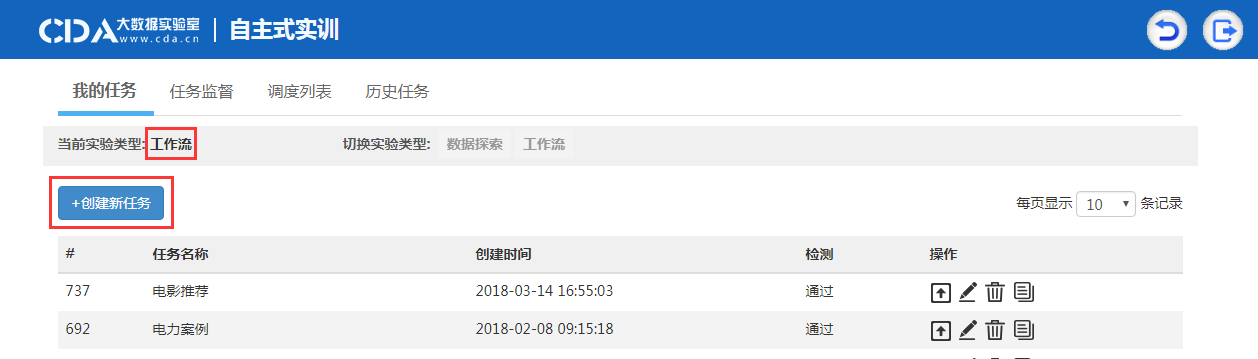
**工作流操作说明之电力聚类**

案例背景：

本案例针对冀北地区中的钢铁行业，运用大数据相关技术，利用相关聚类算法，分析其钢铁行业中的用电企业的用电行为特征，构建冀北大数据辅助决策系统。让供电企业能够及时分析各个用户的行为特征，实现电力的合理分配和配置。结合电力业务，一般情况下，规模较大的钢铁企业，其日负荷较大，其合同容量也较大。因此，本次分析报告中采取的数据归一化方法为，用户的 96 点日负荷数据除以该用户的合同容量，以此得到该用户的 96 点的负荷使用率。根据负荷使用率进行用户的用电行为对比。

1、创建新任务

在“工作流”中点击“创建新任务”，开始编辑任务。



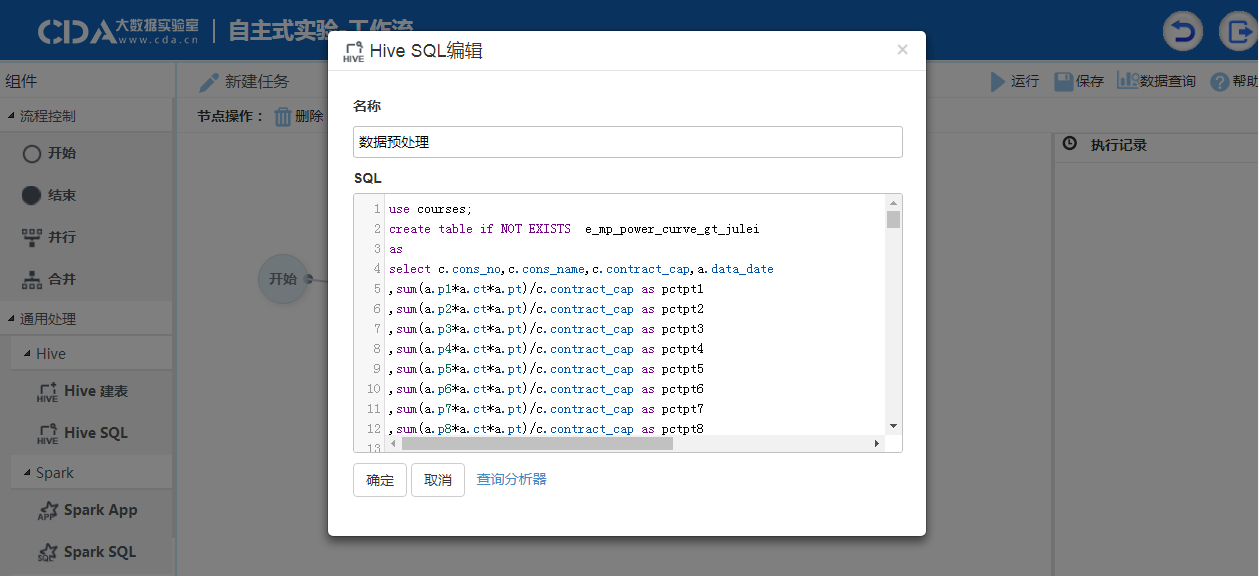
2、创建第一个节点Hive SQL：数据预处理



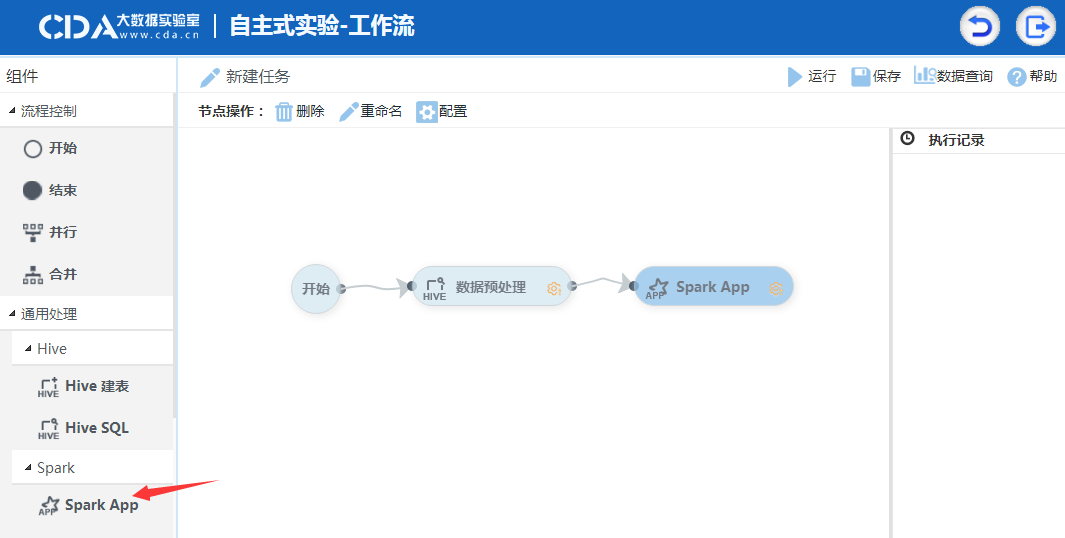
新建Hive SQL节点并进行编辑

标题设置为：数据预处理

编辑相应SQL语句，代码详见“数据预处理SQL语句”文档，然后点击“确定”。



3、创建第二个节点Spark App：聚类模型训练



新建Spark App节点并进行编辑

标题设置为：聚类模型训练

上传jar包：spark-elec.jar；

Main Class ：--class kmeans\_train\_data

参数设置为：jibei\_elec e\_mp\_power\_curve\_gt\_julei 1157014083 3

然后点击“确定”。



4、创建第三个节点Hive SQL：聚类中心信息

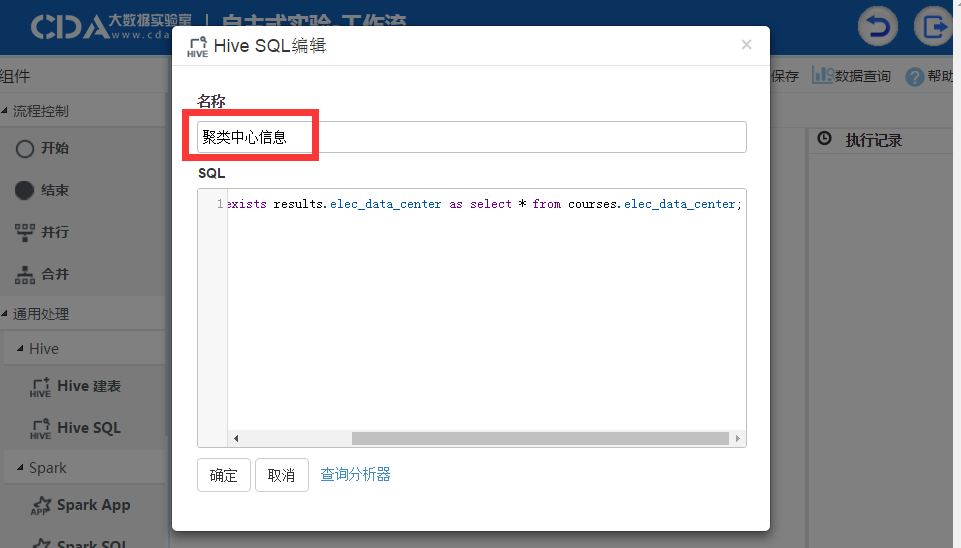


新建Hive SQL节点并进行编辑

标题设置为：聚类中心信息

编辑相应SQL语句（如下），然后点击“确定”

create table if not exists results.elec\_data\_center as select \* from courses.elec\_data\_center;



5、创建第四个节点Hive SQL：导出聚类结果

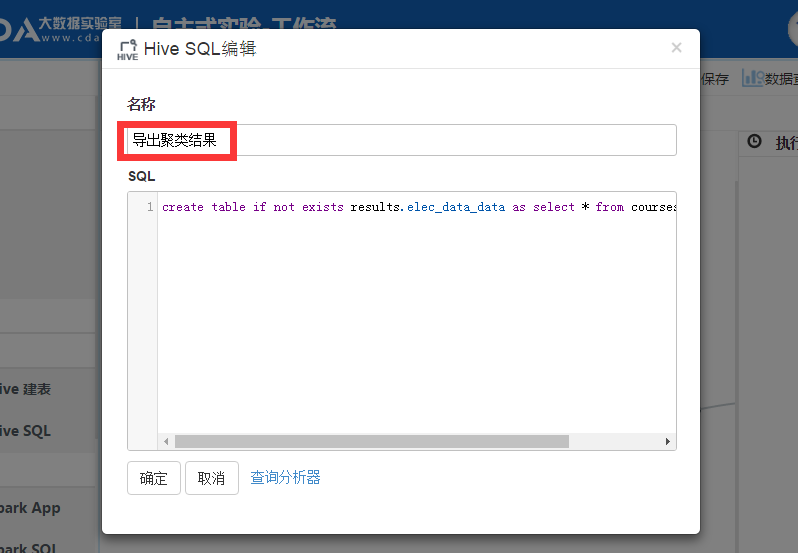


新建Hive SQL节点并进行编辑

标题设置为：导出聚类结果

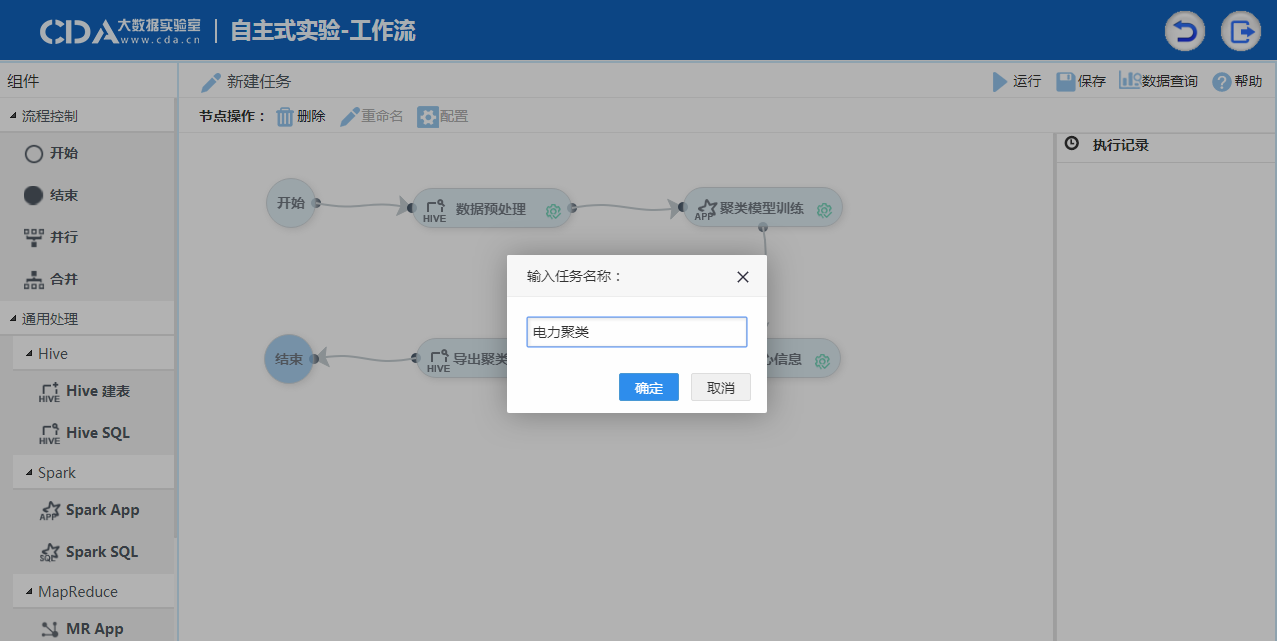
编辑相应SQL语句,然后点击“确定”。

create table if not exists results.elec\_data\_data as select \* from courses.elec\_data\_data;



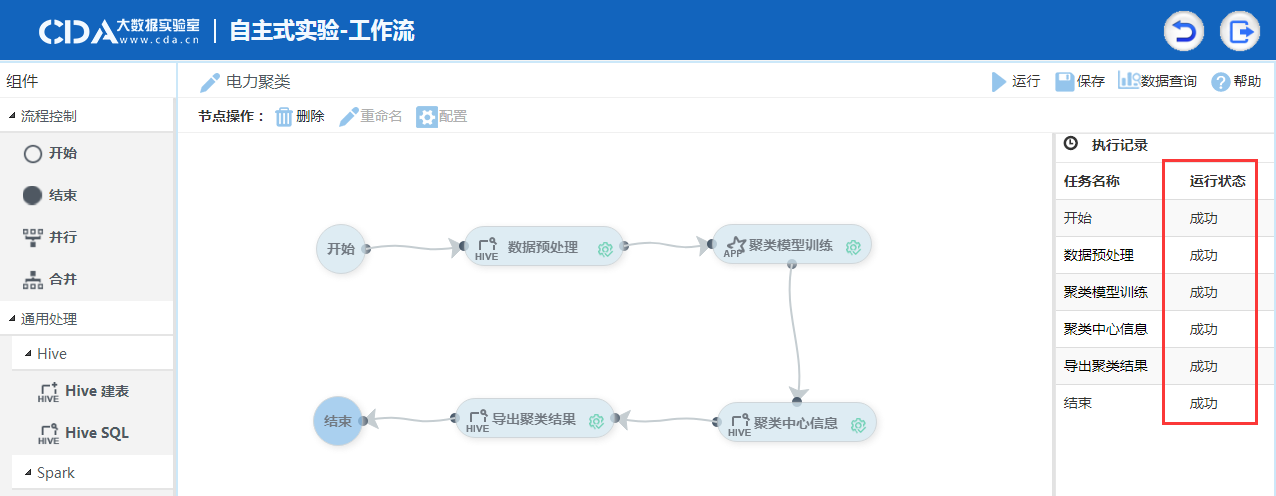
6、保存并提交任务

添加“结束”节点，点击“保存”，修改任务名称为：电力聚类，点击“确定”。



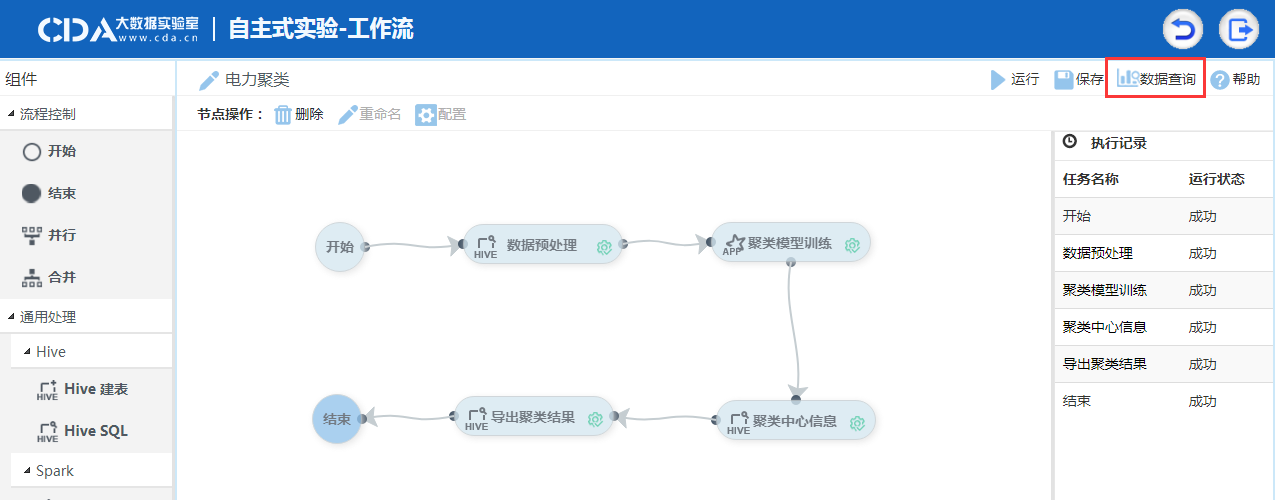
然后点击“运行”，选择“立即执行”即可在右边“执行记录”中看到运行状态。



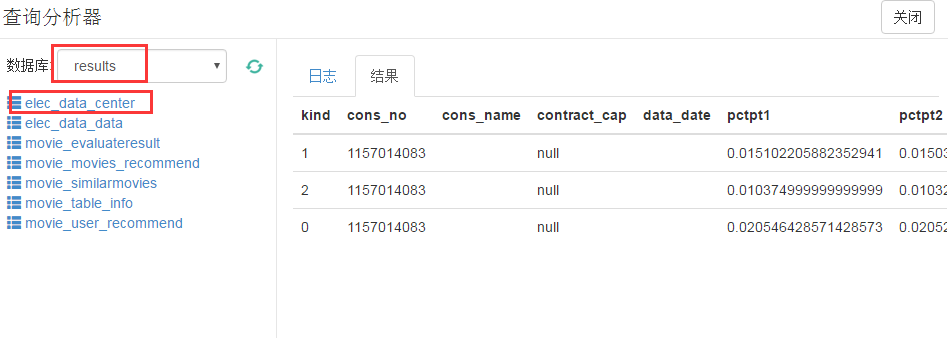


7、查询结果

（1）点击“数据查询”，跳转到“查询分析器”界面。



（2）选择结果数据库：results,其中elec\_data\_center是第三个组件“聚类中心信息”的结果表，点击可查看结果。



（3）选择结果数据库：results,其中elec\_data\_data是第四个组件“导出聚类结果”的结果表，点击可查看结果。

