

5、测试数据挖掘及其组件

- 1. 配置连接信息
- 2. 运行数据挖掘示例
 - 2.1. 测试数据挖掘执行引擎
 - 2.2. 测试数据挖掘服务引擎
 - 2.3. 测试数据挖掘的Python计算

本节介绍smartbi连接数据挖掘、Spark、Hadoop、Python以及测试服务是否正常运行。

文档环境

单机部署数据挖掘组件环境如下：

服务器IP	主机名	组件实例	部署目录
10.10.204.248	10-10-204-248	数据挖掘	/data
10.10.204.249	10-10-204-249	Spark, Hadoop	/data
10.10.204.250	10-10-204-250	Python	/data

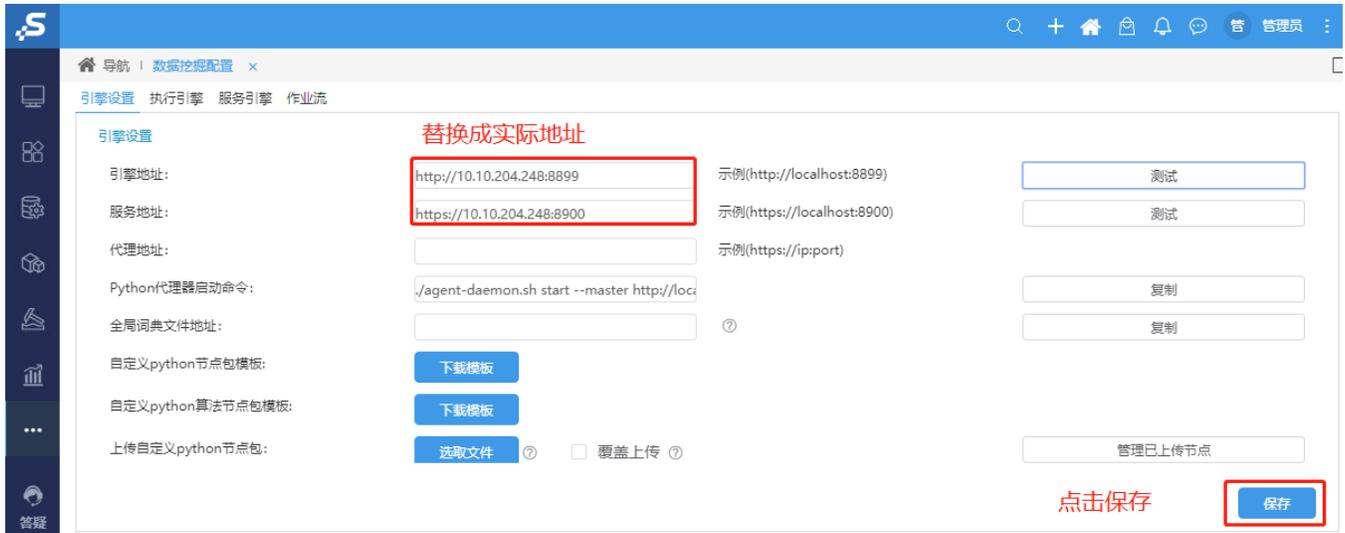
请根据实际部署环境替换相关的配置信息。

1. 配置连接信息

前提条件

配置数据挖掘连接信息前，请确认数据挖掘已正常部署和启动。部署文档参考：[部署数据挖掘](#)

① 浏览器访问Smartbi，打开系统运维 - 数据挖掘配置 - 引擎设置，引擎地址和服务地址设置成正确的数据挖掘引擎地址，并[点击保存](#)



注意事项

修改引擎和服务地址后，需要点击保存，否则执行引擎和服务引擎可能会显示空白页

配置数据挖掘地址后，首次测试引擎地址，会提示“平台到引擎连接成功，引擎到平台连接失败，请检查引擎配置”

http://10.10.204.248 ✖ 平台到引擎连接成功, 引擎到平台连接失败, 请检查引擎配置 ✕ 测试

https://10.10.204.248:8900 示例(https://localhost:8900) 测试

示例(https://ip:port)

首次测试服务地址则会提示“平台到服务连接成功, 服务到平台连接失败, 请检查服务配置”

http://10.10.204.248 ✖ 平台到服务连接成功, 服务到平台连接失败, 请检查服务配置 ✕ 测试

https://10.10.204.248:8900 示例(https://localhost:8900) 测试

示例(https://ip:port)

② 打开系统运维 - 数据挖掘配置 - 执行引擎 - 引擎配置, 参考下图修改smartbi地址、hadoop地址(根据实际环境修改), 修改完成后点击保存:



前提条件

配置Hadoop连接信息前, 请确认Hadoop节点已正常部署和启动。部署文档参考: [部署Hadoop](#)

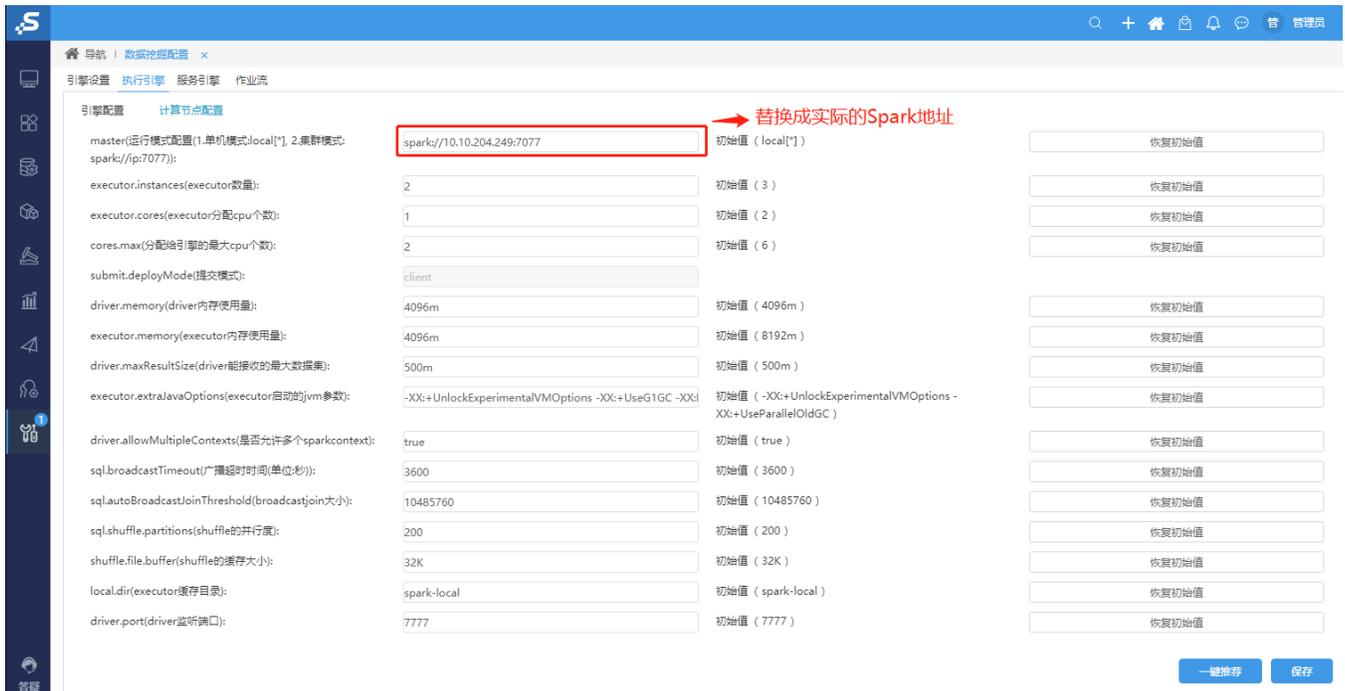
③ 打开系统运维 - 数据挖掘配置 - 执行引擎 - 计算节点配置, 参考下图设置, 修改完成后点击保存



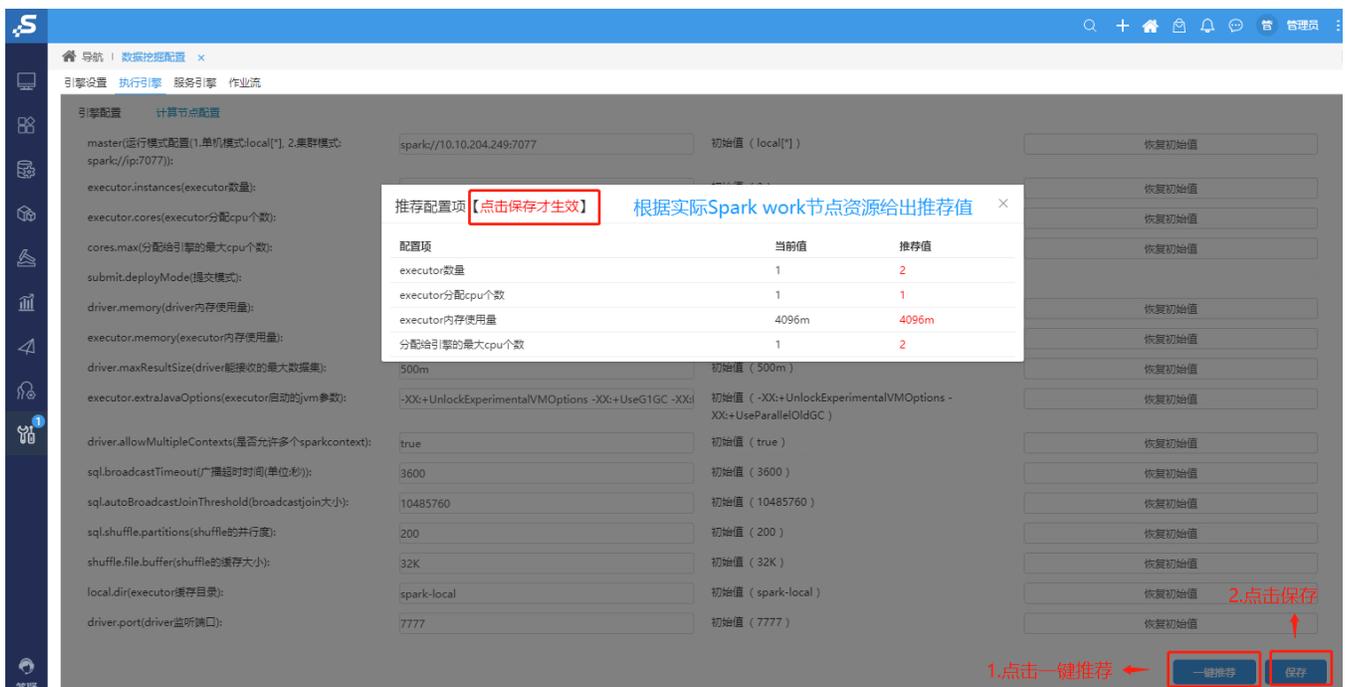
前提条件

配置Spark连接信息前, 请确认Spark节点已正常部署和启动。部署文档参考: [部署Spark3.1](#)

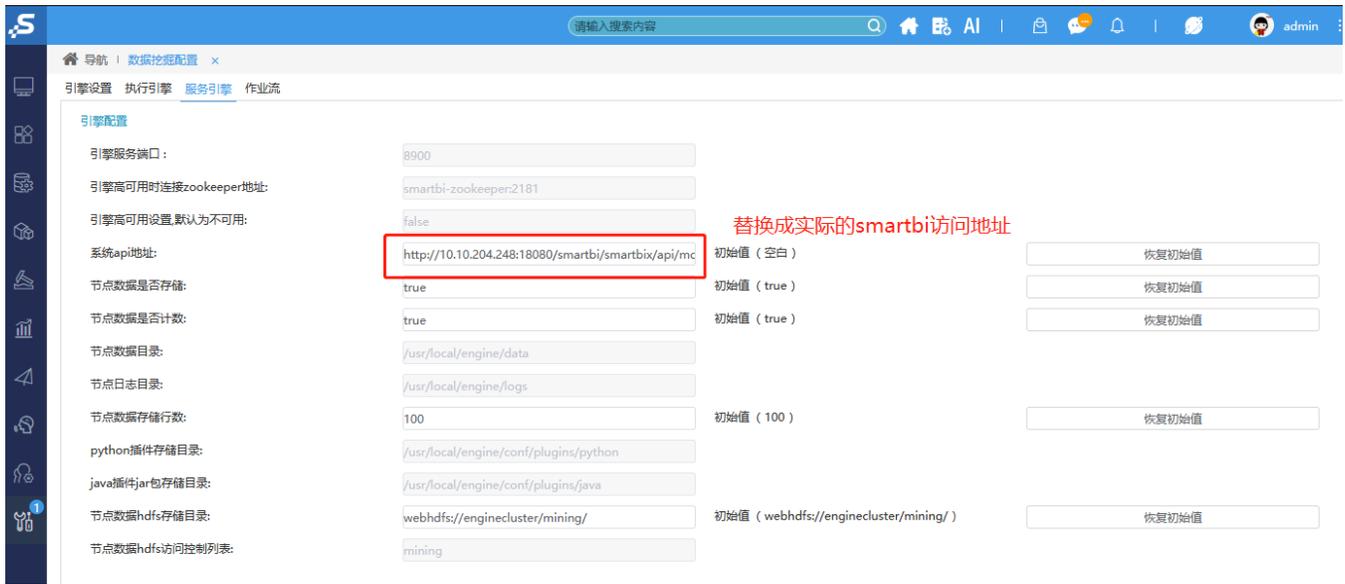
配置spark计算节点:



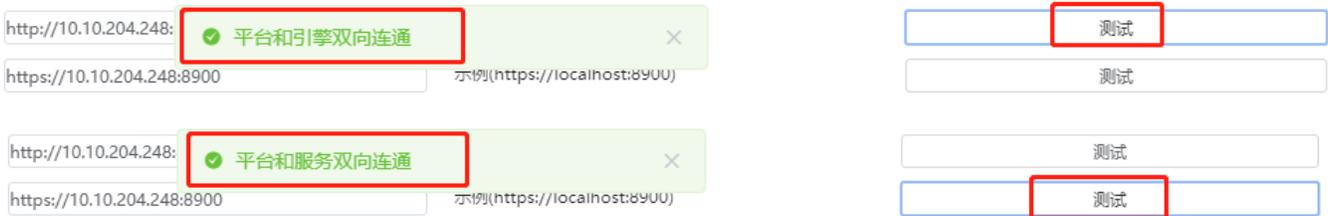
配置Spark节点资源，点击一键推荐，系统会根据Spark work节点的服务器资源，生成推荐的配置(如果使用推荐值，记得点击保存，否则配置不生效)：



④ 打开系统运维 - 数据挖掘配置 - 服务引擎，参考下图修改smartbi地址(根据实际环境修改)，修改完成后点击保存：



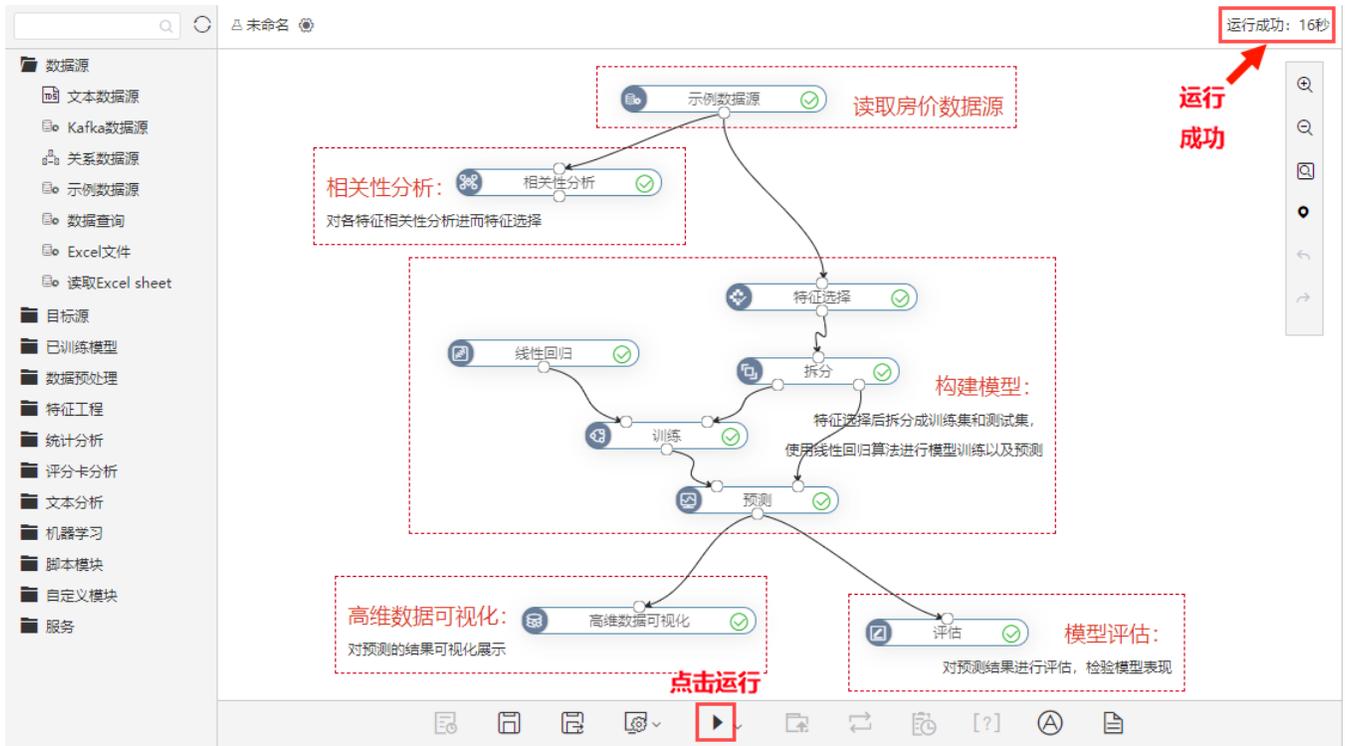
⑤ 执行引擎和服务引擎完成设置、保存后，返回 **系统运维 - 数据挖掘配置 - 引擎设置**，重新点击测试，如下提示表示配置成功：



2. 运行数据挖掘示例

2.1. 测试数据挖掘执行引擎

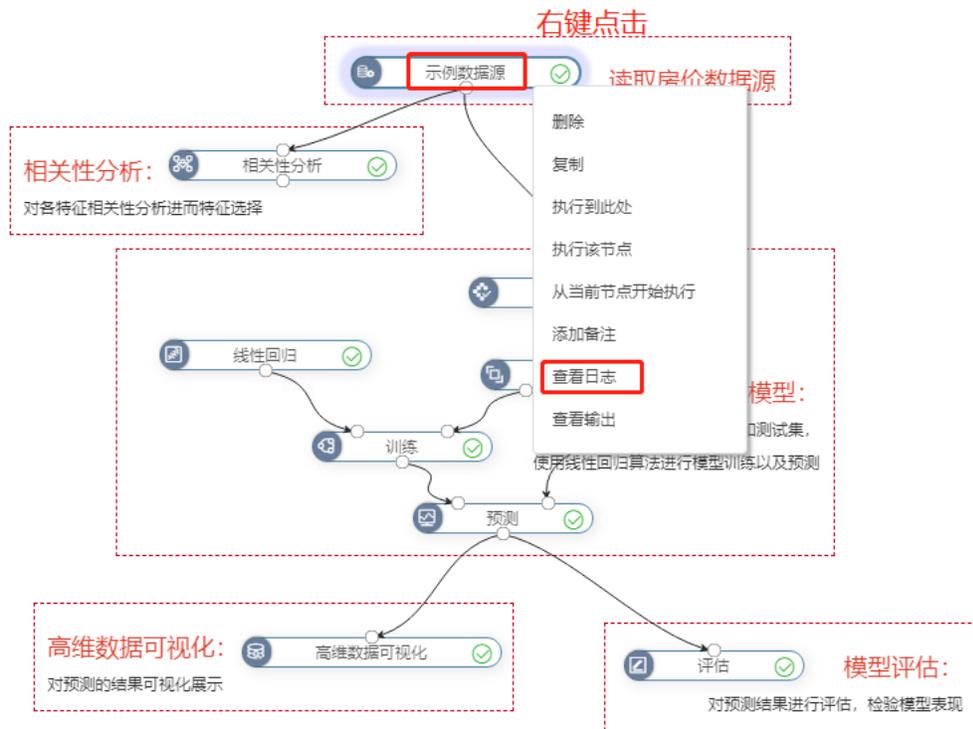
打开**数据挖掘 - 案例一波士顿房价预测**，点击运行，如下图显示运行成功即可



如果某个节点运行失败，可以右键点击节点，选择查看日志分析错误原因：

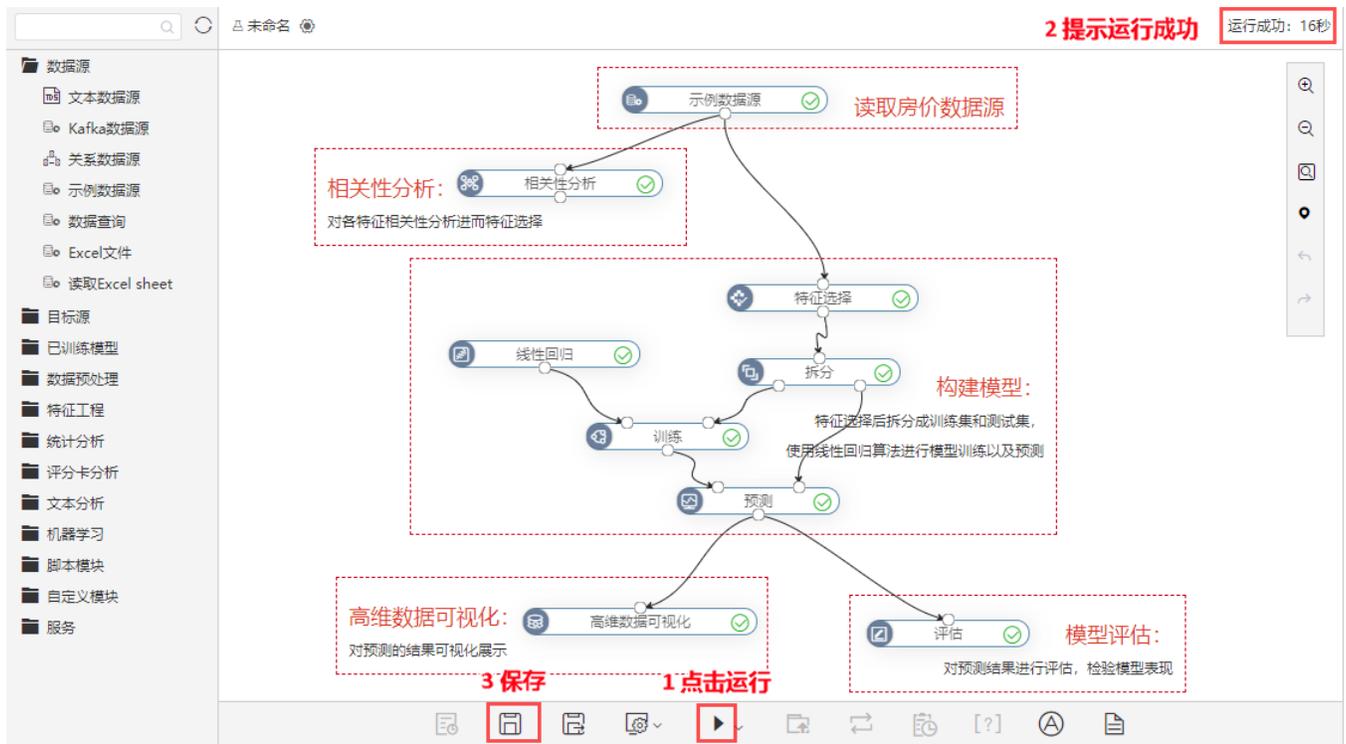
未命名

运行成

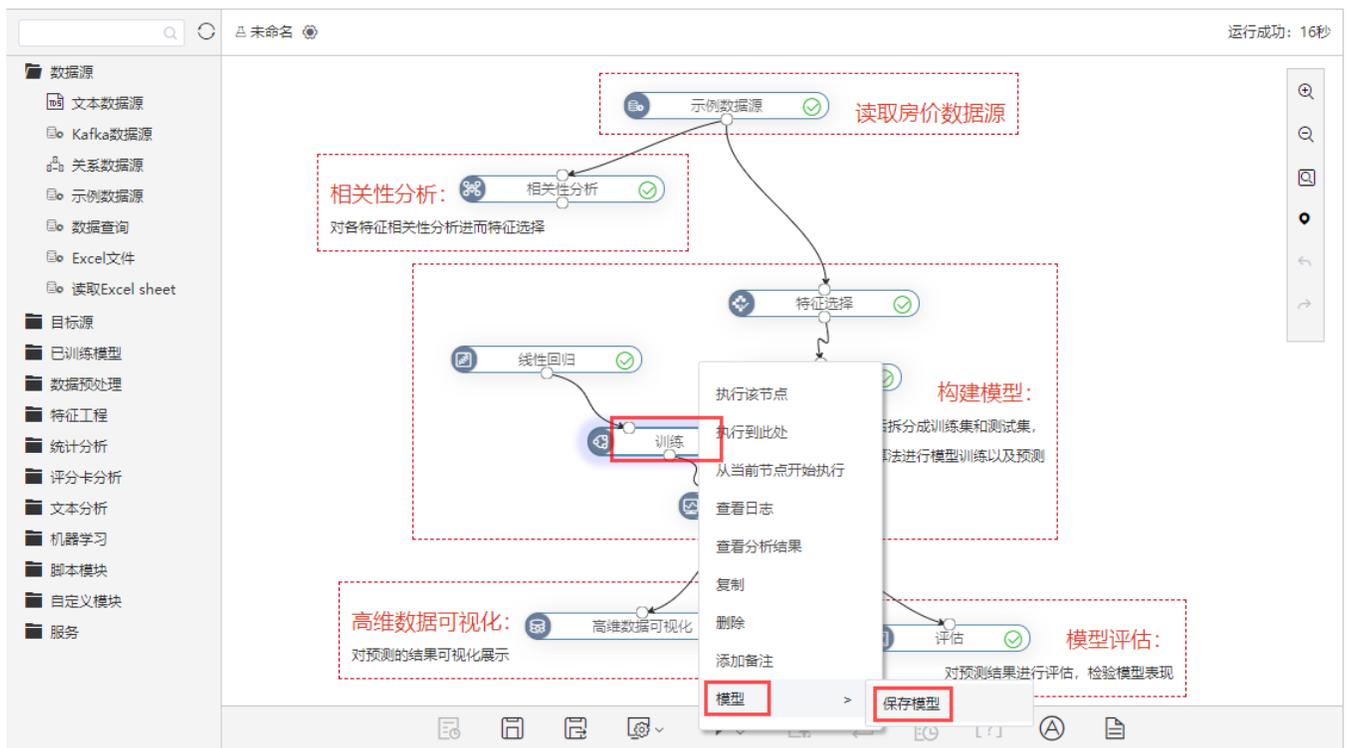


2.2. 测试数据挖掘服务引擎

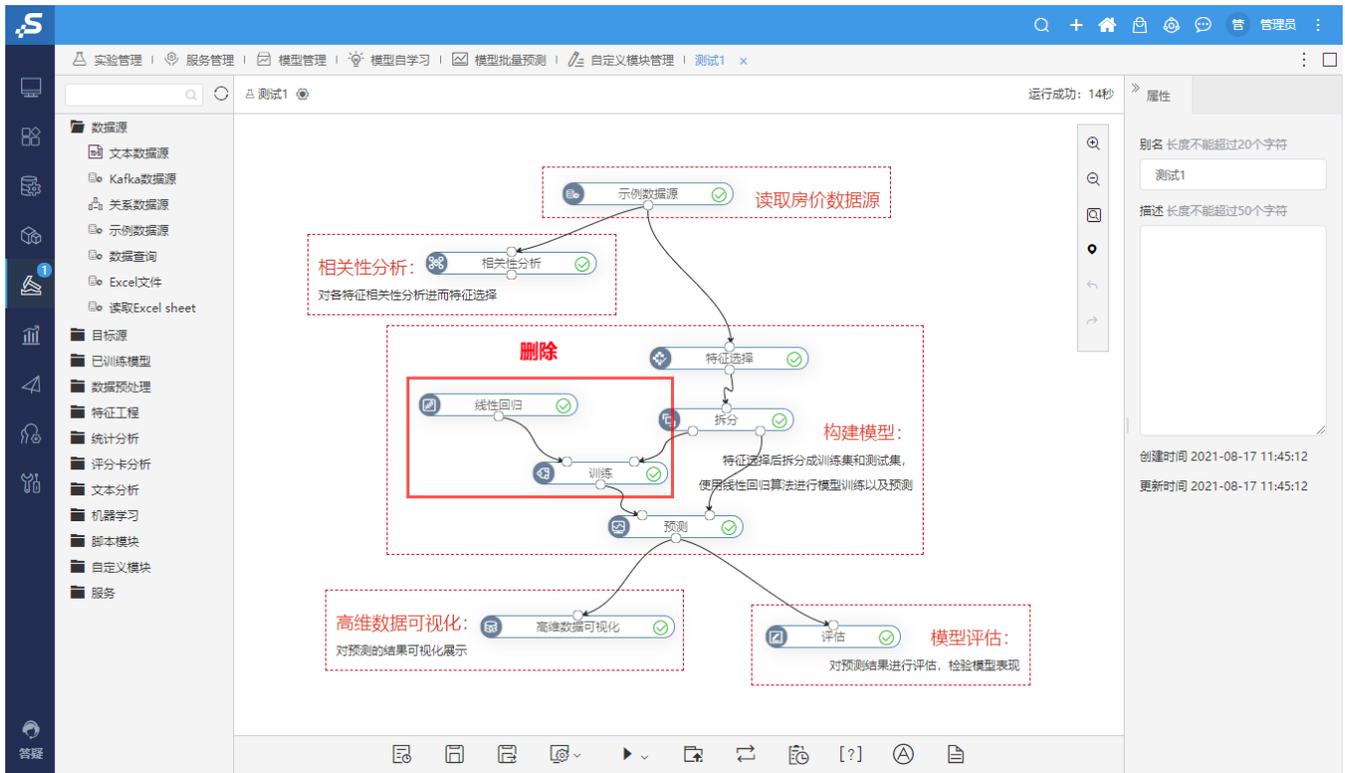
① 打开数据挖掘 - 案例一波士顿房价预测，点击运行，如下图显示运行成功，并点击保存案例



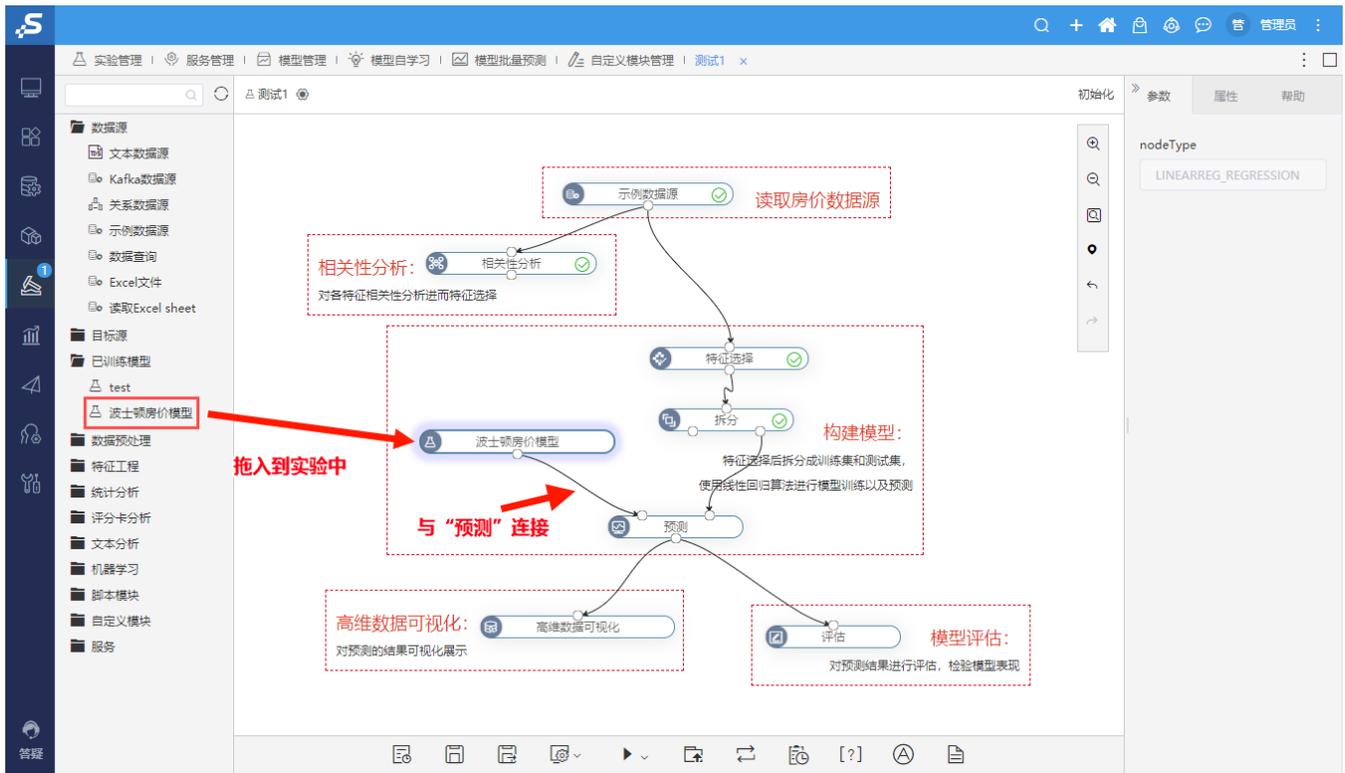
② 右键点击 **训练—模型—保存模型**，如下图所示



③ 保存的模型可以在左侧导航栏的“已训练模型”中查看。右键删除“线性回归”，“训练”。



④ 将保存的模型拖入，并与”预测”连线。



⑤ 导航栏-服务-拖入 服务输入、服务输出，并连线，运行成功后点击部署服务

2.删除“示例数据源”节点

3.左侧导航栏>服务>拖入“服务输入”，连线

相关性分析：对各特征相关性分析进而特征选择

读取房价数据源

构建模型：特征选择后拆分成训练集和测试集，使用线性回归算法进行模型训练以及预测

4.左侧导航栏>服务>拖入“服务输出”，连线

1.删除“评估”节点

高维数据可视化：对预测的结果可视化展示

模型评估：对预测结果进行评估，检验模型表现

5.点击运行

6.运行成功后点击部署服务

⑥ 显示部署成功

服务配置 服务测试

服务部署成功!

服务ID: l8a8a4c7801794b844b84f2be017954191028000f

服务名称: 测试一

服务别名: 测试一

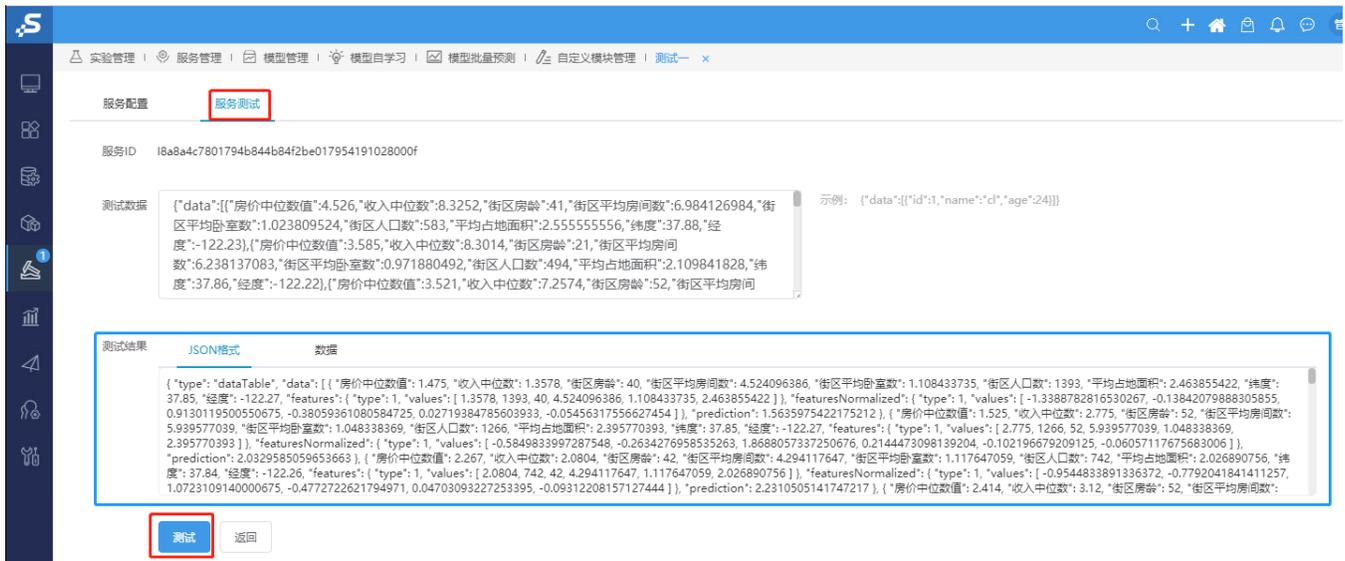
内部调用地址: https://10.10.204.248:8900/api/v1/services/l8a8a4c7801794b844b84f2be017954191028000f

相关实例: l8a8a4c78017959035903e2bb01795903e2bc0000

服务描述:

修改 返回

⑦ 点击服务测试，会显示测试结果



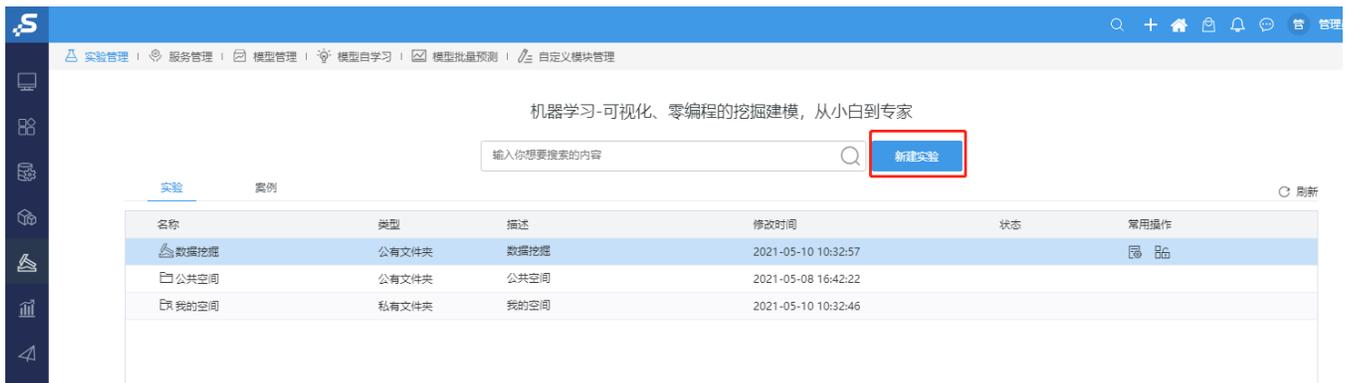
2.3. 测试数据挖掘的Python计算



前提条件

测试数据挖掘的Python计算时，请确认Python节点已正常部署和启动。部署文档参考：[部署Python节点](#)

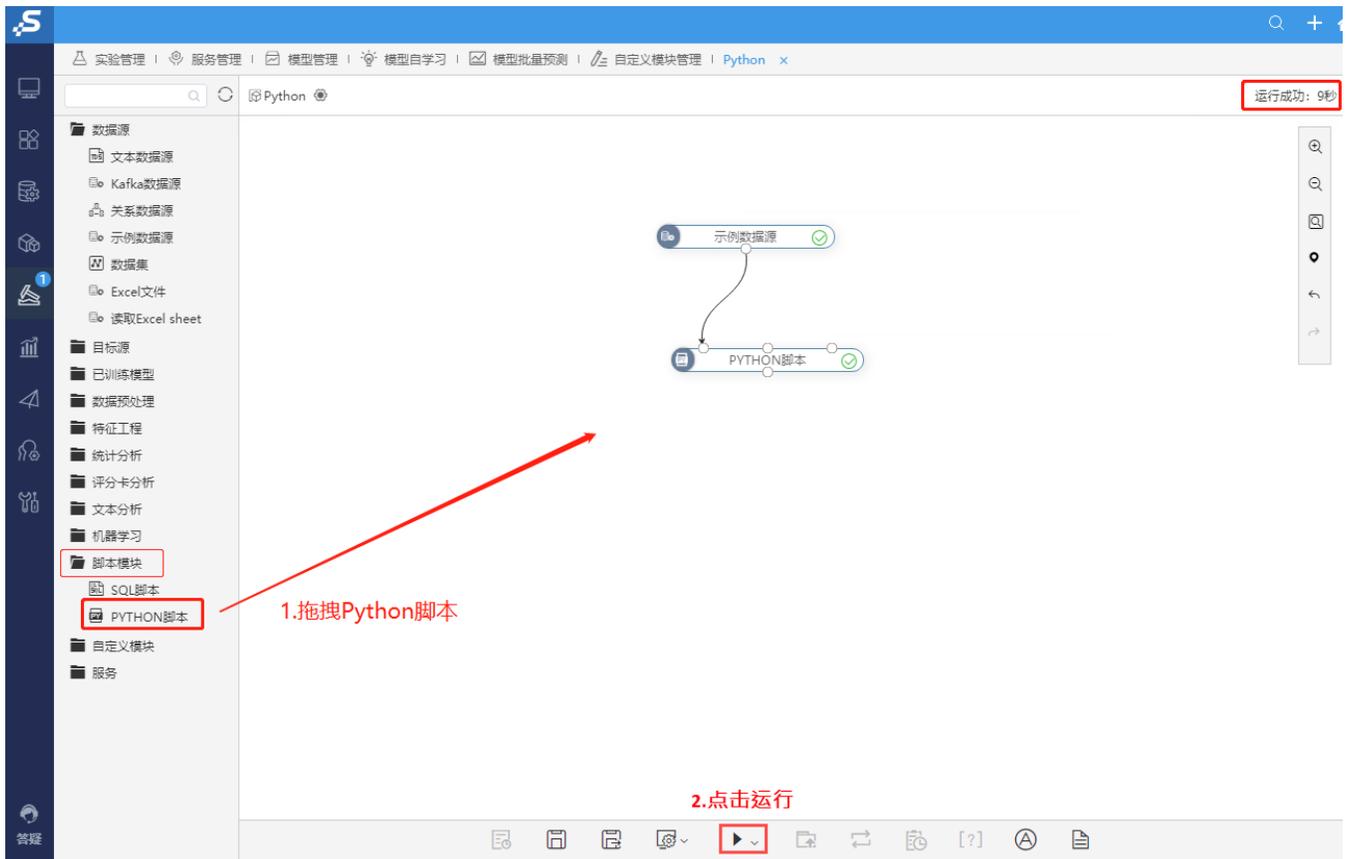
① 打开数据挖掘—新建实验



② 拖拽“示例数据源”，选择数据源 - 保存 - 运行 - 运行成功-保存



③ 拖拽PYTHON脚本，与示例数据源连线，点击运行，如果显示运行成功则表示Python计算节点正常



数据挖掘及其组件测试完成。