**调度中心使用文档**

版本： V2.9

拟制: 朱建军

**修改记录**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **文件编号** | **版本号** | **拟制人/修改人** | **拟制/修改日期** | **更改理由** | **主要更改内容** |
|  | V1.0 | 朱建军 | 2017-04-09 |  |  |
|  | V1.1 | 朱建军 | 2017-04-13 | 完善 | 对调度时间和依赖规则进行说明 |
|  | V2.0 | 朱建军 | 2017-06-02 | 完善 | 主要新增二期的新功能 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| 注1：每次更改归档文档时必须填写该表 注2：一次归档时，“更改理由”和“主要更改内容”填无 | | | | | |

**目录**

**[目录](#_Toc484189197)** [3](#_Toc484189197)

[1 说明 4](#_Toc484189198)

[1.1 基本概念 4](#_Toc484189199)

[2 使用步骤 4](#_Toc484189200)

[2.1 进入调度中心菜单 5](#_Toc484189201)

[2.2 选择应用 5](#_Toc484189202)

[2.3 表查询 6](#_Toc484189203)

[2.3.1 新增表 6](#_Toc484189204)

[2.3.2 删除和清空 7](#_Toc484189205)

[2.4 新建dag 7](#_Toc484189206)

[2.4.1 新建一个新目录 8](#_Toc484189207)

[2.4.2 在新建的目录下创建dag任务 8](#_Toc484189208)

[2.4.3 填写dag的调度配置 9](#_Toc484189209)

[2.5 拖拽节点完成dag 11](#_Toc484189210)

[2.5.1 修改任务（Task）节点 12](#_Toc484189211)

[2.6 提交dag 13](#_Toc484189212)

[2.7 查看运行状态 13](#_Toc484189213)

# 说明

调度中心是为了方便用户执行数据分析任务，底层基于airflow调度系统，实现任务的有效管理和调度。

## 基本概念

任务（Task）节点：一次完整的hive查询sql就是一个任务节点

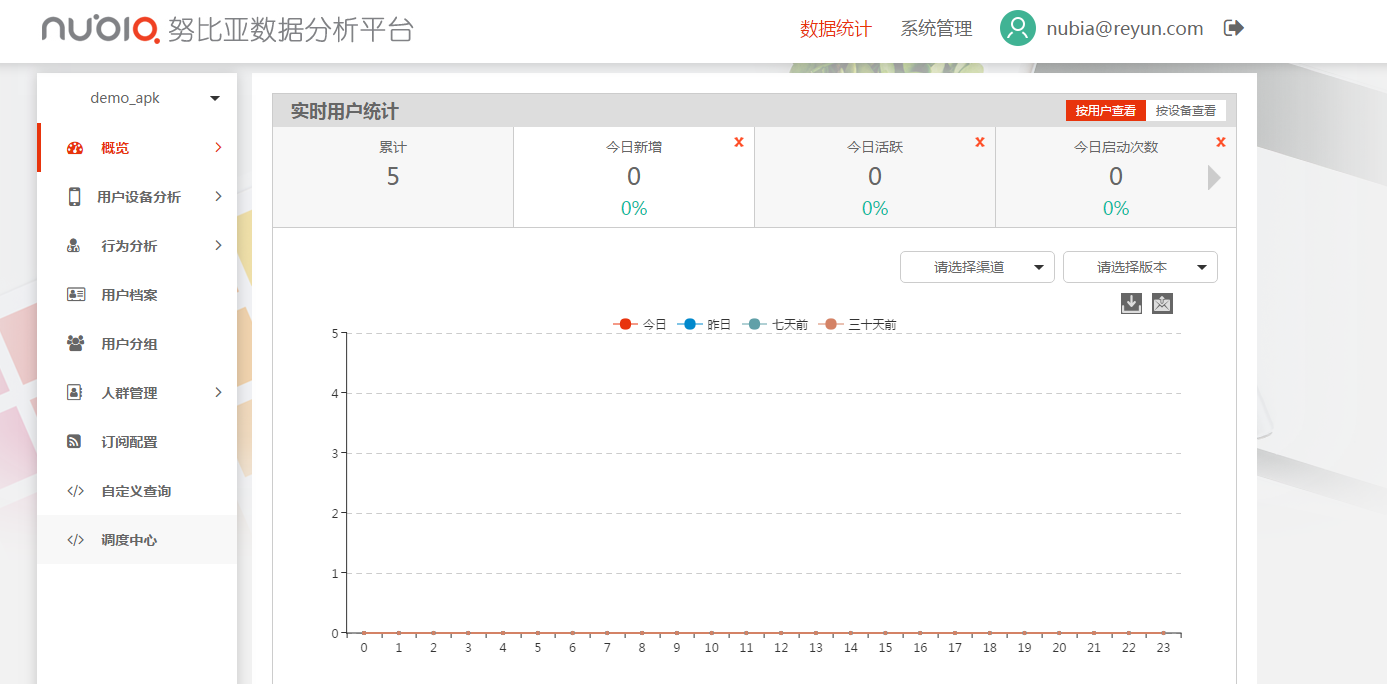
Dag（有向无环图）：几个任务节点有时间上或者功能上的相互依赖，用线（依赖关系）连起来组成一个dag

生效时间：调度生效的开始时间和结束时间

调度周期：在开始和结束时间里面执行dag的间隔叫调度周期

# 使用步骤

使用账户密码进入努比亚数据分析平台，点击左边菜单栏的调度中心



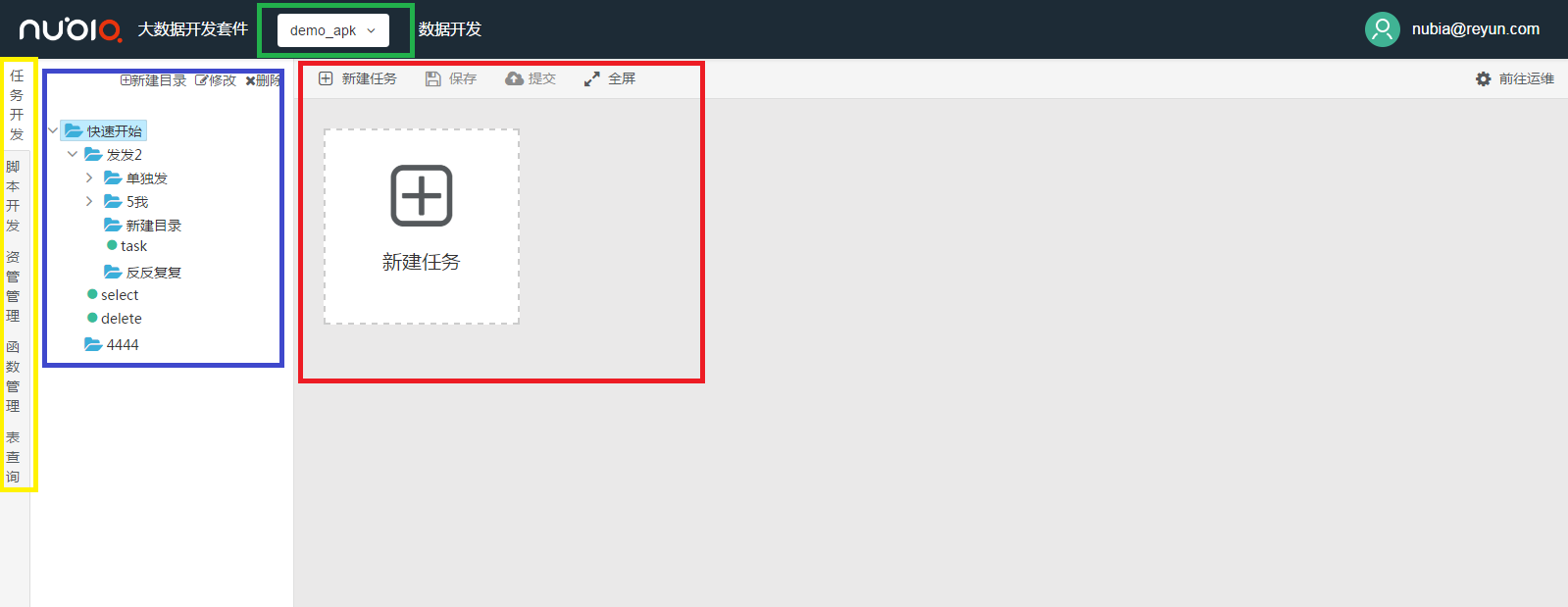
## 进入调度中心菜单

点击调度中心菜单后进入调度中心页面，最左边纵向tab页（黄色框）有任务开发、表查询、脚本开发（待开发）、资源管理（待开发）、函数管理（待开发）、表查询（待开发）；暂时只完成任务开发功能。

上面的下拉菜单（绿色框）可以选择具体的应用。

左边的树形目录（蓝色框）是应用下面的文件目录。

中间的操作页（红色框）可以拖拽任务节点并连线组成dag



## 选择应用

点击下拉框选择要操作的具体应用，树形文件目录会自动映射更新

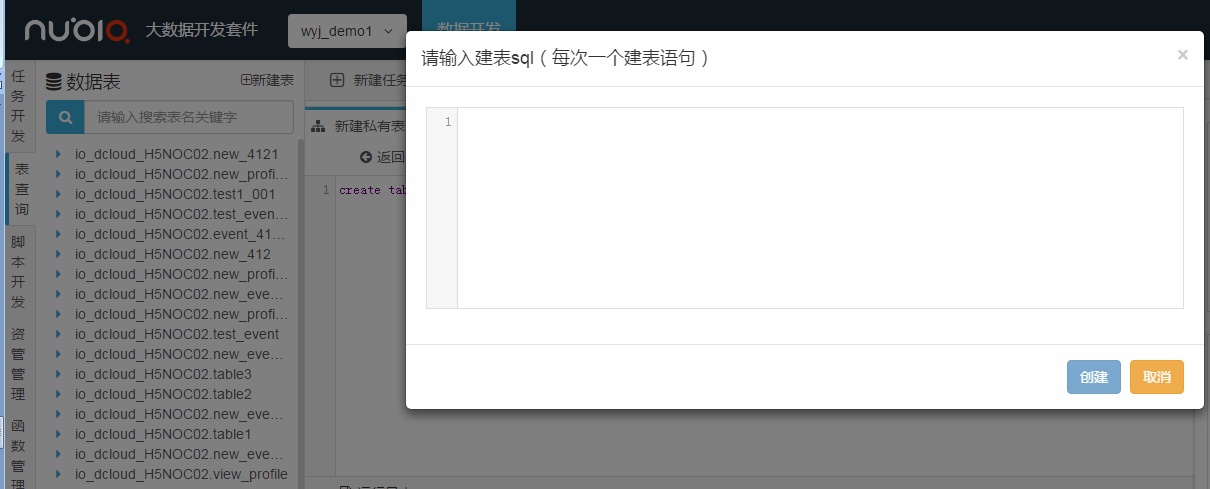


## 表查询

点击左侧的表查询tab页，可以增加、删出、查看、清空表数据等操作。

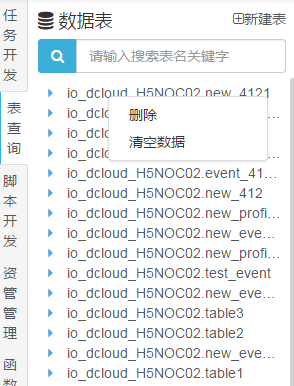
### 新增表

点击新建表弹出建表对话框，输入建表sql后点击创建即可

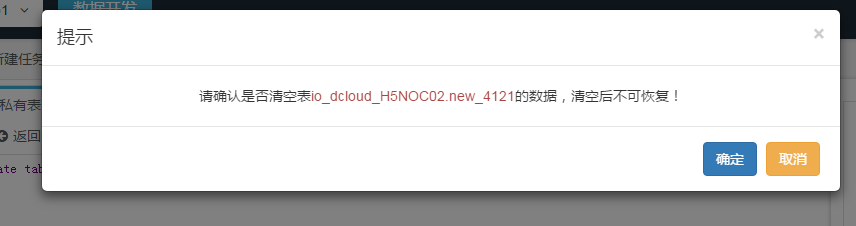


### 删除和清空

在要操作的表名上右键，弹出操作对话框，点击操作选项（删除或清空） 。

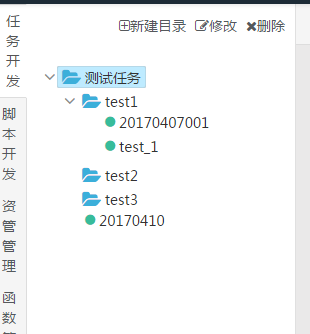


弹出确认对话框，点击确认即可完成操作。请谨慎操作！



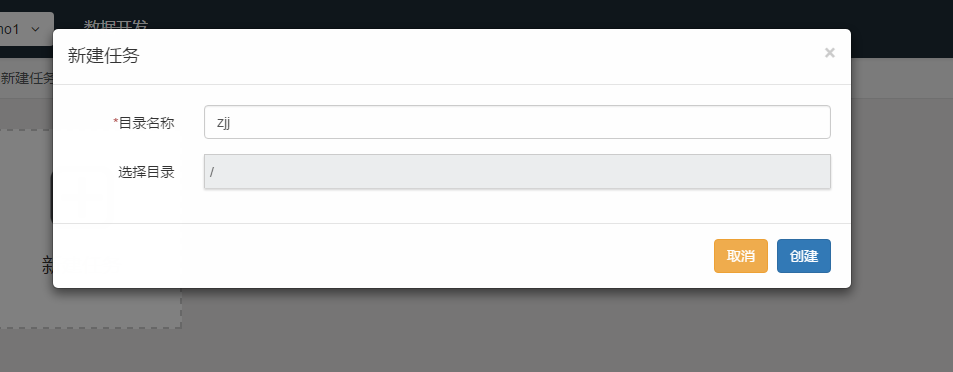
## 新建dag

可以新建、修改、删除、移动dag的目录



### 新建一个新目录

点击新建目录会弹出新建目录对话框。填写名称，然后点创建



### 在新建的目录下创建dag任务

点击中间的新建任务，填写任务名称、描述信息和选择具体的目录，点击创建**。**



创建任务主要区分周期和手动，周期为按预先定义好的时间周期重复的执行，而手动为需要每次点击运行才执行，适合Strorm、Spark、Hadoop MapReduce等提交一次即一直运行的任务。

创建成功后树形目录上可以看到创建的test\_6,并且中间出现两个节点组件，一个是ODPS\_SQL（hive sql），一个是虚节点（什么都不做）

****

### 周期调度Dag填写的调度配置

一、生效日期和调度周期，调度周期一般是按天调度的有两种写法：

（1）按时间（时：分）：00:00（每天的零点零分）

（2）按cron表达式（分 时 日 月 年）：0 0 \* \* \*（每天的零点零分）

**注意：**如果当前是2017-4-11 ，开始日期是2017-4-8，结束时间是2017-4-12，调度周期是

0 0 \* \* \*（注意表达式格式，有空格的五位）

那么，调度会默认是开始时间为昨天（2017-4-10，这是为了防止调度回填），调度周期：2017-4-10 00:00，不会执行2017-4-11 的，2014-4-11的批次会在2017-4-12号凌晨执行；

结束时间不能比当前时间晚，比如结束日期不能是2017-4-10，因为调度系统会任务，你这个时间段已经过期了；

如果想明天早上就跑，那么你的开始日期就是为今日，airflow中只会调度当前日期的以前的批次。

二、依赖属性：默认是不依赖上一调度周期

**注意：**默认就是前一次调度和当前调度没有依赖关系，相互独立的；反之，就是当前调度批次的所有任务都是依赖上一次调度该任务的成功与否的，上一次任务1执行成功，这一次任务1就执行，不然就不执行

三、调度状态

勾选表示提交后不调度，不勾选表示提交后立马进行调度

****

### 手动调度Dag填写的调度配置

基本属性、dag参数同周期调度，不同的是调度属性中只有“点击执行一次”的触发按钮；

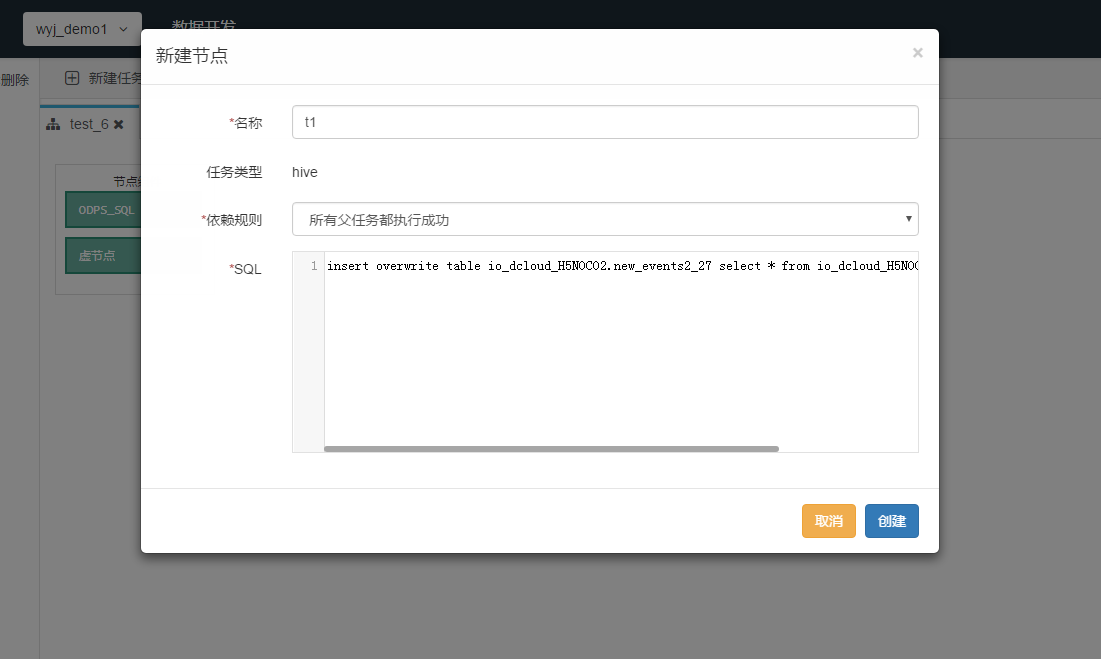
另外需要注意的是，手动调度的dag任务在数据统计中有且只有一“RUNNING”状态，在上次提交未执行完成的前提下，再次提交会提示“该DAG尚有未完成的运行任务，请稍后再试”。

## DAG节点组件使用

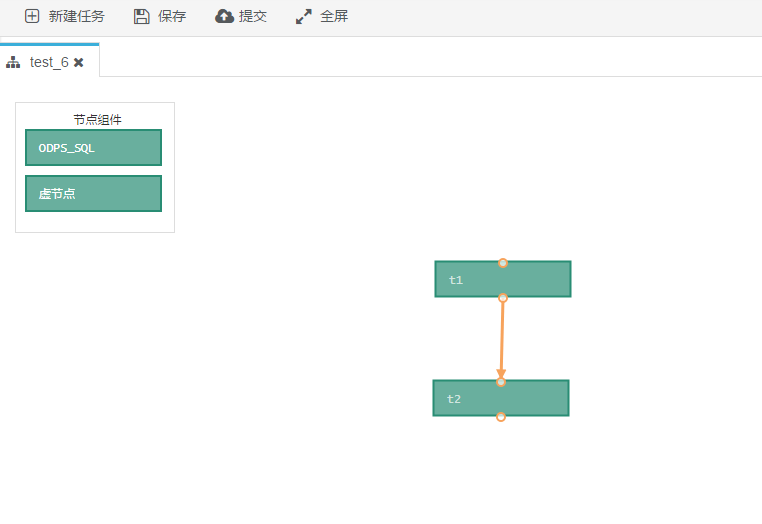
### ODPS\_SQL节点组件

拖拽ODPS\_SQL节点，弹出新建节点对话框。填写完成任务名称（数字英文下划线），选择依赖规则，填写要执行的sql语句(请按规范写sql)，点击创建。

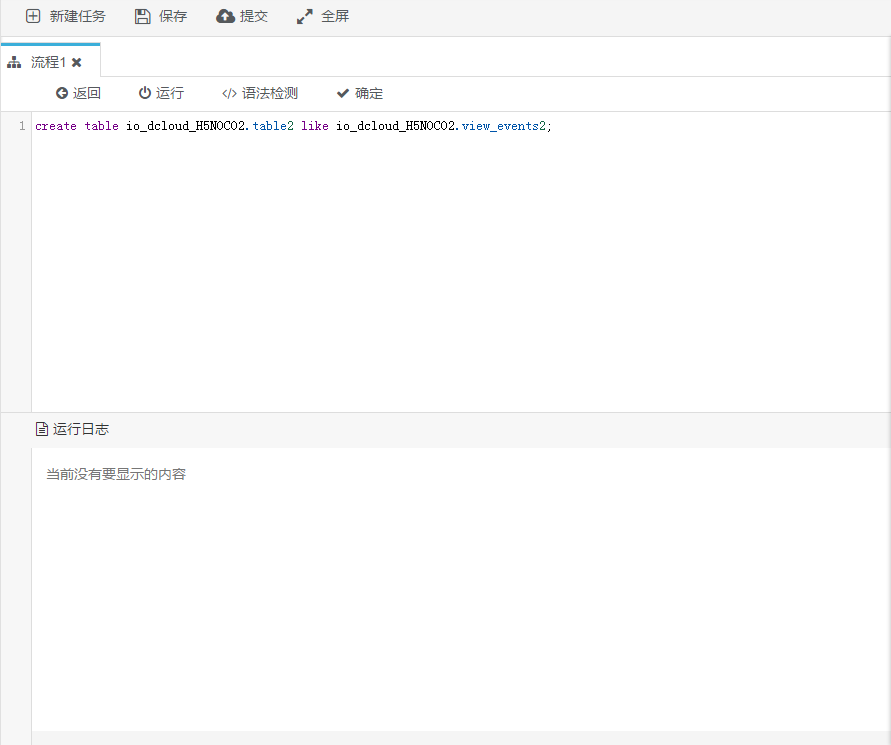
**注意：**依赖规则的意思这个dag的节点和其直接父节点的直接的依赖关系。默认是所有父任务都执行成功，表示当前任务节点执行不执行看其直接父节点的是否是执行成功，有一个不成功，这个节点就不执行，跳过；如果选择的是无所谓，这个要注意，无所谓的时候，父子节点是没有什么时序关系的，可以同时执行的，要注意。



用线将任务之间按照时间依赖连接起来，点击保存

****

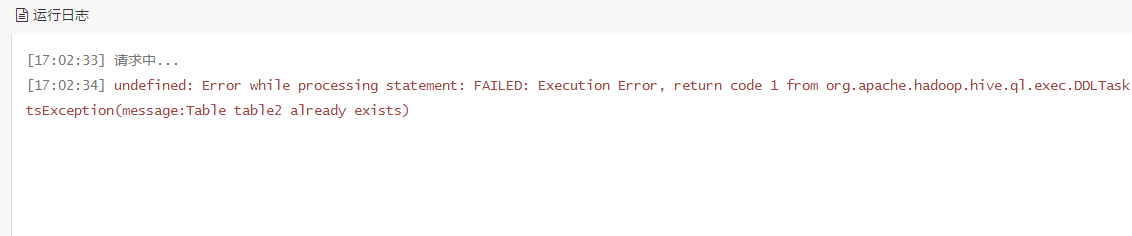
在要修改的节点上右键，点击编辑节点，跳到节点编辑界面

****

可以修改编辑SQl语句，然后可以点击语法检测进行SQL语法检测，检测结果在下面显示

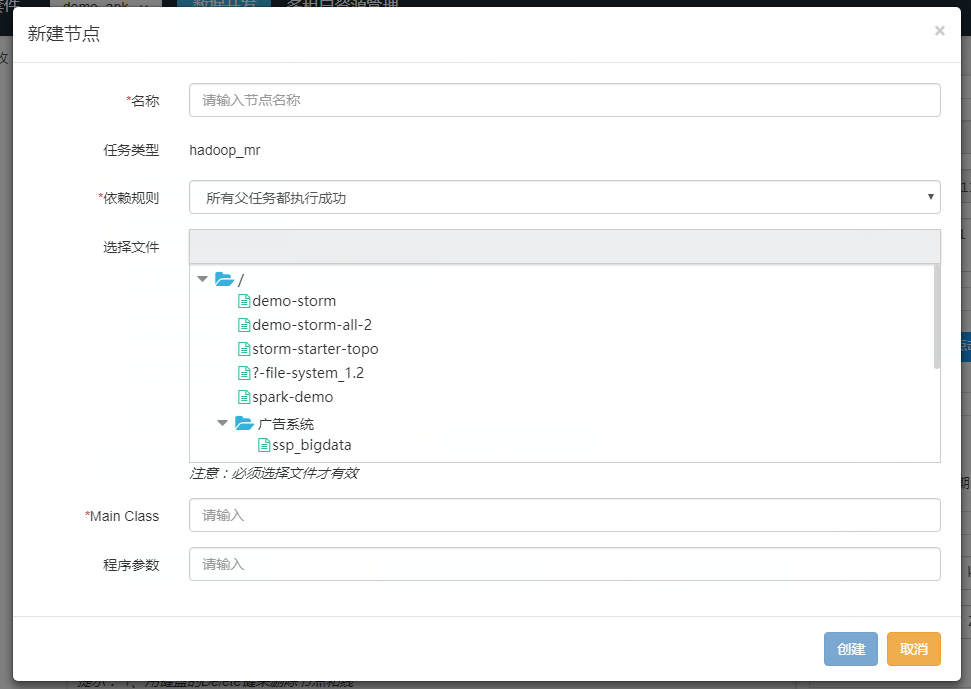
****

可以点击运行，测试SQL是否能正确执行，运行日志在下面同步显示

****

点击返回放弃修改，点击确定确认修改

### ODPS\_MR节点组件



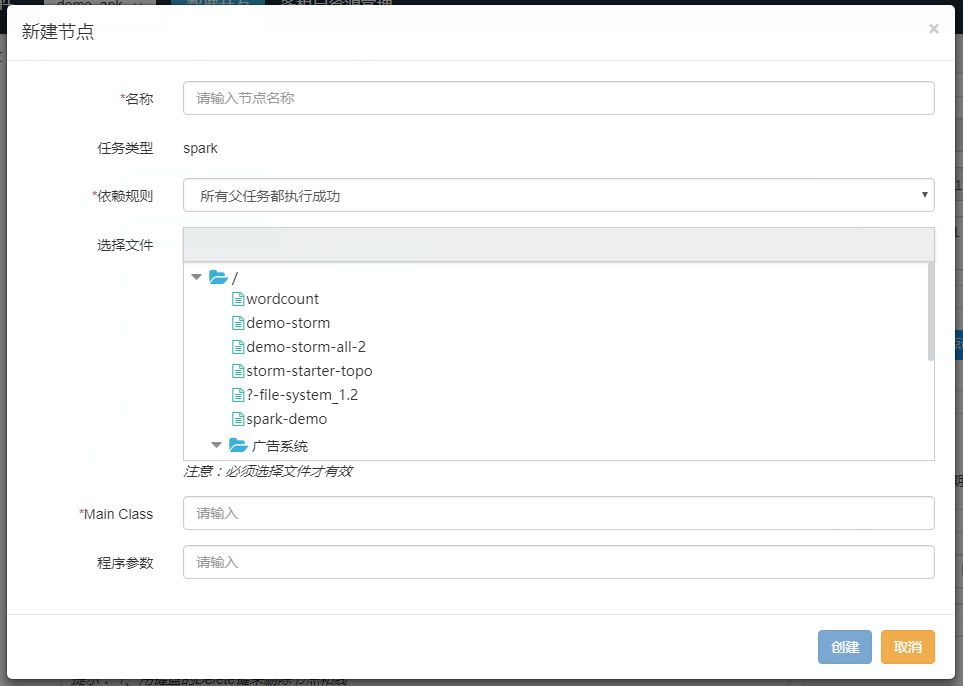
主要参数说明

选择文件：即选择打包的MapReduce程序jar文件

Main Class：MR任务的启动Main类

程序参数：需要传入Main类的参数，多个参数使用空格分隔，如 “arg1 arg2”

### ODPS\_SPARK节点组件



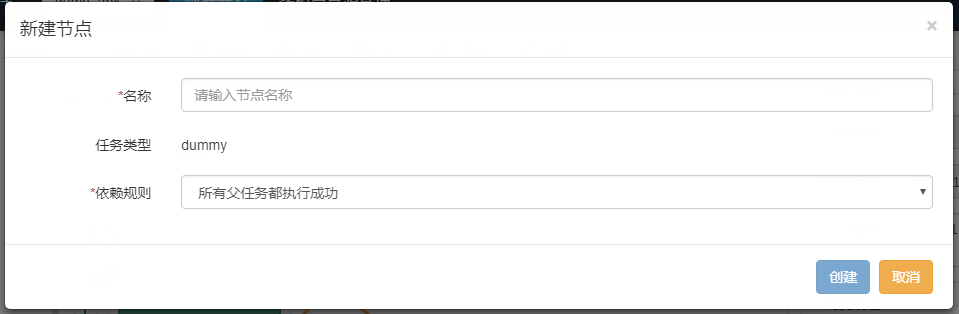
主要参数说明

选择文件：即选择打包的Spark可执行文件，如.jar

Main Class：任务的启动Main类

程序参数：需要传入Main类的参数，多个参数使用空格分隔，如 “arg1 arg2”

### ODPS\_DUMMY节点组件



该节点为虚节点，及没有实际的运行，主要用于dag的任务中的合并和分支；

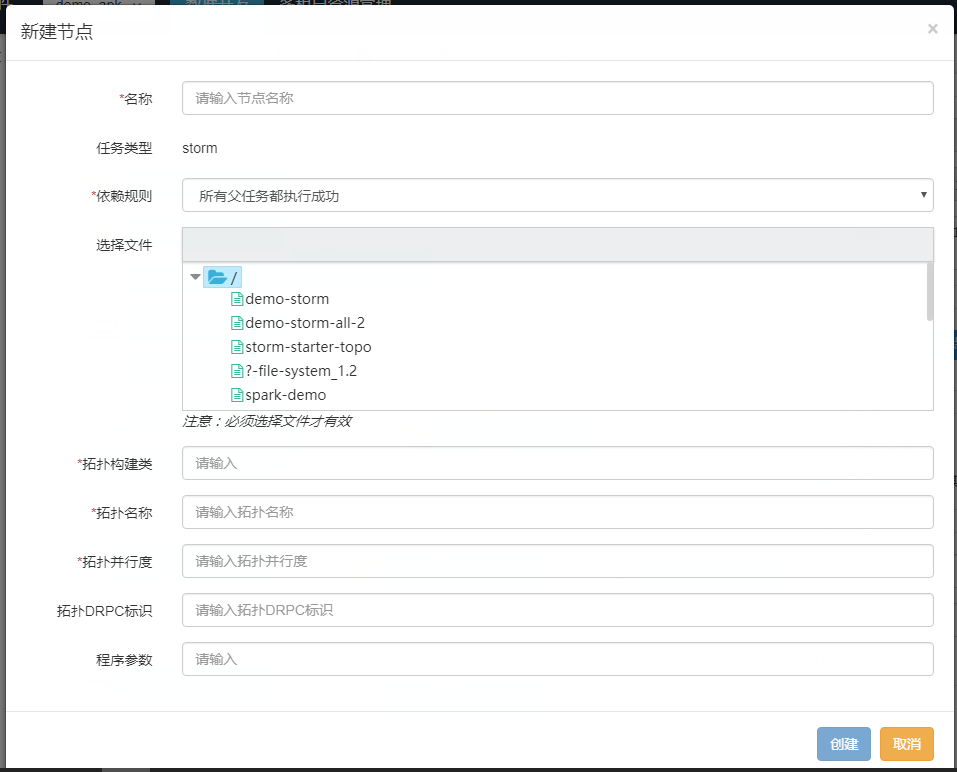
### ODPS\_SHELL节点组件



该任务支持以shell命令方式执行，注意不是shell文件，如果shell中带有环境不支持/未安装的类库和组件，请联系数据统计的负责人；

比如指令为： echo “abc”

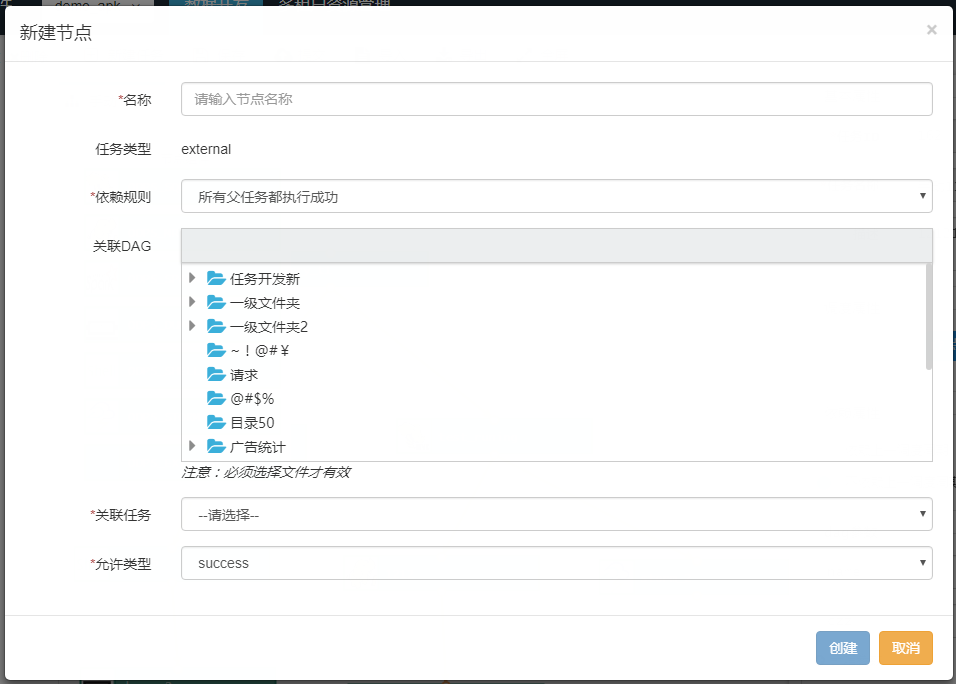
### ODPS\_STORM节点组件



支持Storm类型的任务的提交，需要接入使用Storm的项目及研发同学请详细阅读数据统计的参考《storm流式计算SDK-API使用说明文档V1.0(彭行雄).docx》；

本组件支持DRPC，即可以实时的的查询Storm中Topology中的实时数据，记得授权实时计算资源的接口访问权限。

### ODPS\_EXTERNAL节点组件



主要参数说明

关联DAG：依赖DAG

关联任务：依赖DAG中的某个任务

允许类型：依赖DAG中的某个任务执行的成功/失败触发条件

该节点主要支持任务依赖调度，假设这样的需求：有DagA 和DagB2个任务，DagB的该类型节点b2需要DagA中的某个任务a1成功后才开始执行。该节点组件适用于按天调度的任务，手动调度的DAG不支持该组件；

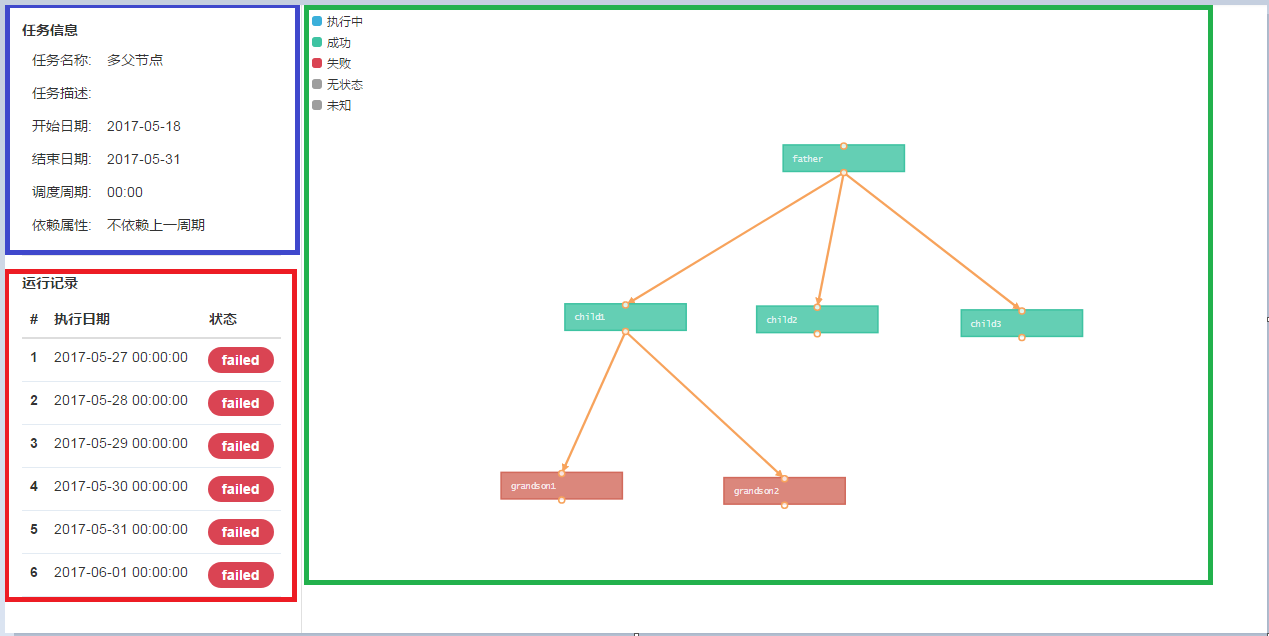
## 提交dag

点击提交，dag会经历如下步骤：

保存🡪环检测**🡪**SQL语法检测**🡪**解析成调度文件**🡪**python语法检测**🡪**airlow进行调度**🡪**离线查询执行

## 查看运行状态

点击当前DAG的右上角的前往运维，跳转到dag状态页面



1. **蓝色框是DAG信息**

包括名称、描述、生效时间、调度周期和依赖属性等等。

1. **红色框是当前DAG的调度状态列表**

每一条记录表示一次调度，包含执行日期、状态。状态分为三种：

绿色：success（成功）

红色：failed（失败）

黄色：up\_for\_retry（挂起重试）

PS：1、如果dag未执行则为空

2、状态列表只取近七天中有执行的DAG展示

**三、绿色框是DAG每次调度执行状态展示区**

1、任务节点状态分为执行中（天蓝）、成功（绿色）、失败（红色）、无状态（灰色）、未知（灰色）五种状态

2、点击DAG图上的节点，可以弹出节点执行状态的详细信息



对于长SQL当前无法显示的，可以点击查看SQL查看完整的SQL语句。