

**BDTC** 2017 中国大数据技术大会  
Big Data Technology Conference 2017

# 大数据教学与人才培养

王元卓

中国计算机学会大数据专家委员会

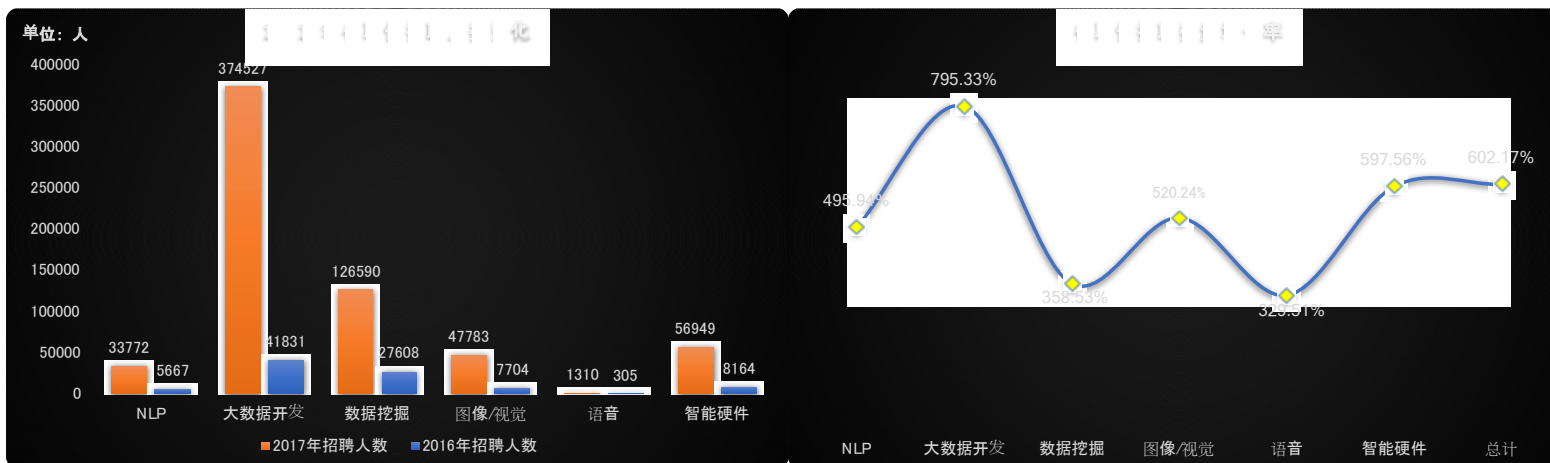
大数据分析系统国家工程实验室

中国科学院计算技术研究所

2017年12月

## 大数据人才缺口

- 2016年全国大数据人才仅46万，未来3-5年内大数据人才缺口将高达150万<sup>1</sup>；
- 截止2017年12月，大数据及人工智能岗位人才需求量比2016年猛增6倍多，大数据开发类增长幅度达到795%<sup>2</sup>；
- 大数据开发、数据挖掘相关人才流动大，平均在职时间不超过2年<sup>2</sup>。



大数据人才需求特点：

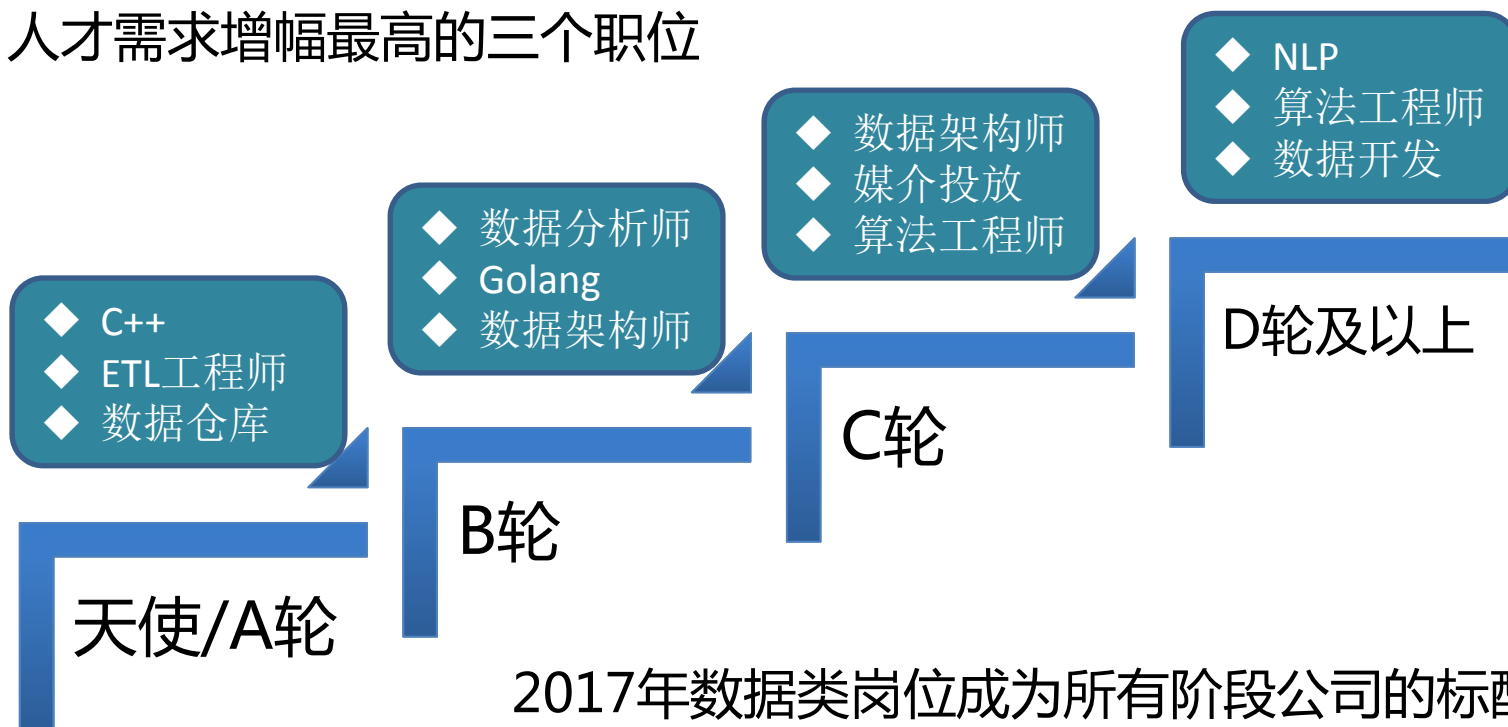
需求量大、薪资水平高、呈上升趋势

<sup>1</sup>数据来源：数联寻英2016年发布《大数据人才报告》

<sup>2</sup>数据来源：e成科技和Dt大数据产业创新研究院联合发布《2017年大数据及人工智能人才发展报告》

# 企业的大数据人才能力差异化需求

2017年春季各融资阶段公司  
人才需求增幅最高的三个职位



2017年数据类岗位成为所有阶段公司的标配，  
C、D轮之后的公司正式进入算法和AI人才的争夺

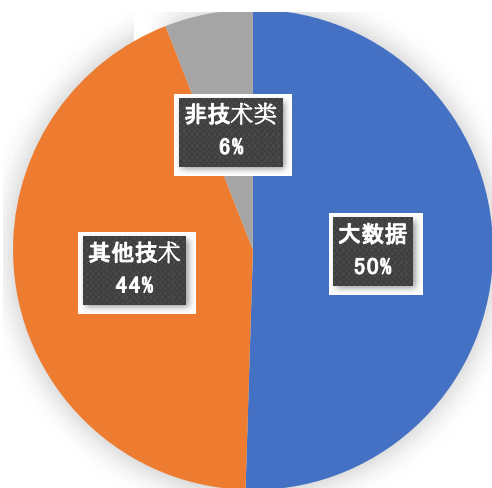
# 大数据核心人才





## 企业大数据人才需求问题

2017年春季在职大数据人才的上一份工作<sup>1</sup>



数据类人才流失率不高，但几乎全部来自学界、业界或其他技术岗转岗，**新鲜血液供给严重不足**



### 合格人才短缺，招聘困难

- 人才供给大幅增加，但人才质量参差不齐，符合需要的人才数量依然有限



### 理论基础强，但缺乏应用实践能力

- 基础理论技术能力强，实际工作解决问题时缺乏业务与技术灵活结合能力



### 人才培养渠道有限，培养周期长

- 工作5年以上的大数据及AI人才是行业主力，大量存在于学界和巨头公司



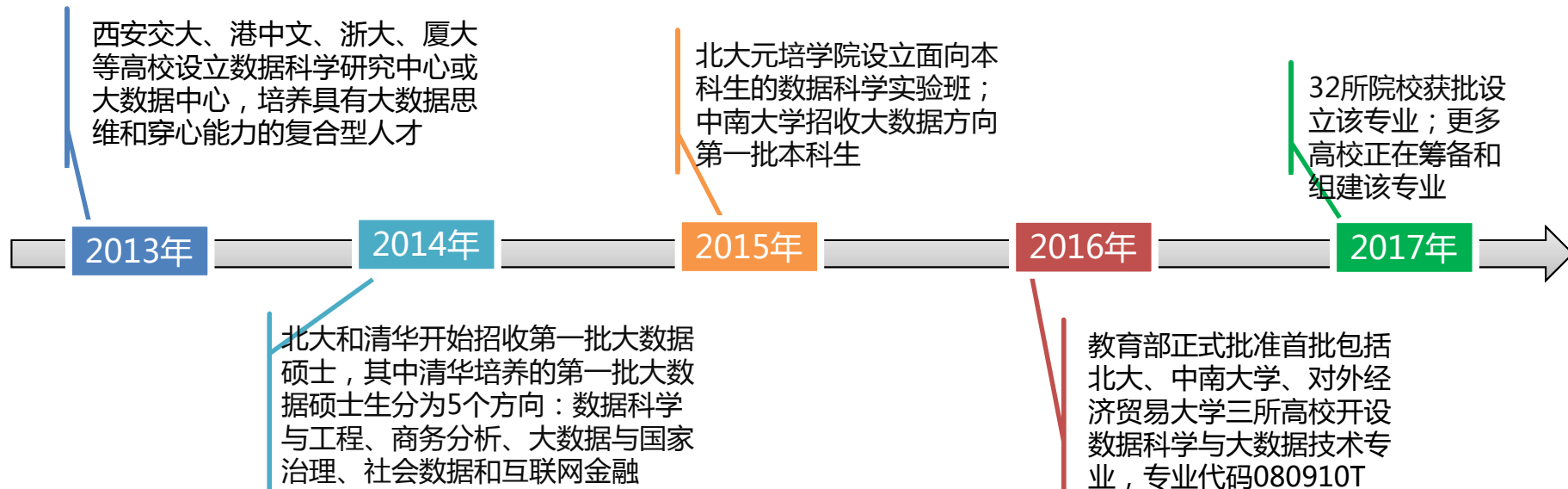
### 市场人才争夺激烈，内部人才流失风险高

- 平均工作1年，市场工资翻倍
- 平均在职时间不超过2年

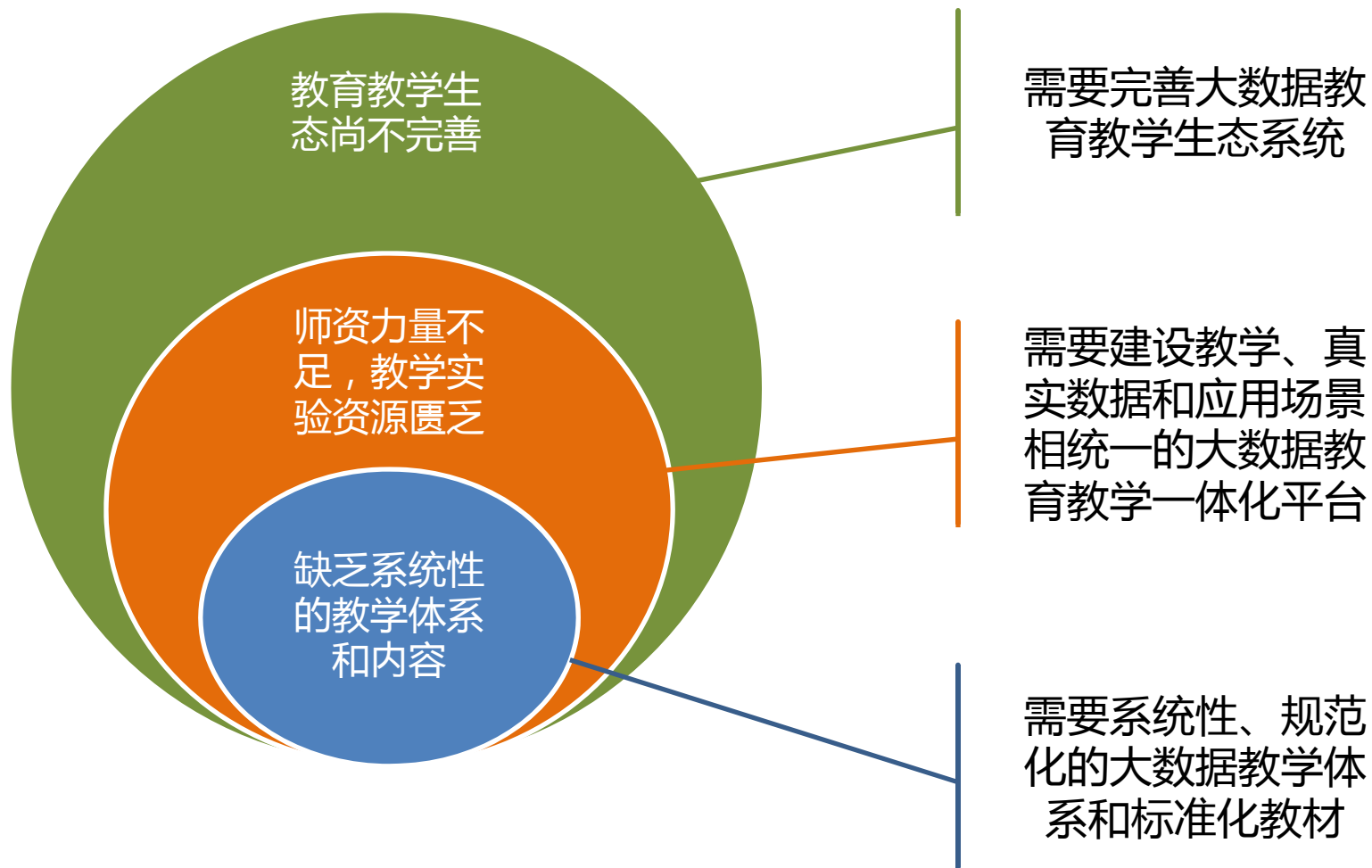
<sup>1</sup>数据来源：BOSS直聘《2017年春季互联网人才趋势报告》

## 大数据相关专业人才培养现状

- 2015年8月国发[2015]50号文《促进大数据发展行动纲要》
  - 提出“……建立健全多层次、多类型的大数据人才培养体系。鼓励高校设立数据科学和数据工程相关专业，重点培养专业化数据工程师等大数据专业人才。鼓励采取跨校联合培养等方式开展跨学科大数据综合型人才培养，大力培养具有统计分析、计算机技术、经济管理等多学科知识的跨界复合型人才。”



# 大数据人才培养面临的问题



## ■ 高校已经开始大数据人才的培养

- 截至目前，35所高等院校（首批三所、第二批32所，见附表1）已经获批设立了“数据科学与大数据技术”专业，2017年263所（其中工学190所，理学73所）正在申报、等待审批；
- 16所学校已设立大数据研究院培养大数据方向的硕士和博士；
- 2016年9月，“大数据技术与应用”专业开始列入高职高专专业目录，目前全国有70多所高职高专院校设有大数据专业；

## ■ 多校合作的协同创新、联合培养模式

- 如中国人民大学、北京大学、中国科学院大学、中央财经大学(招生办)和首都经贸大学5所高校组建了一个协同创新平台，以“应用统计专业硕士”为载体培养大数据分析方面的人才。

## ■ 企业开始重视并着手大数据人才的培养

- 如BAT等皆有自己的大数据人才培养模式

中国“数据科学与大数据技术专业”高校名单

学校名称	专业名称	专业代码	学位授予门类	修业年限	学校地址
长春理工大学	数据科学与大数据技术	080910T	工学	四年	吉林
北京大学	数据科学与大数据技术	080910T	理学	四年	北京
北京信息科技大学	数据科学与大数据技术	080910T	工学	四年	北京
北京邮电大学	数据科学与大数据技术	080910T	工学	四年	北京
对外经济贸易大学	数据科学与大数据技术	080910T	工学	四年	北京
中国人民大学	数据科学与大数据技术	080910T	工学	四年	北京
晋中学院	数据科学与大数据技术	080910T	工学	四年	山西
中北大学	数据科学与大数据技术	080910T	工学	四年	山西
福建工程学院	数据科学与大数据技术	080910T	工学	四年	福建
宿州学院	数据科学与大数据技术	080910T	工学	四年	安徽
复旦大学	数据科学与大数据技术	080910T	理学	四年	上海
华东师范大学	数据科学与大数据技术	080910T	工学	四年	上海
上海工程技术大学	数据科学与大数据技术	080910T	工学	四年	上海
上海纽约大学	数据科学与大数据技术	080910TH	工学	四年	上海
浙江财经大学	数据科学与大数据技术	080910T	理学	四年	浙江
北京师范大学-香港浸会大学联合国际学院	数据科学与大数据技术	080910TH	工学	四年	北京
佛山科学技术学院	数据科学与大数据技术	080910T	工学	四年	广东
广东白云学院	数据科学与大数据技术	080910T	工学	四年	广东
广西科技大学	数据科学与大数据技术	080910T	工学	四年	广西
黄河科技学院	数据科学与大数据技术	080910T	工学	四年	河南
湖北经济学院	数据科学与大数据技术	080910T	工学	四年	湖北
中南大学	数据科学与大数据技术	080910T	工学	四年	湖南
安顺学院	数据科学与大数据技术	080910T	工学	四年	贵州
贵州大学	数据科学与大数据技术	080910T	工学	四年	贵州
贵州理工学院	数据科学与大数据技术	080910T	工学	四年	贵州
贵州商学院	数据科学与大数据技术	080910T	工学	四年	贵州
贵州师范大学	数据科学与大数据技术	080910T	工学	四年	贵州
成都东软学院	数据科学与大数据技术	080910T	工学	四年	四川
电子科技大学	数据科学与大数据技术	080910T	工学	四年	四川
电子科技大学成都学院	数据科学与大数据技术	080910T	工学	四年	四川
昆明理工大学	数据科学与大数据技术	080910T	工学	四年	云南
云南财经大学	数据科学与大数据技术	080910T	理学	四年	云南
云南师范大学	数据科学与大数据技术	080910T	工学	四年	云南
重庆理工大学	数据科学与大数据技术	080910T	工学	四年	重庆
宁夏理工学院	数据科学与大数据技术	080910T	工学	四年	宁夏

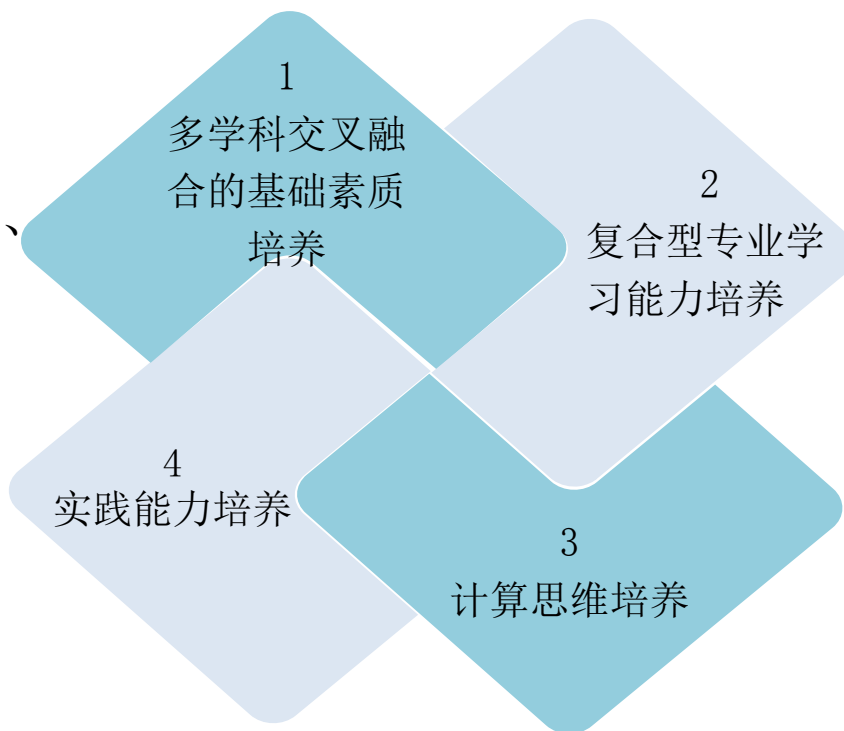


- 国内开设数据科学与大数据技术专业或开设大数据相关课程的学校越来越多，对大数据教材的需求越来越大；
- 目前国内高校的大数据教学尚处在摸索阶段，无统一教材，有的选用教师自己的教案；
- 高校或科研院所对大数据教材的需求主要分为两类：
  - 一类是开设了大数据专业的学校需求一整套完整的大数据教育丛书；
  - 另一类是未开设大数据专业的学校单独开设一至两门大数据相关课程时，需要知识体系完整又不空泛的导论类教材及实验类教材

# 培养兼具行业应用专业知识和数据技术的复合型人才

兼具行业应用和大数据工程知识的通识教育、语言能力、法律道德、计算机伦理、数学基础等

案例教学、工程实践等



行业应用基础理论、数据科学基础理论、行业应用数据技术、学习方法等

建模能力、求解能力、数据工程基础理论与技术等

- **教材受众定位**：面向本科生与研究生，辐射职业教育。教材章节中含初级部分与进阶部分，不同水准的受众可以选取合适的部分进行学习。
- **教材内容标准**：编委会对教材的结构、目录、内容进行严格把关，完成一套高质量、高水准的大数据教材。教材既包括理论和技术，也包括领域应用；并且领域应用部分会根据产业的发展同步进行更新。
- **教材出版计划**：大数据系列教材中的基础部分和关键技术部分于2018年8月前出版，高校秋季开学投入使用；领域应用部分的教材陆续出版。

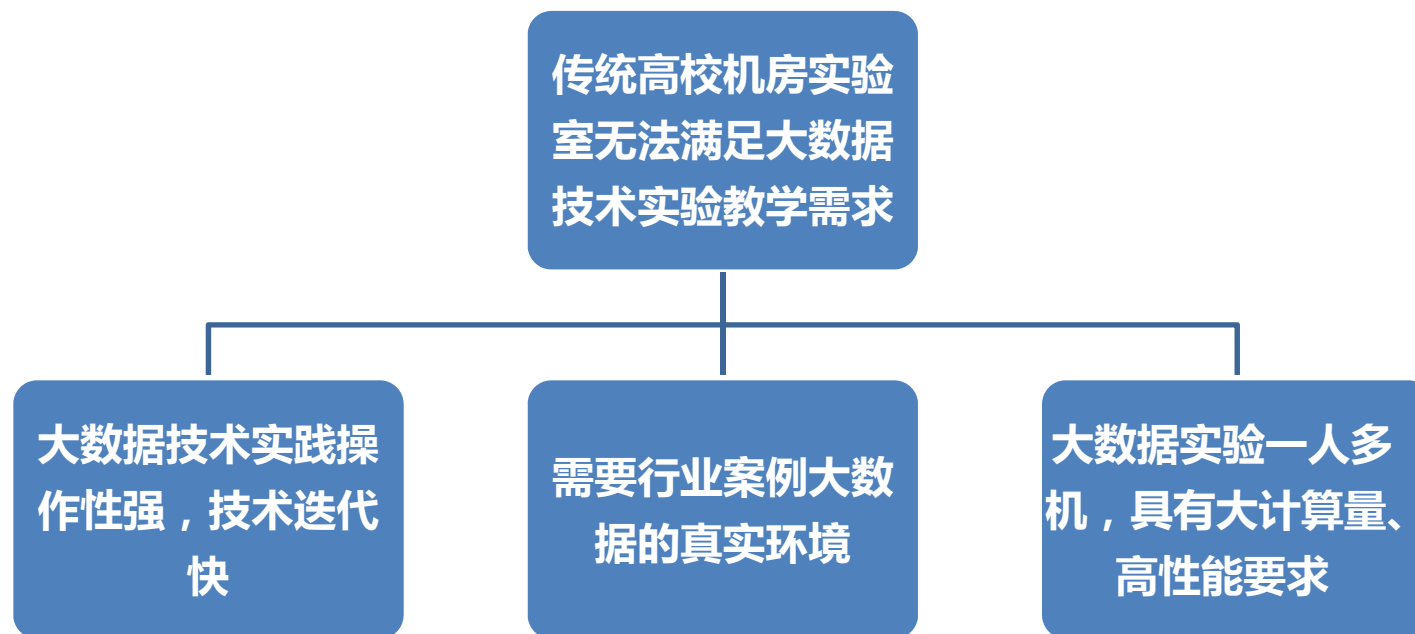
# 师资力量不足



传统计算机、统计、数学、信息系统等专业师资力量充足，但系统掌握大数据技术、具有行业大数据技术经验师资不足



# 大数据实验教学资源匮乏



具有行业应用紧耦合真实数据环境、满足实验实践教学需求的一体化大数据实验教学平台

# 大数据教学一体化实验平台

## 大数据实验室教学软件平台

- 大数据教学管理系统
- 云计算管理系统
- 大数据集群监控系统
- 大数据数图可视化系统

## 大数据实验室硬件环境

- 大数据一体机
- 可视化液晶拼接显示墙



## 大数据基础实验课程包

- 大数据Hadoop基础课程包
- 大数据离线分析课程包
- 大数据实时计算课程包
- 大数据高级进阶课程包

## 大数据行业项目案例课程包

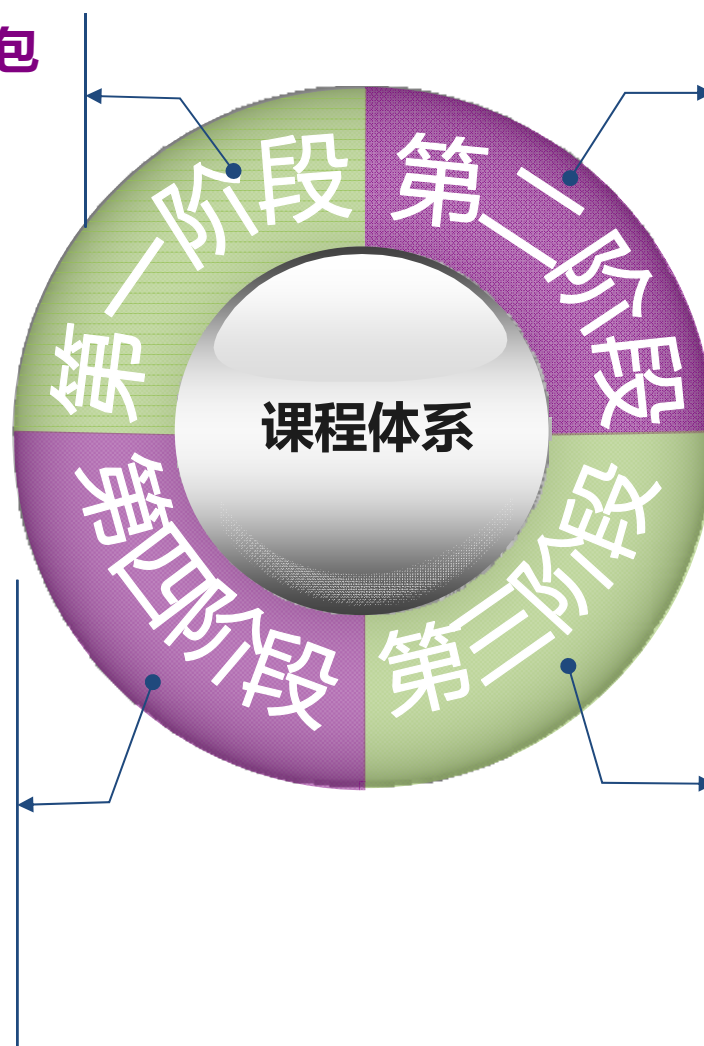
# 大数据基础课程体系

## 大数据Hadoop基础课程包

- Hadoop安装部署
- HDFS分布式文件系统
- MapReduce计算框架

## 大数据离线分析课程包

- Hive基础
- Hive数据定义与操作
- Hive数据查询
- Hive内置函数和UDF
- Sqoop ETL工具
- Azkaban工作流引擎



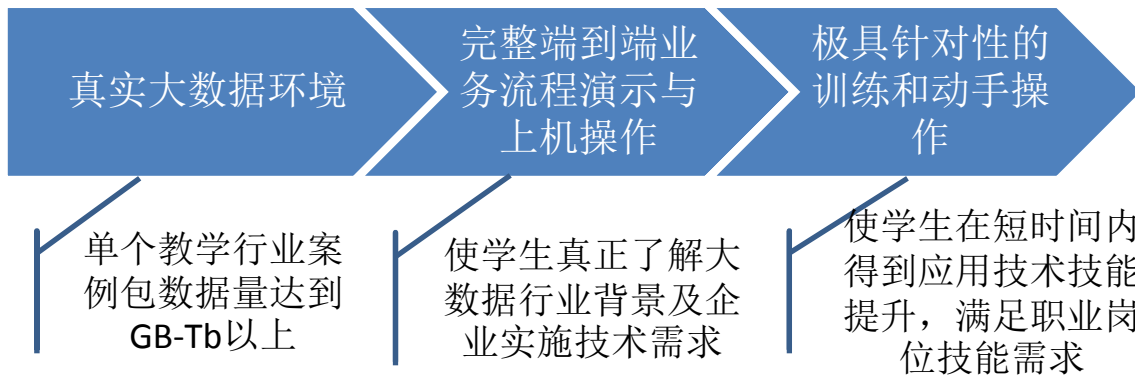
## 大数据高级进阶课程包

- Scala基础编程
- Scala核心概念
- Scala高级编程
- Spark简介和安装部署
- Spark核心机制
- Spark SQL基础
- SQL接口编程和调优
- Spark Streaming开发
- MLlib开发和Graphx

## 大数据实时计算课程包

- ZooKeeper分布式协调系统
- HBase分布式数据库
- HBase核心功能模块和Shell客户端
- HBase原生Java客户端
- HBase批量操作与WEB UI工具
- HBase核心知识点
- HBase高级特性
- Storm基础
- Storm进阶
- Kafka分布式发布订阅消息系统
- Flume海量日志采集系统

# 行业应用真实大数据场景



单个教学行业案例包数据量达到GB-Tb以上

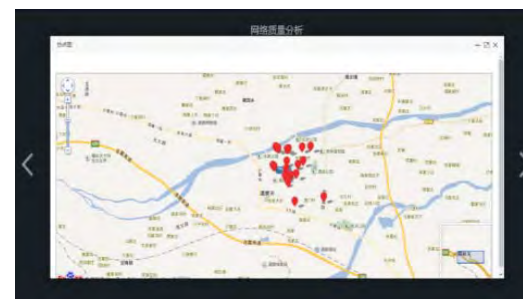
使学生真正了解大数据行业背景及企业实施技术需求

使学生在短时间内得到应用技术技能提升，满足职业岗位技能需求

## 应用案例：运营商在线服务大数据

- ◆ 电子渠道提供足不出户的便捷业务办理方式
- 全区服务统计：各服务区域内的申告量、处理量、回单量等
- 投诉风险：通过申告次数、呼叫次数和用户情绪这三个维度评定用户等级，预防高危用户流失，降低投诉次数，提高用户体验
- 区域服务效能：通过申告量、专席人数来对各区域服务团队进行监控
- 热点故障区域分布：在地图上展现多发故障的位置以及故障信息

★数据描述：来源于服务商业务系统：工单历史数据1000W条；专席通话记录数据2000W条；27个视频，共计时长560分钟。





# 大数据教学平台

## 课程/实验

- 高清课程视频
- 无需安装实验环境
- 实验文档
- 实验报告提交及打分
- 课程资源池
- 课程教案创建
- 班级管理
- 课程安排
- 学生实验界面监控

## 考试/作业

- 丰富的试题题库
- 多种试题类型
- 在线考试系统
- 作业发布及打分
- 错题智能汇总
- 考试评分讲解
- 在线作业编辑

## 数据可视化

- 批量导入数据
- 智能数据入库
- 多种图表类型
- 拖拽式图表创建
- 图表智能推荐
- 自定义项目UI风格
- 移动端项目展示
- 丰富的UI风格库
- 项目分享

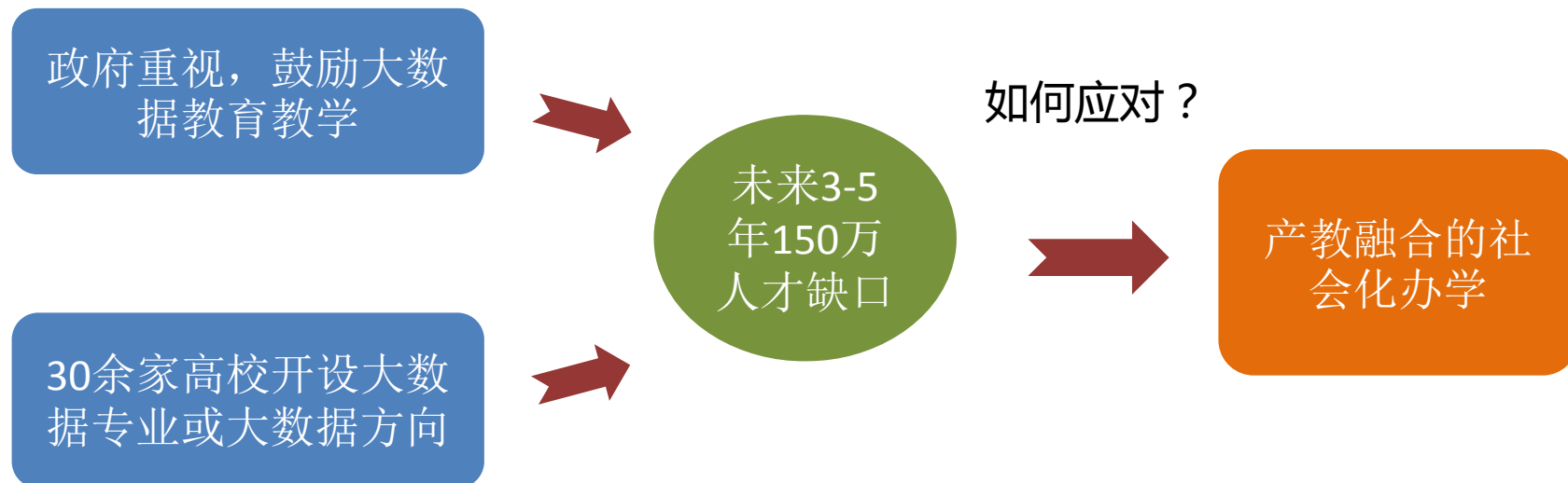
## 学习行为分析

- 学习进度报告
- 实验成绩分析
- 考试作业报告
- 用户反馈报告
- 用户操作统计
- 问卷调查报告

## 云管理平台

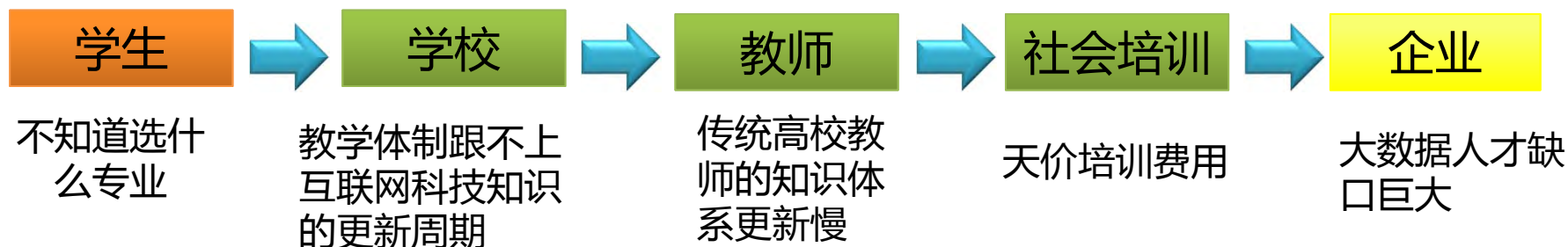
- 服务器资源监控
- 虚拟机管理
- 主机配置管理
- 镜像管理
- 组织结构管理
- 用户管理
- 系统管理

# 引入社会力量办学

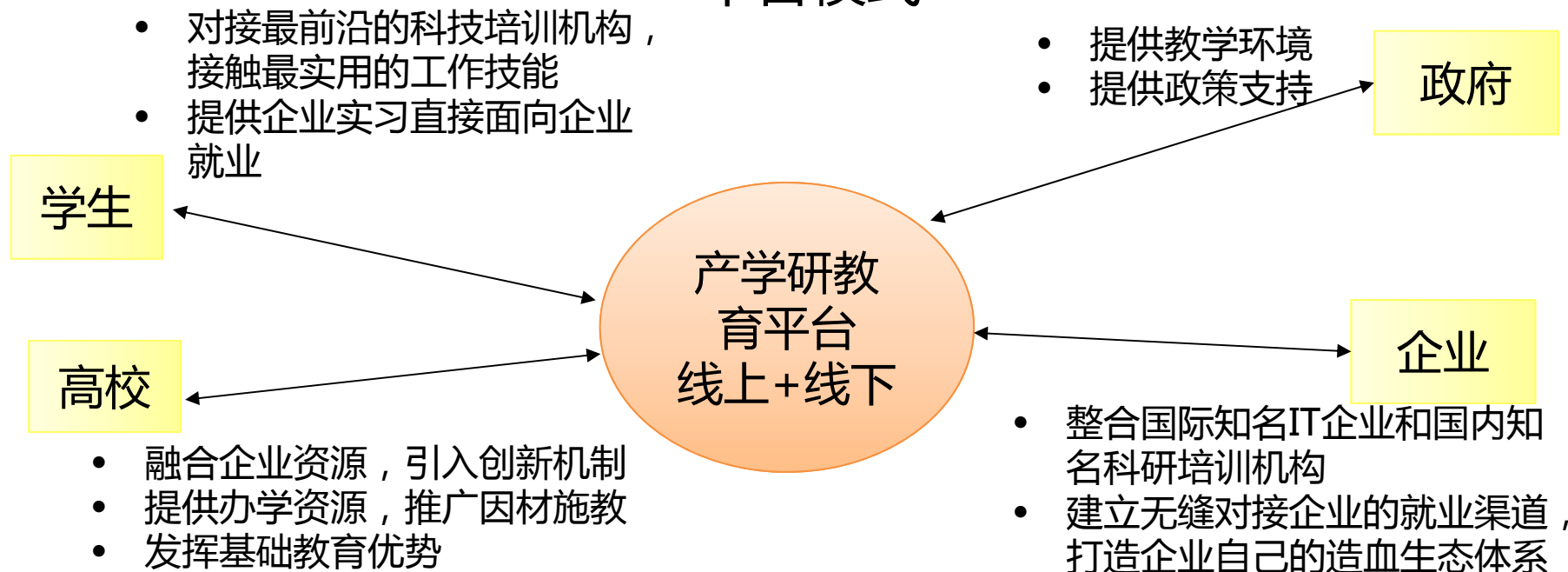


引入社会力量，推动产教融合的社会化办学，探索  
大数据人才教育创新模式

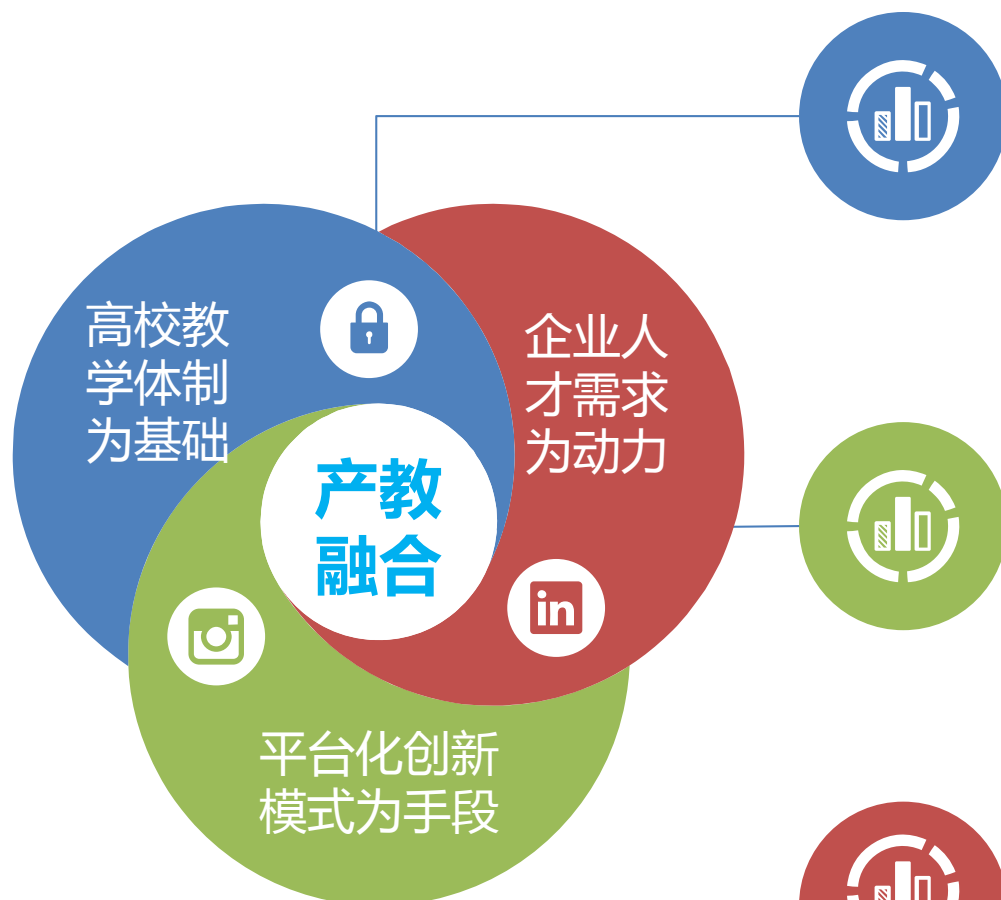
# 产学研教育平台模式



## 平台模式



# 产教融合的平台模式



## 产

引入国内在信息技术研发、行业应用有着强实力和丰富经验，并且在大数据行业具有前瞻和引领能力的IT公司大型的IT企业

## 教

与高校联合办学，深度整合高校资源，接轨最前沿的技术和教学理念，发挥体制优势，突破体制短板

## 平台

能够全方位引入国际化教学资源 and 教学理念；建立大数据教育教学体系、大数据一体化实验教学平台



## 打造大数据教育生态系统

- 组织大数据教材作者走进高校系列活动
  - 系列活动提供教材作者与授课教师、学生面对面交流的机会
  - 为授课教师、学生带来大数据行业领域的前沿进展
  - 为教材的修订收集建议
- 与企业合作推动大数据技术培训
  - 寻求并引入更多知名企业与高校合作，进一步推动产教融合
- 依托CCF大数据学术会议，组织全国高校大数据学院院长高峰论坛
  - 搭建专业的大数据教育教学交流平台，汇聚全国大数据教育精英，共同探讨大数据专业、课程体系建设
  - 进一步推动全社会对大数据教育教学的重视，共同打造我国大数据教育的创新型模式和生态系统