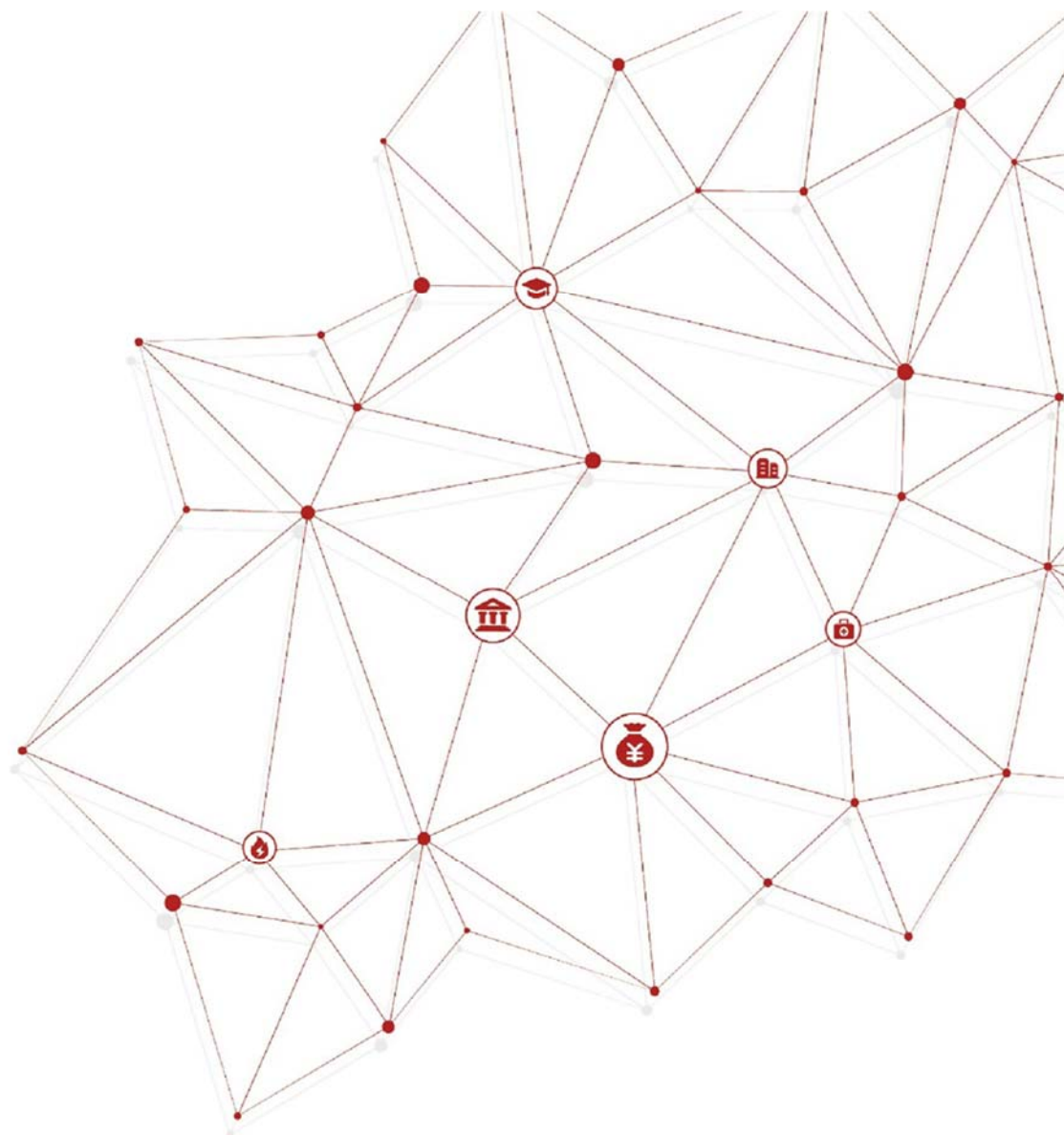


教育大数据平台 BI 简介



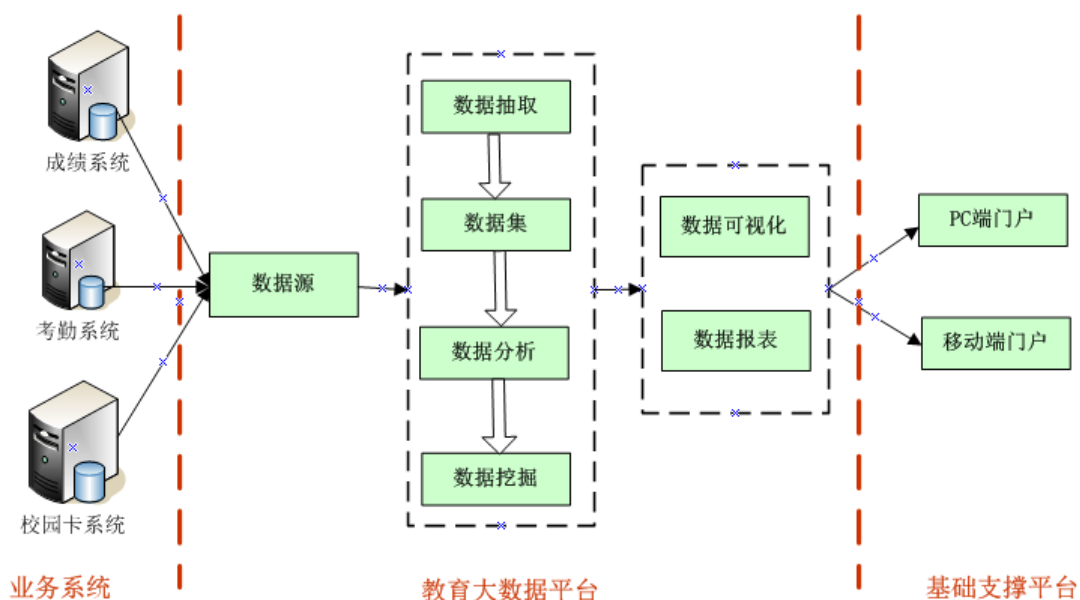
深圳市牛娃教育科技有限公司

2019 年 12 月

1. 产品简介

教育大数据平台 BI 产品是一套动态收集学校业务系统数据、对数据进行智能分析、辅助学校以数据决策的系统。功能包含设定多业务系统数据来源，从多业务系统动态抽取数据，建立数据模型，进行智能分析，分析结果以可视化数据展示和 office 报表输出。依据学校具体数据进行分析，无须写代码，全程操作采用可视化、拖拽式完成，适合所有学校使用。

教育大数据平台架构图



2. 系统架构组件

系统架构包含数据接入层、数据加工层、分析展示层、平台展示层，数据接入层为直接获得业务系统数据，数据加工为对原始数据清洗并获得有效数据，分析展示为依据数据模型分析并选择可视化展示，平台展示为将可视化结果与平台整合。

支持 CAS 或 LDAP 单点登录, 支持 webURL 方式办事大厅集成。



图:产品架构及组件

3、智能分析步骤

用户可以轻松地连接业务数据进行数据准备，通过可视化拖拽方式，轻松完成数据处理，并且使用自助分析工具，透视分析、自助仪表盘等功能进行可视化和分析，实现数据成果分享

(1) 数据准备:

设置数据来源，通过拖放的方式进行二次数据加工处理。如创建表关系、设置维度、数据过滤、数据抽取等，快速完成业务数据的准备工作。

名称: *	teacher_score
别名:	教务系统教师任课数
驱动程序类型: *	MySQL <input type="checkbox"/> 大数据量兼容
驱动程序类: *	com.mysql.jdbc.Driver
连接字符串: *	jdbc:mysql://10.10.10.204:3306/teacher_score_temp?useOldAliasMetadataBehavior=true&useUnicode=true&characterEncoding=GBK
验证类型:	<input checked="" type="radio"/> 静态 <input type="radio"/> 动态
用户名:	root
密码:	*****
数据库源字符集:	GBK
数据库目标字符集:	GBK
最大连接数:	100
校验语句:	获取连接时测试 <input type="text" value="SELECT 1 FROM DUAL"/>
事务隔离级别: *	JDBC默认值
引用标识符: *	<input type="checkbox"/> 应用于自助数据集
描述:	

测试连接(T) 保存(S) 关闭(C)

图:从教务系统抽取数据

名称:	teacher_info
别名:	人事系统教师基本信息
驱动程序类型:	MySQL <input type="checkbox"/> 大数据量兼容
驱动程序类:	com.mysql.jdbc.Driver
连接字符串:	jdbc:mysql://10.20.30.177:3306/teacher_info_temp?useOldAliasMetadataBehavior=true&useUnicode=true&characterEncoding
验证类型:	<input checked="" type="radio"/> 静态 <input type="radio"/> 动态
用户名:	root
密码:	*****
数据库源字符集:	GBK
数据库目标字符集:	GBK
最大连接数:	100
校验语句:	获取连接时测试 <input type="text" value="SELECT 1 FROM DUAL"/>
事务隔离级别:	JDBC默认值
引用标识符:	<input type="checkbox"/> 应用于自助数据集
描述:	

测试连接(T) 保存(S)

图:从人事系统抽取数据

教师基本信息主题

显示名称 显示别名 主题名: 教师基本信息主题 主题别名: 教师基本信息主题 描述:

人事系统教师基本信息

过滤器

教职工基本信息表

Ab 工号

Ab 姓名

Ab 性别

Ab 科目

Ab 学历

Ab 职称

出生年月

年龄

属性区 表关系视图

数据格式: <字符串-默认值>

参与分组: 自动

聚合方式: (只对透视分析有效)

关联属性: 请选择... 清除 (只对透视分析和组合分析有效)

时间计算: 请选择... 清除 (只对透视分析有效)

时间层次: 年 (只对透视分析新建时有效)

转换规则: 否

表达式: `convert(date_format(now(), '%Y')- date_format(出生年月, '%Y'), char)`

是否生成SQL: 是 否

地理维: 是 否

作为透视分析计算字段使用: 是 否

确定(O) 取消(C)

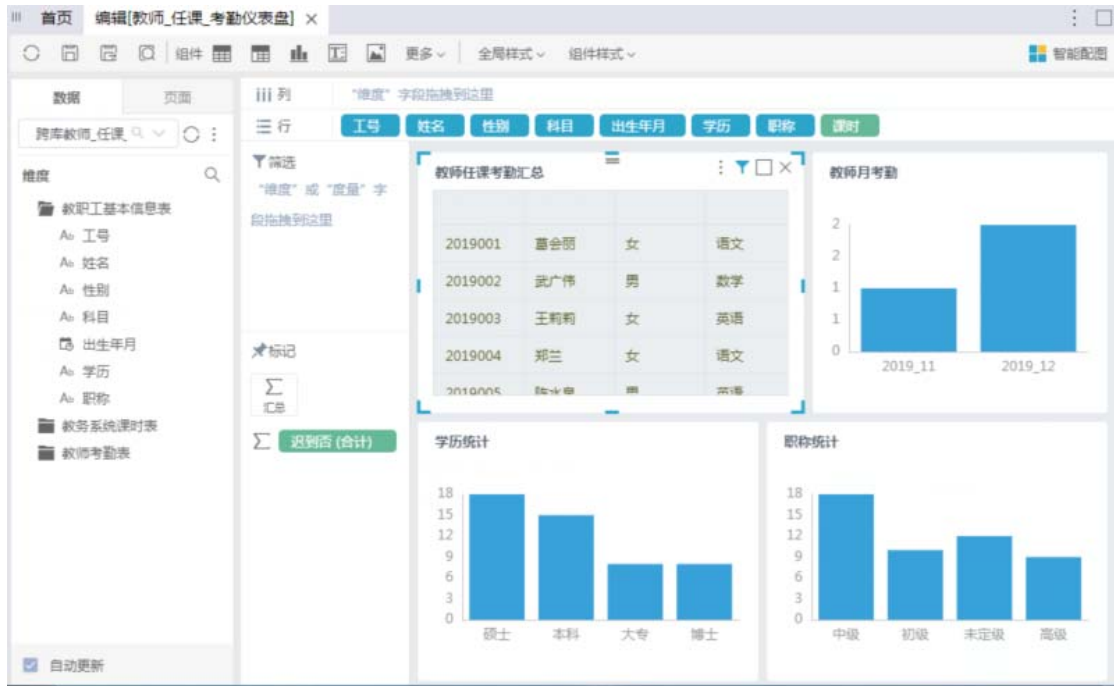
保存(S) 关闭(C)

依据出生年月计算年龄

图:数据过滤与设置维度

(2) 数据分析

利用已准备好的数据，支持 SQL 语言、可视化设置和算法等，对数据汇总、平均、排名、环比、累计、前期值等操作，提供丰富的图形资源库，报表图形任意切换。



图：自助仪表盘

(3) 数据展示

数据分析结果通过 PC、移动端提供不同用户查阅。



图：PC 结果展示

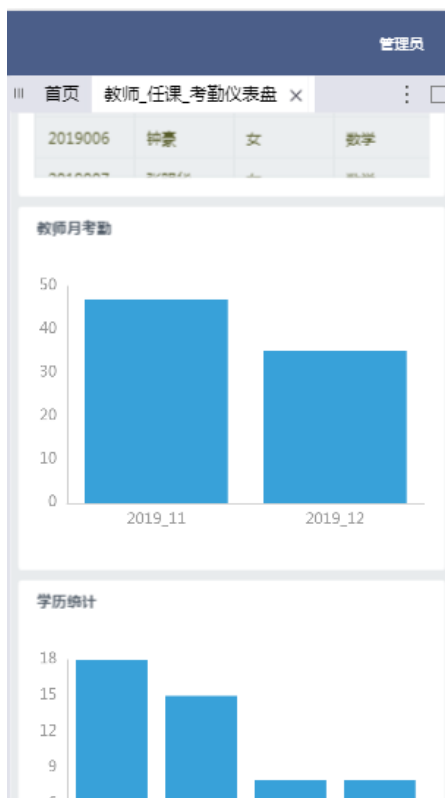


图:移动端展示

名称	别名	描述
Admins	管理角色	系统管理员
GroupAdmins	组管理角色	组管理角色
PowerUsers	高级用户角色	高级用户角色
SpreadsheetAuditing	ssr审核数据	
SpreadsheetPublish	ssr发布数据	
SpreadsheetSubmit	ssr校验数据	
Users	普通角色	普通角色
部门领导	部门领导	
财务主任	财务主任	
校级领导	校级领导	

图:用户查阅权限

4、数据挖掘，预测未来

数据挖掘平台为用户和学校提供拖拽式操作和可视化操作界面。用户只需要拖拽和组合各种数据源、数据预处理操作、机器学习算法、训练、预测和评估组件，即可完成复杂的机器学习任务。因此该平台可轻松地帮助学校快速把机器学习技术应用到业务系统中。

汇集 50+种数据挖掘算法组件，灵活建立业务模型流程，主要包含基本的数据特征处理、分类、聚类、关联、回归、深度学习算法，以及支持 Java 和 Python 算法扩展。



图:数据挖掘过程

(1)、机器学习的预测和评估

机器学习的预测是针对分类和回归问题，根据训练的模型，对测试集数据进行预测。对于得到的预测结果，对于不同类型的问题，使用不同的评估方法进行评估，以此评估模型的性能及有效性。

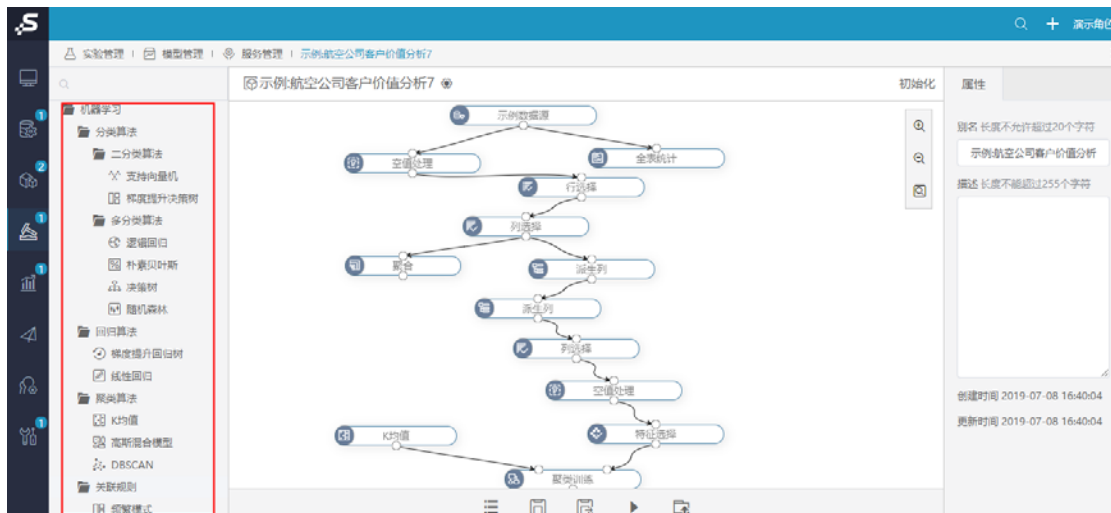
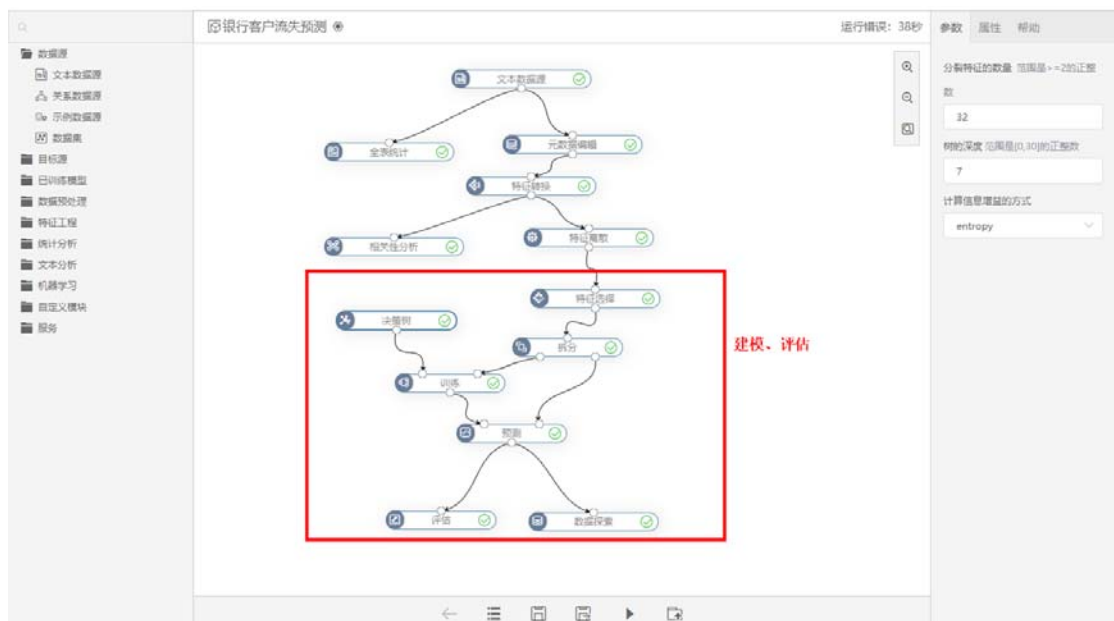


图:建模流程

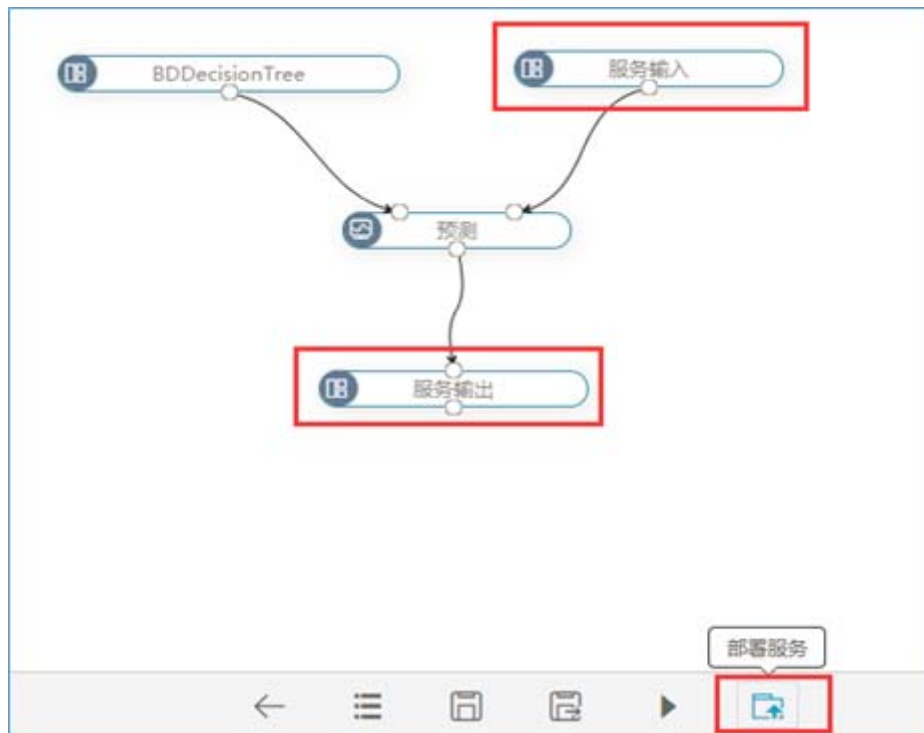
(2)、数据挖掘可视化建模

快速实现各种类型的数据挖掘应用，为学校所做的决策提供预测性智能。

平台具备流程化、可视化的建模界面，内置实用的、经典的统计挖掘算法和深度学习算法，业务人员可通过轻松拖拉拽组件的操作，进行可视化建模，完成模型流程的搭建，并能将模型发布管理。



图：托拉拽建模示例



图：模型部署示例

5、主要功能界面

(1)、基础功能

统一服务平台基础功能涵盖了服务器引擎、数据查询引擎、元数据引擎等一系列系统的底层服务，是系统平稳运行的基础。



图:功能菜单

(2)、数据源管理

丰富的数据源支持、强大的数据集管理和灵活的语义层。丰富的数据源支持极大地减轻了用户的应用难度，方便了用户的使用。



图：不断扩展的主要数据源支持清单



图：支持数据源

(3)、业务主题

业务主题是数据分析师基于业务场景来进行逻辑建模，将维度、指标、表关系进行定义，将数据库中晦涩难懂的字段与表关系转换成业务人员也可以看得明白的知识。并且这部分也是拖拽式完成无需编写代码并且可以控制权限。例如：基于成绩分析的业务主题，包含了总分、平均分、增长率、下降率等业务属性。教务人员能使用这个成绩分析主题，进行探索分析。



图：业务主题示例

(4)、数据集

数据集是定义报表和图形的基础，透视分析、Echarts 图形、电子表格都是基于数据集创建的。可以基于单数据源或者多维数据源创建的一类数据集，可以采用原生 SQL 语言、存储过程等，面向用户提供数据查询和抽取服务，以产生图形展示或者报表输出。

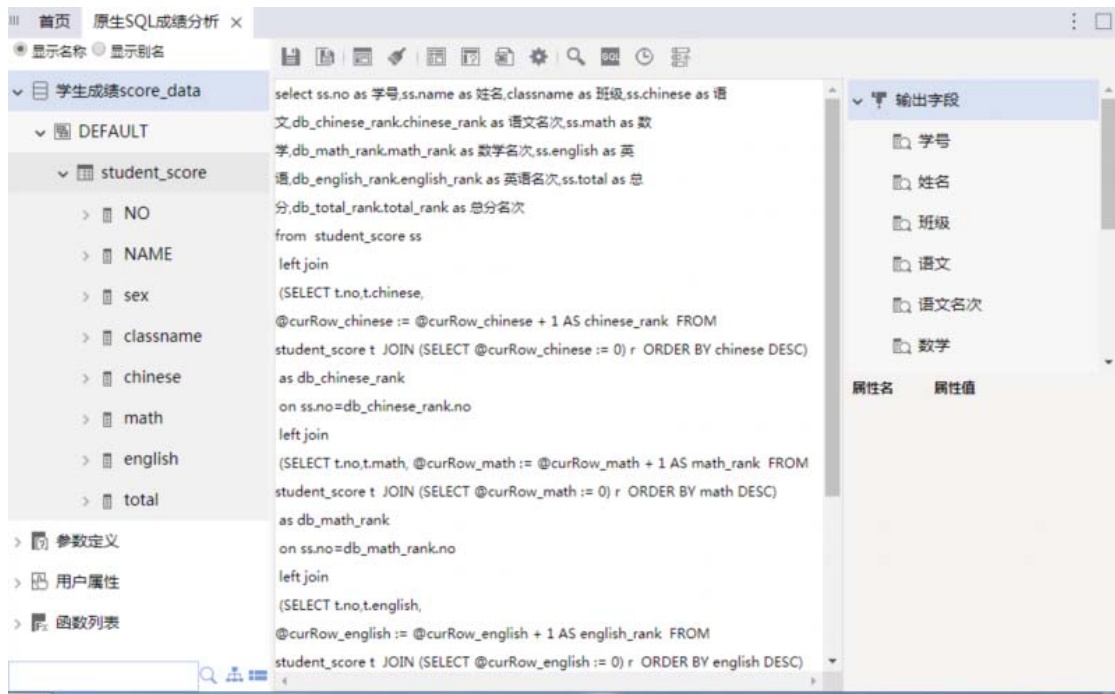


图: SQL 语言构建数据集

(5)、数据展示

BI 作为学校的非核心业务系统，在实际应用部署时，往都会涉及到图表资源在学校门户系统集成的需求。



图：登录门户示例

移动驾驶舱

随着信息化时代的来临，时间成本对于学校的重要性也越来越高，在重要时刻的及时反应甚至可以帮助学校获得更多的回报率，而“移动办公”也已经成为了学校管理者乐于见到的主流办公方式。



图：移动端打开报表

数据报告

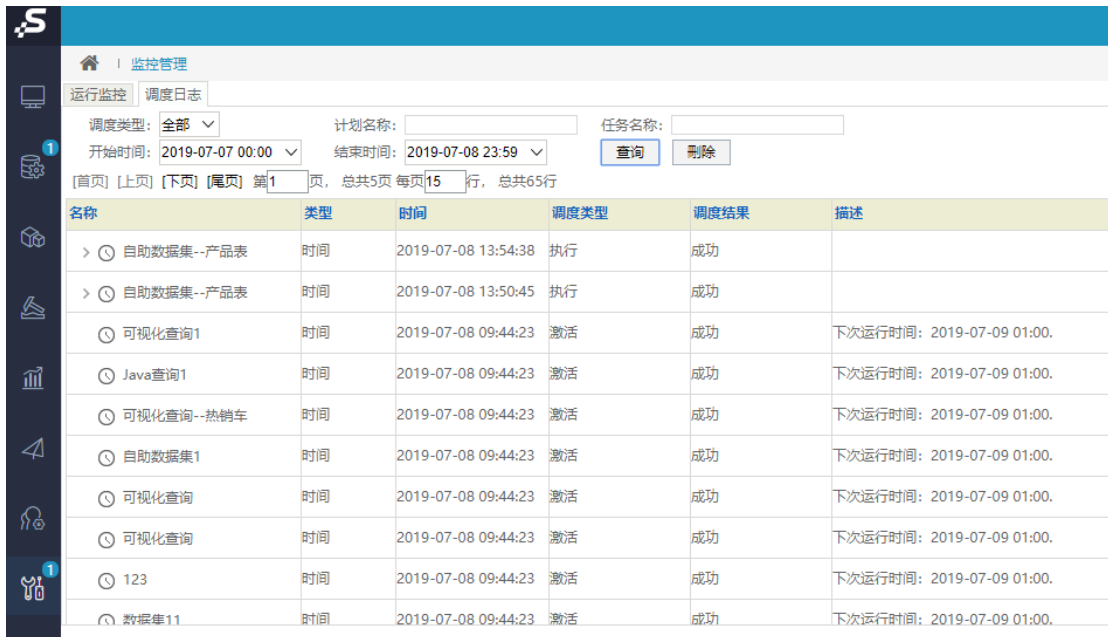
提供丰富的图形资源库，报表图形任意切换，且不受维仪式、度量的限制。基于业务主题，直接创建自助仪表盘。



图：自助仪表盘示例（一）

（6）、计划任务

在日常工作中，经常需要制作业务周期报告，如每月教职工和学生出勤报告，并发送给领导审阅；又如业务系统一般每天进行数据的抽取和导入，所以每天预定时间，学校各级管理人员都需要一份分析报告。BI 提供的计划任务功能，就是帮助用户解决这些问题的。利用该功能，可以将需要发出的报表定时进行运行和发送，支持移动端发送。



名称	类型	时间	调度类型	调度结果	描述
> 自助数据集--产品表	时间	2019-07-08 13:54:38	执行	成功	
> 自助数据集--产品表	时间	2019-07-08 13:50:45	执行	成功	
可视化查询1	时间	2019-07-08 09:44:23	激活	成功	下次运行时间: 2019-07-09 01:00.
Java查询1	时间	2019-07-08 09:44:23	激活	成功	下次运行时间: 2019-07-09 01:00.
可视化查询--热销车	时间	2019-07-08 09:44:23	激活	成功	下次运行时间: 2019-07-09 01:00.
自助数据集1	时间	2019-07-08 09:44:23	激活	成功	下次运行时间: 2019-07-09 01:00.
可视化查询	时间	2019-07-08 09:44:23	激活	成功	下次运行时间: 2019-07-09 01:00.
可视化查询	时间	2019-07-08 09:44:23	激活	成功	下次运行时间: 2019-07-09 01:00.
123	时间	2019-07-08 09:44:23	激活	成功	下次运行时间: 2019-07-09 01:00.
数据集11	时间	2019-07-08 09:44:23	激活	成功	下次运行时间: 2019-07-09 01:00.

图：监控管理调度日志示例

(7)、权限管理

需要值得注意的是，学校里面数据安全是一个必须关注的点，在学校的数据分析应用门户中，每一个用户需要在自己相应的权限范围内查看和获取数据。BI 权限管理支持控制用户功能权限、数据访问权限、资源访问权限，为学校数据安全保驾护航。



资源	拥有的权限	应用到	操作
> 自助ETL	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	该目录、子目录及文件	授权
> 数据集	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	该目录、子目录及文件	授权
> 数据挖掘	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	该目录、子目录及文件	授权
> 分析报表	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	只有该目录	授权
> 资源定制	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	该目录、子目录及文件	授权
> 数据管理	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	该目录、子目录及文件	授权
> 资源发布	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	该目录、子目录及文件	授权
> 公共设置	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	该目录、子目录及文件	授权
> 计划任务	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	该目录、子目录及文件	授权
用户管理	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	该目录、子目录及文件	授权
系统运维	<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	该目录、子目录及文件	授权

图：资源权限管理

6、 教育应用场景

对已经有一定数据积累的大专院校、中小学、教育局等单位，可以采用本大数据平台分析工具，对数据进行发掘和分析，通过数据辅助学校决策。具体略举如下：

(1). 数据汇聚展示

将学校多系统数据汇聚展示，提供学生一张表、教职工一张表、领导校情一张表。实现整体与个体数据一目了然。

(2). 学生异常重点关注

依据学生成绩数据、考勤数据、消费数据、日常评语数据等进行分析，预测异常学生行为，为学生管理人员提供预警支持，对异常学生及早关注，避免意外发生。

(3). 教学诊断

依据学校师资数据、经费数据、图书数据、实验室及学生实践数据、课程设置、科研数据等进行教学诊断与改进分析，弥补教学与管理薄弱环节，为领导决策提供数据依据。

(4). 教师评价

依据教师所教学生考试成绩数据、教师考勤数据、教师科研数据等进行大数据分析，对教师进行绩效评价，杜绝“人情”评优。

(5). 高考预测

依据学校历届考试成绩数据，通过数据的分析，以数据预测高考成绩，对不及预期的科目，及早采取补救措施，以数据决策，避免粗象估计。

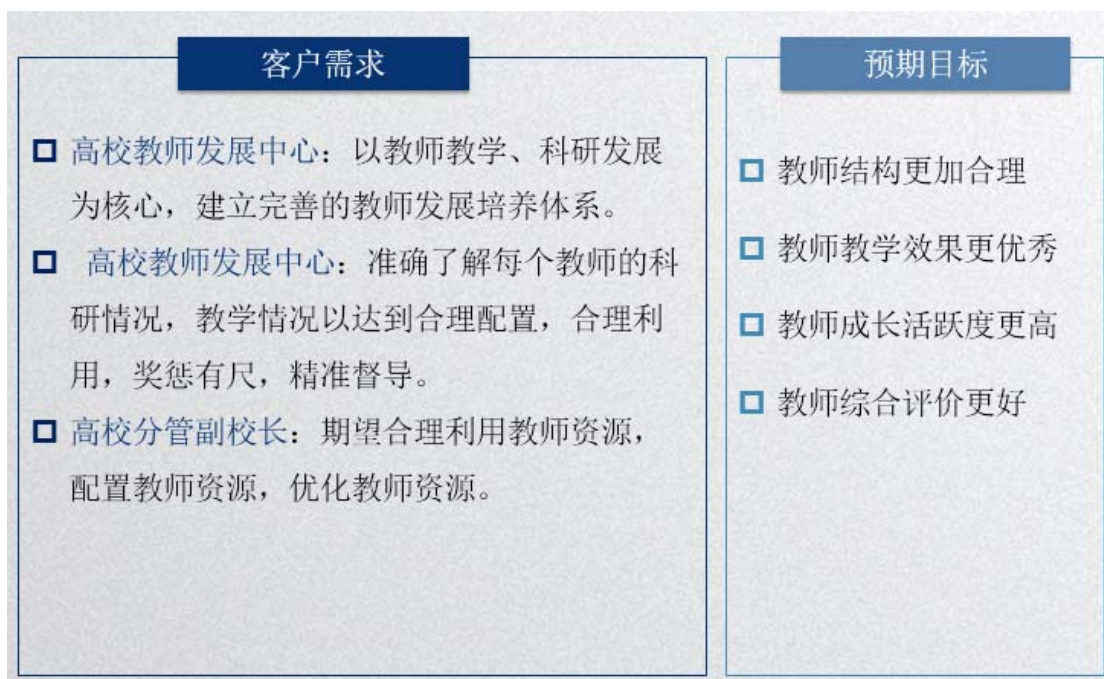
(6). 专业设置

依据学生就业反馈数据（如：就业率、工资待遇等）进行大数据分析，制定学校招生计划（专业设置、专业招生人数），避免专业设置与市场需求脱节。

7、 实施案例:教师画像

7.1、 确定需求

与学校沟通有哪些数据，与学校讨论依据数据能帮用户做哪些决策，确定需求和目标。



7.2、 确定建设思路



7.3、抽取数据

10.10.10.204 teacher_info_temp 运行 停止 解释

```
1 SELECT * from teacher_info
```

人事系统教职工基本信息

NO	NAME	sex	course	birthday	education	title
2019001	葛会丽	女	语文	1975-11-02 00:00:00	本科	高级
2019002	武广伟	男	数学	1985-12-01 00:00:00	硕士	中级
2019003	王莉莉	女	英语	1992-12-01 00:00:00	博士	初级
2019004	郑兰	女	语文	1965-12-01 00:00:00	大专	未定级
2019005	陈水泉	男	英语	1975-11-02 00:00:00	本科	中级
2019006	钟豪	女	数学	1985-12-01 00:00:00	本科	中级
2019007	张明华	女	数学	1992-12-01 00:00:00	硕士	中级
2019008	李玉洁	男	英语	1965-12-01 00:00:00	本科	初级
2019009	晏德安	女	英语	1975-11-02 00:00:00	硕士	高级
2019010	刘俊	男	语文	1985-12-01 00:00:00	博士	中级
2019011	漆尔祥	男	英语	1992-12-01 00:00:00	大专	初级
2019012	李金标	男	数学	1965-12-01 00:00:00	硕士	未定级
2019013	于海艳	男	数学	1980-12-01 00:00:00	硕士	中级
2019014	李建娜	女	数学	1980-12-01 00:00:00	硕士	中级
2019015	王素江	女	语文	1975-11-02 00:00:00	本科	高级
2019016	陈青	男	英语	1985-12-01 00:00:00	硕士	中级
2019017	马梦涵	女	数学	1992-12-01 00:00:00	博士	初级

10.20.40.101 teacher_score_temp 运行 停止 解释

```
1 SELECT * from teacher_score
```

教务系统教师任课数量

Id	NO	NAME	score_count
1	2019001	葛会丽	8
2	2019002	武广伟	9
3	2019003	王莉莉	10
4	2019004	郑兰	12
5	2019005	陈水泉	13
6	2019006	钟豪	14
7	2019007	张明华	8
8	2019008	李玉洁	9
9	2019009	晏德安	10
10	2019010	刘俊	10

10.20.30.177 teacher_kaoqin_temp 运行 停止 解释

```
1 SELECT * from teacher_kaoqin
```

考勤系统教师考勤

信息	结果 1	剖析	状态	
Id	NO	NAME	kaoqindate	status
1	2019001	葛会丽	2019-12-01 00:00:00	0
2	2019002	武广伟	2019-12-01 00:00:00	0
3	2019003	王莉莉	2019-12-01 00:00:00	1
4	2019004	郑兰	2019-12-01 00:00:00	1
5	2019005	陈水泉	2019-12-01 00:00:00	0
6	2019006	钟豪	2019-12-01 00:00:00	0
7	2019007	张明华	2019-12-01 00:00:00	0
8	2019008	李玉洁	2019-12-01 00:00:00	0

7.4、数据分析

(1)、分析思路



(2)、多纬度分析：



(3)、分析算法

教师学缘结构分析

1 指标算法

1 指标算法	1.1指标解释	教师学缘范围分布
	1.2来源表单	1-6-1教职工基本信息
	1.3数据过滤	专任教师

2 数据分析

学缘构成即学缘结构是指全体人员完成某一级学历的学校的构成状态。学缘结构包括毕业高校院所、所学专业等在类型、层次、分布等方面的构成情况。根据高校的实际情况，学缘可以细分，毕业院校，专业类型，学科门类等多个指标体系。

学校学缘分布如下：

学缘情况	人数	比例
本校	[Value]	[Value]
外校（境内）	[Value]	[Value]
外校（境外）	56	[Value]

合理的学缘结构应该是‘远缘杂交’，也就是说，高校的教师应该来自于多个高校，而不应该出自于一个多几个高校。教师来源结构的复杂性有利于不同学术风格和思想的相互渗透和竞争，活跃学术气氛。学缘结构与学校的文化观念有很大的相关性。长期以来，国内有部分院校不重视学缘的结构组成，导致学缘结构单一，‘近亲繁殖’现象非常严重。

根据学校的教职工相关数据，学科门类学缘分布如下：

教职工学科类别	专任教师数量	学科类别所含校内本科专业数量	本科生总数	本科生和同学科类别教师数比例
管理学	3	10	26	7.30

下钻到学校单位的数据如下：

从下表中的数据可以看出，一共有10个单位高级职称比例低于学校平均水平，有4个单位高级职称专任教师比例低于30%，其中有2个单位无正高级职称专任教师。

序号	单位名称	单位类型	正高级		副高级		中级		初级		未评级	具有高级职称比例(%)
			人数	比例(%)	人数	比例(%)	人数	比例(%)	人数	比例(%)		
1	工程训练中心	学院	0	0.00	2	11.76	9	52.94	6	35.29		0.00
2	航海技术	学院	4	3.36	14	11.76	68	57.14	28	23.53		
3	港口学院	学院	1	3.33	5	16.67	17	56.67	7	23.33		
4	船舶工程学院	学院	0	0.00	10	27.03	12	32.43	3	8.11	12	32.43
5	轮机工程学院	学院	2	3.70	15	27.78	35	64.81	2	3.70	0	0.00
6	计算机科学学院	学院	9	9.57	22	23.40	61	64.89	2	2.13	0	0.00
7	航海工程学院	学院	3	2.88	32	30.77	62	59.62	7	6.73	0	0.00
8	航海工程学院	学院	23	18.11	27	21.26	75	59.06	1	0.79	1	0.79
9	土木工程学院	学院	5	7.46	22	32.84	40	59.70	0	0.00	0	0.00
10	管理学院	学院	11	12.36	27	30.34	51	57.30	0	0.00	0	0.00
11	航海工程学院	学院	5	9.26	20	37.04	24	44.44	5	9.26	0	0.00
12	船舶工程学院	学院	27	20.30	36	27.07	69	51.88	1	0.75	0	0.00
13	航海学院（航海工程学院）	学院	4	6.56	25	40.98	18	29.51	13	21.31	1	1.64
14	航海工程学院（航海工程学院）	学院	17	17.71	31	32.29	39	40.63	8	8.33	1	1.04
15	管理学院	学院	4	7.84	24	47.06	21	41.18	1	1.96	1	1.96
16	航海工程学院	学院	14	20.59	24	35.29	27	39.71	2	2.94	1	1.47

职称解读，各单位的结构问题

(4)、教师活跃度分析



(5)、教师教学分析



7.5、分析结果展示



8、应用场景

8.1、领导驾驶仓



8.2.师资大数据分析



8.3.学生大数据分析



8.4.科研大数据分析



8.5.教务大数据分析



8.6. 招录大数据分析



8.7. 就业大数据分析



9、部署环境

序号	组件	要求
1.	处理器	处理器推荐 8 核或以上
2.	Java 内存	内存推荐 16G 或以上
3.	磁盘空间	200G
4.	操作系统	支持下列操作系统： <ul style="list-style-type: none"> Windows 2003/2008/2012/XP/Win7/Win8/Win10 Linux
5.	支持数据源	支持下列关系数据库：

序号	组件	要求
		<ul style="list-style-type: none">• DB2 V9• Informix• MS SQL Server• MySQL• Oracle• PostgreSQL• Sybase• 等常见数据