



欢迎参加

第 1 届全国高校大数据教学研讨会

2017年5月12日-13日 厦门大学



第1届全国高校大数据教学研讨会 (BDTS2017)  
大会特邀报告



2017年5月12日至13日，第1届全国高校大数据教学研讨会（BDTS2017）在厦门大学科艺中心音乐厅隆重举行。本届研讨会由教育部高等学校计算机类专业教育指导委员会主办，厦门大学、厦门理工学院、贵州师范大学、人民邮电出版社联合承办，旨在搭建专业的大数据教学交流平台，汇聚全国高校大数据教学精英力量，共同探讨大数据专业和课程体系建设，为加快推进全国高校大数据教学发展贡献力量。来自全国300多所院校的400余名教师参加了本次研讨会。

厦门大学谭绍滨校长助理、人民邮电出版社教育中心营销部肖稳副主任，北京大学、中国科学院、厦门大学、华东师范大学、同济大学等重点院校的6位大数据教学知名专家，以及来自国内知名大数据企业的3名业界专家出席会议并做特邀大会报告。厦门大学林子雨助理教授主持会议。

更多内容请访问大会官网：<http://dbl原因lab.xmu.edu.cn/post/bigdata2017/>



中新赛克 卢云川 副总裁 在做大会特邀报告



# 大数据与高校科研教学

— 中新赛克 —

1

大数据科研教学面临的挑战

2

中新赛克大数据科研教学解决思路

3

中新赛克大数据科研教学系列

4

总结

1

## 大数据科研教学面临的挑战



## 大数据系统研发工程师

负责大数据系统研发，包括大规模非结构化数据业务模型构建、大数据存储、数据库构设、优化数据库构架、解决数据库中心设计等。

## 大数据应用开发工程师

负责搭建大数据应用平台以及开发分析应用程序，研发各种基于大数据技术的应用程序及行业解决方案。

## 大数据分析师

主要从事数据挖掘工作，运用算法来解决问题，让数据显露出真相，同时，他们还推动数据解决方案的不断更新。

## 数据可视化工程师

负责在收集到的高质量数据中，利用图形化的工具及手段的应用，清楚地揭示数据中的复杂信息，帮助用户更好地进行大数据应用开发。

## 数据安全研发人才

主要负责企业内部大型服务器、存储、数据安全管理工作，并对网络、信息安全项目进行规划、设计和实施。

## 数据科学研究人才

数据科学研究是一个全新的工作，够将单位、企业的数据和技术转化为有用的商业价值。



1万亿

2017年1月17日，在中国农历春节前夕，工信部公布了《大数据产业发展规划（2016-2020年）》，提出了到2020年的发展目标：大数据相关产品和服务业务收入突破1万亿元，年均复合增长率保持30%左右。

150万

2016年8月数联寻英发布的首份《大数据人才报告》显示，目前全国大数据人才只有46万，未来3到5年人才缺口达150万之多。2015年-2016年是大数据人才最为匮乏的两年，因为已开了大数据专业的高等院校，第一批大数据人才还为毕业；已有的人才里，复合型的人才较少，都是术有专攻。

70万

2016年2月万达集团的一份“求贤帖”颇能说明问题：万达开出50万~70万元年薪吸纳大数据架构师、大数据研发工程师和算法工程师，应者寥寥。

## 实验室设备落后

大多数的院校由于资金的不足，无法更换老化的设备，更新实验室资源，使得教学资源分配不均匀。

## 实践环节重视不够

学生动手能力明显不足，网络技术专业是典型的理论与实践结合的专业，学生光学不练，无法满足用人单位的需求。

## 教学资源分散，共享程度低

高校教学资源较多，且比较分散，大部分软件和硬件资源仍然无法有效共享，使得高校教学资源重复投入，造成一定的浪费。

## 校企合作力度不足

校企合作是提升人才培养效果的最佳实践。



2

## 中新赛克大数据科研教学解决思路



- 公司源于通信巨头
- 与安全中心合作十余年



- 中新赛克先后与CCF、清华大学、复旦大学等科研机构 and 高校建立大数据合作关系，先后申请专利30余件，软件著作权100余个。



- 中国网络安全产业联盟理事单位
- 信息安全管理体系认证
- 江苏省软件企业认定
- 江苏省高新技术企业
- 计算机软件著作权认证
- 计算机系统集成资质
- CMMI3

公司员工超过600人，本科及以上学历员工超过90%，研发人员超过67%；OceanMind大数据教学产品已经在多个高校实验室使用。

源于中兴通讯，专注大数据

## 客户分布

公安、安全、军队、政府、  
教育、金融、保险、制造业

## 特聘顾问专家

周立柱：中国计算机学会数据库专委会主任  
李建中：中国计算机学会数据库专委会副主任  
周傲英：华东师范大学数据科学与工程研究院院长  
王晓阳：复旦大学计算机与软件学院两院院长  
于旭：香港中文大学系统工程管理学系教授  
李青：香港城市大学多媒体软件工究中心主任

## 政府伙伴

江苏经信委、江苏省交通厅、雨花区政府、南京软件谷、江  
苏省交通规划设计院、江苏通管局、湖南通管局

## 企业伙伴

汇智、美亚、太极、中航国际、凌华、研华、  
超微、联想、APM、技嘉、DELL、springer

## 资质证书

中国网络安全产业联盟理事单位  
江苏省大数据产业联盟单位  
2016年中国大数据最佳解决方案奖  
2016年中国数据中心高能存储产品奖  
大数据软件著作权

## 学术合作伙伴

CCF、DSE、清华大学、北京大学、华东师范大学、  
厦门大学、

广泛合作，构建大数据生态链

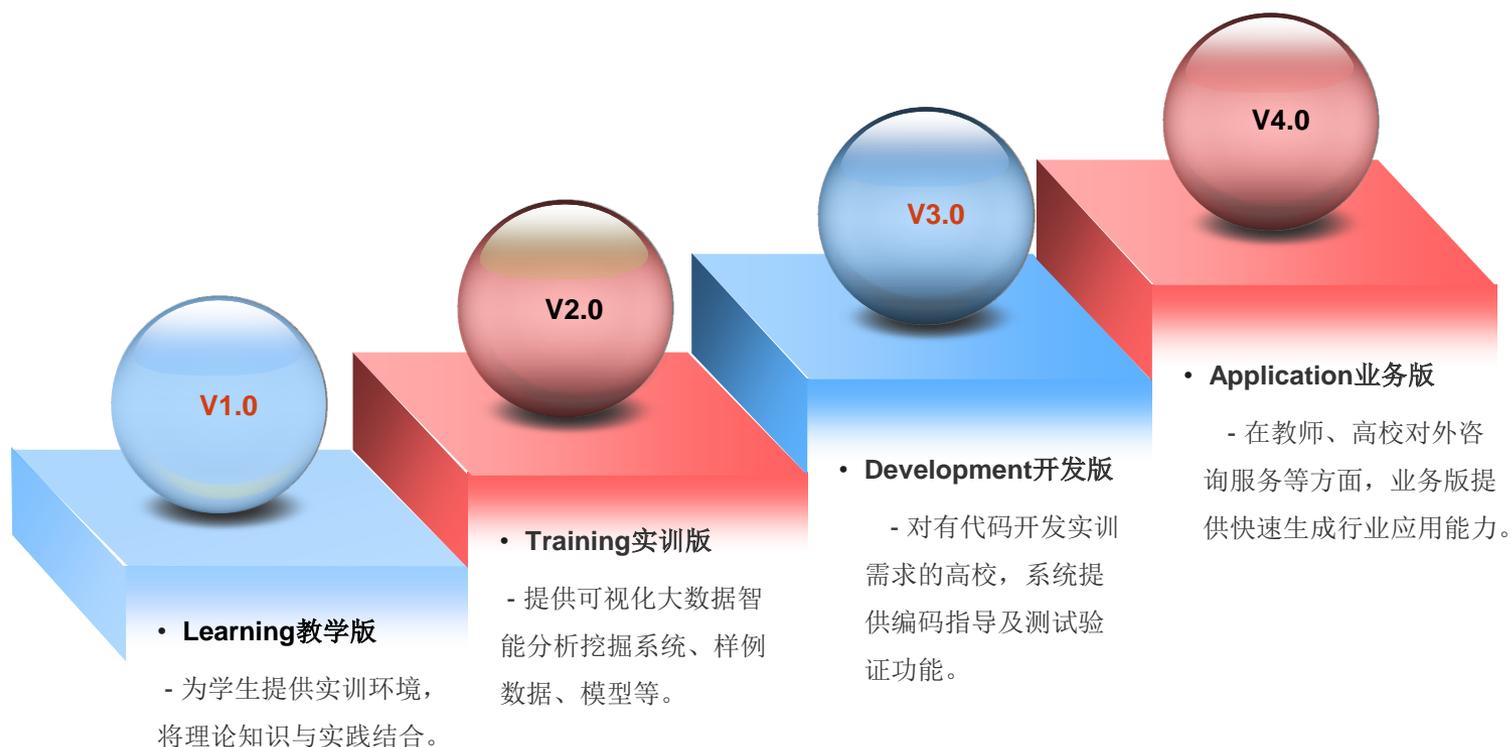
目前，高校在大数据教学方面，不仅需要通过学习理论教材向学生传授大数据知识，还需要配套的大数据实验环境，降低大数据学习和使用门槛，让学生更容易、更快速、更有效地完成大数据专业的课程学习。



大数据教学解决方案架构包括：Ocean Mind大数据基础平台、Ocean Mining智能分析挖掘平台，以及上层各种教学应用环境，为高校教学量身定制。



大数据教学平台是一套因材施教、循序渐进式的大数据教学系列产品，旨在解决高校大数据教学环节不同层面的需求。



3

中新赛克大数据科研教学系列

## 组件

- 提供丰富的Hadoop和Spark生态系统主流组件



## 环境

- 通过虚拟化方式，为每位学生或每组学生提供单独隔离的实验环境



- 提供大数据系统教学实训详细指导手册、视频等  
**指导**

水深	0.0V	0.2V	0.4V	0.6V	0.8V	1.0V	流量	0.0s	0.2s
5.2	0.37	0.36	0.34	0.32	0.28	0.23	22	22	22
6.4	0.57	0.55	0.53	0.49	0.44	0.35	31	31	31
7.3	0.68	0.56	0.53	0.59	0.52	0.40	44	44	44
10.4	0.83	0.80	0.77	0.72	0.64	0.45	51	51	51
12.2	0.94	0.91	0.87	0.61	0.72	0.51	47	47	47
13.0	0.85	0.83	0.79	0.74	0.68	0.45	50	50	50
13.9	1.02	0.99	0.94	0.89	0.78	0.54	60	52	52
13.6	1.04	1.00	0.96	0.89	0.80	0.55	60	50	50
10.4	0.66	0.94	0.80	0.75	0.66	0.49	47	47	47
5.6	0.70	0.66	0.65	0.61	0.54	0.44	47	47	47



- 提供成套的教学数据样例  
**样例**

→ **教学版提供成套教学工具：大数据组件、实训环境、实验指导以及实验样例。**

## 全面详实的手册



Hadoop  
安装指导



Flume  
安装指导



Zookeeper  
安装指导



Kafka  
安装指导



Rhinosql  
安装指导



Scala  
安装指导



Spark  
安装指导



Shark  
安装指导

## 引导式指导

### 一、配置 rhino 用户

新建 rhino 用户并为其设定密码。用户组需要归属于 root 组，才能配置 ssh 无密码登录。

```
[root@rhino001 ~]# useradd -d /home/rhino -m rhino
```

```
[root@rhino001 ~]# passwd rhino
```

### 二、设置 hosts

以 root 用户登录，配置主机名称：

```
[root@rhino001 ~]# vim /etc/hosts
```

```
127.0.0.1 localhost
```

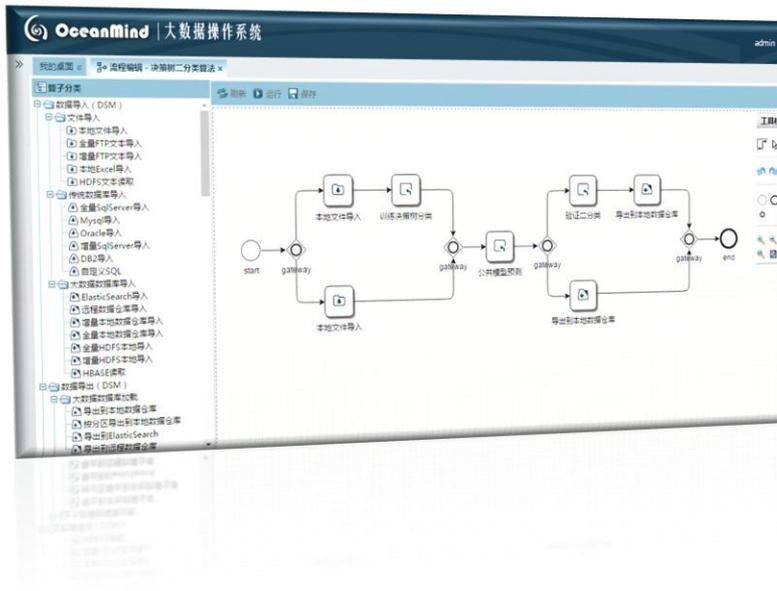
```
192.168.10.1 rhino001
```

### 三、创建无密码访问



通过实验指导手册，让学生轻松完成实训环境搭建，将所学理论用于实践。

Training实训版为学生提供可视化的大数据使用环境，通过拖拉拽的方式即可轻松完成业务模型建立、数据分析、数据挖掘全过程，并可通过行业案例加深对大数据的全面理解。



大数据  
可视化

大数据功能算子化

多个行业  
样例数据

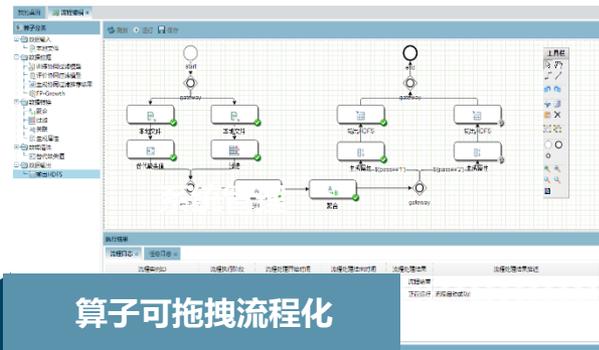
业务实现流程化

数据模型及  
业务分析案例

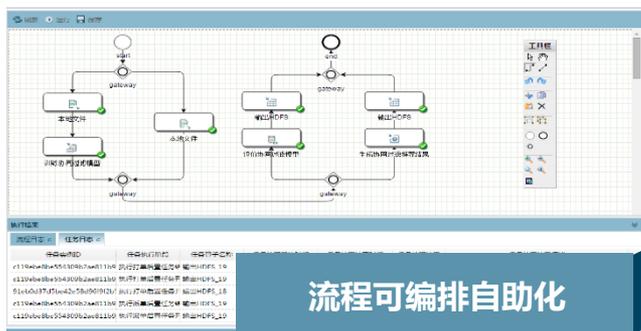
➔ 大数据实验环境搭建完成后，通过实训版操作使用：建立模型、数据分析、数据挖掘。



大数据功能算子化



算子可拖拽流程化



流程可编排自动化

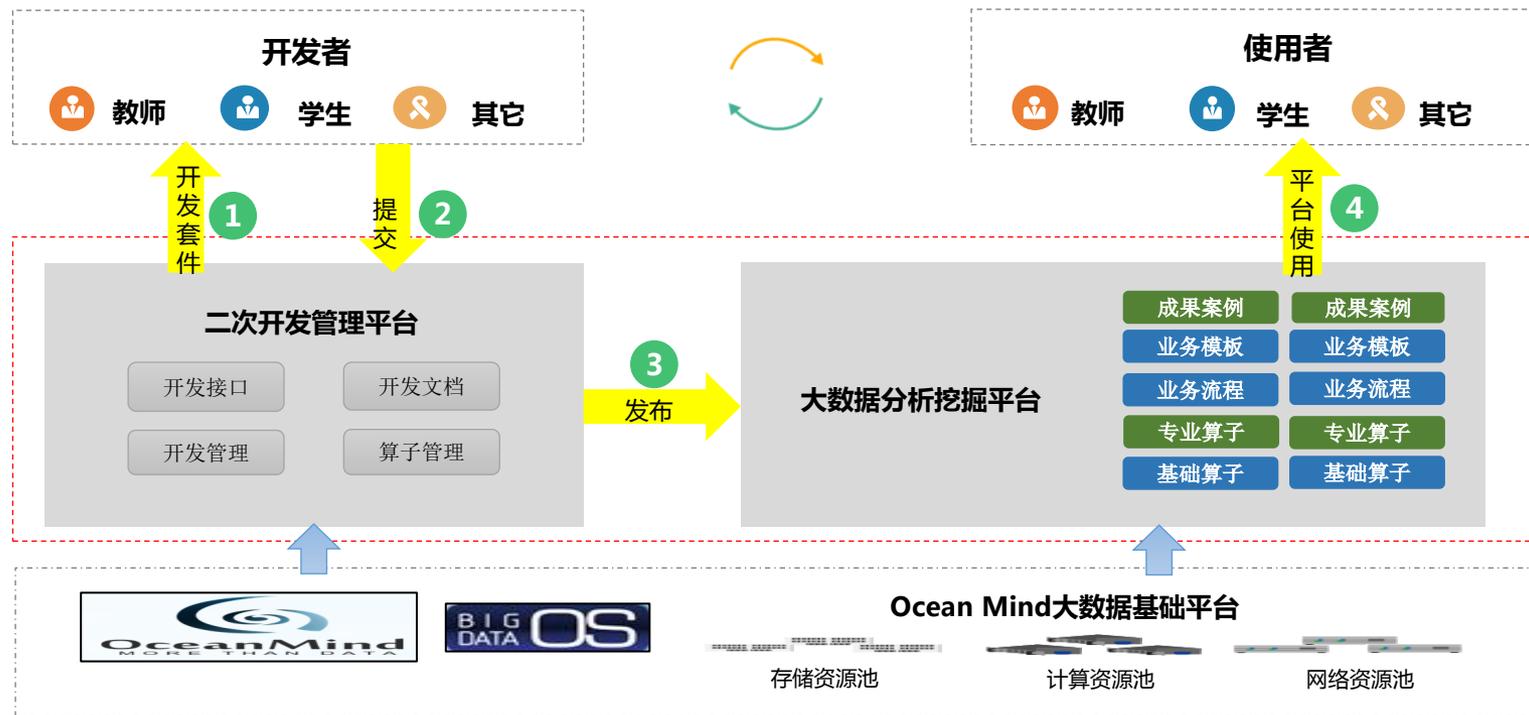


流程模板化可重用



实训版提供大数据可视化操作环境，学生上手快、功能全面、学习效果好。

Development开发版为有开发需求的高校提供一整套大数据开发和验证环境，通过开发版，学生可以将自己开发的功能算子发布至平台，并在平台进行运行测试验证，全面提升学生大数据开发能力。





我的算子 发布新算子

上传 算子信息 发布范围

1 2 3

1 创建算子

上传: 添加文件

下一步

流程管理

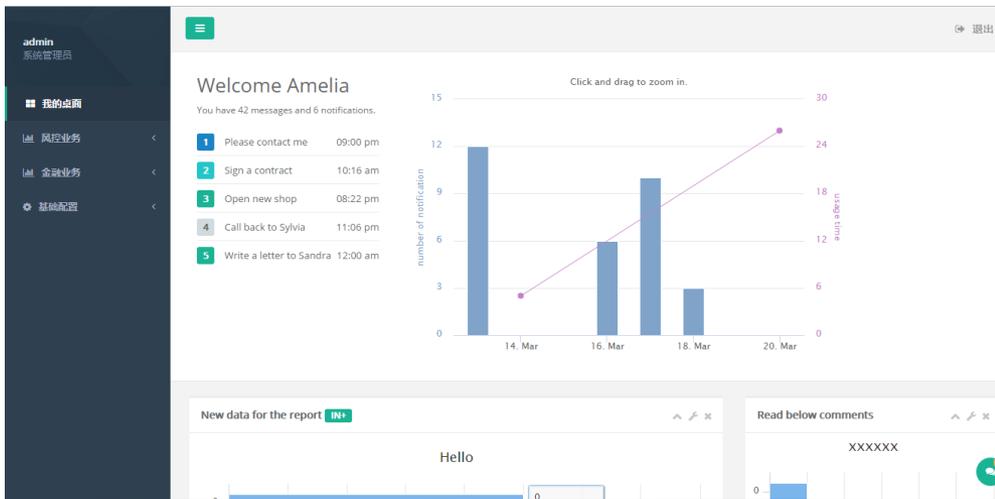
- 算子分类
  - 数据导入 (DSM)
    - 文件导入
    - 传统数据导入
    - 大数据数据导入
  - 数据导出 (DSM)
    - 传统数据导出
    - 大数据数据导出
  - 数据转换 (DAE)
    - 属性名称操作
    - 属性类型操作
    - 属性选择
    - 属性生成
    - 数据操作
    - 表操作
  - 数据清洗 (DAE)
    - 去重
    - 正则化
  - 数据控制 (DME)
    - 数据过滤
    - 数据项集
  - 分类
    - 训练均值聚类
    - 测试均值聚类
    - 分类和回归
    - 验证二分类
    - 验证多分类
    - 验证回归
    - 训练决策树分类
    - 决策树回归
    - 朴素贝叶斯分类
    - 训练线性回归
    - Logistic分类
    - 公共模型预测
  - 异常检测
    - Lof异常检测
    - cof异常检测
  - 统计
    - 训练核函数
    - 预测核函数
  - 实用工具 (UTIL)

运行

start 全量HDFS目录浏览 HDFS文本读取 数据分组 gateway 分组选择 训练协同过滤 gateway 推荐结果 验证协同过滤 HDFS本地导出 HDFS本地导出 end gateway

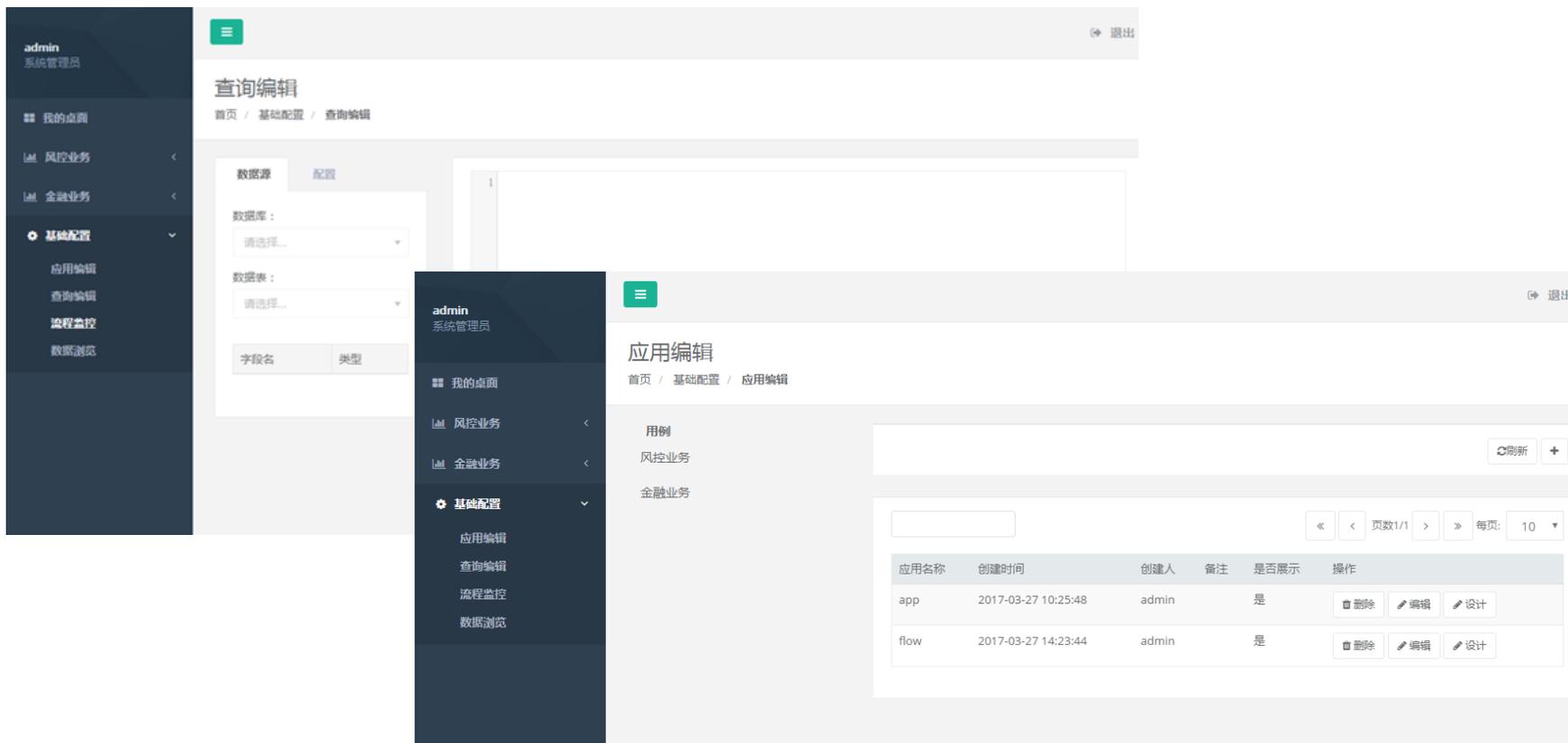
➔ 开发版为学生提供算子代码开发环境，可发布至平台，通过流程来测试验证。

Application业务版提供报表及页面生成工具，可以将业务模型的运行结果数据通过报表进行直观呈现，节约业务生成时间。



- 数据源灵活可选；
- 菜单生成自定义；
- 多种图表可选；
- 快速生成应用页面；

业务版为学生提供业务快速生成环境，加深对大数据应用的理解。

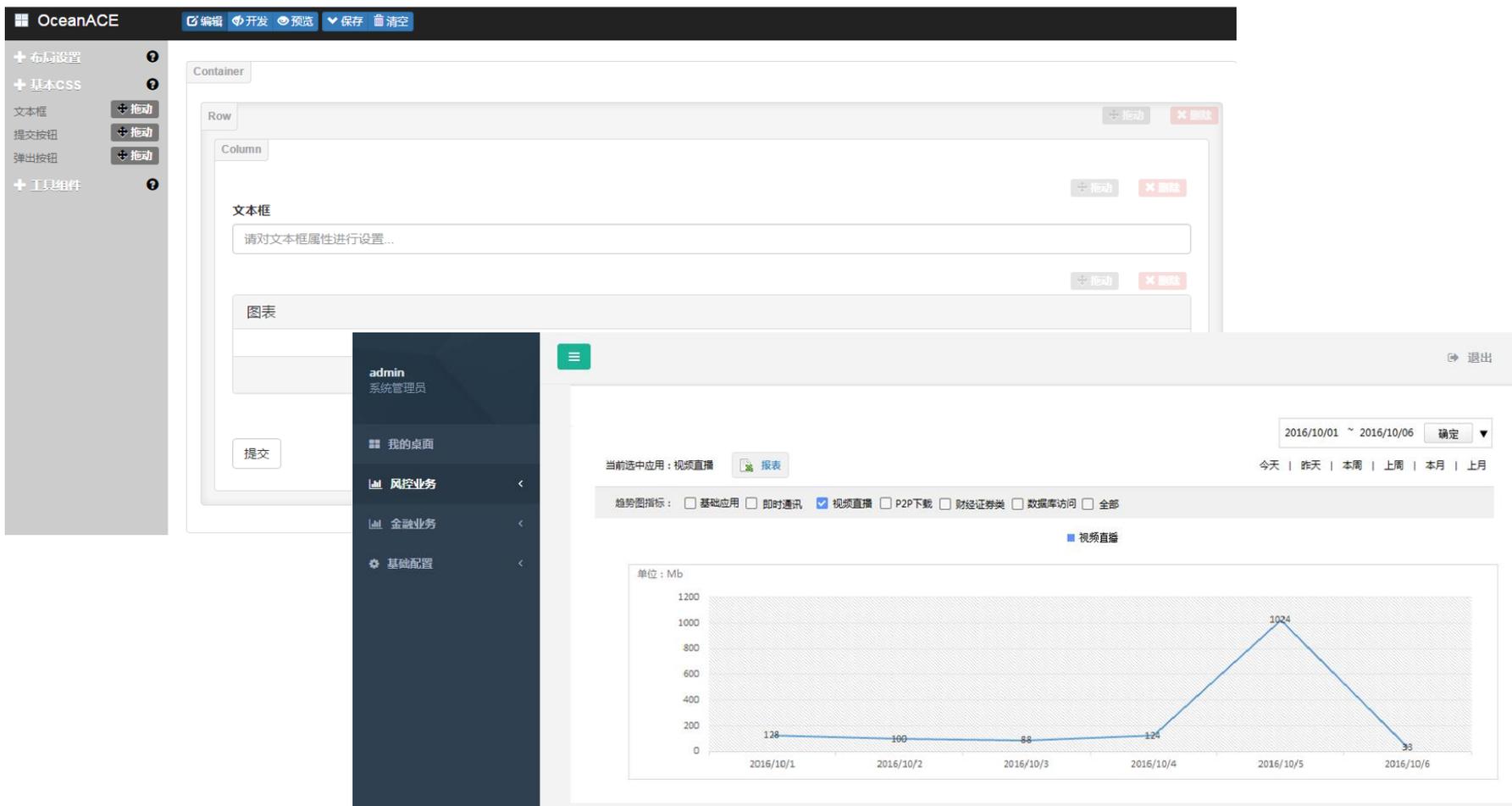


The image displays two screenshots of a web application interface. The top screenshot shows the 'Query Editing' (查询编辑) page, which includes a sidebar menu with options like '我的桌面', '风控业务', '金融业务', and '基础配置'. The main content area has a '数据源' (Data Source) section with dropdown menus for '数据库' (Database) and '数据表' (Table), and a '配置' (Configuration) section with a table for '字段名' (Field Name) and '类型' (Type).

The bottom screenshot shows the 'Application Editing' (应用编辑) page. It features a sidebar menu with '基础配置' expanded to show '应用编辑', '查询编辑', '流程监控', and '数据浏览'. The main content area has a '用例' (Use Case) section with '风控业务' and '金融业务' listed. Below this is a table of applications with columns for '应用名称' (Application Name), '创建时间' (Creation Time), '创建人' (Creator), '备注' (Remarks), '是否展示' (Whether to Display), and '操作' (Operations). The table contains two rows: 'app' and 'flow'. Each row has buttons for '删除' (Delete), '编辑' (Edit), and '设计' (Design).

应用名称	创建时间	创建人	备注	是否展示	操作
app	2017-03-27 10:25:48	admin		是	<a href="#">删除</a> <a href="#">编辑</a> <a href="#">设计</a>
flow	2017-03-27 14:23:44	admin		是	<a href="#">删除</a> <a href="#">编辑</a> <a href="#">设计</a>

➔ 通过应用编辑功能模块，灵活自主进行行业应用编辑。



The screenshot displays the OceanACE interface, which is used for generating business applications. It features a drag-and-drop editor on the left and a preview window on the right. The editor includes a sidebar with components like 'Text Box' (文本框) and 'Chart' (图表), and a main workspace with a 'Container', 'Row', and 'Column' structure. The preview window shows a dashboard with a line chart for 'Video Streaming' (视频直播) from 2016/10/1 to 2016/10/6. The chart shows a peak of 1024 Mb on 2016/10/5.

日期	视频直播 (Mb)
2016/10/1	128
2016/10/2	100
2016/10/3	88
2016/10/4	124
2016/10/5	1024
2016/10/6	33

➔ 通过业务生成功能模块，可自定义页面内容、报表格式等。

## 深耕教育行业，为高校提供全方位的解决方案



建设高校研究平台，  
促进产学研深度合作，  
促进成果转化。

+



优化整合学校各种管理信息系统和  
协同办公系统，完善高校决策支持  
系统的模型库、数据库和知识库建  
设，通过大数据分析，实现学校发  
展的智能决策。

+



实时监测校园网络，  
动态感知网络运行状  
态，预防和控制网络  
风险。



OceanMind  
MORE THAN DATA



4

总结



# THANKS



M O R E T H A N D A T A  
不 止 是 数 据



© SINOVATIO Corporation. All rights reserved.