

# 数据科学与人工智能实验室 产品白皮书

北京奇观技术有限责任公司. 北京. 中国

2019年3月

# 目录

<b>1. 产品背景</b> .....	<b>1</b>
1.1. 需求现状.....	1
1.2. 人才需求.....	1
1.3. 设计目标.....	1
1.4. 人才培养前景.....	2
<b>2. 产品概述</b> .....	<b>3</b>
2.1. 产品简介.....	3
2.2. 产品架构.....	4
2.3. 产品组成.....	5
2.3.1. 硬件组成.....	5
2.3.2. 软件组成.....	6
2.4. 产品优势.....	7
2.4.1. 基于容器集群技术提供高并发、高性能、高可靠的实践环境.....	7
2.4.2. 平台集成丰富的课程内容资源包.....	7
2.4.3. 提供边看边做的实验操作体验.....	8
2.4.4. 提供可视化的教学监控和教学分析，保证实践教学环节的效果.....	8
2.4.5. 实验室平台降低学习成本.....	9
<b>3. 产品功能</b> .....	<b>9</b>
3.1. 学生端.....	10
3.2. 教师端.....	12
3.3. 管理员端.....	19
3.4. 课程资源库.....	21
<b>4. 售前技术服务</b> .....	<b>23</b>
4.1. 技术培训.....	23
4.2. 售后技术服务.....	24
<b>5. 典型案例</b> .....	<b>24</b>
<b>6. 公司简介</b> .....	<b>24</b>

---

## 1. 产品背景

### 1.1. 需求现状

近年来，随着大数据与人工智能专业人才的急缺，越来越多的高校开始申请建设大数据与人工智能专业，招收和培养相关领域的专业技术人才。2018年3月21日，教育部公布了《2017年度普通高等学校本科专业备案和审批结果》，共有862所高校新增了2311个专业。其中最热门的专业当属“数据科学与大数据技术”，共有248所高校申请获批，占新增专业申请的高校数超过1/3，全国申请获批“数据科学与大数据技术”专业的高校总数达到283所。同样，人工智能也成为了各大高校竞相追捧的一块“香饽饽”，从2017年5月中国科学院大学成立人工智能学院开始，后续有西安电子科技大学、清华大学、上海交通大学、南京大学、哈尔滨工业大学等33所高校建立人工智能学院。

### 1.2. 人才需求

预计2020年，我国大数据市场规模将超过8000亿元，有望成世界第一数据资源大国。但数据开放度低、技术薄弱、人才缺失、行业应用不深入等都是产业发展中亟待解决的问题。

根据领英发布的《2016年中国最热职位人才报告》显示，数据分析人才的供给指数最低，仅为0.05，属于高度稀缺。

《2017中国大数据发展报告》显示，目前中国大数据人才依旧稀缺，尤其是数据分析、系统研发等技术类人才更甚。智联招聘调查统计显示，短期内中国大数据人才缺口将超过130万。

中国商业联合会数据分析专业委员会资料显示，未来3至5年，中国需要180万数据人才，但截至目前，中国大数据从业人员只有约30万人。同时，大数据行业选才的标准也在不断变化。

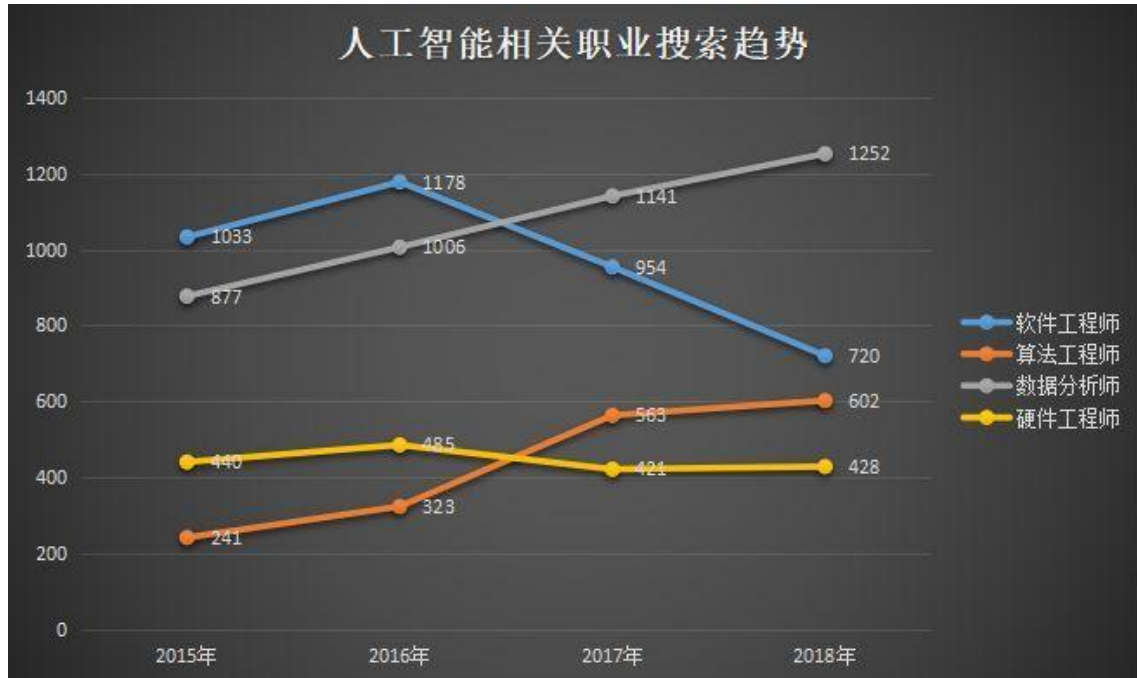
### 1.3. 设计目标

数据科学与大数据实验室是一套完整的面向大数据、人工智能领域的实践教学

平台，内置丰富的课程资源库，针对大数据专业旨在培养具有将领域知识与计算机技术和大数据技术融合、创新的能力，能够从事大数据研究和开发应用的高层次人才。能在统计部门、税务海关、公司企业以及金融保险机构等企事业单位从事大数据分析、大数据应用开发、大数据系统开发、大数据可视化以及大数据决策等工作。针对人工智能专业旨在培养掌握人工智能领域坚实的基础理论和宽广的专业知识；掌握解决工程问题的先进技术方法和现代技术手段；具有创新意识和独立担负工程技术或工程管理工作的能力，能够对复杂问题进行抽象建模，然后针对模型进行算法设计、分析与实现，最后对模型和算法进行不断优化完善。

## 1.4. 人才培养前景

根据 2018 届互联网校招高薪清单显示，众多知名互联网企业技术类岗位的年薪水平，动辄就是 30 万元以上的出价。互联网企业中涉及大数据和人工智能等岗位都非常火热，在他们看来，校招年薪 25 万人民币只是“白菜价”。同时随着人工智能概念的持续火爆，也使得大批求职者主动向人工智能相关岗位靠近，下图为大数据、人工智能专业职位搜索趋势图：



大数据、人工智能专业热门职位技术需求如下：

### ➤ 数据科学家

1、在移动通信和移动互联网领域，利用数据科学专业知识进行通信与互联网机器学习，人工智能等相关产品的研发工作；

---

2、运用量化分析、数据挖掘和数据可视化方面的专业知识，为我们提供超越数字层面的专业洞察；

3、运用数据和业务规则解决大规模网络中系统和数据基础架构中潜在的问题；

4、在跨功能域/跨职能部门的工作中定义问题陈述，收集数据，构建分析模型并提出建议；

5、构建和维护数据驱动和优化模型，实验，预测算法和容量约束模型；

6、建立分析模型来识别和输出推动网络系统关键决策；

7、能熟练运用 R, Python, Spark, Hadoop 和 SQL, Tableau 等工具来提高分析效率；

➤ AI/机器学习工程师

1、掌握 LinearRegression、LogisticRegression、SVM、GBDT/RandomForest、MaximumEntropy、HMM、CRF 等传统机器学习算法的原理以及主流机器学习语言，拥有实际应用的经验。

2、熟悉 CNN、RNN，encoder-decoder 等深度学习算法的原理和应用。

3、熟练使用 TensorFlow、PyTorch、Keras、MxNet 等其中一种或多种深度学习框架。

4、具有较强的数据建模，数据分析与挖掘能力。

➤ 算法工程师

1、负责特定应用场景下机器学习的模型搭建和模型调优。

2、参与构建与推进相关应用场景下机器学习平台的持续演进和落地。

3、利用机器学习，数理统计或最优化等技术和方法对实际业务大数据及其应用场景进行多维度分析和研究，帮助优化实际运营及提供业务决策。

4、参与数据科学和业务分析相关新技术和应用方案的研究和开发。

5、熟悉至少一种以上机器学习框架（Caffe/TensorFlow/MlLib 等）。

6、精通机器学习领域的相关算法；

7、具备良好的数学基础和编程能力，熟悉 Python、Scala、R 等开发语言。

8、有机器学习产品化应用开发经验和产品化应用设计能力者优先。

## 2. 产品概述

### 2.1. 产品简介

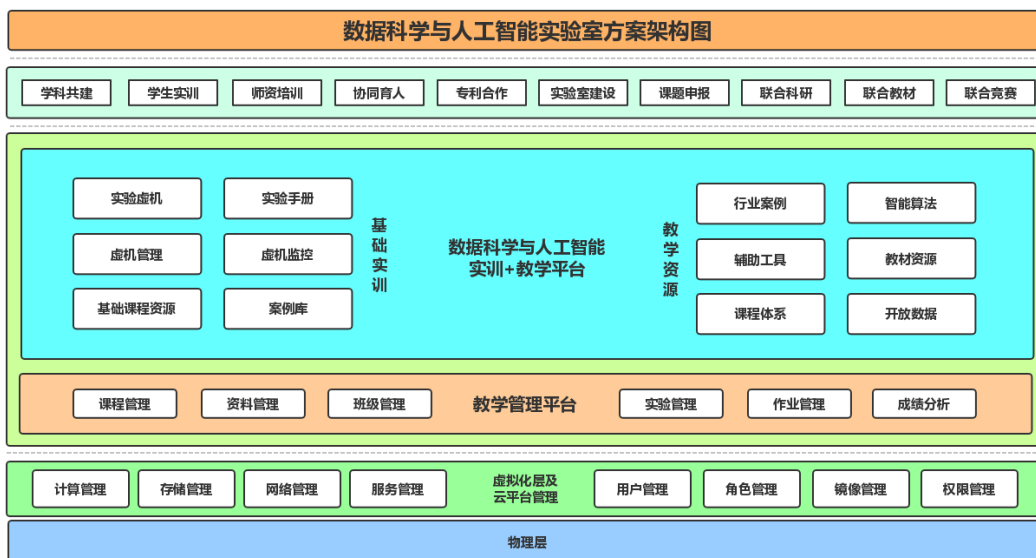
大数据人工智能（DSAI）实验室是一套集成了完整云计算、大数据、人工智能

等课程资源的实践教学系统平台，针对当前云计算大数据、人工智能等专业实验室建设的痛点，采用 Docker 容器化技术一站式提供性能可靠、环境完整、界面交互良好的实验和案例操作环境，容器内置了所有实验和案例的软件开发包和项目源码，配套详细的实验指导说明书，DSAI 实验室以理论+实践、平台+资源的全新模式为教师和学生提供立体化、全周期的实践教学环节的支撑和辅助。



## 2.2. 产品架构

数据科学与人工智能实验室方案架构分为 5 个部分，物理层、虚拟化和云平台管理层、教学管理层、实验实训平台层、产教融合模块。其中物理层有硬件服务器构成，提供实验室平台的计算、存储、网络等支撑服务；虚拟化和云平台管理层主要借助虚拟化技术和容器技术提供虚拟资源池和容器资源池；教学管理层则主要实现教务管理，包括课程管理、班级管理、成绩管理、实验管理等；实验实训平台层是数据科学与人工智能的业务核心层，主要包括学生试验机的调度、分配、监控，课程资源库的管理等；产教融合模块则是基于数据科学与人工智能实验室平台提供多种模式的校企合作。

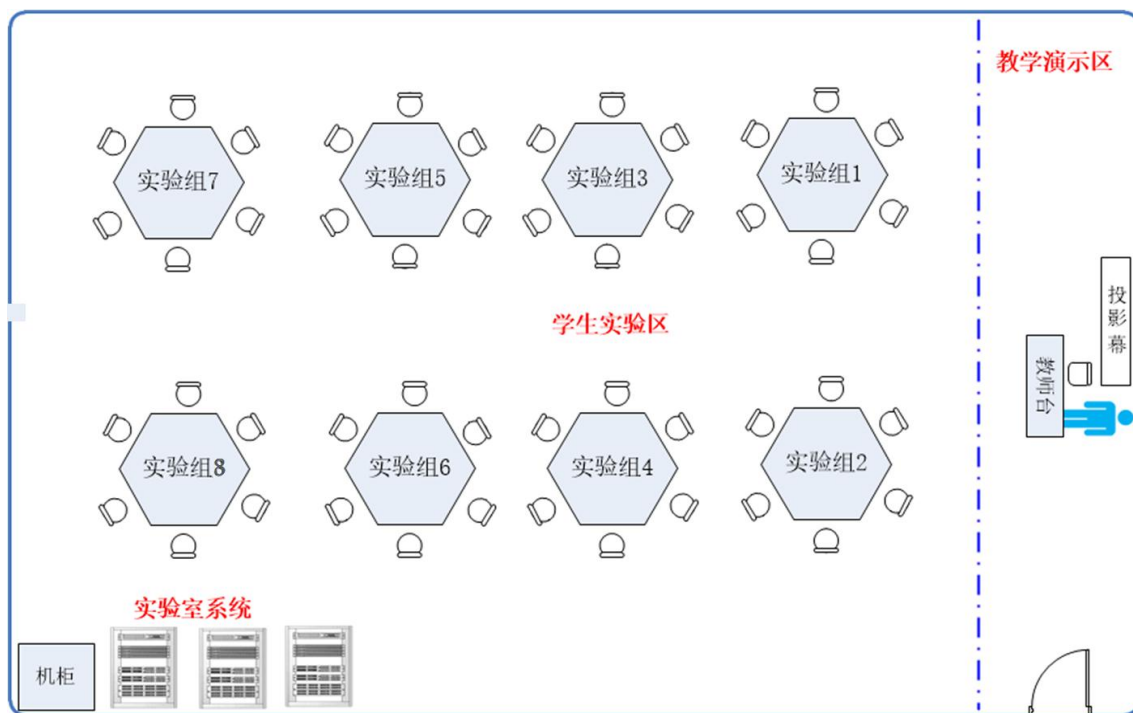


## 2.3. 产品组成

数据科学与人工智能实验室包括服硬件部分、平台软件部分、课程资源及行业案例库部分构成，其中硬件部分有学生用主机和平台服务器组成，平台软件部分主要是数据科学与人工智能实验室相关模块构成，课程资源及行业案例库则包含课程培养方案、大纲、课件、实验文档、案例代码、相关工具等。

### 2.3.1. 硬件组成

数据科学与人工智能实验室平台部署运行在 DSAI 实验室管理控制及计算单元设备上，设备以集群的方式提供硬件支持，终端学生用户使用 DSAI 实验室终端用户设备来访问平台服务，DSAI 实验室平台应用在高校专业机房的硬件拓扑图如下所示：



### 2.3.2. 软件组成

数据科学与人工智能实验室平台分为实验学习模块、教学管理模块、课程资源包模块、基础数据平台模块、可视化分析模块，平台功能模块图如下所示：

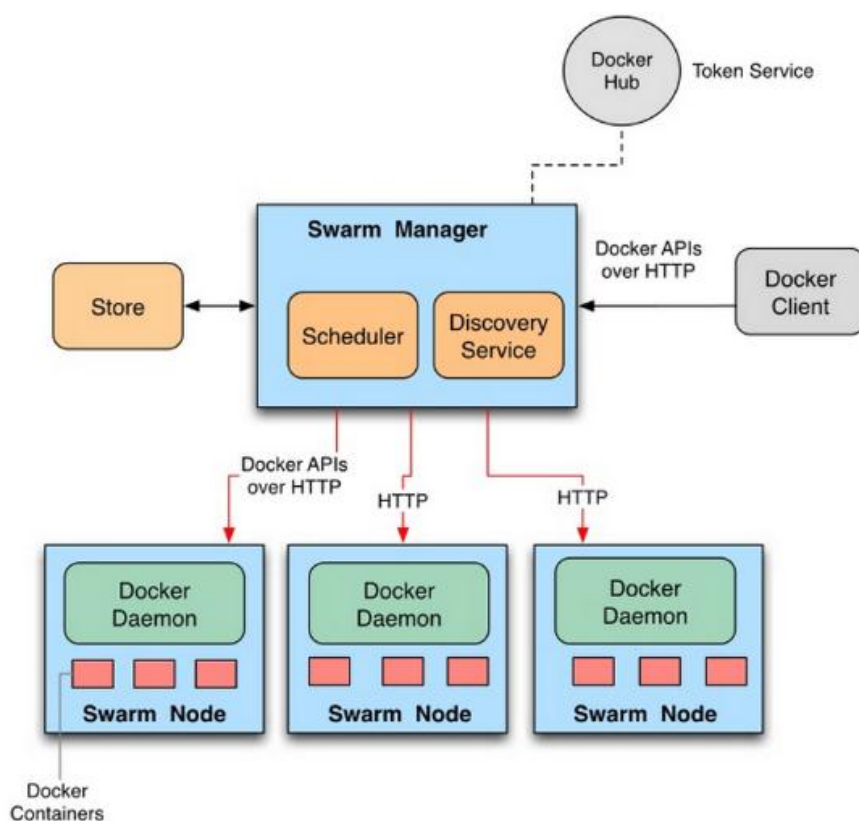




## 2.4. 产品优势

### 2.4.1. 基于容器集群技术提供高并发、高性能、高可靠的实践环境

系统底层采用 Docker 容器化技术构建面向大数据、云计算、人工智能相关实验和案例的试验机资源池，资源池根据实验和案例的复杂程度灵活、动态分配相应性能的容器资源，保证实验和案例的流畅运行。



### 2.4.2. 平台集成丰富的课程内容资源包

系统平台中集成了不同专业类别的课程资源包，其中包括大数据工程类、大数据分析挖掘类、大数据可视化类、科学科学计算类、人工智能与机器学习类的相关行业数据和实验案例资源，累计实验案例 100+，课程课件 300+，实验文档 200+，行业数据 50+。



### 2.4.3. 提供边看边做的实验操作体验

数据科学与人工智能实验室在提供实验机操作界面的同时，配套展示详细的实验操作指导说明书，学生可以灵活调整实验操作界面和指导说明书查阅界面的比例，支持全屏的实验步骤操作、全屏的文档查阅和实验步骤操作与文档查阅同屏交互三种模式，提供学生极致的实践操作体验。



### 2.4.4. 提供可视化的教学监控和教学分析，保证实践教学环节的效果

学生完成实验和案例的过程中，平台实时监控试验机中每一步操作，并将监控数据发送到任课教师端，教师可以直观地了解学生实操掌握情况。同时，基于学

生实验操作数据的分析，平台在不同维度展示教学指标的完成情况，包括学生实验投入时间、学生实验完成数量、学生平台使用记录、班级成绩分析等，通过可视化的教学监控和教学分析功能，及时掌握授课过程中的学生完成情况，提升教学环节的的效果和质量。



## 2.4.5. 实验室平台降低学习成本

学习过程系统化、流程化、平台化，无需安装环境，即可直接实验操作，并且平台可以建立快照保存，实现了学生在任何时间、任何地点不间断地进行实验的功能，大大减少了学习时间，提升学习效率。

课程编号	课程名称	课程简介	文档权重
1	大数据基础部署篇		0.6
2	大数据基础实操篇		0.121
3	大数据核心1部署篇		0.12
4	大数据核心1实操篇		0.12
5	大数据核心2部署篇		0.12
6	大数据核心2实操篇		0.12
7	大数据核心3部署篇		0.12
8	大数据核心3实操篇		0.123
9	大数据综合实训篇		0.12
123		21	0.01

## 3. 产品功能

数据科学与人工智能实验室在功能模块上主要分为学生端模块、教师端模块和管理员端模块，采用 B/S 架构，三端用户通过登录界面进入，访问各自的功能页面，平台底层使用容器化集群技术来支撑实验机环境的运行，同时平台集成丰富的课

程资源库。

### 3.1. 学生端

学生端主要功能包含我的课程、实验成绩、学习资料、个人中心模块。

#### (1) 我的课程

我的课程包含教师给学生分配的所有课程，可以展示课程所包含的实验和案例的个数等基本信息以及学生所完成的进度信息。每门课程包含的实验以及案例按照时间轴的方式进行展示，对应学生当前完成的进度。学生可进行开始实验，进入实验操作界面。根据提供的实验指导文档信息，在虚拟机中完成对应的实验操作，同时编写实验操作文档，最终提交文档完成实验。在实验过程中，系统会对虚拟机操作界面进行自动截图，会随实验文档提交保存在后台。在实验虚拟机中提供复制、截图、重置、延时、关闭虚拟机等功能。部分效果图如下图所示。





### (2) 实验成绩

实验成绩包括学生所完成实验的得分情况，也可以在线查看实验文档信息。部分效果图如下图所示。



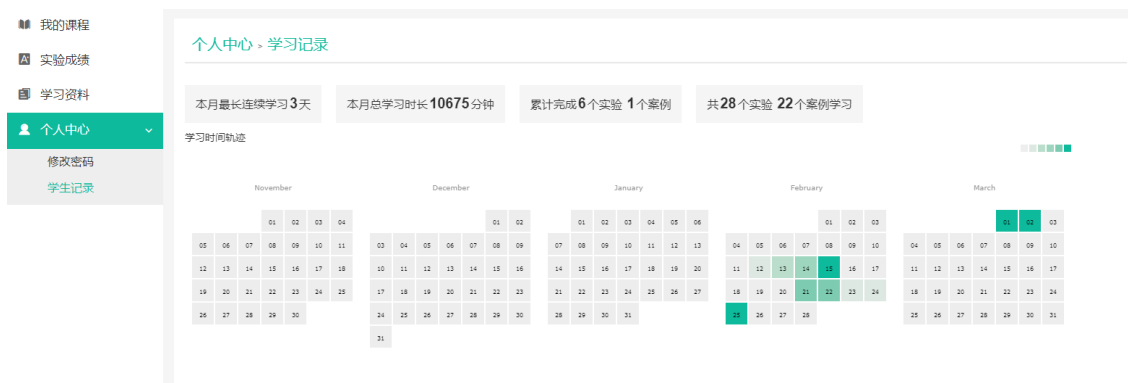
### (3) 学习资料

学习资料包含教师在平台上对应不同课程所上传的学习资料，学生可以在线查看以及下载。部分效果图如下图所示。



#### (4) 个人中心

个人中心包括修改密码以及学习记录等信息。可对密码进行修改操作，同时可对自己学习情况进行统计并以图表形式展示。部分效果图如下图所示。

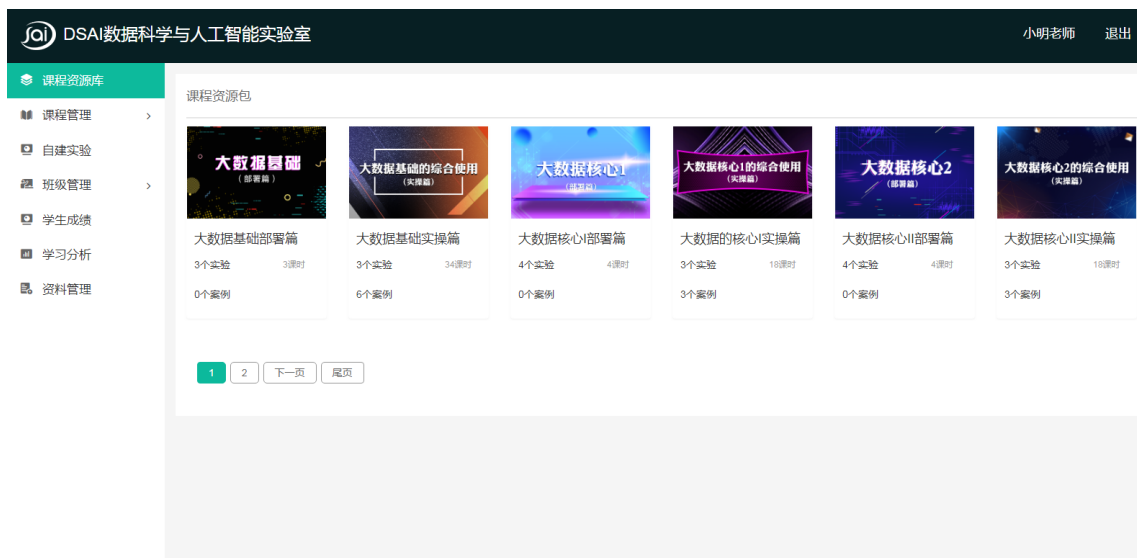


## 3.2. 教师端

教师端主要包括：课程资源库、课程管理、自建实验、班级管理、学生成绩、学习分析、资料管理，主要功能具体描述如下：

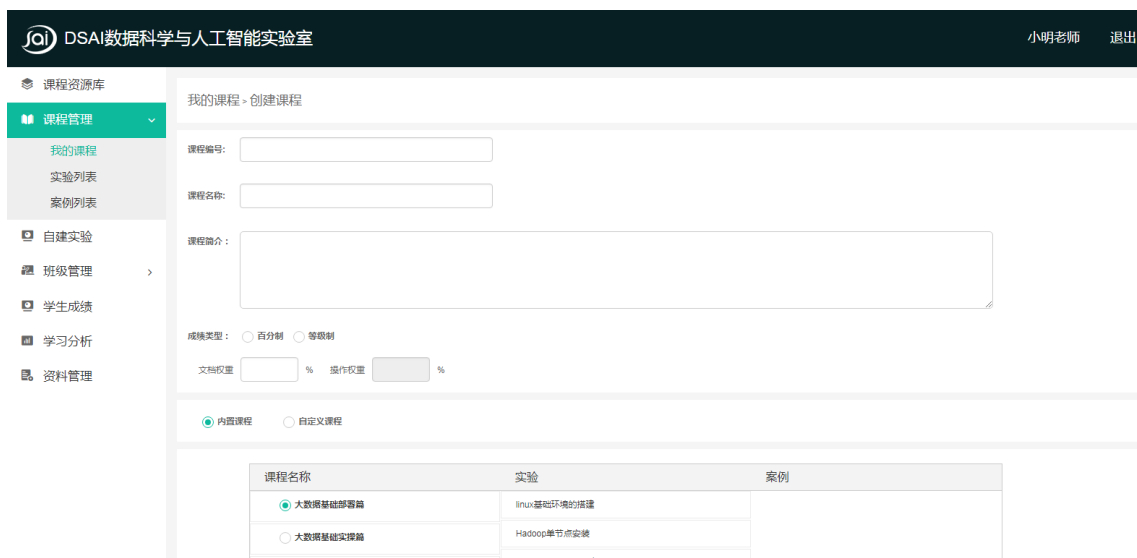
#### (1) 课程资源库

课程资源库包含系统内置的所有课程资源。每门课程中都包含多个实验或案例。部分效果图如下图所示。



## (2) 课程管理

课程管理包含我的课程、实验列表以及案例列表。我的课程展示是教师自己创建的课程，教师可以自己创建课程，可根据内置课程进行创建或者根据所有实验以及案例来自定义组合来创建课程。部分效果图如下图所示。



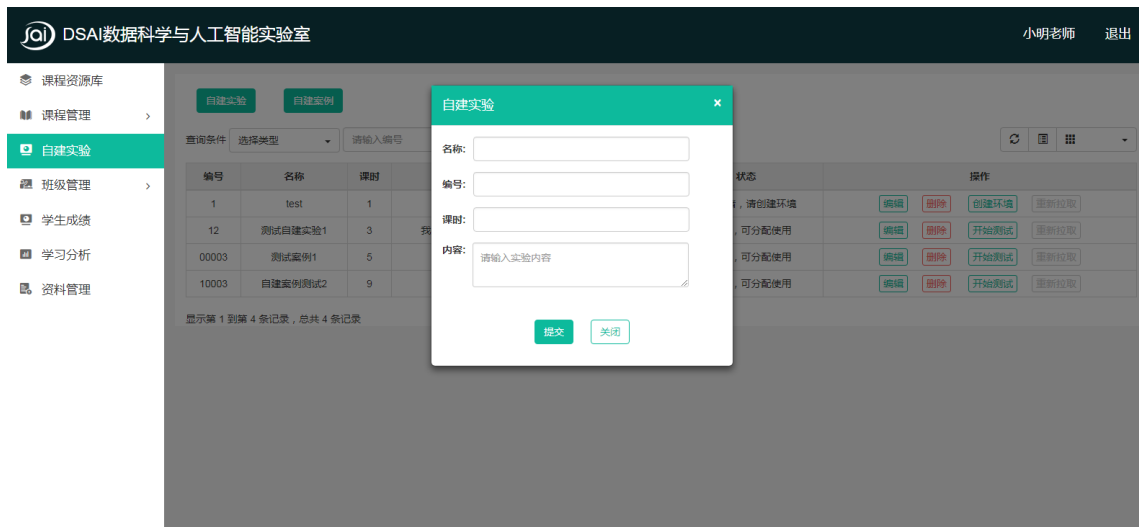
实验列表以及案例列表展示的是教师创建课程中的所有实验以及案例，教师也可以进入虚拟机进行实验。部分效果图如下图所示。



### (3) 自建实验

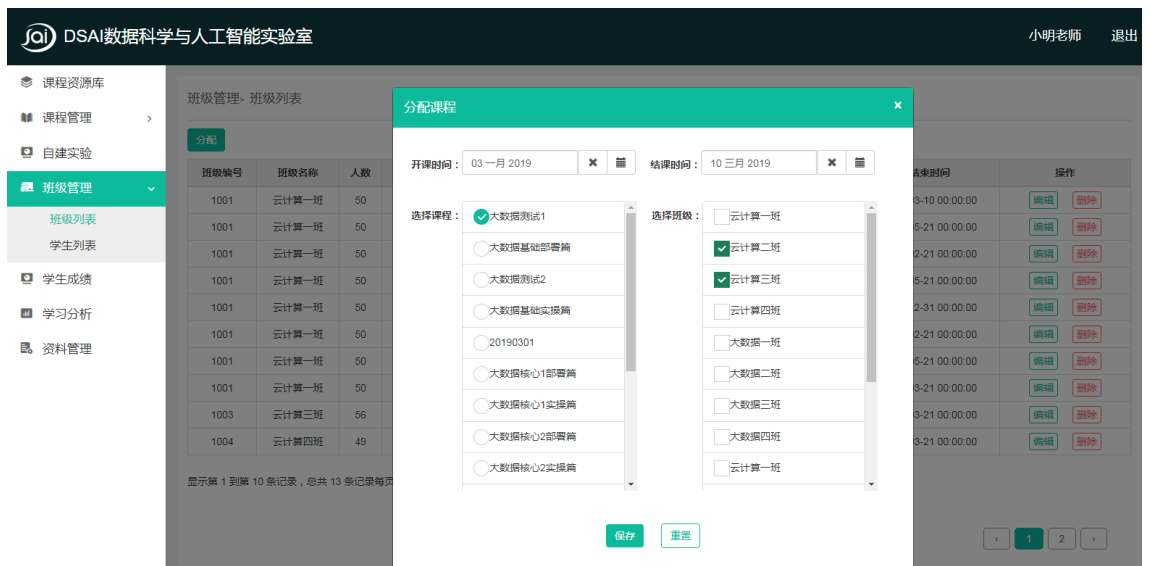
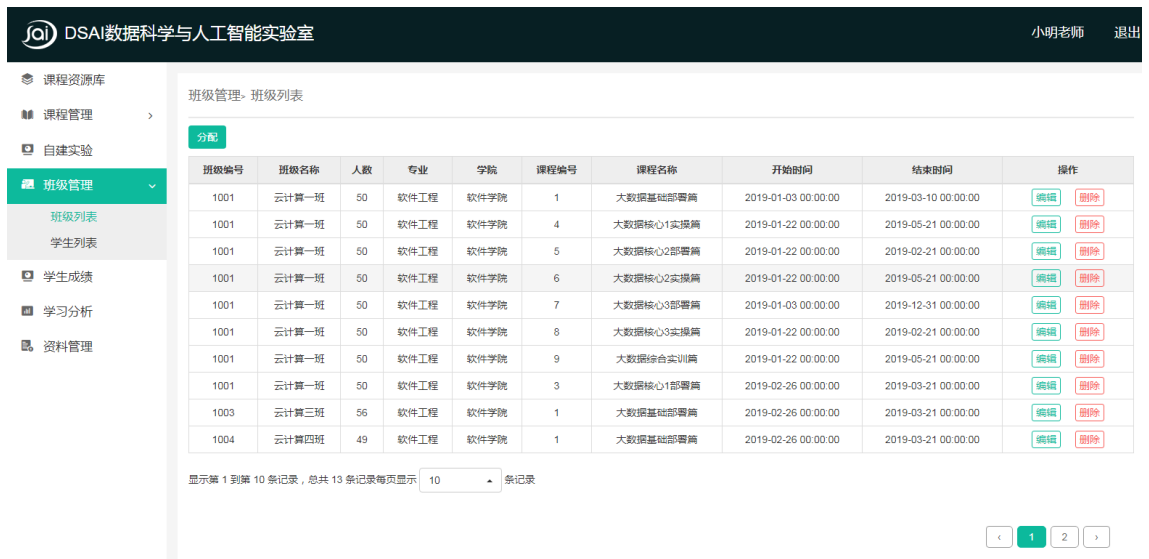
自建实验模块的主要作用是提供教师进行自创实验或者案例。创建成功的实验或案例会在自建课程的实验或案例列表中显示，供老师创建课程使用。部分效果图如下图所示。





#### (4) 班级管理

班级管理包含班级列表以及学生列表。班级列表展示的是教师所教班级的信息，可以对不同班级课程开课结课时间进行调整。同时可以把自己的课程分配给不同的班级。部分效果图如下图所示。



学生列表展示教师所教班级的所有学生信息，同时可以对学生信息修改、密码重置、删除等功能。部分效果图如下图所示。

DSAI数据科学与人工智能实验室 小明老师 退出

班级管理 > 学生列表

查询条件: 选择班级

学号	学生姓名	学生性别	专业	系别	学院	联系方式	邮箱	操作
10001	JM	女	云计算一班	软件工程	软件学院	-	-	<a href="#">编辑</a> <a href="#">重置密码</a> <a href="#">删除</a>
10002	小白	男	云计算一班	软件工程	软件学院	12231231	163@qq.com	<a href="#">编辑</a> <a href="#">重置密码</a> <a href="#">删除</a>
10003	奥利	女	云计算一班	软件工程	软件学院	135666	999555@qq.com	<a href="#">编辑</a> <a href="#">重置密码</a> <a href="#">删除</a>
10004	小周	女	云计算一班	软件工程	软件学院	dsad	sadasd@qq.com	<a href="#">编辑</a> <a href="#">重置密码</a> <a href="#">删除</a>
10005	王五	女	云计算一班	软件工程	软件学院		1111@qq.com	<a href="#">编辑</a> <a href="#">重置密码</a> <a href="#">删除</a>
10006	小小	女	云计算一班	软件工程	软件学院		111qq11@qq.com	<a href="#">编辑</a> <a href="#">重置密码</a> <a href="#">删除</a>
1001001	张三1	男	云计算一班	软件工程	软件学院	13888888878	xx@xx1.com	<a href="#">编辑</a> <a href="#">重置密码</a> <a href="#">删除</a>
1001002	张三2	男	云计算一班	软件工程	软件学院	13888888888	xx@xx.com	<a href="#">编辑</a> <a href="#">重置密码</a> <a href="#">删除</a>
1001003	张三3	男	云计算一班	软件工程	软件学院	13888888888	xx@xx.com	<a href="#">编辑</a> <a href="#">重置密码</a> <a href="#">删除</a>
1001004	张三4	男	云计算一班	软件工程	软件学院	13888888888	ss	<a href="#">编辑</a> <a href="#">重置密码</a> <a href="#">删除</a>

显示第 1 到第 10 条记录, 总共 38 条记录每页显示  条记录

< 1 2 3 4 >

### (5) 学生成绩

学生成绩展示的是学生实验所得分数情况、可以对实验过程中实操截图进行浏览以及对实验文档进行批改操作。同时支持文档下载。部分效果图如下图所示。

DSAI数据科学与人工智能实验室 小明老师 退出

学生成绩

查询条件: 选择课程

[导出表格数据](#)

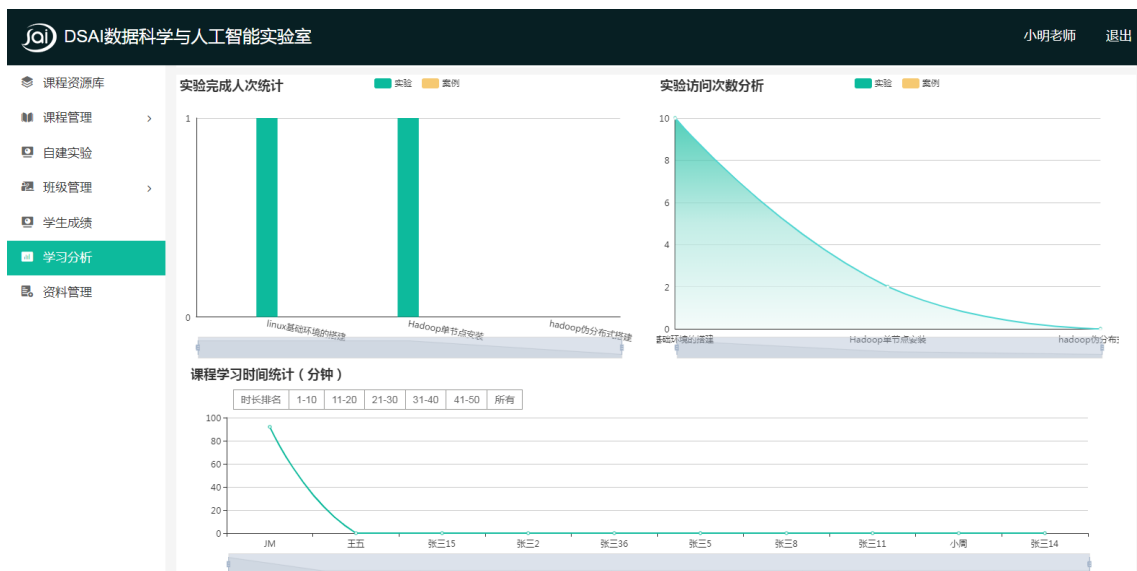
学生姓名	课程名称	实验案例名称	类型	提交时间	报告成绩	操作成绩	总成绩	实操截图	操作
JM	大数据基础部署篇	热闻分析	实验	2019-01-30 13:21:07	6	40	46	<a href="#">预览</a>	<a href="#">批改</a> <a href="#">下载</a>
JM	大数据基础部署篇	地区金额实时统计	案例	2019-01-30 17:25:05	1.2	17.6	18.8	<a href="#">预览</a>	<a href="#">批改</a> <a href="#">下载</a>
JM	大数据基础部署篇	Hadoop单节点安装	实验	2019-03-01 11:02:42	30	25.6	55.6	<a href="#">预览</a>	<a href="#">批改</a> <a href="#">下载</a>
JM	大数据核心1部署篇	zookeeper单节点部署	实验	2019-03-03 16:36:15	10.68	69.52	80.2	<a href="#">预览</a>	<a href="#">批改</a> <a href="#">下载</a>
JM	大数据基础实操篇	HDFS的Shell操作	实验	2019-02-27 11:17:56	4.43	23.46	27.89	<a href="#">预览</a>	<a href="#">批改</a> <a href="#">下载</a>
JM	大数据基础实操篇	HDFS的API操作	实验	2019-02-27 13:29:45	0.96	2.64	3.6	<a href="#">预览</a>	<a href="#">批改</a> <a href="#">下载</a>
JM	大数据基础实操篇	Hadoop自带的WordCount操作	实验	2019-03-03 14:54:13	8.34	79.11	87.45	<a href="#">预览</a>	<a href="#">批改</a> <a href="#">下载</a>

显示第 1 到第 7 条记录, 总共 7 条记录



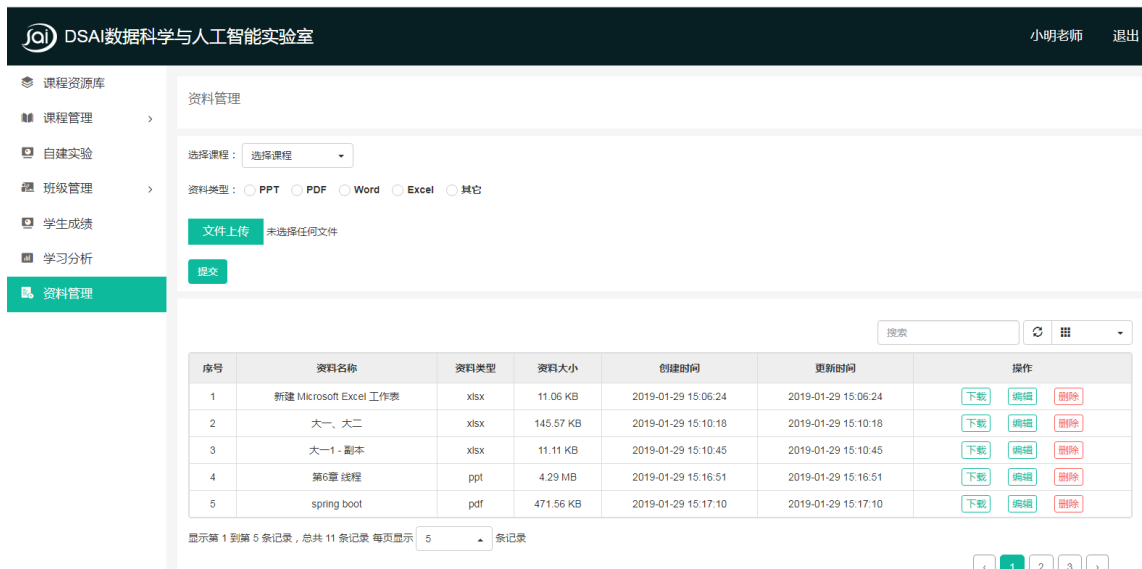
### (6) 学习分析

学习分析模块是分析统计学生实验完成情况以及学生学习时长并以图表的方式展现。部分效果图如下图所示。



### (7) 资料管理

资料管理提供老师上传学习资料的功能，可以对不同的课程上传对应的学习资料，供学生补充学习。部分效果图如下图所示。

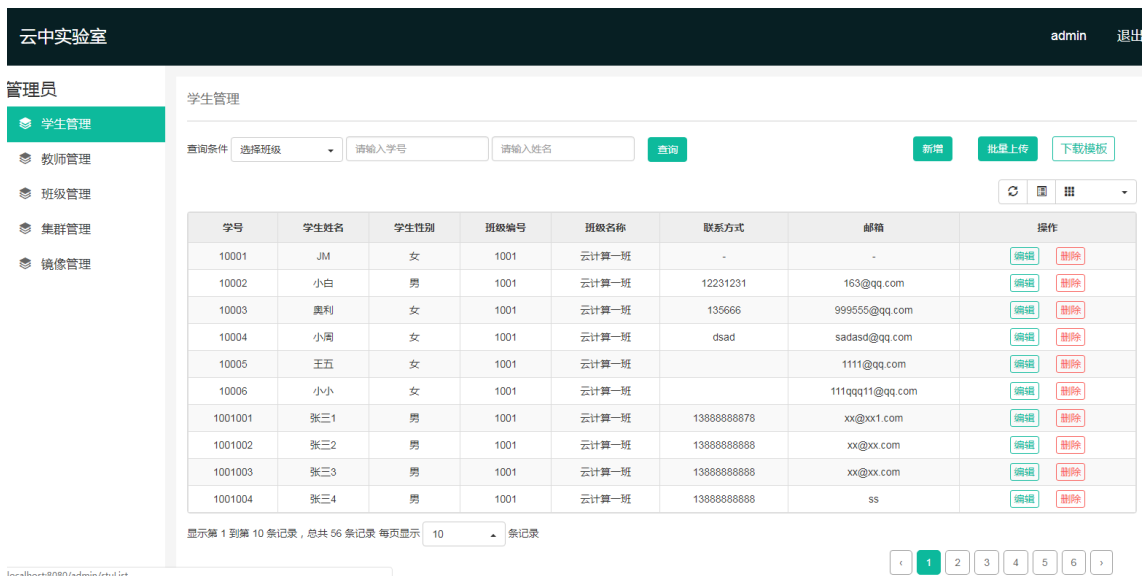


### 3.3. 管理员端

管理员端提供学生、教师、班级基本信息的批量导入、增删改查等基本操作，同时包含集群镜像管理功能。

#### (1) 学生管理

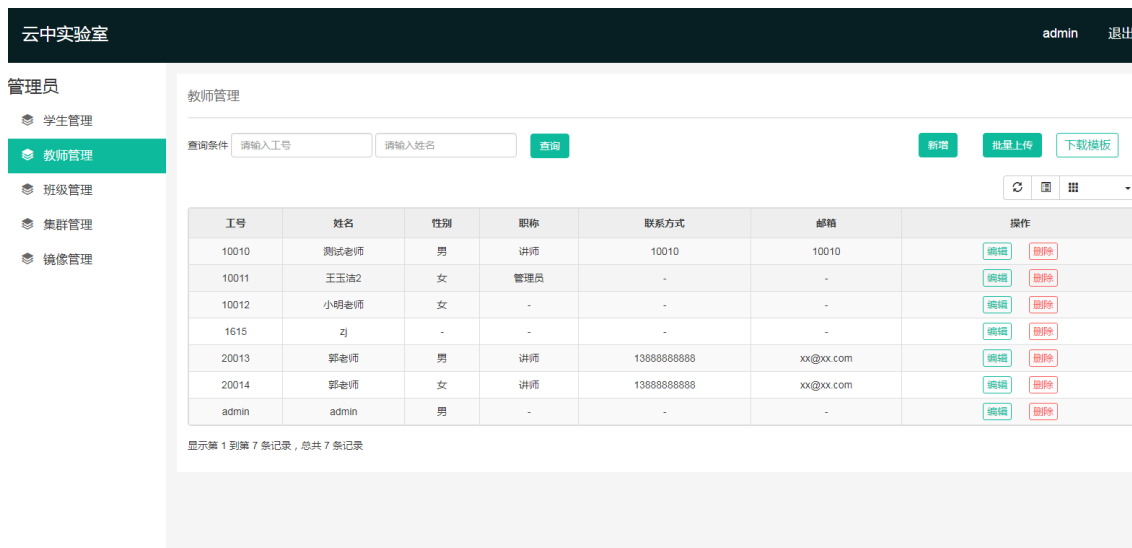
学生管理包含所有学生的基本信息展示，提供批量导入、增删改查等基本功能。部分效果图如下图所示。



#### (2) 教师管理

教师管理包含所有教师的基本信息展示，提供批量导入、增删改查等基本功能。

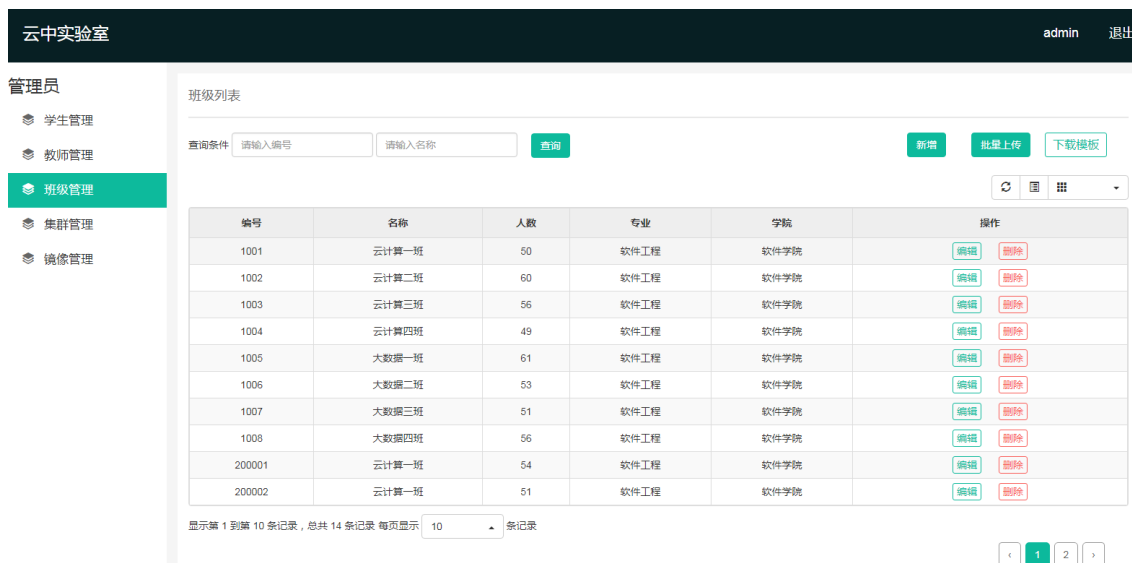
部分效果图如下图所示。



### (3) 班级管理

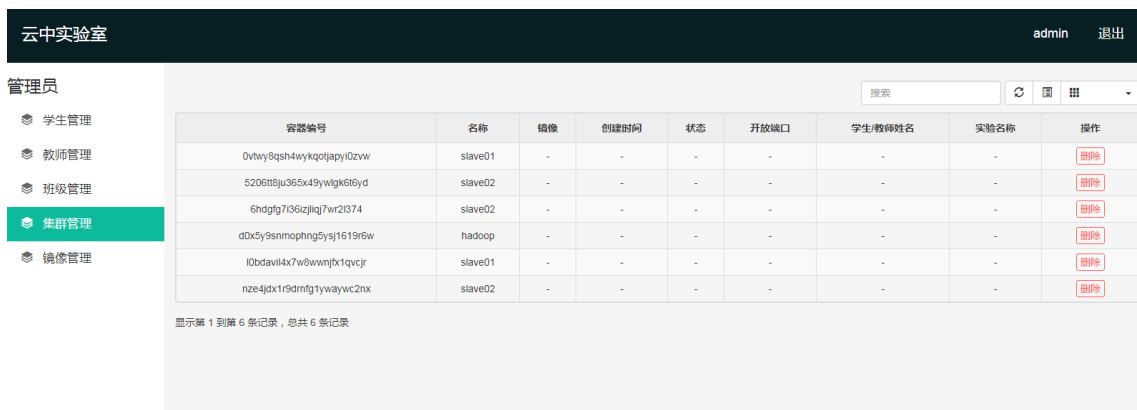
班级管理包含所有班级的基本信息展示，提供批量导入、增删改查等基本功能。

部分效果图如下图所示。



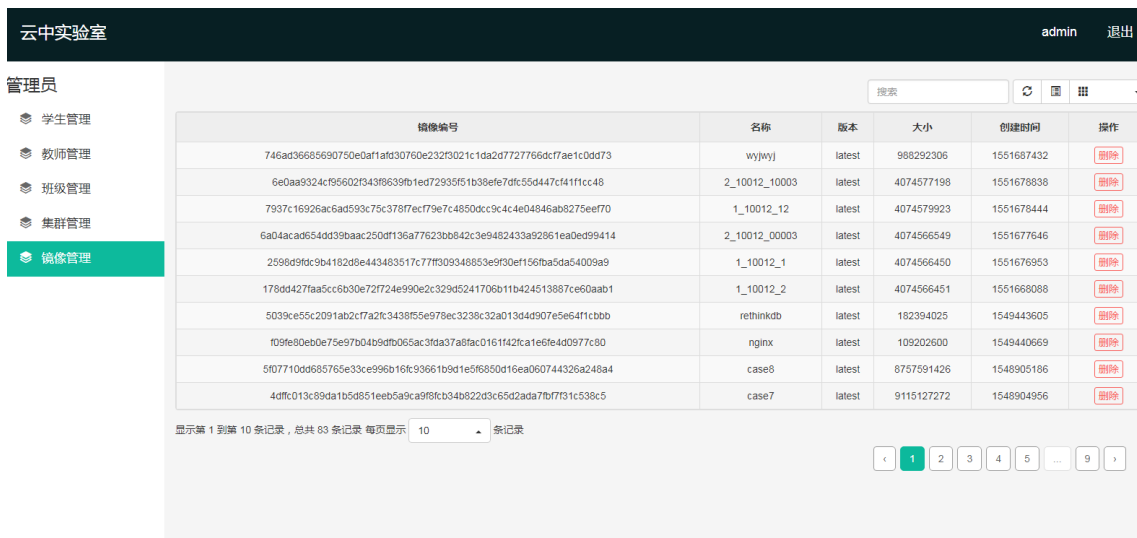
### (4) 集群管理

集群管理模块是展示的系统所有容器的信息，包含名称、端口、使用者、状态等信息，可对容器进行操作。部分效果图如下图所示。



### (5) 镜像管理

镜像管理模块是对系统的所有镜像进行统一的管理和监控，可显示容器名称、版本、大小、创建时间等信息。部分效果图如下图所示。



## 3.4. 课程资源库

数据科学与人工智能实验室不仅提供功能，并且提供给学生和教师足够的实验案例与实验计划等软性资源。课程信息如下表所示。

表 1 《大数据基础》课程简表

课程内容	教学课件	教学视频	实验案例
大数据的数据存储	√	√	√
HDFS	√	√	√
Hbase	√	√	√

其他数据存储方式	√	√	√
Zookeeper	√	√	√
MapReduce	√	√	√
大数据分析及可视化	√	√	√

表 2 《Hbase 核心技术》课程简表

课程内容	教学课件	教学视频	实验案例
HBase Shell 基础	√	√	√
HBase Shell 命令	√	√	√
HBase Shell 脚本	√	√	√
Hbase 程序设计	√	√	√

表 3 《MapReduce 核心技术》课程简表

课程内容	教学课件	教学视频	实验案例
MapReduce 入门	√	√	√
MapReduce 计算模型	√	√	√
MapReduce 编程	√	√	√
MapReduce API 详解	√	√	√
MapReduce 程序设计	√	√	√
MapReduce 作业机制	√	√	√

表 4 《大数据综合案例》简表

课程内容	教学课件	教学视频	案例
热词分析	√	√	√
地区金额的实时统计	√	√	√
用户通话记录分析 (Hive)	√	√	√
用户通话记录分析 (MapReduce)	√	√	√
SoGou 用户搜索记录行为分析	√	√	√
电影推荐系统	√	√	√
实时分布式日志流处理	√	√	√
电信用户行为分析	√	√	√



---

上述表中所显示的是奇观技术的简约课程信息，在课程资源中可以根据用户的实际情况进行定制，遵守高校课程资源制定的标准，满足实验和理论课程的分配。充分体现的实验课程学时安排，案例标准，核心要点等重要数据，同时体现大数据实验的核心内容与知识点。

## 4. 售前技术服务

项目实施前，公司售前技术人员会根据用户需求提供一套系统建设方案，并在与用户分析现有的技术环境及长期建设目标的基础上，为用户设计出最佳解决方案。

此外，根据用户的具体情况，售前技术人员必要时可搭建演示环境，提供相关解决方案产品的演示，以使用户更深了解项目建设情况；同时，售前技术人员还为用户提供相关的技术咨询和方案设计，为用户提供新建系统与已有系统的整合方案，保护用户已有的投资。售前技术人员作为用户项目团队的一个重要组成部分，其掌握的用户信息将一直共享到项目的最终实施，以保障项目的成功。



### 4.1. 技术培训

我们将在建设项目移交给用户运行前对最终用户进行相关的技术培训，此外，每次的系统升级或者根据用户的要求，我们将提供不定期的技术培训服务。该类技术培训主要分为以下两部分：

#### ➤ 用户操作培训

针对系统日常运行操作进行集中培训，培训对象主要是授课老师，内容为日常教学操作步骤等。

#### ➤ 系统管理用户培训

针对系统日常管理、维护、安装进行培训，培训对象为技术管理人员，内容为系统常见技术问题处理方式等。

## 4.2. 售后技术服务

### ➤ 协助计划

在项目实施之前,售后(售前)技术服务人员向用户提供项目实施的前提条件,包含网络环境要求、软硬件要求等。

对于用户购买的硬件和软件,在设备到达安装现场之后,售后技术人员将协助用户进行设备的安装、调试及验收。

### ➤ 现场服务

根据用户需求及合同规定的时间,售后技术人员将在用户现场进行项目实施,包含软硬件设备的安装、调试及验收,帮助用户顺利掌握系统,尽量减少项目实施周期,实现系统的无缝交接。

对有效合同内的软件故障,用户应将故障现象及出错信息通过电邮、信件或传真等通知我们,我方根据故障等级做出不同时间级别的相应,并给予解答。不能远程解决问题时,我方将派技术人员到达用户现场进行实地解决。

## 5. 典型案例



## 6. 公司简介

北京奇观技术有限责任公司地处北京中关村软件园区,专注于云计算、大数

---

据相关领域技术的研究与发展，具有强大的研发实力和持续的创新能力。奇观技术于 2015 年，获得了深圳时代伯乐创业投资管理有限公司的风险投资，公司业务获得了爆炸式发展。目前已经为 Oracle、日本理光、赛尔网络和神州数码提供了专项云计算技术和大数据技术服务，并与上述公司开展了深层次的商业化项目合作。

作为桌面云领域的专家，奇观技术推出的拥有自主知识产权的 Marvel Sky 云平台，采用自主研发 FTC（Fast Transport Cloud）协议的实现了高效的云桌面数据传输。同时，基于底层虚拟化技术和容器化技术研发了一系列教育类产品：数据科学与人工智能实验室、在线评测与考试平台和信息安全翻转课堂教学云平台，并与本科、高职院校开展师资队伍建设、实验室建设、专业共建等合作。

奇观技术属于技术创新型公司，总部设在北京，以商务运作为主。公司计划将研发中心与北京分离，充分发挥和高校结合的这一优势，将研发部门逐步在河南、成都、上海、深圳等地的高校内开设。2016 年至今，公司先后在太原、合肥、武汉、郑州、石家庄、南阳（北京中关村南阳产业园区）等地陆续建立与高校合作的研发中心，与高校展开深层次的合作，促进公司研发以及云计算技术和大数据技术的可持续进步。