



*CDA*

**大数据实验室**

**介绍**

## CDA 大数据实验室介绍

三年来，共有 283 所高校获批数据科学与大数据技术专业专业，虽然很多高校已经申请并获批大数据专业，也有大量院校计划申请该专业，计划进行大数据学科建设，但是在全国范围内还有很多院校无法开设大数据相关课程，培养大数据分析人才。目前建设大数据实验专业最广泛且有效的方式便是校企合作。

依据大数据发展的现状和机遇，响应产、学、研、用协同教学的创新模式，CDA 数据分析师特提出本实验室建设方案。本实验室在技术层面紧贴大数据发展的前沿，采用中国国内先进的大数据设备和软件开发平台组成。能够给学生提供一个真实的大数据应用实训环境，使学生在大数据系统搭建、数据分析、数据处理、平台数据维护和优化、行业数据应用方面等一系列学习与实际工作接轨，便于熟练掌握各种海量数据的存储、处理和应用能力，积累实际工作所需技能和经验。

- 产

CDA 利用多年积累的资源，将企业实际项目问题以案例的形式加入到大数据实验室中，再进行实战案例教学的同时，CDA 给高校师生提供项目服务和企业咨询的机会。

- 学

结合 CDA 多年的大数据教学经验和教学资源，通过教材、专业方案、讲师培养等模块以大数据实验室的形式分享到高校教学中。

- 研

CDA 大数据实验室通过技术创新，大大降低了大数据的门槛，不仅方便教学，同时也给高校的科研提供支持。

- 用

通过 CDA 数据分析师、高校、企业三方的共同合作，将师生的研究成果投入到企业实际应用当中。

为高校解决如下四大痛点：

- a) 学习门槛高：数据科学是一门交叉学科，计算机知识+数学统计基础+熟悉行业背景=数据分析人才的学习模式无疑使得学生无从下手。大数据的技术门槛较高，需要对计算机、统计、数学、行业等多个专业了解，老师任务重，开发课程周期长。

- b) 实验环境缺失：目前大数据的教学大多数基于理论，若要实操，需搭建集群、分布非常复杂，传统机房难以支持。
- c) 可借鉴经验不足：大数据行业兴起不久，数据科学是新兴学科，今年是数据科学与大数据技术本科专业获批的第三年，大数据人才培养体系不完整，绝大多数高校都没有成熟的数据科学人才培养方案，高校对于人才培养方案以及企业用人需求的可借鉴很少，研讨和尝试十分费力。
- d) 动手机会少：高质量的数据和适用于大数据场景的平台都是进行大数据人才培养的必要条件，然而，多数高校缺少企业项目实战案例和可以用于研究的商业数据，使得学生难以做到学以致用。部分高校缺乏企业项目实战案例，要做到学以致用是大数据教学中的又一大难点。无真实数据难以模拟教学。

CDA 大数据实验室包括 CDA 大数据实训平台和 CDA 大数据教学管理系统两部分。



CDA 大数据教学实训平台涵盖自主式实训、向导式实训、数据采集、教师管理系统、答疑平台、考试测评、视频教学 7 大模块。用户使用拖拽式操作即可完成实验，极大程度的降低了大数据分析门槛，使没有编程基础的用户也可以进行大数据分析。通过实训平台，教师可以查看学生的实验报告，可以对学生进行考试测评，也可以让学生自主学习视频。

- 自主式实训

自主实训平台供老师进行开放式教学、学生进行自主式学习，该平台中包括数据探索和工作流 2 种实验类型。学生可以自由查询任务执行状态，查看实验数据、实验日志和实验结果。

数据探索中包含数据源、数据处理、统计分析、机器学习、数据导出、可视化等功能组件，可以实现聚类、分类、回归、推荐等机器学习算法的可视化使用，这些都是已经封装好的可以直接使用的组件，有助于老师教学和学生学习。

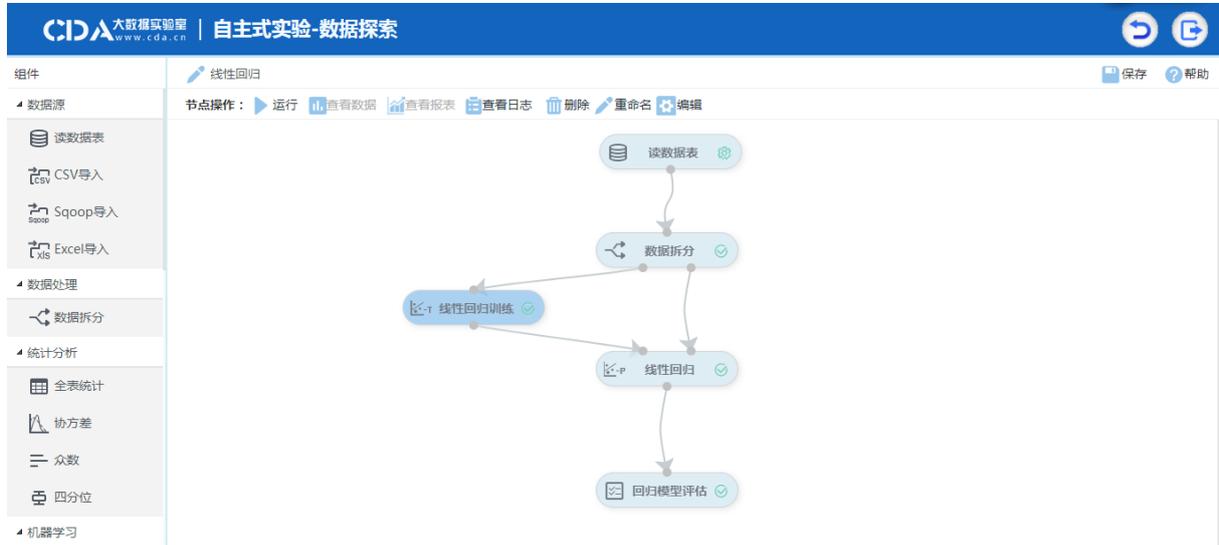


图 数据探索

workflow 主要用于自主编程实验，包含了流程控制、通用处理、导入导出等功能组件，在工作流中可以实现 Hive、Spark、MapReduce 等自主式代码编程，以及当前流行的 Sqoop 导入/导出代码实现。

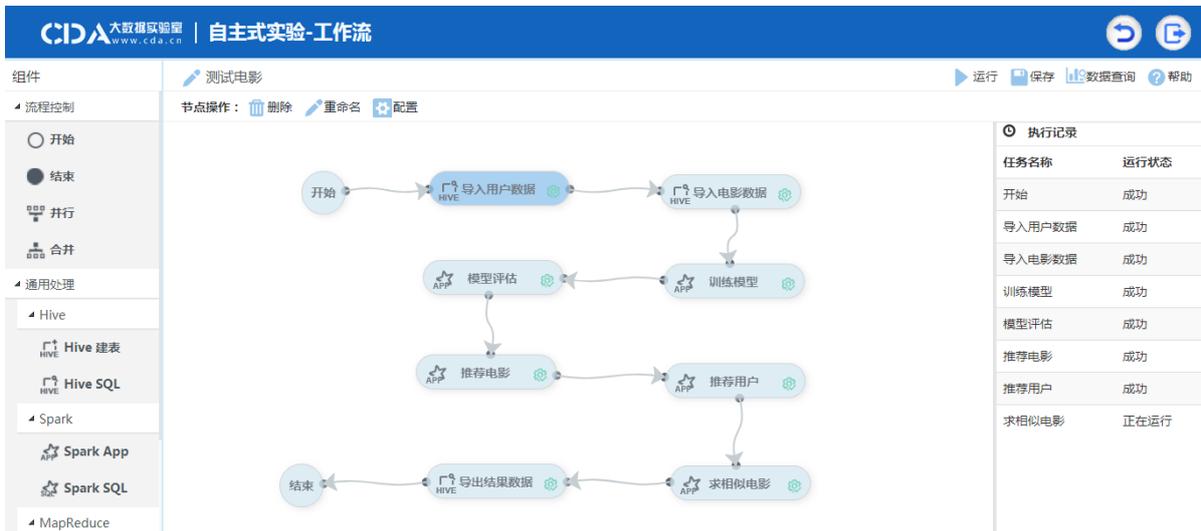


图 工作流

- 向导式实训

向导实训平台中包含了许多企业真实大数据项目案例，我们将这些案例按步骤拆开，代码内置于模块中，每个实验都能成功运行，并且配备了完整详细的实验前导、教学内

容、操作步骤说明。学生可以在本模块中独自学习商业大数据案例，学习每个知识点的使用以及一套完整的大数据分析流程，感受不同行业的大数据应用环境。学生可以对每个实验的操作过程进行修改，因为我们在其中设置了重置实验功能，随时可以还原到任务的最初状态。并且随着和学校的不断合作，这些企业案例库会越来越丰富。



- 数据采集

在数据采集模块，勾选想要爬取的数据源，填写搜索关键词和采集数量，点击开始采集，即可实现关键词的定向爬取，获得想要的数据库。爬虫功能可以解决高校老师或学生手头缺乏可用的有效数据的问题，同时爬虫功能还可以激发学生的学习兴趣，使得学生独立进行学习和提升。

- 答疑平台

答疑平台可以有效地解决学生在学习中遇到的各种难题和困惑。公司旗下的经管之家现在已经成为国内最活跃和最具影响力的在线交流平台，每天都有 50-60 万的访问量，在这里，不管是商业数据分析，还是大数据技术，更甚者是当前人工智能等机器学习技术的相关问题，到本模块中提出问题或寻找答案。学生遇到的困惑在这里完全可以得到解决，也方便了老师进行答疑教学。

- 视频教学

学生在课下可以独自在该平台进行学习，也可以将这些视频设置为通识选修课来辅助教师进行教学，充分调动学生的积极性，增强学生对相关知识点的学习和理解。

- 考试测评

为了加强学生的学习效果，针对各个知识点，我们专为学生开发了考试测评系统，其中包括 SPSS、SAS、R、Python、Hadoop、Excel、数据挖掘、统计概率等。学生可以

选择题型、涉及的相关软件、理论 or 应用、课程分类、学科分类、行业及应用等方向，再选择试卷的难易程度（易、中、难），便可以进行专项知识点的练习和学习。

CDA 大数据教学管理系统提供管理者、教师、学生三种角色的登录。



管理者登录操作系统，可以进行信息批量导入、班级配置、信息删改等操作，还能对各个用户的权限进行管理。

教师登录操作系统，可以实现对自身课程教学任务的增删改查，可以将任务发布给学生，同时也可以查看其它教师的课程教学任务。通过成绩管理，教师可以查看学生的任务报告，依据任务报告的完成质量，给予评分。通过资源管理，教师可以上传自己课程任务的资源。

学生登录操作系统，可以查看、下载、提交任务，可以接收教师反馈的成绩。

除了适用于大数据分析的操作平台，我们还为您提供

- 平台用户使用手册；
- 实验室使用培训；
- 大数据课程培养方案；
- 大数据课程系列教材；
- 教师教学教案；
- 课上展示 PPT；
- 丰富的题库和学习视频；
- ...

欢迎您试用 CDA 大数据实验室并提出宝贵意见!

我们的官网是 <http://www.cda.cn/lab.html>

欢迎联系

上官老师: 15210507544

邮 箱: [sgcx@pinggu.org](mailto:sgcx@pinggu.org)