

## 一、采购清单

序号	货物名称	数量	单位	是否核心产品
1	数据分析处理实验平台	1	套	是
2	大数据分析应用实验平台	1	套	是
3	管理控制设备	1	台	否
4	云资源计算设备	3	台	否
5	云接入设备	1	台	否
6	机架装置设备	1	台	否

## 二、技术要求

序号	货物名称	规格、技术参数
1	数据分析处理实验平台	<p>一、门户管理模块</p> <p>▲1、要求支持门户的自定义管理功能，在同一页面下门户模块包括轮播图、快捷访问、搜索栏、快速导航、自选资源、筛选结果、镶嵌页面、图文信息、热门推荐等；支持自定义调整模块位置；支持自选资源后课程模块在首页显示；（提供软件功能截图并加盖投标人公章予以佐证）</p> <p>★2、要求支持个性化界面设计，允许教师根据个人喜好快速选择和切换空间主题配色，当选定主题配色后，支持自动同步至相关界面元素；支持内置预设空间主题配色，主题配色至少包括天青蓝、石榴红、马尔斯绿、金橘色、普鲁斯蓝等；（提供软件功能截图并加盖投标人公章予以佐证）</p> <p>▲3、要求空间公告以模块的形式展示，每条公告显示出公告标题与内容，支持设置热门公告与公告发布的功能；设置成热门公告在首页置顶，并显示出热门图标；（提供软件功能截图并加盖投标人公章予以佐证）</p> <p>4、要求支持新建公告，填写名称、封面、公告内容等信息，公告内容以 Markdown 形式进行编辑，支持同时直接添加文字、图片、视频、代码、表格等；</p> <p>二、教学规划模块</p> <p>1、要求添加教学规划时，支持添加授课、实训事项；在添加授课与实训内容时，展示出课程列表，选择列表后展示出课程小节，可同时添加多门课程与多个小节，支持一键全选功能；创建教学规划后，授课内容、实训内容、竞赛内容同时在该教学规划中展示；</p> <p>▲2、要求查看教学规划时，具体事项列表展示，在授课事项下，教师端显示备课、授课功能，学生端显示预习、学习功能；在实训事项下，教师端与学生端均显示开始实训功能；在竞赛事项下，教师端显示监控中心功能，学生端显示开始比赛功能；在授课完成后，教师端显示课堂数据、作业批改功能，学生端显示课堂回顾、查看视频功能；（提供软件功能截图并加盖投标人公章予以佐证）</p> <p>▲3、要求数据概览展示出班级人数、学习内容、知识点分布、学习人数等信息；要求知识点分布以雷达图的形式展示；学习人数以日历热力图的形式展示，热力图总体以方格矩阵排列，横向为月份，纵向为日历天，支持采用渐变颜色映射来展示每日数据在时间跨度中的变化。当鼠标放置某一方格矩阵时，可查看该日学习总人数；（提供软件功能截图并加盖投标人公章予以佐证）</p> <p>★4、要求班级数据包括课堂信息、课堂人数变化趋势、出勤情况、学生列表等信息；课堂信息包括课堂名称、累计授课时长、创建时间、参与学习人数、学生平均在线时长；课堂人数变化趋势支持以折线图形式展示，横坐标为时间轴，时间刻度精确到分钟，纵坐</p>

标为课堂人数；

5、提交情况总览支持以可视化的形式展示，展示效果为环状图配百分比，支持统计出总人数、已提交人数、未提交人数信息；系统支持自动统计出未提交学生名单及已提交学生名单，支持教师端对未提交学生进行催交作业；已提交学生名单支持展示姓名、提交时间等信息；要求提交时间展现格式为年-月-日 时:分:秒；

6、批改情况总览支持以可视化的形式展示，展示效果为环状图配百分比，支持统计出已提交人数、已批改人数、未批改人数信息；系统支持自动统计平均分、最高分、最低分；批改详情信息支持显示排名、姓名、得分等信息；

### 三、课程授课模块

☆1、要求 Web 版 PPT 支持在同一页面下翻页讲解；理实一体化课程的 Web 版 PPT，支持在同一页面下与实操演示环境，一键相互切换；支持课堂自主练习，系统可自动检测练习结果的正确性，并给予提示；支持每一个课程小节具有辅助教学功能框，在功能框下具有画板、公告、举手、全屏、投影、课堂邀请等子功能；支持实时课堂功能，可查看人员名单、问答详情、学生举手等；（提供现场功能演示）

▲2、要求授课时支持三屏互动教学方式，包括学生区、教师区、投屏区；学生区、教师区、投屏区可同步画面，也可以锁定画面；支持在授课时选择投屏，系统会自动生成课程投屏链接；（提供软件功能截图并加盖投标人公章予以佐证）

3、要求自主练习数据分析模块显示全部完成人数，支持考核步骤的详情分析，分析内容包括姓名、任务进度、提交次数与完成时间；任务进度以可视化的形式展示；

4、要求教师授课时长包括实际授课时长、建议授课时长；支持统计理论讲解、实操演示、课堂问答、自动练习等环节的授课时长，时间单位精确到分钟级别；

5、要求支持课堂信息数据包括课堂授课时间范围、课堂状态、所属班级、数据导出与历史课堂；支持数据导出格式为 Excel；

6、要求支持查看互动记录功能，可查看内容包括公告信息、提问信息与讲解信息；

### 四、课程实训模块

▲1、要求实训指导书包括任务整体介绍、实训步骤、任务考核要求、步骤完成状态、任务解析；支持教师端设置任务解析是否公布；支持系统自动检测实训结果的正确性，并给予提示；（提供软件功能截图并加盖投标人公章予以佐证）

2、要求任务分析完成进度以进度条的形式展示，以环状图的形式展示完成人数、参与练习人数百分比；以百分比的形式展示考核条件平均通过率，支持平均提交检测次数统计与未完成学生名单；

3、要求实验报告以列表的形式展示，内容包括成员名称、账号、完成任务、完成率、学习时长、最新提交时间、操作等信息；支持批量下载功能，可将报告下载到本地，报告导出格式为 PDF；

▲4、要求实验报告显示总览，包括练习时长、完成任务、系统得分；支持教师评分及教师评价输入；支持自动将任务考核步骤附加在报告中，内容包括题目名称、条件得分、操作环境等；支持步骤考核的通过情况以不同颜色进行区分；支持下载报告功能，报告下载格式为 PDF；（提供软件功能截图并加盖投标人公章予以佐证）

### 五、课程管理模块

1、要求系统提供的 Web 版 PPT 支持插入文本、插入图片、插入形状、插入线条、插入代码、插入表格、插入视频等功能，支持设置动画片段与删除当前页面的快捷键；支持整体的预览功能；

2、要求支持视频上传功能，至少支持 .mp4 .webm .mov 等视频规格，支持设置视频封面；

▲3、要求支持添加视频要点功能，支持选择视频时间轴上的时间点作为视频要点播放节点，支持以 Markdown 格式填写要点详解；支持自动排序功能；（提供软件功能截图并加盖投标人公章予以佐证）

☆4、要求新增实训项目支持设置基本信息与实训内容，基本信息包括名称、简介；实训内容支持设置任务整体介绍，任务整体介绍支持添加多个要点介绍；任务列表支持添加多个考核任务，每个任务均填写任务名称、任务要求与任务解析，对于每个任务可添加智能考核，支持添加一个或多个考核检查项，考核条件的填写内容包括考核条件名称、得分设置、考核验证执行模板、超时时间设置、错误提示；（提供现场功能演示）

5、要求问答内容面板设置名称、简介、考核题，考核题支持选择题、填空题、简答题；选择题内容设置包括题目描述、答案选项与题目解析；填空题内容设置包括题目描述、正确答案与题目解析，支持答案字母区分大小写的功能，支持添加正确答案选项；简答题内容设置包括题目描述、正确答案与题目解析，支持按关键字判断正确性；

#### 六、资源管理模块

1、要求支持新增考核模版功能，设置新模版的名称、脚本参数、脚本、错误提示、描述、适用环境等；虚拟机适用环境至少包括 Windows、Linux、Unix 等三种；容器支持 Linux 类型的环境；

2、要求考核模版具有检查命令执行结果不包含某个关键词、判断文件是否存在、检查命令执行结果包含某个关键词、判断文件是否存在关键字等；支持常见的考核条件至少 60 个；（提供软件功能截图并加盖投标人公章予以佐证）

☆3、要求新增自主练习资源，基本信息包括名称、操作环境、连接方式、简介、任务介绍与知识点介绍；考核内容支持添加多个考核任务，每个任务均填写任务名称、任务要求与任务解析，对于每个任务可设置一个或多个考核条件，考核条件的填写内容包括考核名称、考核模版、考核判定方式、检测超时时间、错误提示；（提供现场功能演示）

#### 七、实训课程内容

##### 1、Python 程序设计基础与应用课程资源包

要求 Python 程序设计基础与应用授课课程资源包支持 60 个课时，提供 30 个 Web 版 PPT，Web 版 PPT 支持在同一页面下翻页讲解，提供 30 个讲解视频；提供 15 个课程理实一体化小节，支持在同一页面下一键切换 Web 版 PPT 与实操演示环境；提供 60 个考核题目；课程章节包括 Python 语言概述、运算符、表达式和内置对象、Python 序列结构、选择结构与循环结构、函数、面向对象程序设计、字符串、正则表达式、文件内存操作、文件与文件夹操作、异常处理结构、异常处理结构、异常处理结构、Python 数据分析与处理、数据可视化等。

##### 2、SPSS 数据预处理课程资源包

要求 SPSS 数据预处理实训课程支持 10 个小节，提供 20 个任务考核步骤，支持自动检测步骤结果；课程章节包括 SPSS 统计软件基本操作、SPSS 数据的预处理、数据选取、数据清洗、数据转换等。

##### 3、MySQL 数据库实训课程资源包

要求 MySQL 数据库实训课程支持 20 个小节，提供 40 个任务考核步骤，支持自动检测步骤结果；课程章节包括 MySQL 基础、操作数据库、操作数据表、数据管理、单表查询、多表查询、索引、视图、事务、存储例程、触发器、数据备份与恢复等。

##### 4、信息技术基础授课课程资源包

要求信息技术基础授课课程资源包支持 100 个课时，提供 100 个 Web 版 PPT，Web 版 PPT 支持在同一页面下翻页讲解，提供 100 个讲解视频；提供 50 个课程理实一体化小节，支持在同一页面下一键切换 Web 版 PPT 与实操演示环境；提供 200 个考核题目；课程章节

包括认识信息技术应用与发展、了解信息系统组成与工作机制、学习和使用信息技术设备、认识操作系统、管理信息资源、系统软件与应用软件、系统安全与维护、认识网络技术与应用、认识浏览器、配置网络、获取并识别网络资源、网络交流、物联网、认识图文编辑并进行基本操作、编辑文本、制作演示文稿、绘制表格、实战-制作座位表、编排图文、实战-制作宣传海报、认识数据处理、采集数据、处理数据、数据分析与可视化、实战-制作学生成绩透视图、初识大数据、程序设计理念、搭建编程环境、基础语法、基本程序设计结构、初识数字媒体技术、获取数字媒体文件素材、处理数字图像、实战-制作动态交互式演示课件、处理数字音频与视频、制作计算机动画、实战-制作荡秋千的小女孩、初识虚拟现实与增强现实技术、信息安全技术、防范信息系统恶意攻击、初识人工智能、认识机器人等；

#### 5、网络信息采集技术课程资源包

要求网络信息采集技术授课课程资源包支持 30 个课时，提供 30 个 Web 版 PPT，Web 版 PPT 支持在同一页面下翻页讲解，提供 30 个讲解视频；提供 15 个课程理实一体化小节，支持在同一页面下一键切换 Web 版 PPT 与实操演示环境；提供 60 个考核题目；课程章节包括网络爬虫入门、网页前端基础、请求模块 urllib3、请求模块 Requests、re 正则表达式模块、XPath 解析、数据解析模块 BeautifulSoup、动态渲染网页爬取、数据存储、Scrapy 爬虫框架等、案例 1-爬取百度首页 Logo、案例 2-爬取豆瓣电影 Top250 榜、案例 3-爬取糗事百科数据并写入数据库、案例 4-爬取中国日报新闻数据。

#### 6、Sqoop 数据传输工具实训课程资源包

要求 Sqoop 数据传输工具实训课程支持 5 个小节，提供 10 个任务考核步骤，支持自动检测步骤结果；课程章节包括 Sqoop 环境搭建、Sqoop 数据导入、Sqoop 数据导出、淘宝用户行为数据分析、大众点评网热门糖水店铺用户评论数据分析等。

#### 7、Spark 技术与大数据应用授课课程资源包

要求 Spark 技术与大数据应用授课课程资源包支持 60 个课时，提供 50 个 Web 版 PPT，Web 版 PPT 支持在同一页面下翻页讲解，提供 50 个讲解视频；提供 25 个课程理实一体化小节，支持在同一页面下一键切换 Web 版 PPT 与实操演示环境；提供 100 个考核题目；课程章节包括 Spark 程序结构、RDD 编程、案例 1-统计学科网页访问次数、数据的读取、Spark SQL-DataSet、Spark SQL-DataSet、Kafka 概述、Spark Streaming、Flume 简介、Redis 简介、案例 2-用户在基站停留时间、Spark Mllib、Spark GraphX 等；要求提供正规出版社出版的配套 Spark 教材，投标时写明教材名称及教材书号，保证所投产品与教材完全匹配，且至少需要提供一位教材主编签字声明文件，声明文件要附签字主编身份证复印件。

#### 8、Python 数据分析与可视化授课课程资源包

要求 Python 数据分析与可视化授课课程资源包支持 60 个课时，提供 60 个 Web 版 PPT，Web 版 PPT 支持在同一页面下翻页讲解，提供 60 个讲解视频；提供 30 个课程理实一体化小节，支持在同一页面下一键切换 Web 版 PPT 与实操演示环境；提供 120 个考核题目；课程章节包括数据分析基础、创建数组、数组的基本操作、NumPy 矩阵的基本操作、NumPy 常用统计分析函数、NumPy 文件读写操作、Pandas 入门、导入导出外部数据、DataFrame 的基本操作、数据清洗、数据的排序与排名、数据计算、数据转换与合并、日期和时间序列处理、图表的常用设置、常用图表的绘制、数据统计分析案例 1-网站 App 平台注册用户分析、数据统计分析案例 2-二手房房价分析等；

#### 9、R 语言统计计算与可视化授课课程资源包

要求 R 语言统计计算与可视化授课课程资源包支持 40 个课时，提供 40 个 Web 版 PPT，Web 版 PPT 支持在同一页面下翻页讲解，提供 40 个讲解视频；提供 20 个课程理实一体化小节，支持在同一页面下一键切换 Web 版 PPT 与实操演示环境；提供 80 个考核题目；课

程章节包括 R 语言基础、R 语言语法基础、数据导入导出、数据预处理、数据分析、数据可视化、数据分析综合实验 1-财政收入分析预测、数据分析综合实验 2-航空公司客户价值分析、数据分析综合实验 3-预测用户留存率、R 集成大数据等；

#### 10、数据获取与采集类

数据获取与采集类包括 Python 网络爬虫实训课程、数据采集工具选择与应用课程、Flume 日志收集系统实训课程、Sqoop 数据传输工具实训课程。

要求 Python 网络爬虫实训课程支持 10 个小节，提供 20 个任务考核步骤，支持自动检测步骤结果；课程章节包括 Python 环境搭建、XPath 解析、数据解析模块 Beautiful Soup、re 正则表达式模块、Ajax 数据爬取、Selenium 模拟浏览器、Scrapy 爬虫框架等；

#### 11、统计分析类

统计分析类包括数据统计算法实践(Python)、统计分析实训(SPSS Statistics)。

要求数据统计算法实践(Python)支持 20 个小节，提供 40 个任务考核步骤，支持自动检测步骤结果；课程章节包括数据描述、参数估计、假设检验、非参数检验、方差分析、判别分析、因子分析、时间序列分析、相关分析、回归分析等；

要求统计分析实训(SPSS Statistics)支持 40 个小节，提供 80 个任务考核步骤，支持自动检测步骤结果；课程章节包括 SPSS 统计软件基本操作、SPSS 数据的预处理、SPSS 描述性统计分析、SPSS 参数检验、SPSS 非参数检验、SPSS 方差分析、SPSS 相关分析和回归分析、SPSS 信效度检验、SPSS 进阶分析、SPSS 高阶分析、某高校大学生就业项目案例分析、京东消费者信任因素项目案例分析、某保险公司客户满意度项目案例分析、城镇居民消费支出项目案例分析等；

#### 12、PowerBI 商业智能可视化实训课程资源包

要求 PowerBI 商业智能可视化实训课程支持 10 个小节，提供 20 个任务考核步骤，支持自动检测步骤结果；课程章节包括初识 PowerBI、数据处理、编辑交互、数据可视化、2006 年-2015 年各省市 GDP 分析、北京、广州、杭州第一季度销售数据分析、豆瓣电影数据分析、大学生问卷调查数据分析等。

#### 13、机器学习算法与实践授课课程资源包

要求机器学习算法与实践授课课程资源包支持 40 个课时，提供 40 个 Web 版 PPT，Web 版 PPT 支持在同一页面下翻页讲解，提供 40 个讲解视频；提供 20 个课程理实一体化小节，支持在同一页面下一键切换 Web 版 PPT 与实操演示环境；提供 80 个考核题目；课程章节包括机器学习概述、数据处理、聚类、降维、分类、线性回归、支持向量机、决策树、集成学习、关联分析、推荐系统、深度学习等；

#### 14、机器学习类包括机器学习算法与实践实训课程、机器学习项目实战。

要求机器学习算法与实践实训课程支持 20 个小节，提供 40 个任务考核步骤，支持自动检测步骤结果；课程章节包括机器学习概述、数据处理、聚类、降维、分类、线性回归、支持向量机、决策树、集成学习、关联分析、推荐系统、深度学习等；

要求机器学习项目实战支持 5 个实训项目。包括泰坦尼克号乘客生存概率分析、信用卡欺诈风控、支持城市道路通行时间预测、京东购买意向分析、粮农数据分析、电力泄漏识别分析、商品销售额预测、纽约交通事故数据分析等；

#### 15、Hadoop 技术与大数据应用课程资源包

要求 Hadoop 技术与大数据应用授课课程资源包支持 90 个课时，提供 40 个 Web 版 PPT，Web 版 PPT 支持在同一页面下翻页讲解，提供 40 个讲解视频；提供 20 个课程理实一体化小节，支持在同一页面下一键切换 Web 版 PPT 与实操演示环境；提供 80 个考核题目；课程章节包括 Hadoop 体系介绍、Hadoop 集群搭建、HDFS 基础入门、HDFS 核心设计、HDFS 原理解析、HDFS 的 Java API 操作、MapReduce 基础入门、MapReduce 流量统计项目案例、

MapReduce 典型编程案例进阶、MapReduce 自定义输入输出组件、Hadoop 数据压缩、YARN 资源调度器、倒排索引案例、寻找微博共同好友案例、用户购买行为分析、手机品牌销售洞察、手机画像等。

#### 16、Excel 数据预处理实训课程资源包

求 Excel 数据预处理实训课程支持 5 个小节，提供 10 个任务考核步骤，支持自动检测步骤结果；课程章节包括基础知识、数据收集。

#### EXCEL 数据可视化实训课程资源包：

要求 EXCEL 数据可视化实训课程支持 4 个小节，提供 8 个任务考核步骤，支持自动检测步骤结果；课程章节包括制作气象站气温变化折线图、制作大学生调查分析柱形图、制作销售部年度利润面积图、制作某公众号粉丝占比图表化等。

#### 17、神经网络与深度学习类

要求神经网络与深度学习原理实现实训课程支持 30 个小节，提供 60 个任务考核步骤，支持自动检测步骤结果；课程章节包括深度学习概述、搭建简易神经网络、卷积神经网络、经典卷积神经网络、循环神经网络、基于 CNN 实现细胞识别分类等；

要求深度学习项目支持 5 个实训项目。包括猫狗识别、对话机器人、行为识别、Fashion-MNIST 服装分类、基于 Seq2Seq 实现文本摘要、使用 Fast R-CNN 进行目标检测等；

#### 18、Python 网络爬虫实训课程资源包

要求 Python 网络爬虫实训课程支持 10 个小节，提供 20 个任务考核步骤，支持自动检测步骤结果；课程章节包括 Python 环境搭建、XPath 解析、数据解析模块 Beautiful Soup、re 正则表达式模块、Ajax 数据爬取、Selenium 模拟浏览器、Scrapy 爬虫框架等。

#### 19、自然语言处理类

要求自然语言处理项目实战支持 5 个实训项目。包括基于逻辑回归进行 THUCNews 文本分类、基于贝叶斯模型进行 THUCNews 文本分类、基于 n-gram 模型的中文分词、使用 LSTM 对电影评论进行简单情感分析、Keras 文本分类实战、动手打造自己的输入法、机器人写唐诗、MNT 机器翻译框等；

#### ▲20、商业类综合实践课程资源包

##### 1) 视频网站访问量实时统计分析

数据量：

要求本数据为随机数据，需运行程序自动生成；

知识点：

要求包括 Flume、Kafka、HDFS 数据存储、FineBI 数据可视化操作；

数据字段：

要求包括 IP 地址、访问时间、访问类别、响应状态码、搜索方式、浏览器 User-Agent；

分析流程：

要求本项目包括以下分析流程：

(1) 大数据平台搭建：搭建项目所需 ZooKeeper、Kafka、Flume、HBase、Phoenix、Spark 环境。

(2) 数据生产模块：编写 Python 脚本模拟产生视频网站访问日志。

模拟数据发送 Kafka 主题实现消息缓冲。

(3) 数据采集模块：将实时产生的视频网站访问日志通过 Flume 采集到 Kafka 主题实现消息缓冲。

(4) 数据分析模块：利用 Spark Streaming 技术从 Kafka 获取日志进行清洗并完成业务统计，保存到 HBase 数据库。

(5) 数据展示模块：通过 Phoenix 对 HBase 表进行 SQL 化映射，然后使用 FineBI 大

数据分析工具连接 Phoenix 数据库实现数据可视化。

## 2) 高校数据统计分析项目案例

数据量:

要求提供不少于 100000 条教育网站高校和专业数据;

知识点:

要求包括 HDFS 数据存储、Hadoop 数据清洗分析、Mysql 数据存储、FineBI 数据可视化操作;

数据字段:

要求包括学校、省份、城市、地址、水平层次、办学类别、办学类型、985、211、双一流、归属、开设专业链接、专业类别、专业名称、国家特色专业;

分析流程:

要求本项目包括以下分析流程:

(1) 数据预处理: 对不规则数据进行规范化操作。

(2) 数据分析:

1) 统计各省份的高校数量, 并按照数量由大到小排列。

2) 统计全国计算机类(本)类别不同专业名称的总数量, 并按总数量逆序排列。

3) 高校国家特色专业数量占比统计, 占比数据为 Double 类型输出。

4) 高校水平层次整体占比, 占比数据为 Double 类型输出。

5) 统计全国高校各办学类别的总数情况。

6) 统计 985 高校各个省份分布数量。

7) 统计 211 高校各个省份分布数量。

8) 统计双一流高校各个省份分布数量。

(3) Mysql 数据处理: 本项目采用 Mysql 作为数据存储用来存储最终结果数据。

(4) Sqoop 数据抽取: 项目通过 Sqoop 将 HDFS 结果数据迁移到 Mysql 数据中, 数据分析过程包含排序操作使用排序后的结果数据进行 mysql 数据写入。

(5) 数据可视化: 通过 FineBI 实现各个需求可视化展示。

## 3) SoGou 日志数据分析

数据量:

要求提供不少于 1720000 条搜索引擎查询日志数据;

知识点:

要求包括 HDFS 数据存储、Mysql 数据存储、FineBI 数据可视化操作;

数据字段:

要求包括访问时间、用户 ID、查询词、该 URL 在返回结果中的排名、用户点击的序号、用户点击的 URL;

分析流程:

要求本项目包括以下分析流程:

(1) Flume 日志采集: 本项目通过 Flume 实时监听指定日志文件, 使用 Taildir Source 接受外部数据源, HDFS 作为 Sink。

(2) Spark 数据分析实现:

数据处理: 对不规则数据进行规则化处理。

需求指标 1: 用户 (UID) 访问频次统计。

需求指标 2: 网站 (URL) 访问统计。

需求指标 3: 用户搜索词访问统计。

需求指标 4: 不同排名访问统计。

需求指标 5: 统计搜索语句中关键词访问量。

需求指标 6: 统计各时间段（小时）用户访问量。

需求指标 7: 统计 URL 的最优 Rank 频率。

(3) Mysql 数据存储: 本项目采用 Mysql 作为数据存储用来存储最终结果数据。

(4) 可视化操作: 通过 FineBI 实现各个需求可视化展示。

4) 物联网日志数据分析项目案例

数据量:

要求本数据为随机数据, 需运行程序自动生成;

知识点:

要求包括 Kafka、HBase 存储、数据可视化;

数据字段:

要求包括机器标识 ID、设备类型、信号强度、时间戳;

分析流程:

要求本项目包括以下分析流程:

(1) 环境搭建: 搭建项目所需 JDK、Hadoop、Zookeeper、Kafka、Hbase、Phoenix 环境。

(2) Kafka 主题创建: 模拟数据发送 Kafka 主题实现消息缓冲。

(3) Phoenix 表创建: 数据结果存储于 HBase 中, 通过 Phoenix 进行 SQL 化映射以及连接可视化工具。

(4) 数据实时分析: 通过 SparkStreaming 实时分析 Kafka 数据写入 HBase。

(5) FineReport 可视化: 针对数据结果使用 FineReport 实现数据可视化。

5) 家用电器用户行为分析及事件识别

数据量:

要求提供不少于 18000 条用户用水记录数据;

知识点:

要求包括商务数据分析、数据可视化、机器学习、数据挖掘;

数据字段:

要求包括热水器编号、发生时间、开关机状态、加热中、保温中、有无水流、实际温度、热水量、水流量、节能模式、加热剩余时间、当前设置温度;

分析流程:

要求本项目包括以下分析流程:

(1) 数据收集: 从文件导入案例数据;

(2) 数据处理: 数据字段类型转换, 衍生变量, 数据分组, 特征选择;

(3) 数据可视化: 用水事件阈值;

(4) 数据建模: 神经网络模型进行参数寻优对用户是否洗浴进行预测, 并判断预测结果是否准确。

6) 用户留存率分析

数据量:

要求提供不少于 300 条观测个案数据;

知识点:

要求包括留存率、非线性回归分析、数据建模、数据可视化;

数据字段:

要求包括用户 365 天编号、产品类型 1、产品类型 2;

分析流程:



要求本项目包括以下分析流程：

(1) 数据读取：读取产品留存率数据，查看数据结构和信息；

(2) 数据探索：绘制留存率曲线图查看产品中的类型 1 和类型 2 随时间的变化情况；

(3) 数据建模：利用 nls 函数求出幂指数函数  $y=a*x^b$  的系数 a、b，建立产品类型 1 和类型 2 的用户留存率预测模型；

(4) 模型预测：对类型 1、类型 2 新增用户在接下来 365 日每天的留存率进行预测，并查看预测结果。

(5) 模型调整：类比回归分析加权重系数的方法，添加 T 值对预测值进行调整。将预测值乘以对应的 T 值，便可得到调整后的预测值；

(6) 数据可视化：绘制对应的留存率预测对比图，以观察两者差异。

#### 7) 营销响应模型

数据量：

要求提供不少于 1000 个客户信贷样本数据；

知识点：

要求包括数据分析、数据挖掘、机器学习、数据可视化；

数据字段：

要求包括 1900 个匿名字段；

分析流程：

要求本项目包括以下分析流程：

(1) 数据读取：读入数据，使用 subset 函数删除不必要的变量，统计数据集总的样本数和字段数，随机抽取 100 个数值型字段，设定种子；

(2) 数据清洗：对数据缺失值、重复值进行处理，分离并统计数值型和字符型字段，分离出日期字段，并转化成日期类型。

(3) 数据可视化探索：绘制抽取出的 100 个数值型字段的相关性系数图，分别采用皮尔逊和斯皮尔曼方法计算相关性系数，绘制相关直方图；

(4) 数据建模：建立分类模型，将数据集分割为训练集和测试集，训练梯度 boosting 模型，测试模型性能，得到 ROC 曲线。

#### 8) 汽车燃料效率分析

数据量：

要求提供不少于 39000 条汽车燃料的相关测量数据；

知识点：

要求包括数据清洗、相关分析、时间序列、数据可视化；

数据字段：

要求包括 barrels08 燃料桶年石油消耗量 1 型、barrelsA08 燃料桶中的年石油消耗量类型 2、charge120 在 120V 下为电动汽车充电的时间（以小时为单位）、charge240 在 240V 下为电动汽车充电的时间（以小时为单位）、city08 城市燃料 MPG 类型、city08U 用于燃料的非圆形城市 MPG 类型、cityA08 城市燃料 MPG 2 型；

分析流程：

要求本项目包括以下分析流程：

(1) 数据读取：导入数据，初步认识数据集中的变量情况对数据进行检查；

(2) 数据探索：查看汽车的燃油类型，并以传动方式变量对汽车类型进行分类；

(3) 相关分析：发现近年来汽车销售的趋势或规律，使用汽车燃料效率特征来对数据做分组汇总统计，观察哪组汽车燃料利用率更高，并且可以看到这些数据随时间变化的趋势；

(4) 数据可视化：查看近年来销售的汽车的燃油经济性是否有显著增长，平均 MPG 是否随着时间有一个趋势上的变化，查看 4 缸发动机汽车在不同传动方式下的 MPG 随年份的变化。

#### 9) 房价指数分析预测

数据量：

要求提供不少于 1400 组房价训练数据，1400 条房价测试数据；

知识点：

要求包括多重插补、时间序列、数据挖掘、机器学习；

数据字段：

要求包括

(1) 房屋本身的相关物理特性相关特征（如总体材料和加工质量）

(2) 房屋面积相关的特征（如地下室总面积）

(3) 房屋位置相关变量（如地理位置）；

分析流程：

要求本项目包括以下分析流程：

(1) 数据读取：分别导入数据集为后续分析做准备；

(2) 数据清洗：对缺失的数据进行统计，计算各字段含有缺失值的比例，并对各特征数据缺失做一个可视化，将缺失值进行多重插补；

(3) 相关分析：将数据集中的字符串型和数值型分开，提取数值型变量做相关性分析，给出相关度高于 0.5 的变量名称；

(4) 筛选重要变量：设置种子，通过随机森林筛选出对于因变量最重要的前 20 项；

(5) 探索性分析：对目标变量和重要性变量进行探索性分析，分别绘制特征图；

(6) 数据建模：根据各特征的分布图易得数据的偏度都比较大，并且曲线比较复杂，这里进行偏度修正，近似于正态分布，并且对数据做标准化处理，对因子型数据做独热编码处理，使用 lasso 回归建模，用 10 折交叉验证来调整参数  $\lambda$  以尽量取得最小的 RMSE 值，最终获得房价预测表。

★21、其他要求：供应商提供能够满足 90 人同时在线使用的云资源计算设备 4 台（提供承诺函）

2

大数据  
分析应  
用实验  
平台

#### ● 教学平台功能

1、平台需具备数智化角色工作台功能，该功能提供不少于三类工作台，并依据角色提供智能服务，方便院校不同角色通过使用场景使用相应服务。应包含学生工作台、教师工作台、管理员工作台，并至少包含学生的待办事项，结合当前已开设、正在开设课程项目进行大数据智能推荐提升学生自主学习意识。平台不限制用户数，支持至少 150 人同时在线使用。

▲2、应具备灵活的教学设计服务，教师可在课堂中根据自己的实际情况，实时调整课堂内容，至少包含 5 类不同的教学组件用于教师自定义添加和调整。理论项目支持新增视频任务、文档任务、富文本、网页链接，实战项目支持新增作业任务、测验任务、团队成果；作业任务中，可支持修改作业名称、作业描述、评分方式等，支持上传参考资料与参考答案，设置答案可否下载，评分方式为系统评分可设置关键词；团队成果中，可支持修改团队成果名称、要求描述、评分方式，参考答案和参考模版，设置答案可否下载，评分方式可多选教师主管评分、组间互评、组内互评；随堂测验中，支持修改测验名称、答题限时、页面输入或批量导入测验题；支持添加删除题目、答案、答案解析、勾选正确答案；修改题目排序，保存随堂测验；视频任务中，可编辑任务名称、上传、替换视频，增加视频测验；文档任务支持修改任务名称，重新上传文档，覆盖之前文档；富文本任务支

持修改标题、内容；网页链接任务支持编辑标题和链接。（提供软件功能截图）

▲3、应具备完善的教学过程管理，包含不少于6种教学过程管理功能且不少于5类过程数据。可以实现基于考勤、进度、作业、团队成果、文案任务、项目报告等角度进行过程管理，同时支持学习进度、实战次数、平均学习时长、作业提交进度、团队成果提交进度等。（提供软件功能截图并加盖投标人公章予以佐证）

★4、课程应具备课程简介、本章简介、训练计划、能力项、思维导图、知识图谱、公告等信息，课程内项目应包含名称、难度、任务完成度等信息。（提供软件功能截图）

5、应具备课外学习项目库。学生可在项目广场挑选项目进行学习，应支持不少于5个维度自主筛选项目。筛选维度可包含项目难度、项目技术领域、项目业务领域、项目案例领域、项目类型等。（提供软件功能截图）

☆6、应具有能力分析模块，能够生成学生能力报告，包含不少于5项能力指标。包括：在当前教学班中，应通过折线图展现学生个人的能力成长值，通过柱状图学生不同的工作职责下的初级中级高级能力成长情况，通过雷达图展现工作胜任力完成情况，通过可视化的方式展示工作胜任力得分情况，以及工作胜任力掌握情况（热度）；（提供现场演示）

▲7、应具备多类型教学辅助工具，包含不少于10种教学组件且不少于5种教学互动工具，支撑课堂实践教学以及师生互动。支撑的教学组件类型可包含：文档、视频、富文本、互动式课件、随堂测验、作业、团队成果、仿真试题任务、VR任务、文案任务、实践任务等；教学互动可包含：投票、讨论、问答、弹幕、笔记等，互动过程中系统可自动屏蔽部分网络非文明用语，以\*号显示；（提供软件功能截图并加盖投标人公章予以佐证）

8、教学任务推送具备自由、闯关两种模式。其中，通过闯关模式能有效控制教学过程，任务推送后学生需逐一完成教学任务。

9、支持教师个性化教学。可在教学班内编辑课程，包含但不限于编辑课程基本信息（如编辑课程名称、课程标签、课程概述等）、编辑项目信息（项目简介、项目任务添加删除与修改、项目顺序调整等）、添加删除项目（理论项目、实战项目等）等，其中，可添加或修改的项目任务至少包含8类不同类型的任务。（提供软件功能截图予以作证）

## 二、数据分析系统功能指标

1、消息中心：消息中心能够查看平台所有消息的记录，默认按时间倒序排列。数据连接：支持用户连接各种数据源，拖拽方式搭建表关系，设置筛选条件，建立数据模型；支持常见的关系型数据库MySQL、Oracle、SQL Server、JDBC数据库驱动、ODBC数据库驱动，通过配置实现数据的连接和表的导入，与权限结合，满足数据权限管理的需求；支持本地数据Excel或csv文件上传。

2、数据整理：据整理模块支持用户通过拖拽的方式，所见即所得的进行数据的整合和清洗，同时支持用户通过sql生成复杂的数据集；创建的数据集可以是实时从数据库获取数据，同时支持设置不同的物化规则，把分析数据按照业务规则抽取到执行库中，保证分析数据的执行效率。

3、数据关联：支持对多个数据源，包含企业数据库、填报数据、本地数据（Excel、CSV）的不同表之间设置关联；支持拖拽方式搭建表间关联（内关联、左关联、右关联、全关联）关系；支持拖拽方式追加多张表，构建union数据集；支持多表关联时自动和手动对同名字段进行匹配。

▲4、数据筛选：用户可以通过筛选对数据集进行查询和筛选，支持用户通过多个条件筛选和公式筛选进行数据过滤；通过条件筛选时，选择筛选字段，自动根据字段类型可以选择操作符，筛选值中支持选择动态变量（包含用户名、租户名称、显示名、部门、角色、手机、部门电话）或者枚举值。（提供软件功能截图并加盖投标人公章予以佐证）

▲5、数据物化：支持用户选择实时连接或者物化，若选择物化，则数据分析系统定时从各个业务数据库把数据抽取到数据分析系统，并按照一定频率进行更新；基于物化创建的嵌套数据集自动更新，若底层数据集更新，则与之相关的上层数据集都依次更新，进而可视化和故事板的数据更新；数据物化支持用户设置物化频率到天、时、分、秒，设置物化开始时间和截止时间；同时显示刷新时间，用户可以自动或者手动去刷新数据。（提供软件功能截图并加盖投标人公章予以佐证）

6、内置可视化图表：目前内置有 34 种图表，包括柱状图、阶梯瀑布图、条形图、图案填充条形图、堆叠柱状图、堆叠条形图、饼图、环形图、玫瑰图、折线图、堆叠区域图、中国地图、散点中国地图、世界地图、迁徙图、百度地图、双轴图、仪表盘、表格、指标图、气泡图、漏斗图、热力图、矩形树图、雷达图、桑基图、和弦图。

★7、辅助线和预警线：用户通过辅助线可以快速识别出数据中的异常值；预警线除了识别出异常值，还支持用户设置预警的级别和预警推送的对象，不同层级对象对于同一个指标可以设置不同的预警规则；预警级别三级，不同的预警级别可以设置不同的信息推送媒介，包含系统信息、短信、微信服务号、邮箱，用户可以自己设置。（提供现场演示）

8、故事板（可视化看板）：支持用户选择可视化、创建综合分析界面，设置联动、链接；选择多个可视化创建故事板，把可视化对象在故事板页面中拖拽布局，支持故事板页面分页；支持页面灵活布局，拖拽位置和可视化对象大小；支持设置展示属性，添加全局筛选、文本、图片。

9、故事板控件：包含过滤器、层级过滤器、文本、图片、嵌入 URL、标签控件；支持对多个可视化进行过滤筛选，支持创建数据层级并通过树形对可视化对象进行筛选，支持输入文本文字和插入计算字段的方式形成智能报告，文字和计算字段的样式等均可以设置，支持插入图片，支持 png、jpg、gif 等常见图片格式，支持输入 URL 嵌入第三方界面，通过标签可以实现从当前故事板跳转到其他故事板或者第三方界面。

▲10、函数与变量：在计算字段中可使用字符串函数 10 种、日期函数 7 种、数学函数 11 种、上下文函数 2 种、判断函数 2 种；通过函数与变量可进行诸多复杂数据清洗、数据集成。

### 三、代码编辑器功能指标

1、自主研发，属于国产产品，内置 180 个以上常用库，上课时可直接调用，免去繁琐安装环节；内置常用库包括：pandas, matplotlib, pandas-datareader, scipy, PyMySQL, snownlp, gensim, pytest, mlxtend, pyfolio, turtle 等。

★2、支持“代码模式”，可自定义编写 Python 代码并运行，包括但不限于数据清洗、数据集成、数据可视化、数据转换、数值模拟、统计建模、机器学习等功能；代码模式分为 4 块区域，引导说明区、目录文件操作区、文件操作区、运行结果展示区；支持文件下载功能；文件操作区包含运行脚本、终止运行脚本、重置并恢复预置脚本、清空当前脚本、保存脚本、提交脚本功能。

3、支持“积木模式”，积木模式包含显示字符串内容、数字变量、字符变量、列表类型变量、元祖类型变量、字典类型变量、集合类型变量、数学运算、数字格式、数字类型转换、字符串与字符串拼接、变量与变量拼接、输出列表变量、输出元祖变量、输出字典变量、数据变量、代码注释积木等应用；

▲4、内置支持预览 CSV 文件、图片（PNG、JPG、GIF 等）与在线预览 HTML 文件。编辑区可根据对应文件类型自动代码提示、代码高亮。可对全部文件搜索。支持任务描述、Markdown 操作步骤、操作视频、参考答案的功能展示，并可以自由切换及关闭；支持多页签显示脚本；持响应式交互，运行 input 函数等；支持运行 SQL 数据库；支持左右屏展示操作步骤及代码编辑区域，并可拖拽调节左右区域位置大小。

#### 四、课程模块功能指标

##### (一) 金融大数据分析课程平台 1 套

###### ▲1、商业银行理财产品分析：

- 以某商业银行为背景，支持对商业银行理财产品进行经营分析、客户分析、产品分析以及购买分析。
- 经营分析支持通过代码编辑器对上市银行企业的经营数据采集，支持通过分析工具实现数据关联、经营指标分析、银行网点分布、经营趋势分析、同行业经营情况对比分析。
- 在客户、产品和购买分析中，支持通过分析工具，采用 RFM 模型、帕累托分析、漏斗模型，实现模型的构建及可视化分析。

###### ▲2、中小微企业信贷决策：

- 以“发票贷”业务作为背景，利用中小微企业相关进项、销项发票数据，从数据理解、构建评价指标、信贷评级完成中小微企业信贷决策。
- 基于进项、销项发票数据，支持从企业规模、经营成果、盈利变化和议价能力 4 个方面构建中小微企业评价指标。
- 利用客观赋值法（熵权法）确认评价指标权重，利用综合评分法模型，准确、高效的挖掘、识别出优质信用公司，进而完成对案例当中的中小微企业进行评级及授信。

###### ▲3、投融资大数据分析：

- 内置家电行业的投融资数据，通过分析工具，完成行业分布情况、投融资事件行为分析、投资趋势分析实践任务。
- 行业分布情况分析提供数据拆分、行业分布分析、词云图分析实践任务，实现对智能家居细分行业进行深入分析。
- 投资事件分析提供地区分布、投融资规模、投融资阶段实践任务，实现对家电行业投融资事件进行深入分析。

##### 4、第三方支付风控：

- 包含不限于小微商户、服务商、客户经理、风控经理等角色。
- 第三方支付申请及审核提供客户经理构建入网策略、小微商户产品申请、渠道商商户审核。
- 构建风控策略任务包括不限于异常转账、异常提现、盗卡转账、限额限笔、异常支付（快捷支付、异常流失客户支付、突增判断、高频整数交易、异常关联交易）等。
- 提供风险大盘功能，包含不限于各项事件总笔数、通过商户数量、人工审核数量、拒绝商户数量。

##### (二) 营销大数据实验平台 1 套

###### 1、消费者行为分析

- 在支持开展基于企业视角的消费者行为洞察分析，在了解项目情境的基础上，进行项目数据准备、消费者画像、消费者价值分析、消费者购物篮分析、消费者行为分析报告撰写的多维度营销大数据分析训练。预置企业真实案例背景信息，让学生清晰并明确项目目标。
- 预置真实企业脱敏数据，包括消费者信息数据、商品数据、电商渠道销售订单数据、门店渠道销售订单数据等多维度数据，数据指标不少于 20 个，数据量不少于 29,000 条。
- 支持应用数据分析系统、代码编辑器等工具进行消费者画像分析、消费者 RFM 分析，以及数据可视化；代码编辑器内预置可实现 RFM 模型分析的 Python 代码，支持一键运行；支持对预置 Python 代码的修改、清空、重置、保存等功能。
- 支持自主设计、构建分析模型，应用数据分析系统、代码编辑器等工具完成模型计算。
- 支持应用代码编辑器完成消费者购物篮分析，完成商品关联营销实施方案；代码编辑

器内预置可实现关联规则分析的 Python 代码，支持一键运行；支持对预置 Python 代码的修改、清空、重置、保存等功能。

- 预置消费者行为分析报告模板，支持参照模板完成分析任务，提出营销建议。

### ▲2、商品价格分析

- 在支持开展基于企业视角的商品价格分析报告制定，在了解项目情境的基础上，进行项目数据准备、商品数据分析、竞品价格分析、商品个性化定价策略、商品定价策略分析报告撰写的多维度营销大数据分析训练。

- 预置真实企业脱敏数据，包括销售订单数据、消费者信息数据、商品信息数据、商品成本数据、消费者线上浏览及行为数据、竞品商品数据等多维度数据，数据指标不少于 40 个，数据量不少于 24,000 条。

- 支持应用数据分析系统、代码编辑器、EXCEL 等工具进行商品毛利分析、商品 ABC 分析，定位重点分析商品。

- 支持通过爬虫代码抓取电商仿真网站中的竞品数据信息；支持应用代码编辑器完成爬取数据清洗；代码编辑器内预置可实现爬虫功能的 Python 代码，支持补齐代码后一键运行；支持对预置 Python 代码的修改、清空、重置、保存等功能。

- 电商仿真网站：模拟主流电商交易平台首页、服装类商品列表页和商品详情页；支持按商品分类、商品品牌、价格区间查询或筛选；商品详情页面能够动态显示商品主图、名称、商城价、券后价、尺寸、库存数量、月销量、商品评价以及详情属性等信息；系统内置丰富的服装商品数据，并支持教师依据课程设计自主批量导入、上传、替换、维护商品数据，且可以自主选择自建数据方案、预置数据方案或混合数据方案，满足个性化教学设计创新需求；网站商品数据支持学生运行 Python 代码完成数据爬取，避免使用真实网站带来责任事故。主数据源包含商品类别 15 类以上，商品数据 15000 条以上；数据字段 100 以上。（提供软件功能截图予以作证）

- 支持应用代码编辑器实现基于客户支付意愿的“遗传算法”计算，得出商品个性化定价策略；代码编辑器内预置可实现遗传算法的 Python 代码，支持一键运行；支持对预置 Python 代码的修改、清空、重置、保存等功能。

### ▲3、广告精准营销分析

- 在支持开展基于企业视角的广告精准营销，了解项目情境的基础上，进行项目数据准备、广告投放数据预处理、广告投放效果分析、会员数据分析、目标用户画像分析、竞品数据分析、广告优化试投分析、广告精准营销分析报告撰写的多维度营销大数据分析训练。

- 预置真实企业脱敏数据，包括店铺网站流量监控数据、微信端投放结果数据、微博粉丝通投放结果数据、小红书投放结果数据、抖音投放结果数据、会员画像数据、会员积分数据、目标人群标签爬取数据、竞品评价爬取数据、竞品传播数据、广告盈亏平衡分析数据等多维度数据，数据指标不少于 110 个，数据量不少于 88,000 条。

- 支持应用数据分析系统、代码编辑器完成用户画像分析及数据可视化；代码编辑器内预置可实现用户画像分析的 Python 代码，支持一键运行；支持对预置 Python 代码的修改、清空、重置、保存等功能。

- 支持应用代码编辑器完成竞品数据分析、竞品文本数据（非结构化数据）分析；代码编辑器内预置可实现竞品数据和竞品文本数据分析的 Python 代码，支持一键运行；支持对预置 Python 代码的修改、清空、重置、保存等功能；支持应用数据分析系统完成竞品传播数据分析。

- ABTest 广告优化策略仿真系统：模拟头部社交平台 ABTest 实验流程，支持实验模式与实验目标设定、创建广告实验基础信息；通过设置预算与出价、选择投放位置、创意制作等建立广告投放实验对照组；可以按素材定向、人群定向、预算定向三种方式进行效果

实验，建立对应实验组；依据实验情况可以动态生成实验报告，实验报告包含对照组和实验各组的进组用户数、进组占比、广告点击率、差异绝对值、差异相对值、置信区间、P-Value 值等数据，以及根据实验数据自动生成的对照组和实验各组概率分布正态图。（提供软件功能截图予以作证）

#### 4、网络促销分析

- 在支持开展基于企业视角的网络销售策略，在了解项目情境的基础上，进行项目数据准备、双十一总体情况预测、双十一重点营销客户画像、双十一流量分析、旗舰店大促流量策略、促销组合策略、网络促销分析报告撰写的多维度营销大数据分析训练。

- 预置真实企业脱敏数据，包括历年双十一平台成交数据、企业历史订单数据、店铺运营流量日报数据、双十一在售商品数据、促销活动结果数据、单品预热期订单数据等多维度数据，数据指标不少于 40 个，数据量不少于 414,000 条。

- 支持应用代码编辑器，结合历年双十一成交数据，实现双十一总销量预测及数据可视化；代码编辑器内预置可实现销售预测的 Python 代码，支持一键运行；支持对预置 Python 代码的修改、清空、重置、保存等功能。

- 支持应用数据分析系统，实现双十一重点营销客户画像分析，明确重点营销人群。支持应用数据分析系统，实现双十一流量结构分析、免费流量分析和付费流量结构分析，确定促销流量策略。

- 支持应用数据分析系统，实现预热期的购物车转化率和预热期销售占比分析及数据可视化，得出目标销量下的大促流量需求。支持应用数据分析系统，实现促销组合效果分析，完成组合促销决策。

- 审计大数据实训平台 1 套

##### ▲1、RPA 设计器功能指标

(1) RPA 设计器：通过 RPA 开发能实现对各种应用的自动化，可以在无需改造原有系统的基础上，依据设定的程序完成特定的工作任务，替代繁琐和重复的事务性工作。

(2) RPA 设计器内置 9 大类，37 种以上插件，300 个以上命令，拥有流程设计、数据操作、图像识别、用户交互、文件操作、Office 读写、网络服务、邮件操作、浏览器操作、ERP 类操作等插件命令，可使用各种命令组合，便捷高效地设计机器人流程模板。（提供软件功能截图予以作证）

(3) 平台无缝集成：通过教学管理平台中的教学任务可直接跳转到 RPA 设计器登录页面。

(4) 国产化：中国自主品牌的 RPA 开发软件，符合信创国产化趋势，能够满足院校教学需求。

2、文本分析在审计中的应用。通过项目引例与知识背景，让学生体会关注舆情的重要性，了解舆情的内涵及舆情与审计的关系。导入项目案例，引导学生基于审计目标，通过业务需求与技术需求的梳理，设计项目执行方案，运用大数据技术得出项目分析结果。本项目结合文本挖掘技术的特点及使用方法，基于案例真实应用场景介绍非结构化数据的处理，熟悉 jieba 分词与人工分词的区别，能够对主题分析和语义网络进行结果解读。同时利用文本挖掘法中的词频统计和语义网络结构分析对股民评论进行分析，梳理案例中股民的观点、意见和态度。使用可视化技术实现舆情分析可视化报告的设计与开发，对外部舆情相关数据进行处理后，进行可视化展现，包括折线图、柱形图、K 线图、词云图、热力图及可视化大屏等，依据可视化报告完成舆情报告。通过真实的审计案例，培养学生对非结构化数据进行挖掘与分析的能力。

3、聚类分析在审计中的应用。在风险评估阶段，应用大数据技术辅助完成被审计单位综合风险评估。本项目应用聚类算法（K-means 算法、层次聚类），利用数据挖掘技术

建立模型，快速把财务指标相似的上市公司进行分类，定位被审计单位的可比公司，提高分析效率。通过可视化技术与指标对比分析、趋势分析等分析方法的结合，实现对行业数据、可比公司数据、被审计单位数据的综合分析。培养学生设计大数据应用方案的能力，应用大数据技术生成指标风险评估报告的能力，判断重大错报风险的能力，指引进一步审计程序。

4、**社会网络分析在审计中的应用。**在风险应对实质性程序阶段，应用大数据技术辅助完成收入实质性程序。通过案例导入，引导学生基于审计目标，进行业务需求与技术需求的梳理，设计项目执行方案，运用大数据技术得出项目分析结果。应用数据处理与分析技术及RPA技术，对销售订单详细信息进行全量核查，包括销售政策复核、装运规则复核、物流信息复核等。应用社会网络分析技术对销售存疑客户进行关联关系分析，得出审计结论。培养学生设计大数据应用方案的能力，应用大数据技术发现审计疑点、提升审计效率、降低审计风险的能力。

5、**自动化技术在审计中的应用。**在风险应对实质性程序阶段，应用大数据技术辅助完成应收账款实质性程序。通过项目引例与知识背景，让学生体会函证程序的重要性，了解区块链技术及自动化技术在审计中的应用。通过案例导入，引导学生基于审计目标，进行业务需求与技术需求的梳理，设计项目执行方案，运用大数据技术辅助完成相关审计工作底稿。应用Python自动化技术及RPA技术，自动生成函证。通过Python调用API接口，获取被函证企业地址信息，自动完成函证地址信息复核。通过Python调用API接口，获取客商工商综合信息，进行全量客商信用风险信息复核。培养学生设计大数据应用方案的能力，应用大数据技术进行全量审计，提升审计效率、降低审计风险的能力。

6、**图像技术在审计中的应用。**通过项目引例与知识背景，让学生了解智能分析技术在审计中的应用。通过案例导入，引导学生基于审计目标，进行业务需求与技术需求的梳理，设计项目执行方案，运用大数据技术辅助完成制度执行情况审计。应用图像定位技术、图片识别技术，检查被审计主体存货盘点制度执行情况，识别未严格执行存货盘点计划的主体。培养学生设计大数据应用方案的能力，应用大数据技术发现审计疑点、获取审计证据的能力。

▲7、**异常值检测在审计中的应用。**通过项目引例与知识背景，让学生了解舞弊的内涵及员工舞弊的常见类型。导入项目案例，引导学生基于审计目标，通过业务需求与技术需求的梳理，设计项目执行方案，运用大数据技术辅助完成内部舞弊审查。本项目应用KNN算法，利用数据挖掘技术建立模型，快速识别异常交易，定位异常主体，提高分析效率。通过数据分析与处理技术，建立自动化检测模型，建立常态化预警监控。培养学生设计大数据应用方案的能力，应用大数据技术建立分析模型的能力，判断重大错报风险的能力，指引进一步审计程序。以某集团内部审计为案例背景，针对某分公司物流成本偏高问题进行调查分析。通过业务场景与业务数据的导入，引导学生设计异常值检测方案及业务数据分析模型，并启发学生逐步丰富完善模型。应用大数据技术实现模型自动化，并固化模型，复用模型，快速发现舞弊行为，提高审计效率，改善企业经营管理水平。（提供软件功能截图予以作证）

★8、**其他要求：**供应商能够提供审计大数据实训平台配套教材编写服务，保证该教材能够成功出版；能够提供支撑150人同时在线使用的云资源计算设备2台。（提供承诺函）



3	管理控制设备	<p>平台核心设备，通过管理控制设备实现平台的统一管理功能；</p> <p>1、配置不少于 2 颗 CPU、核心数量不低于 20 核心、40 线程，CPU 主频不低于 2.0GHz；内存不小于 320GB；硬盘采用 NVME 固态硬盘、容量不低于 1.9TB；</p> <p>2. 要求支持通过可视化界面可管理和监控各计算节点的计算服务，支持选择关闭、开启服务。要求支持对主机进行编辑、删除、更新元数据；</p> <p>3. 要求虚拟化平台支持虚拟机镜像及容器 Docker 共同管理和使用，并支持 Docker 与虚拟机同属同一子网，IP 地址段为同一地址段，网络互通；多台虚拟机或容器可以互通；要求支持通过界面化功能，对虚拟化平台的所有镜像进行管理，包括镜像的上传、启动、编辑和删除；要求支持通过界面化功能，对虚拟化平台的所有容器进行管理，包括容器的创建、更新、暂停和删除；</p> <p>4. 供应商免费负责所有软硬件安装并做好与现有设备的电源、网络等连接。</p>
4	云资源计算设备	<p>平台核心设备，通过云资源计算设备实现实训项目计算支撑功能；</p> <p>1、配置不少于 2 颗 CPU、核心数量不低于 20 核心、40 线程，CPU 主频不低于 2.0GHz；内存不小于 320GB；硬盘采用 NVME 固态硬盘、容量不低于 1.9TB；</p> <p>2. 要求计算节点支持队列处理，如果瞬时大量请求，支持计算节点自动启动排队访问功能；要求计算节点具备负载均衡功能，多个云资源计算设备支持负载运行；</p> <p>3. 要求单台计算节点出现故障后，其他计算节点可快速衔接该计算节点的业务；要求支持设定虚拟配置与真实设备配置配额比例，以便当系统临时提高需求时快速增加虚拟配置；</p> <p>★4. 供应商负责所有软硬件安装并做好与现有设备的电源、网络等连接。</p>
5	云接入设备	<p>1. <math>\geq 24</math> 个 10/100/1000 Mbps 电口；支持 VLAN 功能；静音交换机；支持三层管理</p> <p>2. 带宽：<math>\geq 8.8</math>Gbps（无阻塞）；包转发率：<math>\geq 6.5</math>Mpps；缓冲区内存：<math>\geq 256</math>KB；寻址：<math>\geq 48</math> 位 MAC 地址；地址数据库大小：每系统 8000 个 MAC 地址；</p> <p>平均无故障间隔：不少于 300000 小时。</p> <p>★3. 供应商负责安装并负责做好与现有设备的电源、网络等连接。</p>
6	机架装置设备	<p>1、尺寸：宽度 600mm，深度 900mm；机柜高度 2000mm，42U；</p> <p>2、前门：钢化玻璃材质，带锁，可拆卸；后门：全金属材质，散热网孔设计，带锁，可拆卸；侧板：速装卸侧板，全金属材质。</p> <p>★3、供应商负责安装并负责做好与现有设备的电源、网络等连接。</p>