✓ | ✓ | ✓ | |
| | 69 | R语言算法：随机森林分类 | ✓ | ✓ | ✓ | |
| | 70 | R语言算法：贝叶斯分类 | ✓ | ✓ | ✓ | |
| | 71 | R语言算法：KNN分类 | ✓ | ✓ | ✓ | |
| | 72 | R语言算法：SVM分类 | ✓ | ✓ | ✓ | |
| | 73 | R语言算法：K-means聚类 | ✓ | ✓ | ✓ | |
| | 74 | R语言算法：DBSCAN聚类 | ✓ | ✓ | ✓ | |
| | 75 | R语言算法：回归分析 | ✓ | ✓ | ✓ | |
| | 76 | R语言算法：Apriori关联规则实验 | ✓ | ✓ | ✓ | |
| | 77 | R语言算法：时间序列分析 | ✓ | ✓ | ✓ | |
| R语言数据挖掘实战实验 | 78 | R语言实战：回归预测空气指数 | ✓ | ✓ | ✓ | |
| | 79 | R语言实战：按月进行时间序列预测温度 | ✓ | ✓ | ✓ | |
| | 80 | R语言实战：区域经纬度聚类 | ✓ | ✓ | ✓ | |
| | 81 | R语言实战：随机森林分类空气质量 | ✓ | ✓ | ✓ | |
| Linux基础和系统运维实验 | 82 | Linux基础：常用基本命令 | ✓ | ✓ | ✓ | |
| | 83 | Linux基础：文件操作 | ✓ | ✓ | ✓ | |
| | 84 | Linux基础：文本编辑器Vi | ✓ | ✓ | ✓ | |
| | 85 | Shell基础：主机信息检测 | ✓ | ✓ | ✓ | |
| | 86 | Shell基础：流程控制 | ✓ | ✓ | ✓ | |
| | 87 | Linux基础：正则表达式 | ✓ | ✓ | ✓ | |

| 实验分类 | 序号 | 实验名称 | 实验手册 | 实验代码 | 实验数据 | 教学视频 |
|---|---|---|---|---|---|---|
| Linux基础和系统运维实验 | 88 | Linux基础：SED | ✓ | ✓ | ✓ | |
| | 89 | Linux基础：AWK | ✓ | ✓ | ✓ | |
| | 90 | Shell运维实践：用户管理 | ✓ | ✓ | ✓ | |
| | 91 | Shell运维实践：服务器安全 | ✓ | ✓ | ✓ | |
| 网络数据爬取实验 | 92 | Python实战：爬取双色球开奖信息数据 | ✓ | ✓ | ✓ | |
| | 93 | Python实战：爬取电影信息 | ✓ | ✓ | ✓ | |
| | 94 | Python实战：爬取豆瓣图书信息 | ✓ | ✓ | ✓ | |
| | 95 | Python实战：爬取网络小说（1） | ✓ | ✓ | ✓ | |
| | 96 | Python实战：爬取网络小说（2） | ✓ | ✓ | ✓ | |
| Spark综合实战实验 | 97 | Spark综合实战：环境数据读写 | ✓ | ✓ | ✓ | |
| | 98 | Spark综合实战：GM11预测环境数据 | ✓ | ✓ | ✓ | |
| | 99 | Spark综合实战：词频统计 | ✓ | ✓ | ✓ | |
| | 100 | Spark综合实战：聚类算法学习 | ✓ | ✓ | ✓ | |
| | 101 | Spark综合实战：算子学习 | ✓ | ✓ | ✓ | |
| | 102 | Spark综合实战：特征提取 | ✓ | ✓ | ✓ | |
| | 103 | Spark综合实战：分类算法学习 | ✓ | ✓ | ✓ | |
| | 104 | Spark综合实战：日志分析 | ✓ | ✓ | ✓ | |
| | 105 | Spark综合实战：ALS算法推荐 | ✓ | ✓ | ✓ | |
| | 106 | Spark综合实战：中文分词 | ✓ | ✓ | ✓ | |

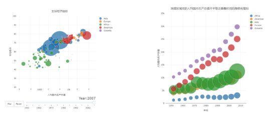
## 典型实验1：Spark垃圾邮件分类

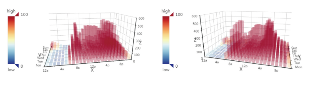
## 典型实验2：最优投资组合



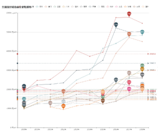
## 典型实验3：全球经济指标动态分析



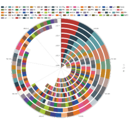
## 典型实验4：一周不同时段的骑行需求



## 典型实验5：
## 全国部分城市房价变化趋势



## 典型实验6：
## 某房源发布网站房价极坐标图



实验配套教材：由清华大学博士、中国信息协会大数据分会副会长刘鹏教授主编，电子工业出版社出版，并被全国高校广泛使用的“高级大数据人才培养丛书”：

中国工信出版集团 電子工業出版社

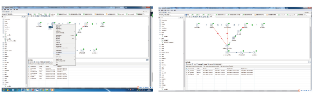
云计算（第三版）
CLOUD COMPUTING, Third Edition
大数据
大数据库
数据挖掘
大数据实验手册
Manual for Big Data Experiments
深度学习
大数据可视化
虚拟化与容器

# BDRack大数据实验平台（高职版）

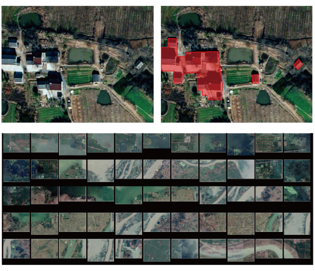
BDRack大数据实验平台（高职版）主要针对高职院校，上线24个涵盖数据清洗与数据标注的高职版大数据实验：

| 实验分类 | 序号 | 实验名称 | 实验手册 | 实验代码 | 实验数据 |
|---|---|---|---|---|---|
| 数据清洗实验 | 1 | Kettle介绍及从文本文件抽取数据到数据库 | ✓ | ✓ | ✓ |
| | 2 | CSV文件数据抽取到数据库 | ✓ | ✓ | ✓ |
| | 3 | Excel文件导入数据库 | ✓ | ✓ | ✓ |
| | 4 | Json文件和XML文件的抽取 | ✓ | ✓ | ✓ |
| | 5 | MySQL数据迁移MongoDB | ✓ | ✓ | ✓ |
| | 6 | 住房数据清洗 | ✓ | ✓ | ✓ |
| | 7 | 银华基金数据清洗实例 | ✓ | ✓ | ✓ |
| | 8 | 数据库增量数据抽取 | ✓ | ✓ | ✓ |
| | 9 | 客户签到数据的清洗转换 | ✓ | ✓ | ✓ |
| | 10 | 基于触发器的数据增删改的增量更新 | ✓ | ✓ | ✓ |
| | 11 | 数据脱敏实例 | ✓ | ✓ | ✓ |
| 数据标注实验 | 12 | 人脸分类标注 | ✓ | ✓ | ✓ |
| | 13 | 物品分类标注 | ✓ | ✓ | ✓ |
| | 14 | 细胞分类标注 | ✓ | ✓ | ✓ |
| | 15 | 行人标框标注 | ✓ | ✓ | ✓ |
| | 16 | 车牌标框标注 | ✓ | ✓ | ✓ |
| | 17 | 人脸标框标注 | ✓ | ✓ | ✓ |
| | 18 | 细胞标框标注 | ✓ | ✓ | ✓ |
| | 19 | 人脸描点标注 | ✓ | ✓ | ✓ |
| | 20 | 车辆多边形标注 | ✓ | ✓ | ✓ |
| | 21 | 细胞多边形标注 | ✓ | ✓ | ✓ |
| | 22 | 道路标志标线区域标注 | ✓ | ✓ | ✓ |
| | 23 | 遥感影像区域标注 | ✓ | ✓ | ✓ |
| | 24 | 医疗影像区域标注 | ✓ | ✓ | ✓ |

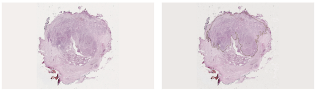
## 典型实验1：客户签到数据的清洗转换



## 典型实验2：遥感影像区域标注



## 典型实验3：医疗影像标注



实验配套教材：由清华大学博士、中国信息协会大数据分会副会长刘鹏教授主编，清华大学出版社出版，并被全国高校广泛使用的“大数据应用人才培养系列教材”。

清华大学出版社

大数据系统运维
数据清洗
大数据实践
R语言
大数据导论
云计算导论
数据挖掘基础

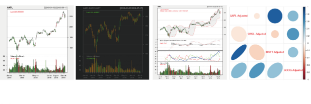
# BDRack大数据实验平台（金融版）

在BDRack大数据实验平台（本科）的基础上，BDRack大数据实验平台（金融版）进一步丰富了实验内容，增加了30个金融类实验。具体实验列表如下：

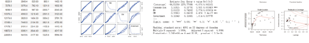
| 实验分类 | 序号 | 实验名称 | 实验手册 | 实验代码 | 实验数据 |
|---|---|---|---|---|---|
| 风险管理 | 1 | 航空燃油的交叉对冲 | ✓ | ✓ | ✓ |
| | 2 | 金融风险管理：VaR 和 ES | ✓ | ✓ | ✓ |
| | 3 | 金融风险管理：Delta-normal方法计算 VaR 和 ES | ✓ | ✓ | ✓ |
| | 4 | 金融风险管理：历史模拟法、蒙特卡罗模拟法计算 VaR 和 ES | ✓ | ✓ | ✓ |
| | 5 | 金融风险管理：分位数回归法计算 VaR | ✓ | ✓ | ✓ |
| 股票数据 | 6 | 分析新浪股票数据 | ✓ | ✓ | ✓ |
| | 7 | 股票数据抓取与分析（一） | ✓ | ✓ | ✓ |
| | 8 | 股票数据抓取与分析（二） | ✓ | ✓ | ✓ |
| | 9 | 分析股票指数的GARCH效应 | ✓ | ✓ | ✓ |
| | 10 | 夏普比率与最大回撤和最大回撤时间 | ✓ | ✓ | ✓ |
| | 11 | 最优风险资产组合（一） | ✓ | ✓ | ✓ |
| | 12 | 最优风险资产组合（二） | ✓ | ✓ | ✓ |
| | 13 | 预测股票走势 | ✓ | ✓ | ✓ |
| 建模分析 | 14 | 时间序列分析-ARIMA模型（一） | ✓ | ✓ | ✓ |
| | 15 | 时间序列分析-指数平滑模型（二） | ✓ | ✓ | ✓ |
| | 16 | 对英国房屋价格建模并预测 | ✓ | ✓ | ✓ |
| | 17 | 多元线性回归研究我国经济增长 | ✓ | ✓ | ✓ |
| | 18 | 建立VAR模型分析联合内生变量的动态关系（一） | ✓ | ✓ | ✓ |
| | 19 | 建立VAR模型分析联合内生变量的动态关系（二） | ✓ | ✓ | ✓ |

| 实验分类 | 序号 | 实验名称 | 实验手册 | 实验代码 | 实验数据 |
|---|---|---|---|---|---|
| 建模分析 | 20 | 基于CAPM模型的预期收益率与实际收益率 | ✓ | ✓ | ✓ |
| | 21 | 航空公司客户价值分析 | ✓ | ✓ | ✓ |
| | 22 | 漏电窃电用户行为分析与事件识别（一） | ✓ | ✓ | ✓ |
| | 23 | 漏电窃电用户行为分析与事件识别（二） | ✓ | ✓ | ✓ |
| | 24 | Hadoop综合实战：贷款风险评估 | ✓ | ✓ | ✓ |
| 其他实验 | 25 | 电商产品评论数据情感分析 | ✓ | ✓ | ✓ |
| | 26 | 实时中美货币转换 | ✓ | ✓ | ✓ |
| | 27 | 利用层次聚类算法进行基于基站定位数据的商圈分析 | ✓ | ✓ | ✓ |
| | 28 | 应用系统负载分析与磁盘容量预测（上） | ✓ | ✓ | ✓ |
| | 29 | 应用系统负载分析与磁盘容量预测（下） | ✓ | ✓ | ✓ |
| | 30 | Hadoop综合实战：推荐系统 | ✓ | ✓ | ✓ |

## 典型实验1：股票数据抓取与分析



## 典型实验2：多元线性回归研究我国经济增长



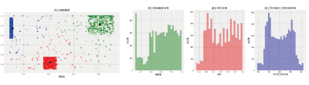
# BDRack大数据实验平台（数学统计版）

在BDRack大数据实验平台（本科）的基础上，BDRack大数据实验平台（数学统计版）进一步丰富了实验内容，增加了53 个数学统计实验。具体实验列表如下：

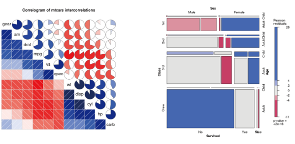
| 实验分类 | 序号 | 实验名称 | 实验手册 | 实验代码 | 实验数据 |
|---|---|---|---|---|---|
| 统计分析 | 1 | R语言数学统计分析-时间序列（一） | ✓ | ✓ | ✓ |
| | 2 | R语言数学统计分析-时间序列（二） | ✓ | ✓ | ✓ |
| | 3 | R语言数学统计分析-时间序列（三） | ✓ | ✓ | ✓ |
| | 4 | R语言数学统计实验-主成分分析和因子分析（一） | ✓ | ✓ | ✓ |
| | 5 | R语言数学统计实验-主成分分析和因子分析（二） | ✓ | ✓ | ✓ |
| | 6 | R语言数学统计实验-分类 | ✓ | ✓ | ✓ |
| | 7 | R语言数学统计实验-分类和聚类（一） | ✓ | ✓ | ✓ |
| | 8 | R语言数学统计实验-分类和聚类（二） | ✓ | ✓ | ✓ |
| | 9 | R语言数学统计实验-功效分析 | ✓ | ✓ | ✓ |
| | 10 | R语言数学统计实验-回归（一） | ✓ | ✓ | ✓ |
| | 11 | R语言数学统计实验-回归（二） | ✓ | ✓ | ✓ |
| | 12 | R语言数学统计实验-回归（三） | ✓ | ✓ | ✓ |
| | 13 | R语言数学统计实验-回归（四） | ✓ | ✓ | ✓ |
| | 14 | R语言数学统计实验-基本统计分析（一） | ✓ | ✓ | ✓ |
| | 15 | R语言数学统计实验-基本统计分析（二） | ✓ | ✓ | ✓ |
| | 16 | R语言数学统计实验-基本统计分析（三） | ✓ | ✓ | ✓ |
| | 17 | R语言数学统计实验-数据平滑 | ✓ | ✓ | ✓ |
| | 18 | R语言数学统计实验-数据的筛选与汇总 | ✓ | ✓ | ✓ |
| | 19 | R语言数学统计实验-数据重构 | ✓ | ✓ | ✓ |
| | 20 | R语言数学统计实验-方差分析（一） | ✓ | ✓ | ✓ |
| | 21 | R语言数学统计实验-方差分析（二） | ✓ | ✓ | ✓ |
| | 22 | R语言数学统计实验-时序数据分析 | ✓ | ✓ | ✓ |
| | 23 | R语言数学统计实验-概率与分布 | ✓ | ✓ | ✓ |
| | 24 | R语言数学统计实验-聚类分析 | ✓ | ✓ | ✓ |

| 实验分类 | 序号 | 实验名称 | 实验手册 | 实验代码 | 实验数据 |
|---|---|---|---|---|---|
| 统计分析 | 25 | R语言数学统计实验-重抽样与自助法（一） | ✓ | ✓ | ✓ |
| | 26 | R语言数学统计实验-重抽样与自助法（二） | ✓ | ✓ | ✓ |
| | 27 | R语言数学统计实验-高级数据管理 | ✓ | ✓ | ✓ |
| | 28 | Python数学统计实验-研究统计学基础（一） | ✓ | ✓ | ✓ |
| | 29 | Python数学统计实验-研究统计学基础（二） | ✓ | ✓ | ✓ |
| | 30 | Python数学统计实验-研究统计学基础（三） | ✓ | ✓ | ✓ |
| | 31 | Python数学统计实验-统计计算 | ✓ | ✓ | ✓ |
| 数据建模 | 32 | R语言数学统计实验-广义线性模型（一） | ✓ | ✓ | ✓ |
| | 33 | R语言数学统计实验-广义线性模型（二） | ✓ | ✓ | ✓ |
| | 34 | R语言数学统计实验-建模实战 | ✓ | ✓ | ✓ |
| | 35 | Python数学统计实验-财政收入影响因素分析及预测模型 | ✓ | ✓ | ✓ |
| 数据可视化 | 36 | R语言数学统计实验-中级绘图（一） | ✓ | ✓ | ✓ |
| | 37 | R语言数学统计实验-中级绘图（二） | ✓ | ✓ | ✓ |
| | 38 | R语言数学统计实验-基本图形 | ✓ | ✓ | ✓ |
| 实战项目 | 39 | R语言数学统计实战实验-房价预测：高阶回归技术应用 | ✓ | ✓ | ✓ |
| | 40 | Python数学统计实验-二氧化碳含量预测 | ✓ | ✓ | ✓ |
| | 41 | Python数学统计实验-估计 | ✓ | ✓ | ✓ |
| | 42 | Python数学统计实验-分析美国对数据科学家专业技能需求（一） | ✓ | ✓ | ✓ |
| | 43 | Python数学统计实验-分析美国对数据科学家专业技能需求（二） | ✓ | ✓ | ✓ |
| | 44 | Python数学统计实验-利用 Python 进行员工流失分析（一） | ✓ | ✓ | ✓ |
| | 45 | Python数学统计实验-利用 Python 进行员工流失分析（二） | ✓ | ✓ | ✓ |
| | 46 | Python数学统计实验-利用 Python 进行员工流失分析（三） | ✓ | ✓ | ✓ |
| | 47 | Python数学统计实验-利用 Python 进行员工流失分析（四） | ✓ | ✓ | ✓ |
| | 48 | Python数学统计实验-新加坡空气污染原因分析（上） | ✓ | ✓ | ✓ |
| | 49 | Python数学统计实验-新加坡空气污染原因分析（下） | ✓ | ✓ | ✓ |
| | 50 | Python数学统计实验-统计分析集成开发软件受欢迎程度（上） | ✓ | ✓ | ✓ |
| | 51 | Python数学统计实验-统计分析集成开发软件受欢迎程度（下） | ✓ | ✓ | ✓ |
| | 52 | Python统计学实战实验-回归分析预测房价（上） | ✓ | ✓ | ✓ |
| | 53 | Python统计学实战实验-回归分析预测房价（下） | ✓ | ✓ | ✓ |

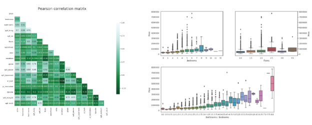
### 典型实验1：利用 Python 进行员工流失分析



### 典型实验2：中级绘图



### 典型实验3：房价预测：回归分析房价



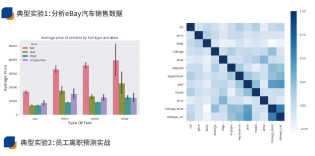
# BDRack大数据实验平台（电子商务版）

在BDRack大数据实验平台（本科）的基础上，大数据实验平台（电子商务版）进一步丰富了实验内容，增加了26个电子商务实验。具体实验列表如下：

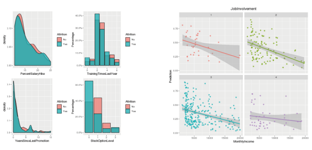
| 序号 | 实验名称 | 实验手册 | 实验代码 | 实验数据 |
|---|---|---|---|---|
| 1 | 员工离职预测实战(一) | ✓ | ✓ | ✓ |
| 2 | 员工离职预测实战(二) | ✓ | ✓ | ✓ |
| 3 | 基于基站定位数据的商圈分析 | ✓ | ✓ | ✓ |
| 4 | 数据分析实战1：用R做柱状图分析销售额减少 | ✓ | ✓ | ✓ |
| 5 | 数据分析实战2：交叉列表统计何种顾客会离开 | ✓ | ✓ | ✓ |
| 6 | 数据分析实战3：AB测试分析哪种广告的效果更好 | ✓ | ✓ | ✓ |
| 7 | 数据分析实战4：多元回归分析如何通过各种广告的组合获得更多的用户 | ✓ | ✓ | ✓ |
| 8 | 数据分析实战5：逻辑回归分析根据过去的行为能否预测当下 | ✓ | ✓ | ✓ |
| 9 | 数据分析实战6：利用k-means聚类选择目标用户群 | ✓ | ✓ | ✓ |
| 10 | 数据分析实战7：利用决策树分析哪些用户是长期用户 | ✓ | ✓ | ✓ |
| 11 | 航空公司客户价值分析 | ✓ | ✓ | ✓ |
| 12 | 分析eBay汽车销售数据 | ✓ | ✓ | ✓ |
| 13 | 分析客户流失(一) | ✓ | ✓ | ✓ |
| 14 | 分析客户流失(二) | ✓ | ✓ | ✓ |
| 15 | 利用Python做淘宝商品的数据挖掘分析 | ✓ | ✓ | ✓ |
| 16 | 利用Python进行市场购物篮分析 | ✓ | ✓ | ✓ |
| 17 | 利用层次聚类算法进行基于基站定位数据的商圈分析 | ✓ | ✓ | ✓ |
| 18 | 水产养殖企业企业水质分析 | ✓ | ✓ | ✓ |
| 19 | 爬虫爬取拉勾网职业信息分析 | ✓ | ✓ | ✓ |

| 序号 | 实验名称 | 实验手册 | 实验代码 | 实验数据 |
|---|---|---|---|---|
| 20 | 用户投诉分析 | ✓ | ✓ | ✓ |
| 21 | 用户欺诈预测 | ✓ | ✓ | ✓ |
| 22 | 电商产品评论数据情感分析 | ✓ | ✓ | ✓ |
| 23 | 电商打折套路解析（上） | ✓ | ✓ | ✓ |
| 24 | 电商打折套路解析（下） | ✓ | ✓ | ✓ |
| 25 | 航空公司客户价值分析 | ✓ | ✓ | ✓ |
| 26 | 销售收入预测 | ✓ | ✓ | ✓ |

### 典型实验1：分析eBay汽车销售数据

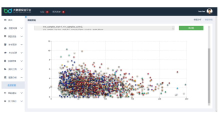


### 典型实验2：员工离职预测实战



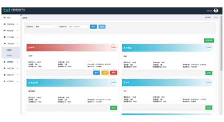
# 平台服务

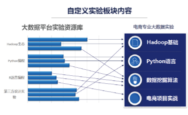
BDRack大数据实验平台集成考试评测系统、第三方实验开发工具集、数据挖掘及可视化工具等应用服务进一步提升用户使用体验。



可视化工具

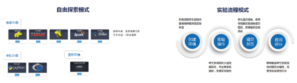
考试测评系统



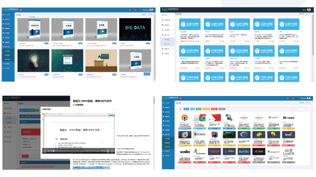

自定义实验板块内容
大数据平台实验资源库
Hadoop生态
Python编程
R语言编程
第三方设计实验
电商专业大数据实验
Hadoop基础
Python语言
数据挖掘算法
电商项目实战

第三方实验开发工具集

对于具体实验而言，BDRack大数据实验平台提供自由探索模式与实验流程模式，用户可根据具体应用需求选择适用的实验模式。


自由探索模式
集群环境
Spark
STORM
单机环境
python
图形环境
实验流程模式
创建环境
实验操作
提交报告
教师评分

BDRack大数据实验平台提供丰富的可供下载的学习资料，包括教学课件、实验视频、思维导图等


BIG DATA
实验三 HDFS实验：读写HDFS文件

# 师资培训

2016年末至今，为了助推高校老师大数据、人工智能人才的培养，云创大数据举办了数十场免费的大数据、人工智能实战培训班，且深受学员好评，在教育领域引发强烈反响。在这些培训班举办过程中，参训学员直接使用云创开发的人工智能实验平台，跑通了多项人工智能实验。

2016年12月-2017年1月，多次举办的高校（高职）大数据师资免费培训班

2017年4月，全国千所高校大数据师资免费讲习班在南京举行

2017
全国千所高校大数据师资讲习班

2018年5月，2018信息技术新工科产学研联盟大数据师资免费培训班举行

2018第三期——
信息技术新工科产学研联盟
大数据技术师资培训班

2018年9-10月，第二届全国高校大数据人工智能师资实战免费培训班连续举办3期

2019年1月，2019全国高校大数据人工智能师资实战免费培训班连续举办2期

# 培养方案

编制与大数据专业配套的本科人才培养方案与专科人才培养方案，对人才培养的全过程进行了描述，具体涵盖招生对象及学制、培养目标、职业岗位群和人才培养规格、毕业条件、课程体系设置及进程安排、课程描述、本专业教学指导委员会、专业教师配备与要求、实践教学配置与要求、培养方案编制说明（教学附录）等详细内容。

大数据科学与技术专业人才培养方案（本科）
高校大数据专业（专科）人才培养方案
数据标注专业（中专）人才培养方案
大数据运维专业（中专）人才培养方案
数据清洗专业（中专）人才培养方案

# 认证证书

云创大数据本着“考、教、学一体化”与“1+X”相对接的原则，为合作院校提供证书认证工作。对于达到考证合格标准的学生，相应授予工信部、SSE（硅石教育集团）以及云创大数据颁发的三级认证证书。

工业和信息化领域
急需紧缺人才培养工程
证书
工业和信息化部人才交流中心
International Certificate Programs
Silicon Stone Education
Certificate of Completion
Statistical Knowledge Today
Silicon Stone Education, Inc.

# 网站资源

业界专业领域排名居前的网站——大数据世界（thebigdata.cn）、云计算世界（chinacloud.cn）等专业网站将提供配套PPT与其他资料，让大数据学习变得更加轻松。

云计算世界 chinacloud.cn
大数据世界 thebigdata.cn
物联网世界 netofthings.cn
智能硬件世界 www.chinaznyj.com
PM2.5世界 www.pm25.org.cn
存储世界 www.chinastor.org
深度学习世界 www.dlworld.cn
机器人世界 chinarobots.cn
智慧城市世界 SmartCityChina.cn

# 专家团队

云创大数据研发团队由清华大学博士、中国信息协会大数据分会副会长刘鹏教授带领具有丰富实战经验的大数据研发人员组成，可为大数据实验室共建以及大数据教学科研提供专业的技术指导与支持。

### 刘鹏教授

清华大学博士毕业，现任中国信息协会大数据分会副会长、中国大数据应用联盟人工智能专家委员会主任。刘鹏教授在大数据和云计算领域具有18年的研究积累，在大数据、人工智能、分布式存储等技术领域卓有建树。

刘鹏教授曾于2002年获得全球数据处理比赛 PennySort世界冠军，于2003年获全国大学“挑战杯”比赛总冠军。曾获全军“十大学习成才标兵”、南京“十大杰出青年”、“中国大数据创新百人”、江苏省中青年领军人才、清华大学“学术新秀”等称号。主持科研项目40多项，发表论文80余篇。主编出版了“高级大数据人才培养丛书”和“大数据应用人才培养系列教材”，被各高校普遍采用。创办了人工智能世界（www.chinaaiworld.cn）、深度学习世界（www.dlworld.cn）、云计算世界(www.chinacloud.cn)、大数据世界(www.thebigdata.cn)等网站。

### 沈大为总监

德克萨斯理工大学（Texas Tech University）计算机科学专业硕士，在美国工作20多年，拥有包括AT&T和Verizon等行业知名企业的从业经历。现任云创大数据万物云物联网大数据服务平台（www.wanwuyun.com）、环境云环境大数据服务平台（www.envicloud.cn）总架构师和项目技术负责人，担任云创大数据平台总监。主持开发的项目包括cCloud虚拟主机云服务平台、大数据实验平台、大数据实验一体机、人工智能实验一体机等。

### 实战经验丰富的研发团队

大数据研发团队成员毕业于国内外知名高校，均为负责大数据实验平台、万物云平台、环境云平台等平台开发的技术骨干，覆盖车联网云平台项目、大数据挖掘平台DataSense项目、地震数据云平台项目等，有着丰富的算法开发、数据建模、数据分析等大数据经验。

# 成功案例

BDRack大数据实验平台已经在贵州大学、郑州大学、西南大学、成都理工大学、重庆工商大学、西北工业大学、重庆师范大学、重庆文理学院、天津农学院、金陵科技学院、郑州升达经贸管理学院、信阳师范学院等数十所涵盖中专、高职、本科层次的院校落地实施。以下为部分应用单位：

云创大数据与合作院校签约（部分）

西工大-云创大数据联合实验室签约暨揭牌仪式
南京云创大数据科技股份有限公司 重庆工商
战略合作签约仪式
签约仪式
第二届亚洲人工智能技术大会开幕式暨
重庆两江人工智能学院成立仪式
重庆两江人工智能学院项目建设合作签字仪式
人工智能校企合作
共建实验室签约仪式
2018年7月4日
淮南师范学院与南京云创大数据科技有限公司合作签约暨客座教授聘任仪式
淮南师范学院
HUAINAN NORMAL UNIVERSITY
“大数据和人工智能”研讨会
大数据应用实践
与人才培养
刘鹏 教授