

NO.1

高校用户
口碑第一

大数据、人工智能 实验、实训室建设经验分享

技术赋能 产教融合 创新教育

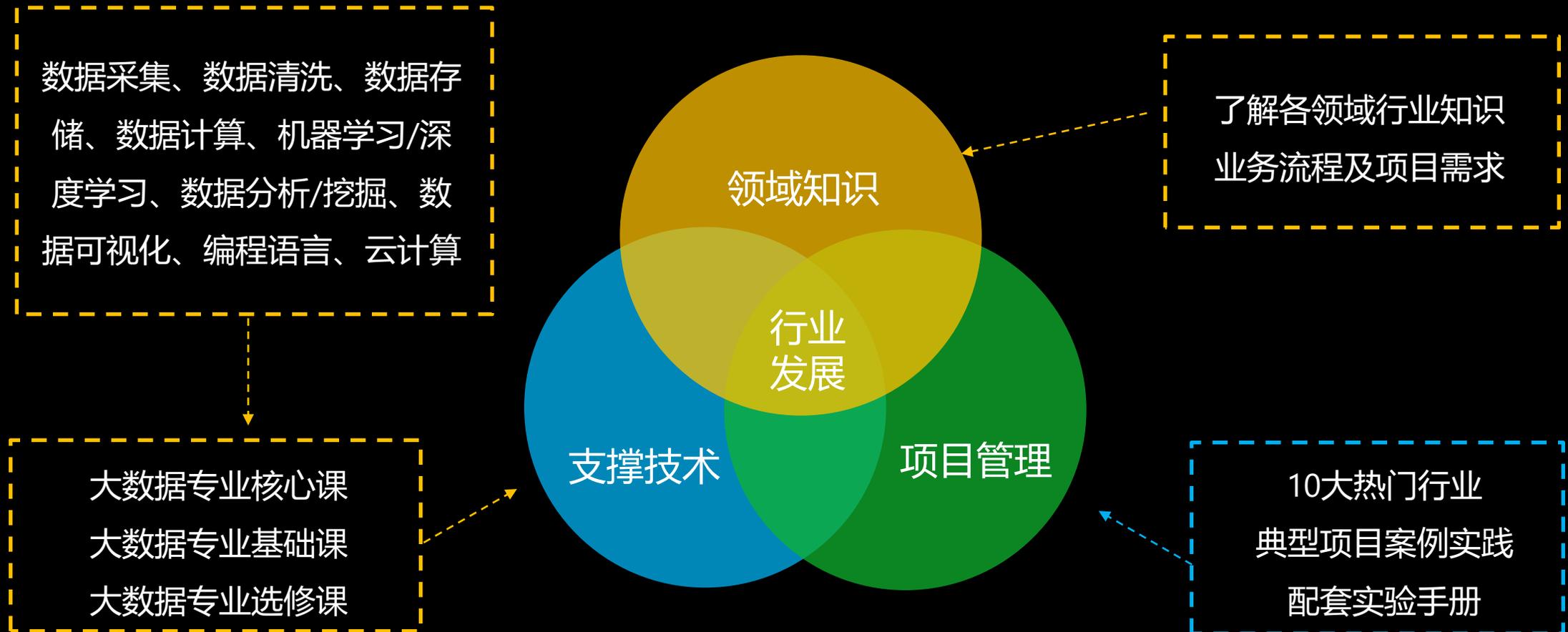
目录

- 一、大数据人才培养理念
- 二、大数据专业建设方案
- 三、大数据专业课程体系
- 四、人工智能专业建设方案
- 五、9类教学资源
- 六、普开简介

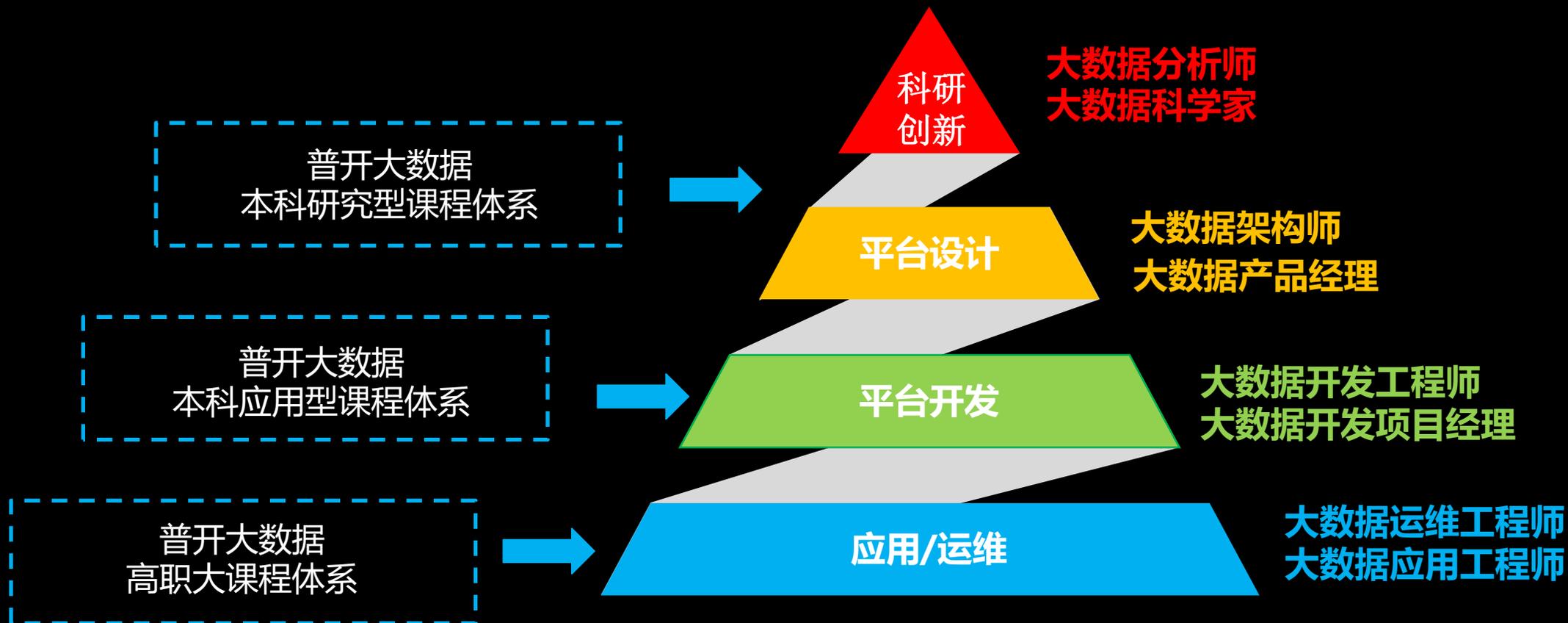
人才培养理念

构建大数据教育生态链

普开数据产教融合新技术人才培养理念



四段位金字塔人才梯队（岗位模型）



**大数据
专业建设方案**

整体架构



应用型本科课程

包含**17+**课程
涵盖大数据开发、
Hive、NoSQL
Spark、Python、
深度学习等

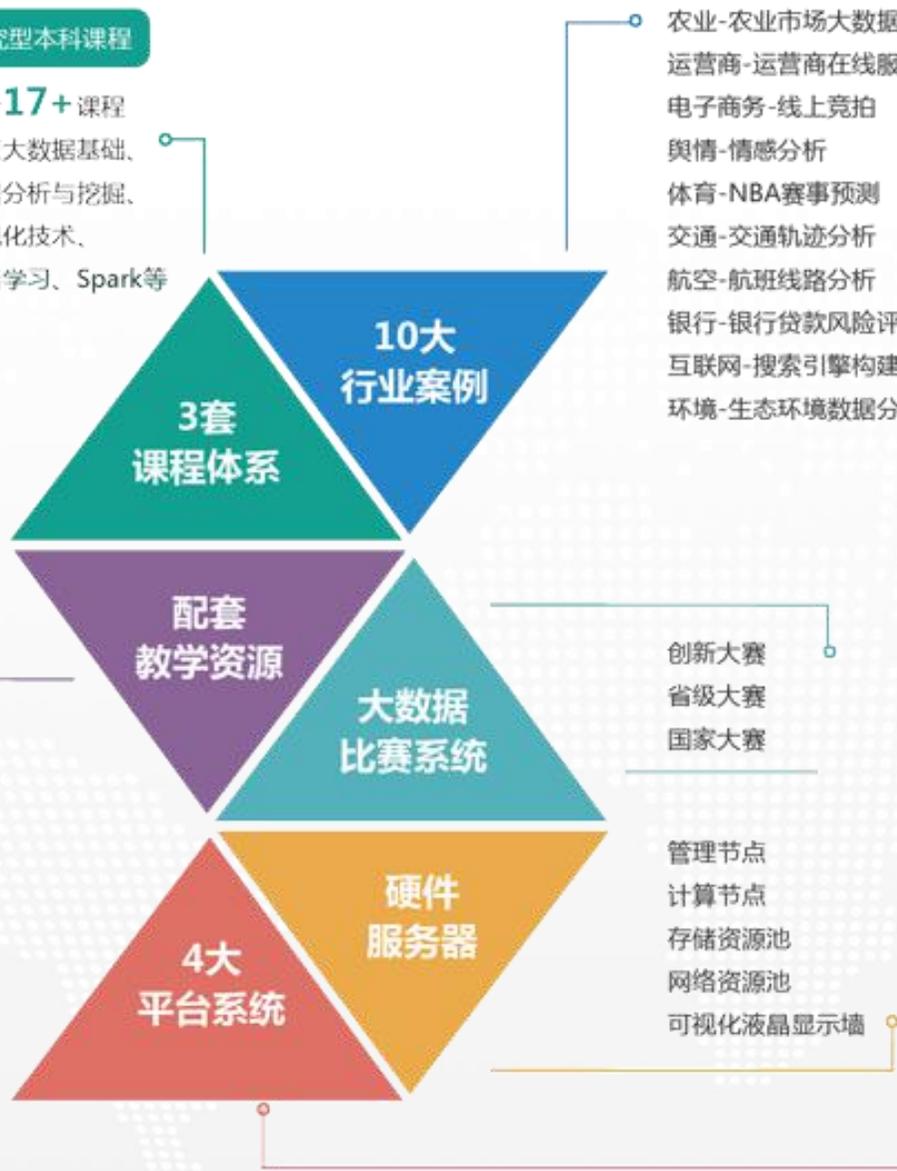
研究型本科课程

包含**17+**课程
涵盖大数据基础、
数据分析与挖掘、
可视化技术、
机器学习、Spark等

高职专业课程

包含**13+**课程
涵盖Hadoop基础、
Hive离线分析、
ZooKeeper、HBase、
Scala语言等。

3200+分钟理论视频
2000+分钟实验视频
200+实验手册
1TB+实验数据
200+教学PPT
40+虚拟机镜像



农业-农业市场大数据分析
运营商-运营商在线服务
电子商务-线上竞拍
舆情-情感分析
体育-NBA赛事预测
交通-交通轨迹分析
航空-航班线路分析
银行-银行贷款风险评估
互联网-搜索引擎构建
环境-生态环境数据分析

创新大赛
省级大赛
国家大赛

管理节点
计算节点
存储资源池
网络资源池
可视化液晶显示墙

云计算管理系统

是大数据教学平台的资源管理系统；
采用 Docker 及 OpenStack 技术，
实现了硬件资源集中调度和管理。

大数据监控管理系统

是大数据教学平台的运维监控系统；
实现大数据实验集群的性能监控。

数图可视化系统

为大数据学习者和开发者设计的大数据
图表可视化产品；解决研发人员对前端
技术不熟悉无法将数据进行可视化的问题。

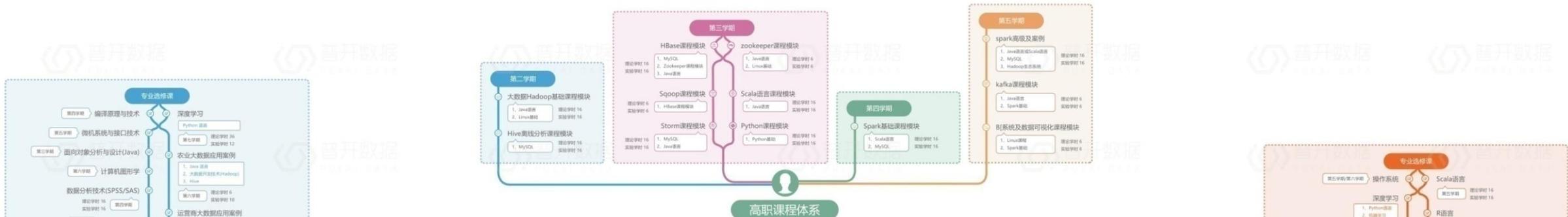
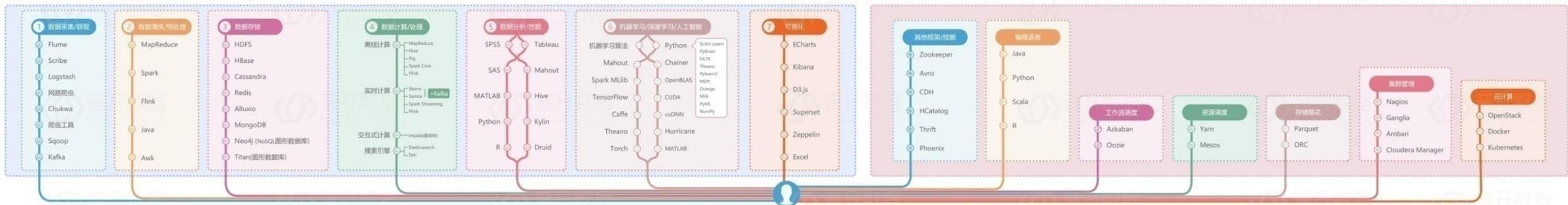
大数据实验管理平台系统

是大数据教学平台的核心教学管理系统；
由课程管理系统、资源管理系统、实验报
告、考评系统、考试系统组成，为实验教
学提供技术支撑。

课程体系

全面覆盖本科、高职

可操作型产教融合大数据课程体系架构图



应用型本科课程体系

研究型本科课程体系

大数据专业课程体系

| 序号 | 服务内容 | 数量 | 形式 |
|----|------|------|--------------|
| 1 | 高职课程 | 13门 | 课件/指导视频/指导手册 |
| 2 | 本科课程 | 17门 | 课件/指导视频/指导手册 |
| 3 | 综合案例 | 13个 | 课件/指导视频/案例手册 |
| 4 | 系统镜像 | 200个 | 单机镜像/集群镜像 |
| 5 | 理论视频 | 7门 | 高清视频 |
| | | | |

大数据专业课程资源清单

| 序号 | 课程名称 | 序号 | 课程名称 |
|----|-------------------|----|---------------|
| 01 | 大数据开发技术(Hadoop) | 12 | 机器学习-基础 |
| 02 | 分布式数据仓库(Hive) | 13 | 机器学习-高级 |
| 03 | 数据库技术(NoSQL) | 14 | 深度学习-主流框架 |
| 04 | 消息发布订阅系统(Kafka) | 15 | 深度学习-典型实例 |
| 05 | Scala编程 | 16 | R语言 |
| 06 | 大数据处理技术(Spark)-基础 | 17 | R语言统计与建模 |
| 07 | 大数据处理技术(Spark)-中级 | 18 | Hadoop-高清视频理论 |
| 08 | 大数据处理技术(Spark)-高级 | 19 | Hive-高清视频理论 |
| 09 | Python编程 | 20 | HBase-高清视频理论 |
| 10 | 数据分析与挖掘 | 21 | Scala-高清视频理论 |
| 11 | 大数据可视化技术 | 22 | Spark-高清视频理论 |

行业案例--大数据行业应用-运营商

运营商在线服务

今日工单量
2.0W

月度工单量
62W

年度工单量
509W

区域申告分布



区域服务效率



区域服务效能



| 云南 | | | 甘肃 | | | 广东 | | | 广西 | | | 新疆 | | |
|--|-------|--------|--|--------|--------|--|--------|--------|--|--------|--------|--|--------|--------|
| 申告量 | 成功率 | 及时率 | 申告量 | 成功率 | 及时率 | 申告量 | 成功率 | 及时率 | 申告量 | 成功率 | 及时率 | 申告量 | 成功率 | 及时率 |
| 2028 | 46.1% | 100.0% | 1767 | 34.86% | 100.0% | 2615 | 32.77% | 100.0% | 3516 | 26.54% | 97.38% | 2615 | 32.77% | 100.0% |
| Legend: 申告量 (Cyan), 成功率 (Red), 及时率 (Green) | | | Legend: 申告量 (Cyan), 成功率 (Red), 及时率 (Green) | | | Legend: 申告量 (Cyan), 成功率 (Red), 及时率 (Green) | | | Legend: 申告量 (Cyan), 成功率 (Red), 及时率 (Green) | | | Legend: 申告量 (Cyan), 成功率 (Red), 及时率 (Green) | | |



1.项目核心知识点： HDFS、Spark、Scala、MySQL、echarts。

2.项目描述：

随着通信行业的普及，以及人们对网络的需求越来越大，因此运营商的一些在线服务需求也越来越大。对于客户体验来说，电子渠道提供了一个足不出户办理业务的便捷方式，对于运营商来说，电子渠道低成本分流了实体渠道的业务压力，将线下渠道的人力资源从低价值的业务办理中释放出来。

3.功能模块：

- | | | | |
|-------------|----------------|-------------|---------------|
| 3.1.申告量统计分析 | 3.2.处理量统计分析 | 3.3回单量统计分析 | 3.4回单率统计分析 |
| 3.5重障量统计分析 | 3.6重障率统计分析 | 3.7及时量统计分析 | 3.8 及时率统计分析 |
| 3.9成功率统计分析 | 4.0用户等级多维度统计分析 | 4.1区域服务团队监控 | 4.2热点故障区域分布监控 |

4.项目实训方案：

a)架构简介： 本系统架构采用模块化设计，分为数据预处理、数据存储模块、数据处理模块、数据ETL模块、结果展现模块。。

b)模块名称：

数据预处理模块： 将运营商给的历史数据进行格式转换。

数据存储模块： 将清洗后的数据存储到HDFS上

数据处理模块： 通过spark-SQL将存储在HDFS上的数据进行处理，通过spark和MapReduce对具体的需求进行实现。

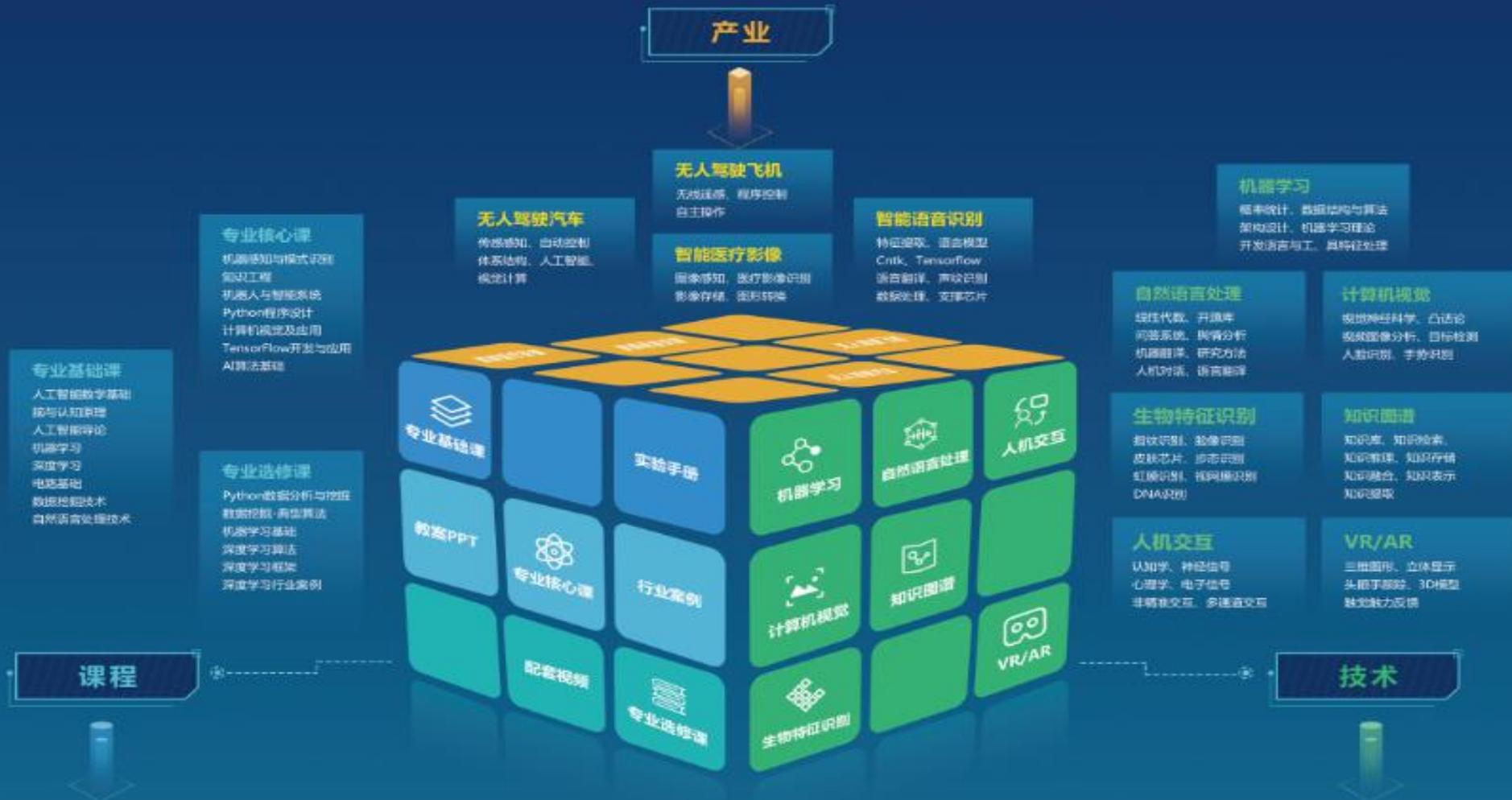
数据ETL模块： 通过shell脚本实现数据的get，入库MYSQL等操作

结果展示模块： 通过利用柠檬数图/superset对结果的展示

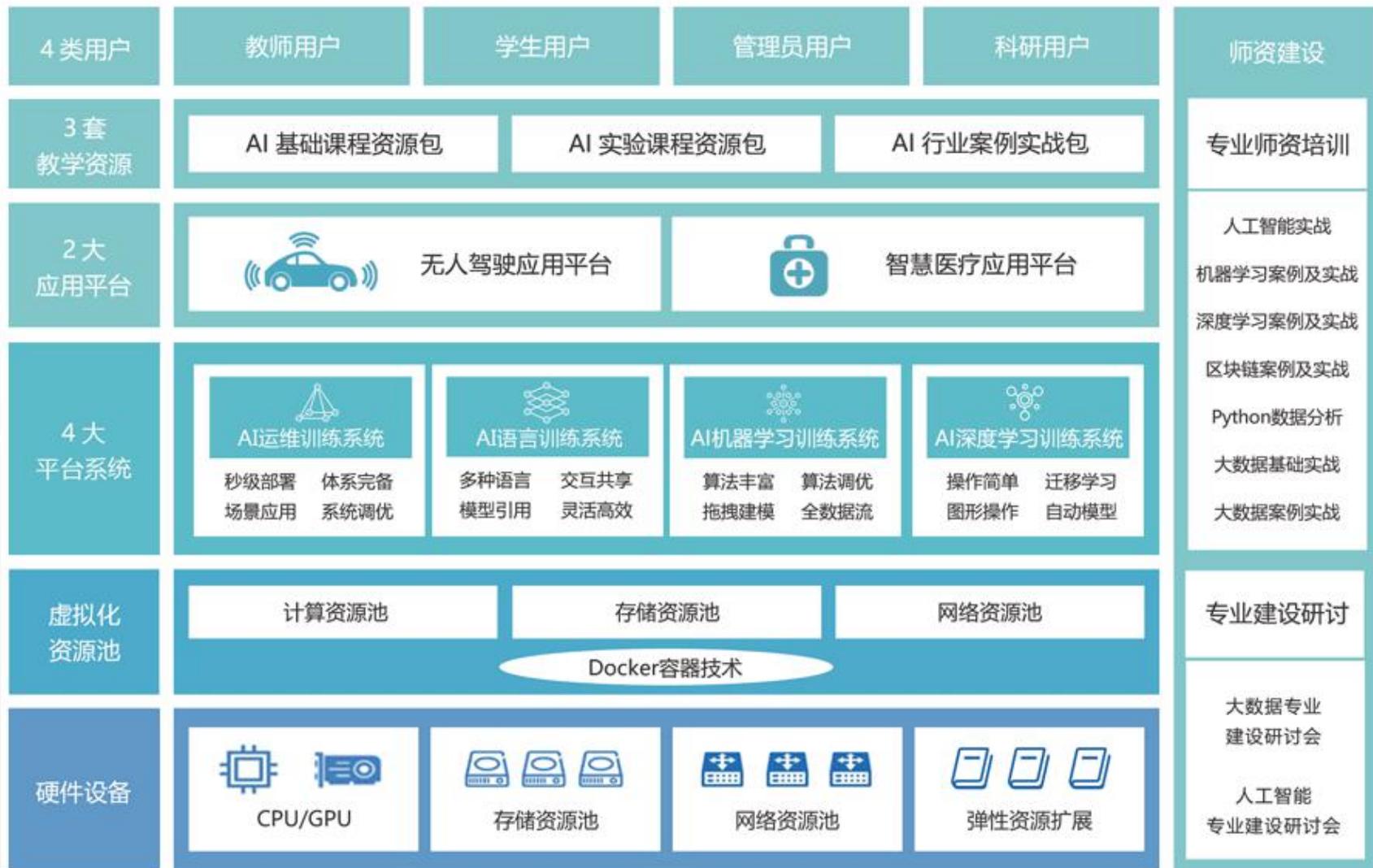
人工智能课程 体系

课程体系——产教融合应用型培养体系

应用型AI人才培养魔方



人工智能专业建设整体架构



课程体系

基础课程17+
行业案例课程16+

AI基础课程

包含**17+**课程
涵盖Python编程、算法、机器学习、深度学习、自然语言处理、计算机视觉等

AI行业案例课程

16+行业案例
涵盖深度学习、机器学习、数据挖掘、人工智能、计算机视觉等。

2套
课程体系

配套
教学资源

200+ 实验手册
1TB+ 实验数据
180+ 教学PPT
46+ 虚拟机镜像
220+ 配套习题

全面的配套资源
实验手册200+
实验数据1TB+
教学PPT180+
虚拟机镜像46+

2大
行业应用

无人驾驶、智慧医疗

虚拟化层

包含docker容器
Rancher编排系统
Kubernetes (K8S)
编排调度系统

4大
平台系统

硬件设备
(算力)

管理节点
计算节点
存储资源池
计算资源池
网络资源池

AI机器学习训练系统

内置有监督学习、无监督学习、强化学习等40多种经典算法，每个算法都是并行的，可以在分布式系统hadoop、spark上执行。

AI深度学习训练系统

使用TensorFlow、PyTorch等框架做图像识别、物体检测、声音识别等；包含AlexNet、FasterRCNN-VGG16等十多种神经网络。

AI语言训练系统

通过Jupyter notebook快速构建包含Python、Java、C等40种语言学习训练。

AI运维训练系统

通过Docker快速容器生成实验环境，满足大数据、人工智能课程训练。

CPU: Intel E5 2600

GPU: Nvidia Tesla P100/P40/V100 显卡

4大训练平台

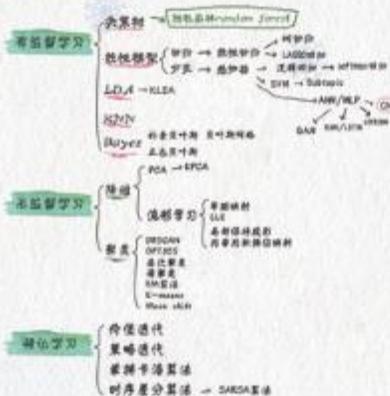
AI机器学习训练系统
AI深度学习训练系统
AI语言训练系统
AI运维训练系统

人工智能可操作型产、教融合课程体系架构图

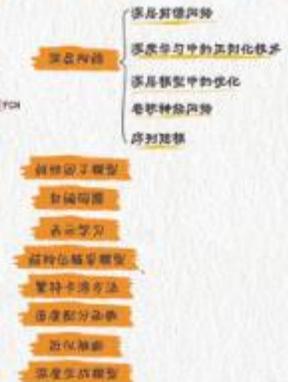
技术生态环境

【语言】
【算法模型】

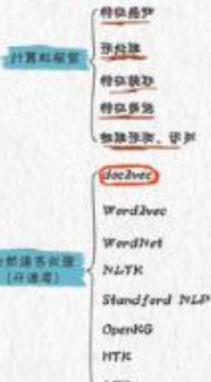
机器学习 (SparkML)



深度学习 (Tensorflow/PyTorch)



进阶技术



教学案例

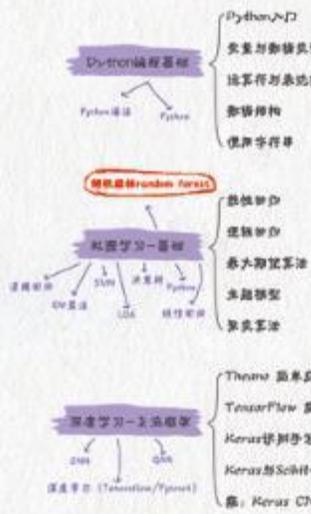


教学平台

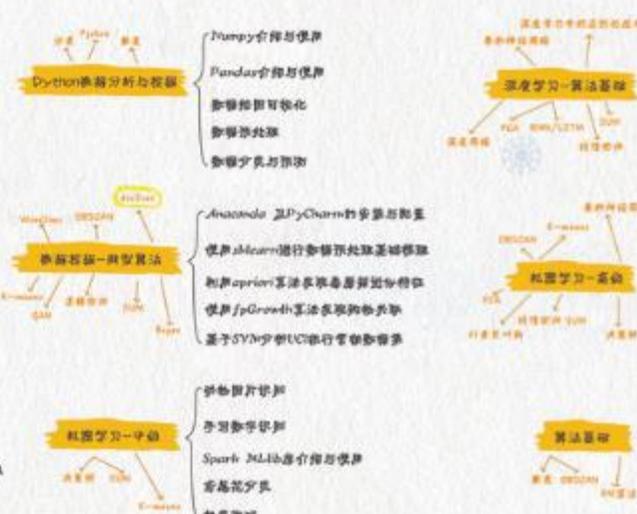


课程

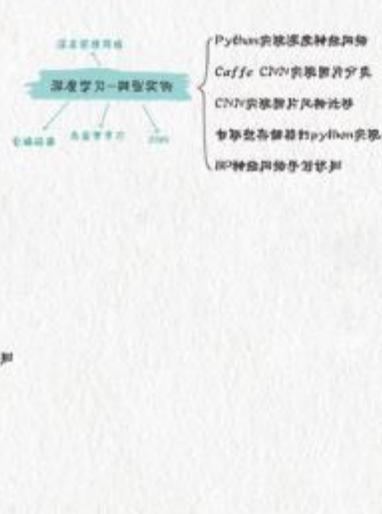
【专业基础课】



【专业核心课】



【专业选修课】



课程地图



挑战1：教学内容——工程与课程结合



典型实验课程

| 序号 | 课程名称 | 实验数量 | 实验手册 | 课后习题 | 实验代码 |
|----|----------------|------|------|------|------|
| 01 | Python 编程基础 | 12 | √ | √ | √ |
| 02 | Python 数据分析与挖掘 | 7 | √ | √ | √ |
| 03 | 数据挖掘 - 典型算法 | 9 | √ | √ | √ |
| 04 | 机器学习 - 基础 | 11 | √ | √ | √ |
| 05 | 机器学习 - 中级 | 11 | √ | √ | √ |
| 06 | 机器学习 - 高级 | 16 | √ | √ | √ |
| 07 | 深度学习 - 算法基础 | 16 | √ | √ | √ |
| 08 | 深度学习 - 主流框架 | 8 | √ | √ | √ |
| 09 | 深度学习 - 典型实例 | 8 | √ | √ | √ |
| 10 | AI 算法基础 | 8 | √ | √ | √ |

挑战2：学生实验——4大训练平台满足不同实验



机器学习

Machine Learning

一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。



虚拟机实验室

Virtual Lab

计算机模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能的学科。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征，以发现数据的分布式特征。



交互式编程

Jupyter Notebook

本质是一个 Web 应用程序，便于创建和共享文学化程序文档，支持实时代码，数学方程，可视化和 markdown。用途包括：数据清理和转换，数值模拟，统计建模，机器学习等等。



深度学习

Deep Learning

计算机模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能的学科。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征，以发现数据的分布式特征。

挑战3：项目教学——案例实践积累实操经验



点击可观看实验视频

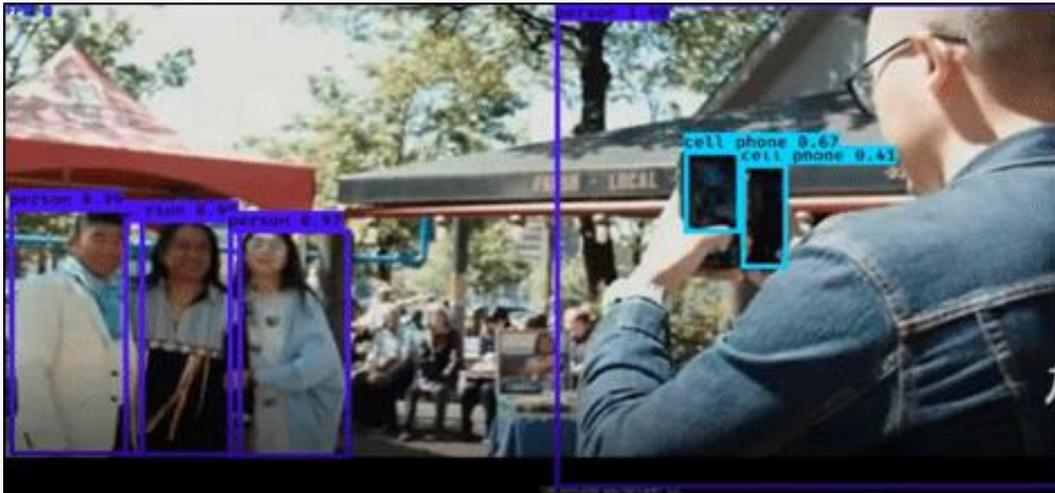
人工智能基础课程 (17门+并持续更新)

| 序号 | 课程名称 | 序号 | 课程名称 | 实验镜像 | 实验手册 | 课后习题 | 实验代码 |
|----|---------------|----|---------|------|------|------|------|
| 01 | Python编程基础 | 10 | AI-算法基础 | √ | √ | √ | √ |
| 02 | Python数据分析与挖掘 | 11 | 移动机器人视觉 | √ | √ | √ | √ |
| 03 | 数据挖掘-典型算法 | 12 | 视频监控 | √ | √ | √ | √ |
| 04 | 机器学习-基础 | 13 | 无人机视觉 | √ | √ | √ | √ |
| 05 | 机器学习-中级 | 14 | 视觉伺服持 | √ | √ | √ | √ |
| 06 | 机器学习-高级 | 15 | 自然语言处理 | √ | √ | √ | √ |
| 07 | 深度学习-算法基础 | 16 | 计算机视觉 | √ | √ | √ | √ |
| 08 | 深度学习-主流框架 | 17 | 知识图谱 | √ | √ | √ | √ |
| 09 | 深度学习-典型实例 | 18 | 持续更新 | √ | √ | √ | √ |

人工智能行业项目实践—目标检测

基于YOLOv3模型的Keras实现目标检测

本实验使用Keras实现YOLOv3模型进行目标检测，YOLOv3相对YOLOv2在小物体检测上进行了提高，对于紧凑密集或者高度重叠目标的检测也有显著效果。



人工智能行业项目实践—人脸表情精准识别

基于dlib库实现人脸特征提取识别表情

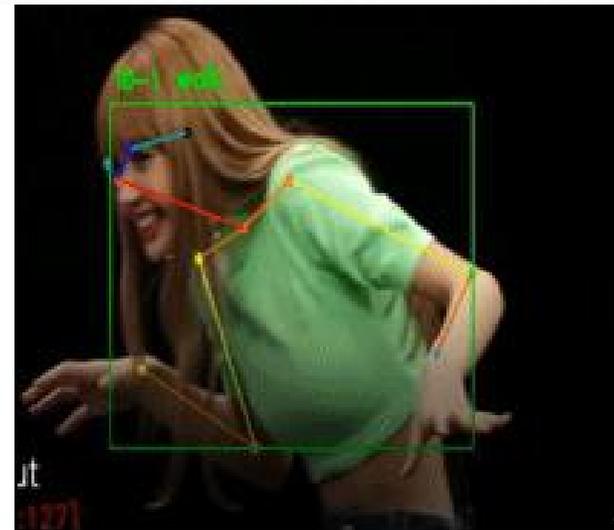
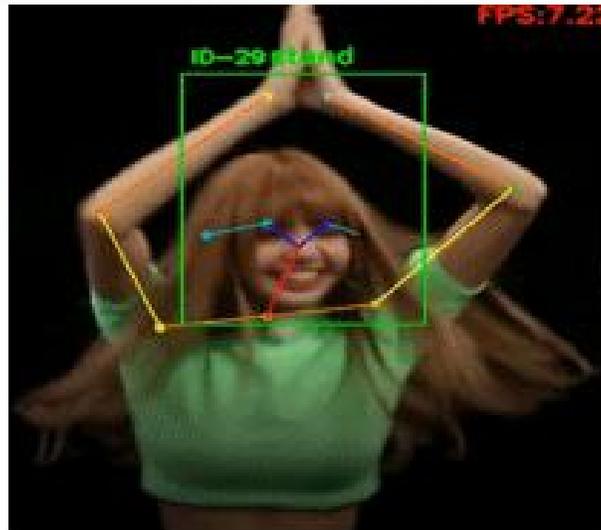
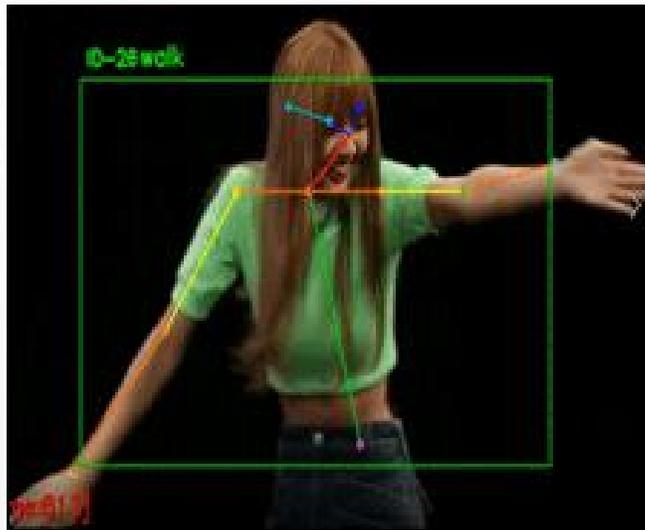
本实验使用dlib库，一个机器学习的开源库，包含了机器学习的很多算法。能够很好的识别人脸，并能标注出人脸的关键点，通过人脸关键点眼睛、眉毛、嘴的变化，然后根据一定的规则区分相应特征变化对应的表情，从而识别出人脸的表情。



人工智能行业项目实践—人体姿态识别

基于OpenPose模型的使用Keras实现

本实验使用Keras实现OpenPose模型，OpenPose是由卡内基梅隆大学认知计算研究室提出的一种对多人身体、面部和手部形态进行实时估计的框架。本实验能够实时识别出人体关键的关节点，并识别出人体的形态是行走、站立还是其他。





**9大类
教学资源**

9大类资源库

资源中心



课件库：135

课件库

查看



视频库：344

视频库

查看



手册库：365

手册库

查看



试题库：38

试题库

查看



试卷库：19

试卷库

查看



虚拟机库：4

虚拟机库

查看



镜像库：167

镜像库

查看



图片库：36

图片库

查看



录像库：0

录像库

查看

最全面的大数据课程资源

7大系列教材+7大配套实验手册+200PPT+100实验案例+300视频+50数据资源包

985.211联合名企共同编写，清华大学出版



PPT教学课件 (图文并茂+流程动画)

访问HDFS

Let's start

访问HDFS

- 1 HDFS分布式文件系统
- 2 HDFS核心设计
- 3 HDFS体系结构

HDFS简介

● HDFS: Hadoop Distributed File System

基于流数据模式访问和处理超大文件的需求而开发的

01 一个分布式文件系统

02

03 适合应用在大规模数据集上

副本存放策略

HDFS核心设计

数据块

- 数据块是HDFS上最基本的存储单位
- HDFS块默认大小为128M
 - 小于一个块大小的文件不会占据整个块的空间
- 对块进行操作会带来好处
 - 一个文件的大小可以大于网络中任意一个节点的容量
 - 使用块抽象而不是文件可以简化存储子系统
 - 块非常适合用于数据备份进而提供数据容错能力和可用性

HDFS特性

- 1 高吞吐量及可伸缩性
- 2 跨平台
- 3 Shell命令接口
- 4 机架感知功能
- 5 负载均衡
- 6 Web界面

HDFS目标

移动计算优于移动数据价值

一个应用需求的计算，离它操作的数据越近就越高效，这在数据达到海量级别的时候更是如此。

可移植性

HDFS在设计时就考虑到平台的可移植性，这其特性方便了HDFS作为大规模数据应用平台的推广。

通信协议

所有的通信协议都是在TCP/IP协议之上的。

HDFS简介

优点

- 处理超大文件
 - 能用来存储管理PB级的数据
- 处理非结构化数据
- 流式的访问数据
 - 一次写入，多次读写
- 运行于廉价的商用机器集群上
 - 可运行在低廉的商用硬件集群上
 - 故障时能继续运行且不让用户察觉到明显的中断

HDFS核心设计

设置集群Block的备份数

- 方法一：配置文件hdfs-site.xml


```
<property>
<name>dfs.replication</name>
<value>3</value>
</property>
```
- 方法二：通过命令修改备份数
 - bin/hadoop fs -setrep -R 1 /

注意：方法二只能在集群初始化完成且数据块已经写入，不要重复执行命令，否则会导致数据块丢失。

HDFS体系结构

Master/Slave架构

HDFS文件权限

与Linux文件权限类似

如果Linux系统用户zhangsan使用hadoop命令创建一个文件，那么该文件在HDFS中就是zhangsan

r: read; w: write; x: execute. 权限x对于文件加锁，对于文件夹表示是否允许访问内容

HDFS的权限目的：防止坏人做坏事，而不是防止坏人做坏事。HDFS的权限，防止坏人做坏事。我认为你是对的。

HDFS目标

检测和快速恢复硬件故障

故障的检测和快速自动恢复是HDFS的一个核心目标

流式数据访问

- 设计成适合进行批量处理
- 重复数据冗余量，而不是数据访问的延迟速度

大规模数据集

- 支持大文件存储
- 一个单一的HDFS实例能支撑数以千万计的文件

简化一致性模型

对文件实行一次性写入，多次读写的访问模式

口碑好

才是真的好



北京邮电大学

项目时间：2016年11月

项目目的：北京邮电大学计算机学院大数据教学平台

实验室规模：同时满足200人上课





北京航空航天大学
BEIHANG UNIVERSITY

项目时间：2017年10月

项目目的：物联网与云计算数据中心公共应用实验室

实验室规模：同时满足100人上课





黔南民族师范学院
Qiannan Normal University for Nationalities

项目时间：2018年4月

项目目的：黔南民族师范学院大数据统计学院教学平台

实验室规模：同时满足60人上课



2018教育部产学合作协同育人唯一获奖大数据项目

普开数据支持的“黔南民族师范学院大数据统计分析实训平台建设”项目



普开大数据平台——超60+所高校的共同选择

| 地区 | 高校名称 | 地区 | 高校名称 |
|----|----------|----|--------------|
| 山东 | 泰山学院 | 广东 | 广东技术师范学院 |
| | 山东临沂大学 | | 广州航海学院 |
| | 山东医科大学 | | 广州华立科技职业学院 |
| 北京 | 北京邮电大学 | 内蒙 | 内蒙古农业大学 |
| | 北京航空航天大学 | | 兴安盟职业技术学院 |
| | 北京石油化工学院 | | 内蒙古医科大学 |
| | 北方工业大学 | | 内蒙古民族大学 |
| 河北 | 河北工业大学 | 广西 | 广西民族大学 |
| | 河北经贸大学 | | 桂林理工大学 |
| | 河北新龙科技集团 | | 广西城市职业学院 |
| 湖北 | 长江大学 | | 广西河池学院 |
| | 中南民族大学 | | 桂林理工大学（雁山校区） |
| | 武汉轻工大学 | | 梧州学院 |

黔南民族师范学院——获教育部产学合作优秀项目

| 地区 | 高校名称 | 地区 | 高校名称 |
|----|----------------|----|--------------|
| 山西 | 山西农业大学 | 四川 | 电子科技大学 |
| | 中北大学 | | 四川大学锦城学院 |
| | 晋中学院 | | 重庆理工大学(两江校区) |
| | 山西财政税务专科学校 | | 绵阳师范学院 |
| 东北 | 哈尔滨广厦学院 | | 重庆邮电大学 |
| | 辽宁工程技术大学 | | 重庆工商大学 |
| | 东北石油大学 | | 上海 |
| 江苏 | 南通大学 | 安徽 | 皖西学院 |
| | 苏州工业园区服务外包职业学院 | | 安徽电子信息职业技术学院 |
| | 苏州科技大学 | | 铜陵学院 |
| 贵州 | 贵州电子信息职业技术学院 | | 安徽电子 |
| | 黔南民族师范学院 | 宁夏 | 宁夏大学 |

**普开
资质荣誉**

专注大数据/人工智能教育

提供**系统化、智能化**的专业建设解决方案

3000人



师资培训

30所/年



高校实验室用户

6场/年



大数据教学研讨会

1次/月



软件/课程升级

关于普开

普开数据是国内第一家专注大数据、人工智能领域的教育、应用开发及资源服务国家高新技术企业；是教育部高教司产学合作协同育人项目合作单位；是工业和信息化技术人才培养服务平台

普开数据竭诚为高校提供以大数据、人工智能及区块链为核心的新一代IT领域教育解决方案，以服务高校、企业和个人为目标以教育资源、实验实训平台为核心，为高校新工科教育提供课程体系、教材、教案、行业应用案例、实验实训平台等产品和服务，协助高校提升相关专业竞争力，培养研究型、应用型等稀缺人才。

- 荣获2018年教育部产学合作协同育人项目唯一大数据优秀项目案例奖
- 荣获2018年度全国高校人工智能与大数据产品技术创新奖

公司愿景

构建大数据教育生态链，做新工科教育的践行者

公司使命

让实践教学更Easy!

价值观

正直诚信、尊重他人、团结拼搏、坚持创新、共赢共享

普开数据

教育部产学合作协同育人指定企业

人工智能与大数据产品技术创新奖

- 教材编写及课程建设 **30+**

- 师资培训 **200+**

- 实践条件建设 **100+**

- 累计投入经费 **1500w+**

源于产业，集高校、研究院、大型企业多渠道研发力量



师资培训用户数量位居行业第一

1500 + **4000** +
高校数量 老师数量

大数据实验室高校用户口碑第一

60

高校已落地

课程体系完备
行业案例丰富
+ 实验手册符合3M认知模型
+ Docker+OpenStack混合架构、节省资源
管理平台灵活便捷
.....

专注新工科建设，让实践教学更Easy!

大数据、AI、区块链等新技术师资班、专业建设研讨会



教育部产学合作协同育人项目



教育部产学合作协同育人项目——70+



安徽师范大学



北京建筑大学



北京石油化工学
院



北京邮电大学



常熟理工学院



巢湖学院



东北电力大学



东北石油大学



广东技术师范学
院



广州航海学院



贵州财经大学



贵州师范大学



哈尔滨金融学院



哈尔滨理工大学



河北工业大学



河北经贸大学



湖北文理学院



湖南人文科技学
院



华东交通大学



辽宁工程技术大
学



辽宁石油化工大
学



临沂大学



岭南师范学院



绵阳师范学院



南京审计大学



南京信息工程大
学



内蒙古农业大学



黔南民族师范学
院



曲阜师范大学



三峡大学



山西农业大学信
息学院



上海工程科技大
学



苏州科技学院



泰山学院



铜陵学院



皖西学院



武汉纺织大学



燕山大学

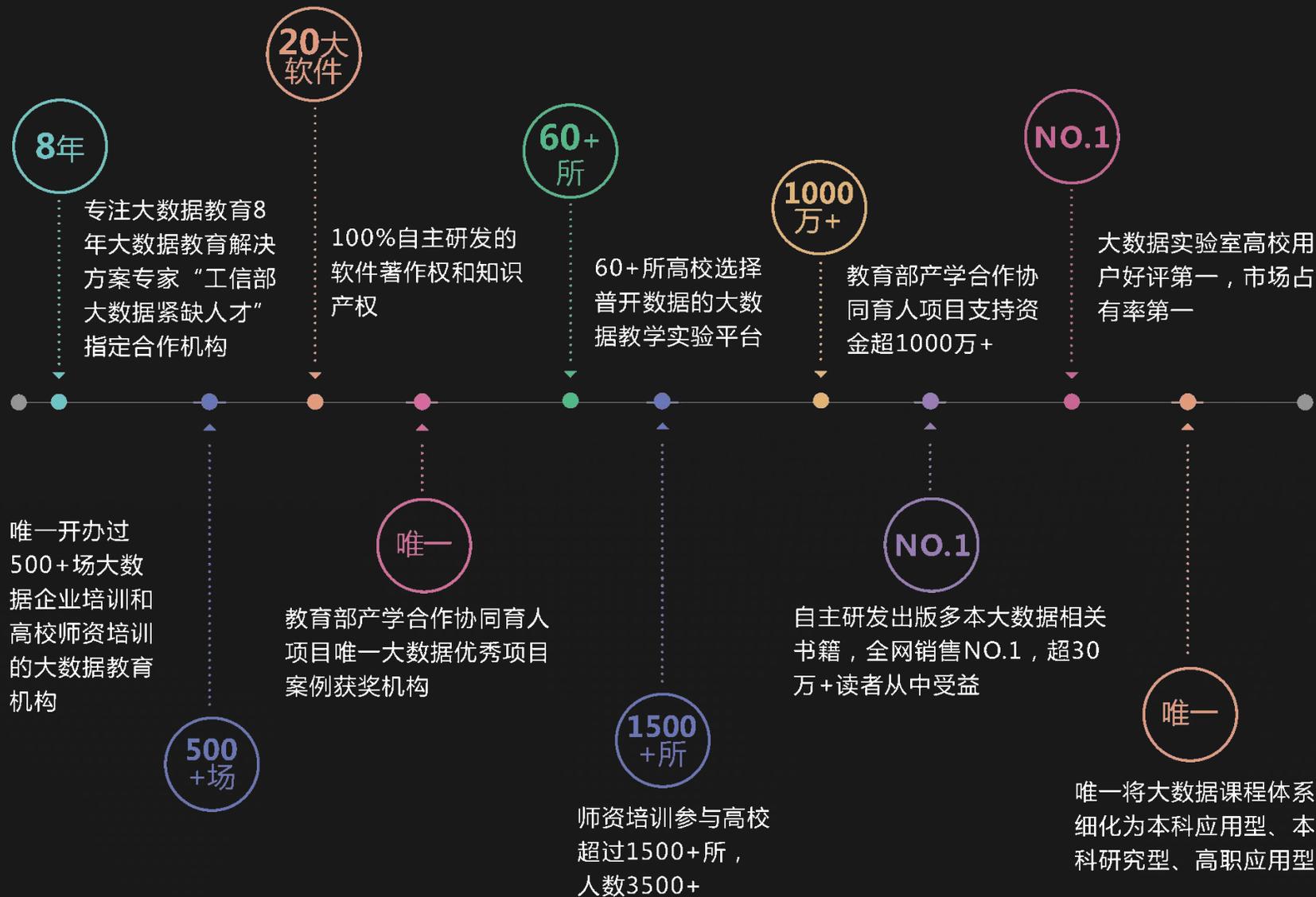


长江大学



中北大学

普开成绩



普开资质



软件著作权——22+





渠道合作



关注普开

THANK YOU