



# 大数据整合平台

TipDM-IB

---

## 产品操作手册

广东泰迪智能科技股份有限公司 版权所有

地址：广州市黄埔区科学城开泰大道 36 号 1 栋 212

网址：<http://www.tipdm.com>

邮箱：[services@tipdm.com](mailto:services@tipdm.com)

邮编：510000

电话：020-22205718

---

## 目录

平台介绍.....	4
1.1 产品简介.....	4
1.2 名词解释.....	4
1.3 技术支持.....	5
2 快速入门.....	5
3 功能说明.....	8
3.1 转换.....	8
3.1.1 输入.....	8
3.1.2 输出.....	13
3.1.3 转换.....	17
3.1.4 应用.....	37
3.1.5 流程.....	38
3.1.6 查询.....	39
3.1.7 连接.....	40
3.1.8 统计.....	44
3.2 作业.....	46
3.2.1 通用.....	46
3.2.2 邮件.....	50
3.2.3 文件管理.....	52
3.2.4 条件.....	57
3.2.1 文件传输.....	59
4 用户权限管理.....	62
4.1 组织架构.....	62
4.1.1 新增组织架构.....	62
4.1.2 删除组织架构.....	63
4.1.3 编辑组织架构.....	63
4.1.4 查看组织架构信息.....	64
4.2 角色管理.....	65

---

4.2.1	新增角色 .....	65
4.2.2	删除角色 .....	65
4.2.3	编辑角色 .....	66
4.2.4	资源绑定 .....	66
4.2.5	查看角色详情 .....	67
4.2.6	设为默认角色 .....	68
4.1	用户管理 .....	68
4.1.1	新增用户 .....	68
4.1.2	编辑用户 .....	69
4.1.3	查询用户 .....	69
4.1.4	角色绑定 .....	70
4.2	资源管理 .....	70
4.2.1	资源绑定 .....	70
4.2.1	资源管理 .....	71
4.3	客户端管理 .....	74
4.4	监控系统 .....	74

广东泰迪智能科技有限公司

# 平台介绍

## 1.1 产品简介

大数据整合平台是由广州泰迪智能科技有限公司自主研发，面向高校数据分析课程教学的数据整合工具。平台使用 JAVA 语言开发，采用 B/S 结构，用户不需要下载客户端，可通过浏览器进行访问。平台将功能进行封装，采用图形用户界面的形式呈现，用户可在没有编程基础的情况下，通过拖拽的方式进行操作，将数据抽取、数据清洗、数据合并等环节通过流程化的方式进行连接，以达到数据整合的目的。

平台以实际数据分析案例为切入点，对老师而言，老师在使用平台进行教学时，可以讲授数据整合基本流程、帮助学生建立数据整合的工程化思维。对学生而言，只需理解每个功能的作用，无需编程，即可实现数据分析，简单易用，并能提高数据处理效率，大大降低了学习数据整合的门槛，让学生对数据整合有了更感性的认识，激发学生的学习兴趣。

## 1.2 名词解释

**组件：**将建模过程涉及到的数据抽取、数据清洗、数据合并等程序分别进行封装，每一个封装好的程序都可称之为组件。

**转换：**转换的过程是以数据流实现的，是指从输入控件（Input）到输出控件（Output）之间的数据流动，针对的是在数据流动过程中的每一行记录、每一列数据的处理，比如增加一个字段、对字段 A 截取前 3 位得到新的字段 B、把字段 A、B 串联成新的字段 C，按字段 A、B 组合，汇总字段 C（SELECT C1, C2, SUM(C3) FROM 表 GROUP BY C1, C2）。转换可以调一个作业，执行结果会变成数据流里的新的一系列数据，强调的是执行结果；转换可以调另一个转换，执行结果会变成数据流里的新的一系列数据，强调的是执行结果；

**作业：**作业的过程是以步骤流的方式实现的，是指完成一项工作的具体步骤。作业的每一个步骤，必须等到前面的步骤都跑完了，后面的步骤才会执行。作业可以调另一个作业，强调的是父作业与子作业；作业可以调一个转换，强调的是其中一个步骤。

## 1.3 技术支持

感谢您选择广州泰迪科技公司的数据挖掘产品，在系统的使用过程中如果遇到问题，请通过如下的方式与我们联系，我们将为您提供周到满意的服务。

主页：<http://www.tipdm.com>

电话：020-22205718

热线：40068-40020

QQ 群：197738983

地址：广州市黄埔区科学城开泰大道 36 号 1 栋 212

邮编：510000

邮箱：[services@tipdm.com](mailto:services@tipdm.com)

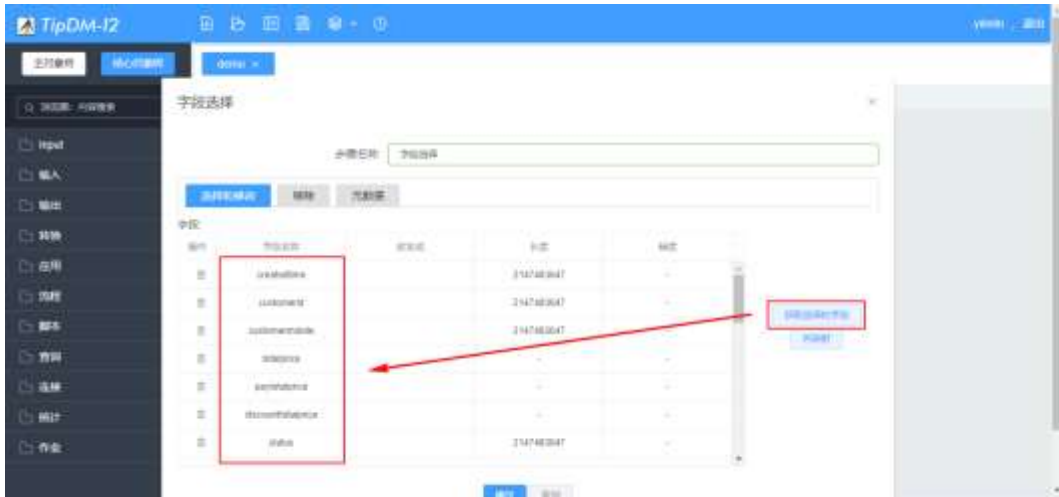
## 2 快速入门

基于自动售货机订单详情数据,使用大数据整合平台构建每月销售金额、订单量、利润、购买人数等指标。

(1) 抽取“订单详情表”。



(2) 选择需要分析的字段。



(3) 构建“每单利润”指标。

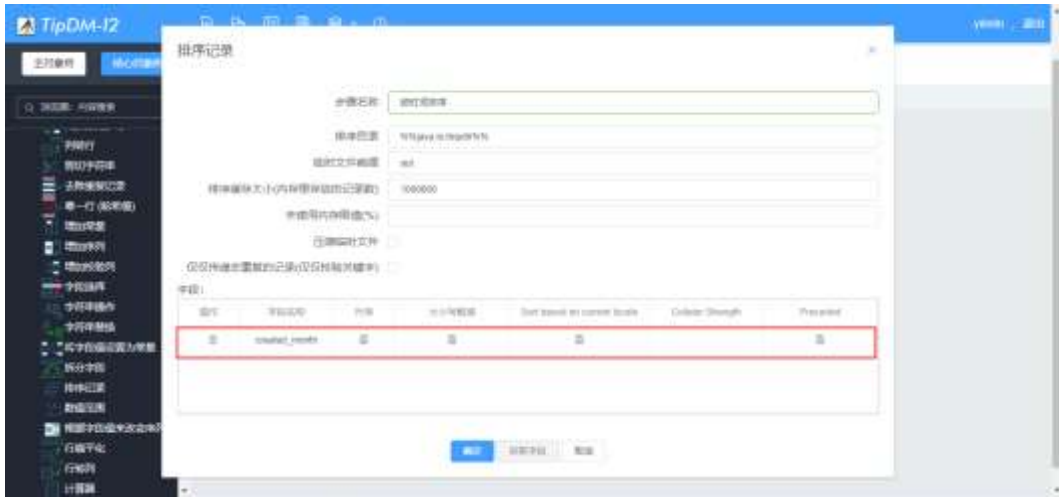


(4) 按月统计数据，构建“月销售金额”“月订单量”“月利润”“月购买人数”指标。

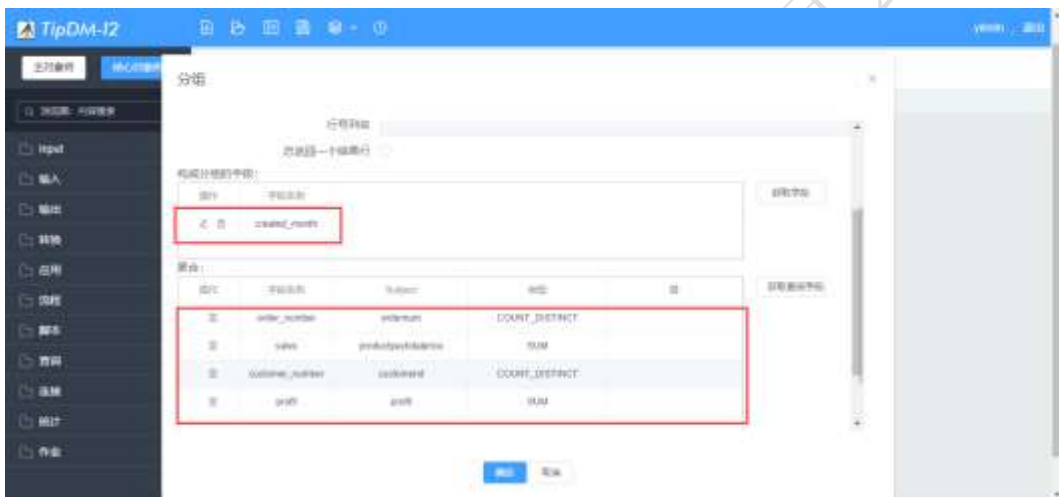
1. 获取月份信息



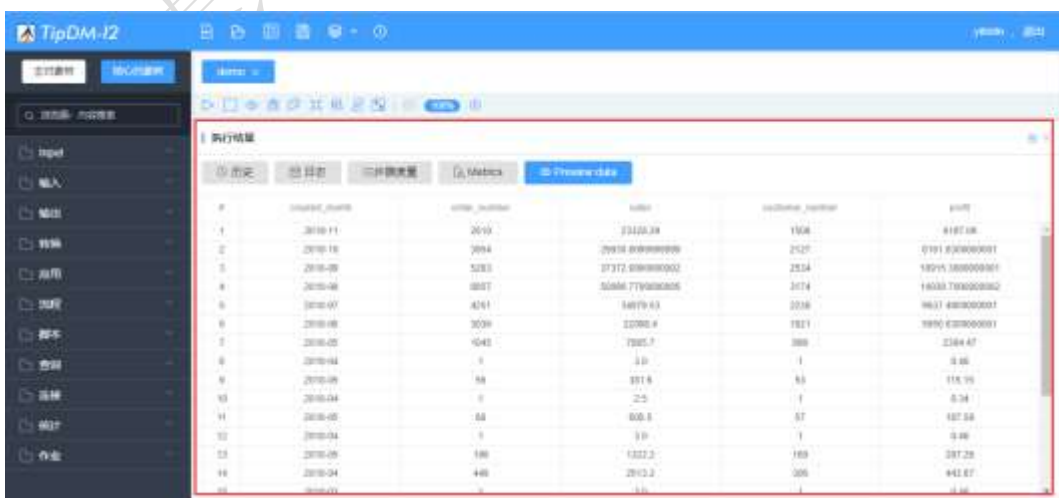
2. 按月份进行记录排序。



3. 分组聚合，构建月度指标。



4. 输出结果。



(5) 将指标数据存放至分析数据库。



### 3 功能说明

#### 3.1 转换

##### 3.1.1 输入

##### 3.1.1.1 CSV 文件输入

图标: 

作用: 读取 CSV 文件中的数据作为输入

参数:

步骤名称: 当前步骤的名称。

文件名: 读取的 CSV 文件的名称。

列分隔符: CSV 文件中用于分隔列的符号。

NIO 缓存大小: 缓冲区中可用元素的大小

简易转换: 是否进行简易转换。

包含列头行: 是否包含列名。

将文件添加到结果文件中: 是否将文件添加至结果文件。

行号字段 (可选): 用作行号的字段名称。

并发运行: 是否并发运行



字段中有回车换行：字段中是否有回车换行

文件编码：CSV 文件的编码



图 1

### 3.1.1.2 Excel 输入

图标:

作用: 将 Excel 文件中的数据作为输入。

参数:

步骤名称: 当前步骤的名称

表格类型 (引擎): 当前文件的类型

文件或目录: Excel 文件的文件名或者多个文件所在的目录

正则表达式: 进行文件筛选的表达式

正则表达式 (排除): 进行反向选择的正则表达式

选中的文件: 显示获取文件的位置



图 2

### 3.1.1.3 生成记录

图标: 

作用: 将生成的记录作为输入。

参数:

步骤名称: 当前步骤的名称

限制: 生成的记录限制

Never stop generating rows : 是否不停生成记录

Interval in ms (delay): 每次生成间隔时间

Current row time field name: 当前生成的时间字段名

Previous row time field name: 预生成的时间字段名

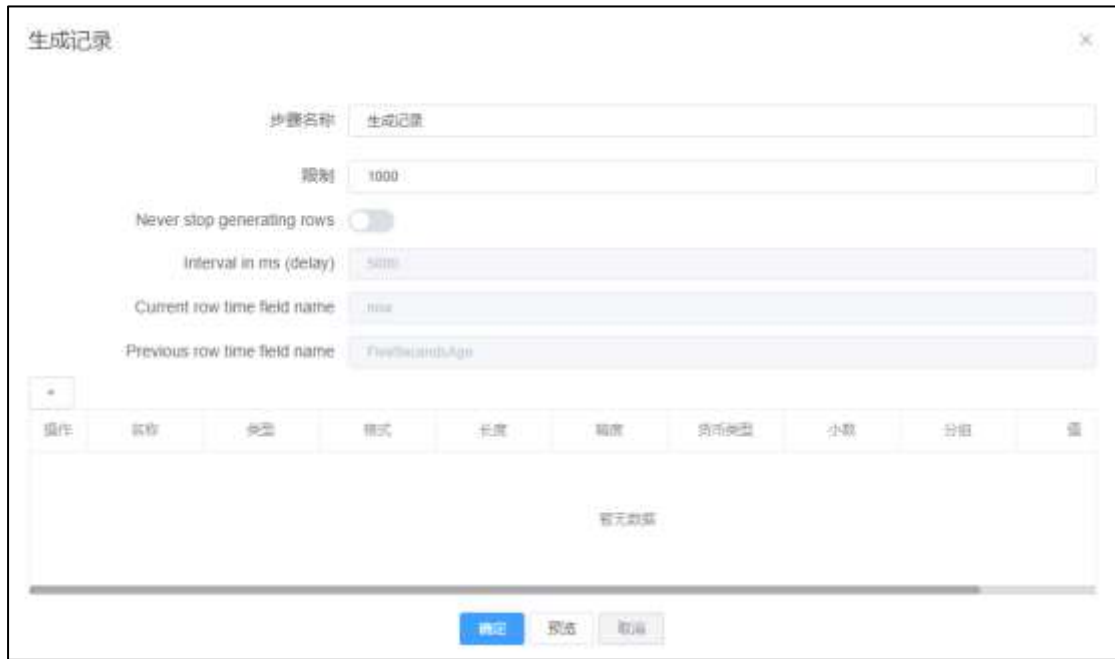


图 3

### 3.1.1.4 生成随机数

图标: 

作用: 将随机数作为输入。

参数:

步骤名称: 当前步骤的名称

字段: 添加一个生成的字段

名称: 添加的字段名称

类型: 生成的随机数类型



图 4

### 3.1.1.5 表输入



**作用:** 将数据库表作为输入。

**参数:**

步骤名称: 当前步骤的名称

数据库连接: 数据库连接信息, 新建则可以新建一个数据库连接

SQL: 进行数据抽取的 SQL 脚本

允许简易转换: 是否允许建议转换

替换 SQL 语句里的变量: 是否替换 SQL 语句中的变量

从步骤插入数据: 选择一个步骤插入数据

执行每一行: 执行 SQL 语句的每一行

记录限制, 抽取的记录限制



图 5

### 3.1.2 输出

#### 3.1.2.1 删除

图标:



作用: 删除部分数据并输出至数据库表

参数:

步骤名称: 当前步骤的名称

数据库连接: 选择一个数据库连接, 新建则可以新建一个数据库连接

目标模式: 输出数据库的模式

目标表: 输出输出数据库的表

提交记录的数量: 输出的记录数目

查询值所需要的关键字: 查询所需关键字



图 6

### 3.1.2.2 Excel 输入



图标:

作用: 将上一个流程的数据输出至文本文件。

参数:

步骤名称: 当前步骤的名称

文件名称: 输出的文件全路径

结果输送至命令行或脚本: 是否将结果输送至一个脚本或命令中

输出传递到 serclet: 是否将结果输出传递到 servlet

创建父目录: 在父目录不存在的情况下是否创建

启动时不创建文件: 启动时是否不创建文件

从字段中获取文件: 是否从字段中获取文件名

文件名字段: 用于做文件名的字段

扩展名: 后缀名

文件名里包含步骤数: 文件名里面是否天剑步骤数

文件名里包含数据分区号: 文件名里是否包含数据分区号

文件名里包含日期：文件名中是否包含日期

指定日期的格式：是否指定日期的格式

时间日期格式：选择一种时间日期的格式

结果种添加文件名：是否在结果中添加文件名



图 7

### 3.1.2.3 表输出

图标：



作用：将上个步骤生成的数据输出至数据表。

参数：

步骤名称：当前步骤的名称

数据库连接：选择一个数据库连接，点击新建即可新建数据库数据库连接

目标模式：输出数据库表的模式

目标表：输出表的名称

提交记录数量：输出的记录数目

裁剪表：是否裁剪表

忽略插入错误：是否忽略插入错误

制定数据库字段：是否指定数据库字段

表分区数据：是否为表分区数据

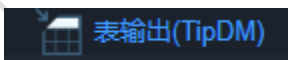
分区字段：分区的字段名称



图 8

### 3.1.2.4 表输出 (TipDM)

图标：



作用：将上一个步骤的数据输出至 TipDM 建模平台的数据源中。

参数：

步骤名称：当前步骤的名称

目标表：输出表的名称

提交记录数量：输出记录的数目

裁剪表：是否裁剪表



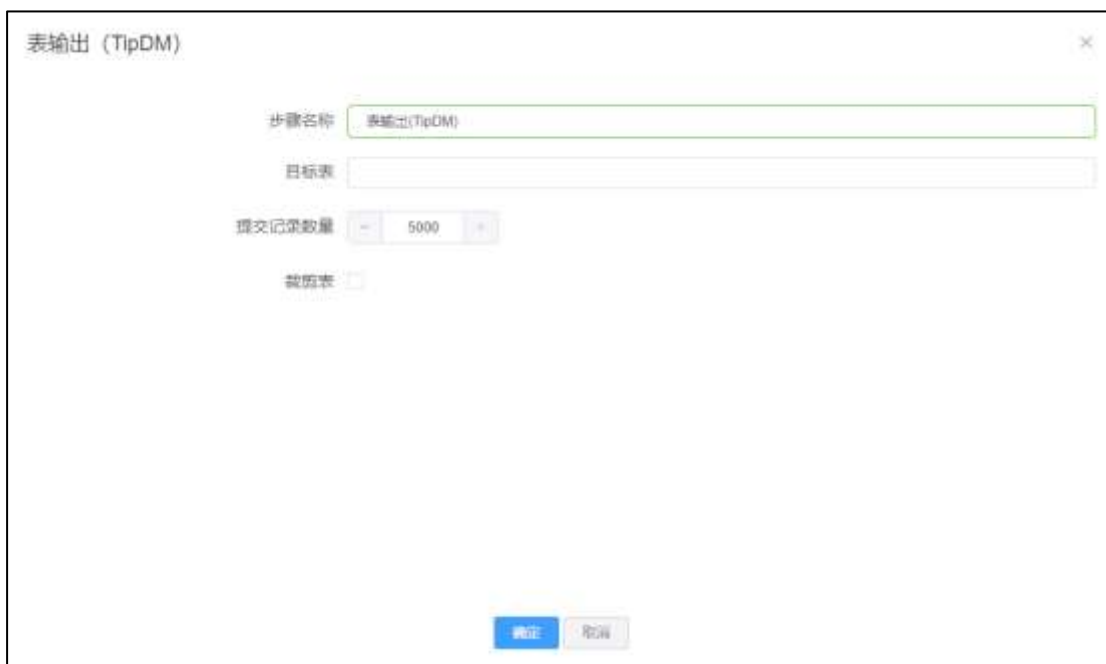


图 9

### 3.1.3 转换

#### 3.1.3.1 值映射

图标:



作用: 将值 A 映射为 B

参数:

步骤名称: 当前步骤的名称。

使用的字段名: 用于映射的字段名

目标字段名: 映射后的字段名。

不匹配时的默认值: 如若不匹配, 默认的值

添加一行: 添加一个匹配规则

源值: 原来的值

目标值: 映射后的值



图 10

### 3.1.3.2 列拆分为多行

图标:

作用: 将一列差分为多行。

参数:

步骤名称: 当前步骤的名称

要拆分的字段: 需要进行拆分的字段

分隔符: 拆分的分隔符

分隔符是一个正则表达式: 分隔符是否为一个正则表达式

新字段名: 新字段的名称

输出中包括行号: 输出中是否包含行号



图 11

### 3.1.3.3 列转行



**作用:** 将生列转为行

**参数:**

步骤名称: 当前步骤的名称

关键字段: 进行分组的主键

构成分组的字段: 进行分组的字段

目标字段: 分组后的字段



图 12

### 3.1.3.4 剪切字符串

图标: 

作用: 以位置剪切字符串

参数:

步骤名称: 当前步骤的名称

字段: 需要进行剪切的字段

输入流字段: 进行剪切的字段

输出流字段: 剪切后的字段名

起始位置: 进行剪切的起始位置

结束为止: 进行剪切的结束位置



图 13

### 3.1.3.5 去除重复记录

图标:  去除重复记录

作用: 去除重复记录。

参数:

步骤名称: 当前步骤的名称

增加计数器到输出: 是否增加计数器到输出

计算器字段: 计数器的字段名


重定向重复记录: 是否重定向重复记录

错误描述: 重定向发送的错误描述语句



图 14

### 3.1.3.6 唯一行（哈希）

图标:  唯一行 (哈希值)

作用: 去除重复记录。

参数:

步骤名称: 当前步骤的名称

使用存储的记录值进行比较: 是否使用存储的记录值进行比较

重定向重复记录: 是否重定向重复记录

错误描述: 重定向发送的错误描述语句



图 15

### 3.1.3.7 增加常量

图标: 

作用: 增加常量字段

参数:

步骤名称: 当前步骤的名称

+: 增加一个常量字段



图 16

### 3.1.3.8 增加序列

图标: 

作用: 增加有序字段

参数:

步骤名称: 当前步骤的名称

值的名称: 增加的序列名称

使用 DB 来获取 sequence: 是否使用数据库来获得序列

数据库连接: 选择指定数据库的连接, 新建可新建数据库连接

模式名称: 对应序列的模式名称

Sequence 名称: 对应序列的名称

使用计数器来计算 sequence: 是否使用计数器来生成 sequence

计数器名称 (可选): 当前计数器的名称

起始值: 生成序列的起始值

增长根据: 生成序列相邻两个值的差

最大值: 生成序列的结束值



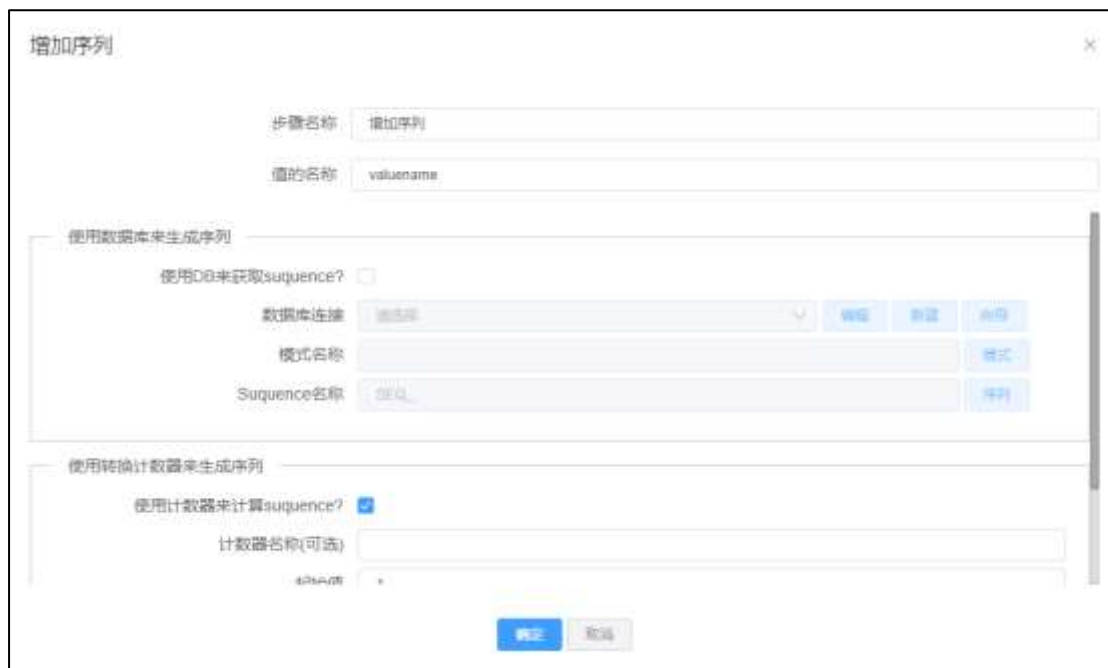


图 17

### 3.1.3.9 增加校验列

图标: 

作用: 增加一列校验结果。

参数:

步骤名称: 当前步骤的名称

类型: 校验的类型

结果类型: 校验结果的类型


结果字段: 校验结果生成的字段

兼容模式: 是否以兼容模式运行



图 18

### 3.1.3.10 字段选择

图标:  字段选择

作用: 选择字段

参数:

步骤名称: 当前步骤的名称

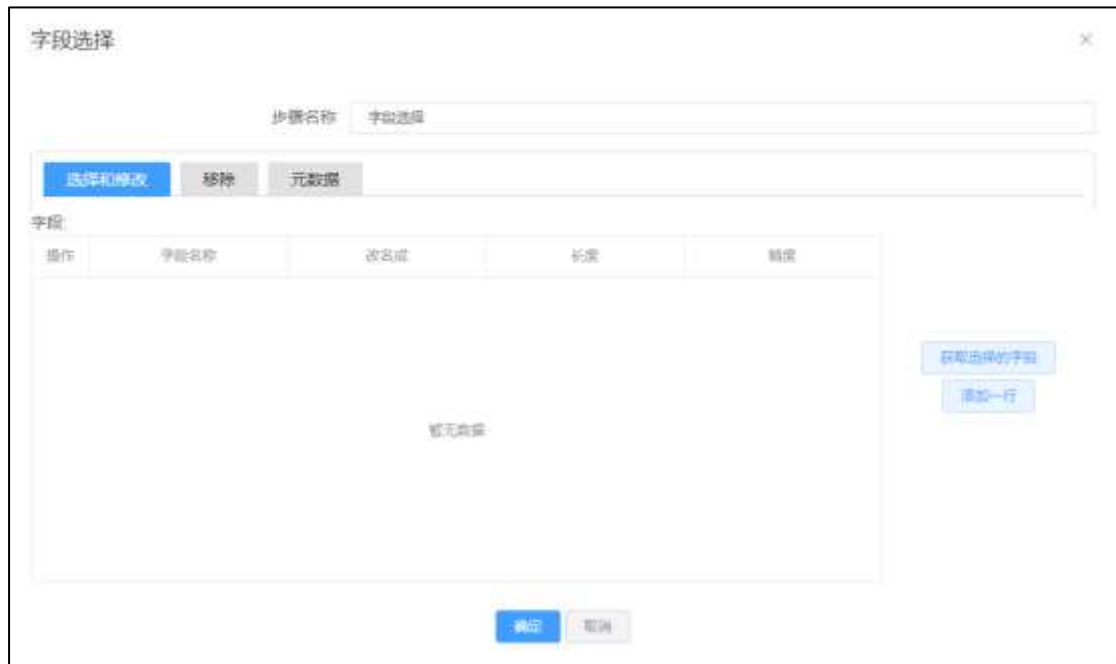


图 19

### 3.1.3.11 字符串操作

图标:



作用: 操作字符串

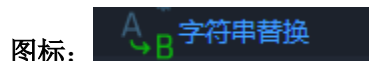
参数:

步骤名称: 当前步骤的名称



图 20

### 3.1.3.12 字符串替换



作用: 将指定字符串替换为另一个

参数:

步骤名称: 当前步骤的名称



图 21

### 3.1.3.13 将字段值设置为常量



作用: 将某个字段的值设置为常量

参数:

步骤名称: 当前步骤的名称

在常量中使用变量: 是否在常量中使用变量



图 22

### 3.1.3.14 拆分字段



**作用:** 去除重复记录。

**参数:**

步骤名称: 当前步骤的名称

需要拆分的字段: 选择需要进行拆分的字段

分隔符: 输入拆分的分隔符

Enclosure: 附加字符串



图 23

### 3.1.3.15 排序记录

图标:



作用: 根据某一列进行排序

参数:

步骤名称: 当前步骤的名称

排序目录: 进行排序的数据目录

临时文件前缀: 生成的临时文件前缀

排序缓存大小: 排序过程中最大的缓存大小

未使用内存限制 (%): 限定未使用内存不小于指定百分比

压缩临时文件: 是否压缩临时文件

仅仅传递非重复的记录 (仅仅校验关键字): 是否仅排序非重复的记录



图 24

### 3.1.3.16 数值范围

图标:

作用: 筛选出符合范围要求的数据

参数:

步骤名称: 当前步骤的名称

输入字段: 输入的字段名称

输出字段: 输出的字段名称

缺省值 (如果没有匹配到范围): 如果没有匹配到范围内的值, 则值为

范围: 字段及其取值范围



图 25

### 3.1.3.17 根据字段值来改变序列

图标:

作用: 根据某一列进行排序

参数:

步骤名称: 当前步骤的名称

Result Filed: 结果字段名

Start at value: 序列的起始值

Increase by: 序列相邻两个数的差, 步长





图 26

### 3.1.3.18 行扁平化

图标:



作用: 将某列的数据进行扁平化处理

参数:

步骤名称: 当前步骤的名称

需要扁平化的字段: 需要进行扁平化操作的字段



图 27

### 3.1.3.19 行转列

图标: 

作用: 根据某一列进行排序

参数:

步骤名称: 当前步骤的名称

key 字段: 行转列的主键



图 28

### 3.1.3.20 计算器

图标:  计算器

作用: 对上个步骤的结果进行简单计算

参数:

步骤名称: 当前步骤的名称



图 29

### 3.1.3.21 设置字段值

图标:

作用: 设置某个字段的值

参数:

步骤名称: 当前步骤的名称



图 30

### 3.1.4 应用

#### 3.1.4.1 替换 NULL 值

图标: 

作用: 将 NULL 值替换为指定值

参数:

步骤名称: 当前步骤的名称

值替换为: 将 NULL 替换的值为

设置空字符串: 是否设置为空字符串

掩码 (日期): 日期的格式

选择字段: 指定进行该操作的字段

选择值类型: 指定进行该操作的值的类型



图 31

### 3.1.5 流程

#### 3.1.5.1 空操作（什么也不做）

图标:



作用: 一个空操作

#### 3.1.5.2 过滤记录

图标:



作用: 根据指定条件过滤记录

参数:

步骤名称: 当前步骤的名称

发送 True 给步骤: 当为 True 时输出的字段

发送 False 数据给步骤: 当为 False 时输出的字段

条件: 筛选条件



图 32

### 3.1.6 查询

#### 3.1.6.1 REST Client

图标:

作用: 连接 API 获取数据

参数:

步骤名称: 当前步骤的名称

URL: 连接地址

Accept URL from field?: 是否从字段中接收连接地址

URL field name: 链接地址字段名

HTTP Method: HTTP 请求方式

Get Method from field: 是否从字段中获取请求方法

Method field name: 请求方法字段名

Body field: 主体字段名

Accept type: 接收数据类型

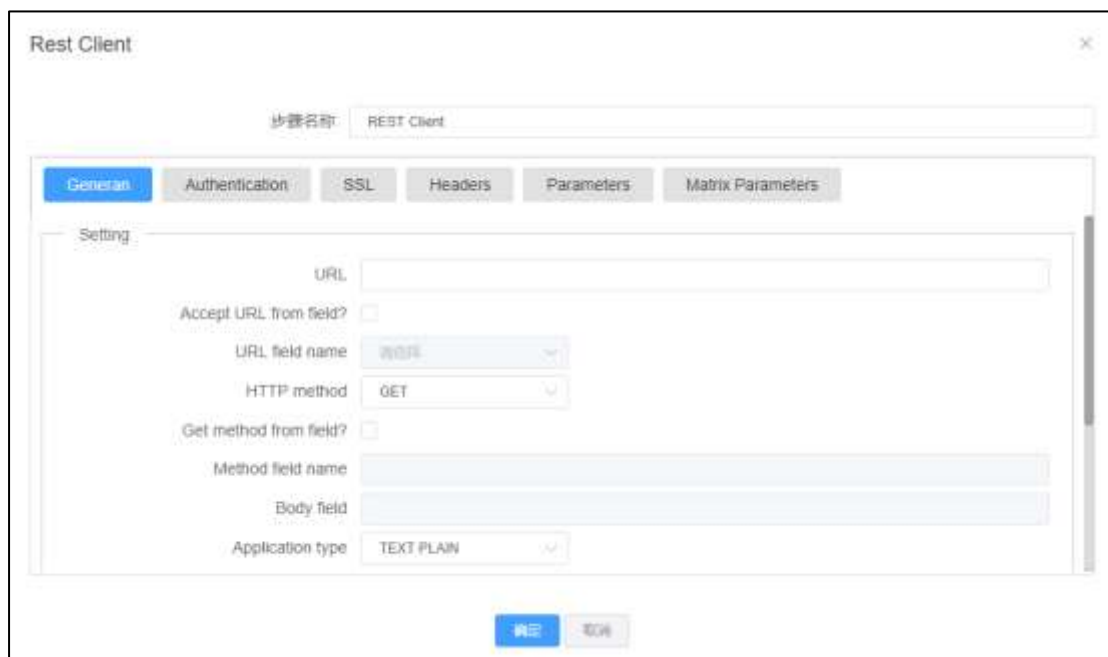


图 33

### 3.1.7 连接

#### 3.1.7.1 合并记录

图标:



作用: 合并两份数据源的数据

参数:

步骤名称: 当前步骤的名称。

旧数据源: 旧的数据源

新数据源: 新的数据源

标志字段: 主键字段

匹配的关键字: 进行合并的主键字段

数据字段: 参与合并的数据字段





图 34

### 3.1.7.2 排序合并

图标: 

作用: 先排序再进行合并

参数:

步骤名称: 当前步骤的名称

字段: 参与排序合并的字段



图 35

### 3.1.7.3 记录关联（笛卡尔输出）

图标: 

作用: 关联记录

参数:

步骤名称: 当前步骤的名称

临时目录: 临时文件存放的目录

临时文件前缀: 临时文件的前缀

最大缓存大小: 缓存最大的大小

Main step read from: 数据取自

条件: 进行合并的条件

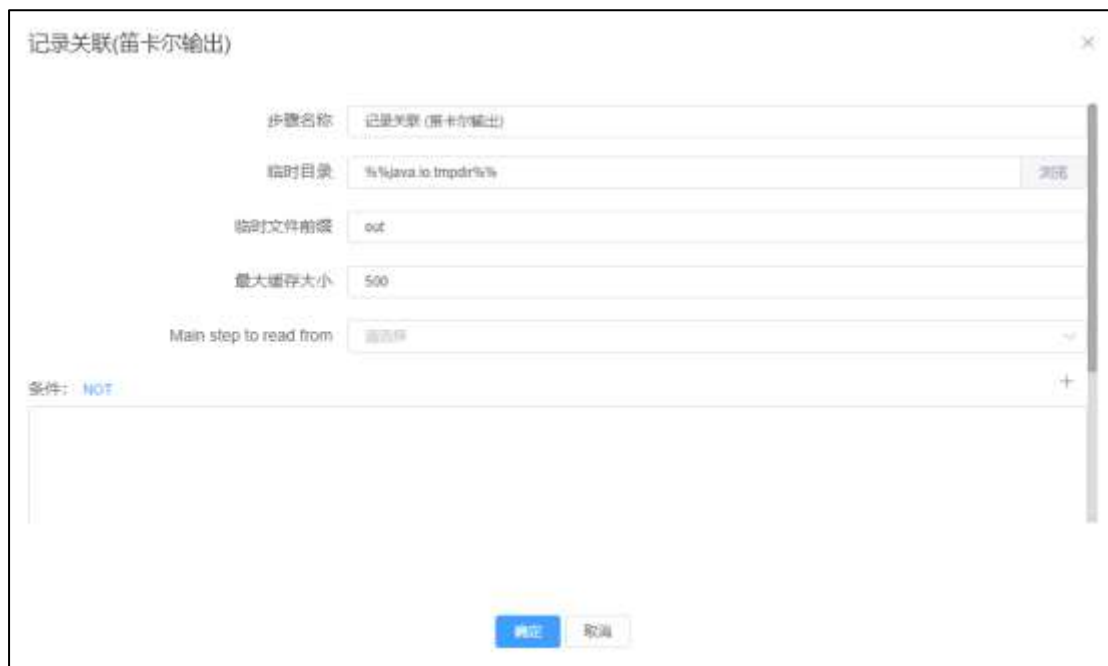


图 36

### 3.1.7.4 记录集连接

图标: 

作用: 主键连接两份数据

参数:

步骤名称: 当前步骤的名称

第一个步骤: 第一个数据源

第二个步骤: 第二个数据源

连接类型: 主键合并的类型

第一个步骤的连接字段: 第一个数据源参与合并的字段

第二个步骤的连接字段: 第二个数据源参与合并的



图 37

### 3.1.8 统计

#### 3.1.8.1 分组

图标:



作用:

参数:

步骤名称: 当前步骤的名称。

包括所有行: 是否包含所有的行

排序目录: 进行排序的临时文件目录

临时文件前缀: 临时文件的前缀

增加行号, 每组重新开始: 是否对每组增加行号

行数列名: 行数列的列名

总返回一个结果行: 是否返回一个结果

构成分组的字段: 参与分组的字段

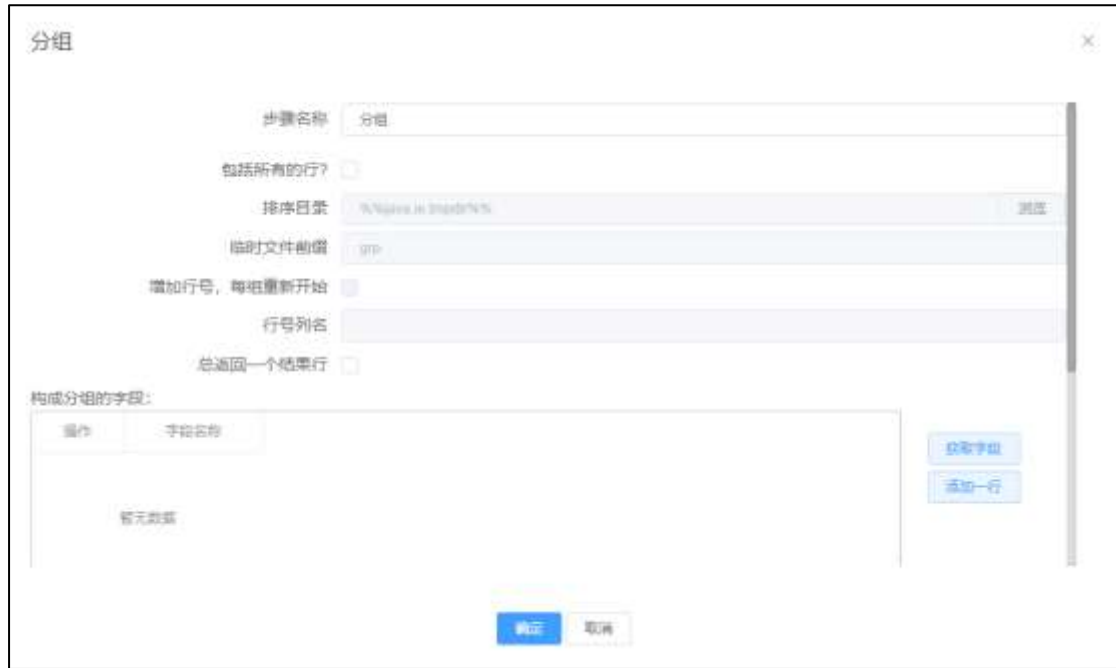


图 38

### 3.1.8.2 数据采样

图标:



作用: 随机抽取数据

参数:

步骤名称: 当前步骤的名称

样本数目: 随机抽取的样本数目

Random seed: 随机种子



数据采样

步骤名称 数据采样

样本行数 100

Random seed 1

确定 取消

图 39

## 3.2 作业

### 3.2.1 通用

#### 3.2.1.1 STRAT

图标: 

作用: 设置作业的定时调度规则

参数:

重复: 是否重复执行

类型: 定时间隔

以秒计算的间隔: 每隔多少秒执行一次

以分钟计算的间隔: 每隔多少分钟执行一次

每天: 每天执行的时间

每周: 每周执行的时间

每月: 每个月执行的时间

### 作业定时调度 ×

重复

类型 不需要定时

以秒计算的间隔 0

以分钟计算的间隔 60

每天 12 0

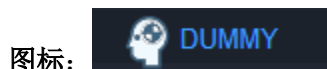
每周 星期一

每月 1

确定
取消

图 40

### 3.2.1.2 DUMMY



作用: 空作业

### 3.2.1.3 作业



作用: 作业中嵌套作业

参数:

步骤名称: 当前步骤的名称

Job: 选择当前用户中已存在的作业

每月: 每个月执行的时间

Run configuration: 运行的程序确认

执行每一个输入行: 是否执行每一个输入行

将作业执行结果发送到服务器上: 是否将作业执行的结果发送至指定位置

Enable monitoring for sub jobs and transforations: 是否可监控子任务和子转换

本地作业终止时远程作业也同时终止：本地作业和远程作业是否同步停止



图 41

### 3.2.1.4 成功

图标:

作用: 判定是否成功

参数:

步骤名称: 当前步骤的名称



图 42

### 3.2.1.5 设置变量





图标:

作用: 设置一个变量

参数:

步骤名称: 当前步骤的名称

属性文件名: 变量所在的文件名

变量有效范围: 变量在哪些地方有效

变量替换: 是否替换变量

变量: 添加的变量



图 43

### 3.2.1.6 转换



图标:

作用: 在作业中添加一个转换作为一个步骤

参数:

步骤名称: 当前步骤的名称

**Transformation:** 选择一个当前用户所拥有的转换

**Run configuration:** 运行的程序确认

执行每一个输入行: 是否执行每一个输入行

将作业执行结果发送到服务器上: 是否将作业执行的结果发送至指定位置

**Enable monitoring for sub jobs and transforations:** 是否可监控子任务和子转换

本地作业终止时远程作业也同时终止: 本地作业和远程作业是否同步停止

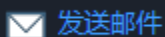


图 44

### 3.2.2 邮件

#### 3.2.2.1 发送邮件

图标:



作用: 发送上个步骤的结果给邮件

参数:

步骤名称: 当前步骤的名称

收件人地址: 收件人邮箱

抄送: 抄送邮箱地址

暗送: 暗送邮箱地址

回复名称：回复用户名

发件人地址：发件人的邮箱

联系人：联系人姓名

联系电话：联系人联系方式

SMTP 服务器：发送服务器

端口号：端口号

用户验证：是否验证

用户名：邮箱地址

密码：邮箱密码

使用安全验证：是否使用安全验证

安全连接类型：安全连接的类型

信息里带日期：是否在信息里提供日期

只发送邮件注释：是否只发送邮件注释

使用 HTML 邮件的格式：是否使用 HTML 邮件的格式

编码：邮件正文的编码

管理优先级：是否管理优先级

优先级：优先级类型

重要：重要程度

Sensitivity：敏感度

主题：邮件的主体

注释：邮件的注释

带附件？：邮件是否带附件

文件类型：附件的文件类型

压缩成统一文件格式：是否将文件压缩至统一文件格式

压缩文件名称：压缩文件的名称

文件名：图片文件名

内容 ID：内部文字 ID



图 45

### 3.2.3 文件管理

#### 3.2.3.1 创建一个目录

图标:



作用: 在指定位置创建一个目录

参数:

步骤名称: 当前步骤的名称

目录名: 创建的目录名称

如果目录存在则失败: 在目录存在的情况创建目录是否失败

创建一个目录

步骤名称 创建一个目录 2

目录名 浏览

如果目录存在则失败

确定 取消

图 46

### 3.2.3.2 删除一个文件



图标:

作用: 删除某个文件

参数:

步骤名称: 当前步骤的名称

文件名: 删除的文件名称

如果文件不存在则失败: 在文件不存在的情况是否失败

删除一个文件

步骤名称 删除一个文件

文件名 浏览

如果文件不存在则失败

确定 取消

图 47

### 3.2.3.3 删除多个文件



图标:

作用: 删除目录中的多个文件

参数:

步骤名称: 当前步骤的名称

包含子文件夹: 是否包含子文件夹中的文件

复制上一步的结果到参数: 是否复制上一步的结果到参数

文件/文件夹: 删除的文件或文件夹

通配符 (正则表达式): 进行通配的正则

文件/文件夹: 符合要求的文件/文件夹



图 48

### 3.2.3.4 删除目录



图标:

作用: 删除某个目录

参数:

步骤名称: 当前步骤的名称

保存上一步的结果到参数: 是否保存上一步的结果到参数

成功条件: 选择成功的条件

数目：允许的错误条数

文件夹：删除的文件夹

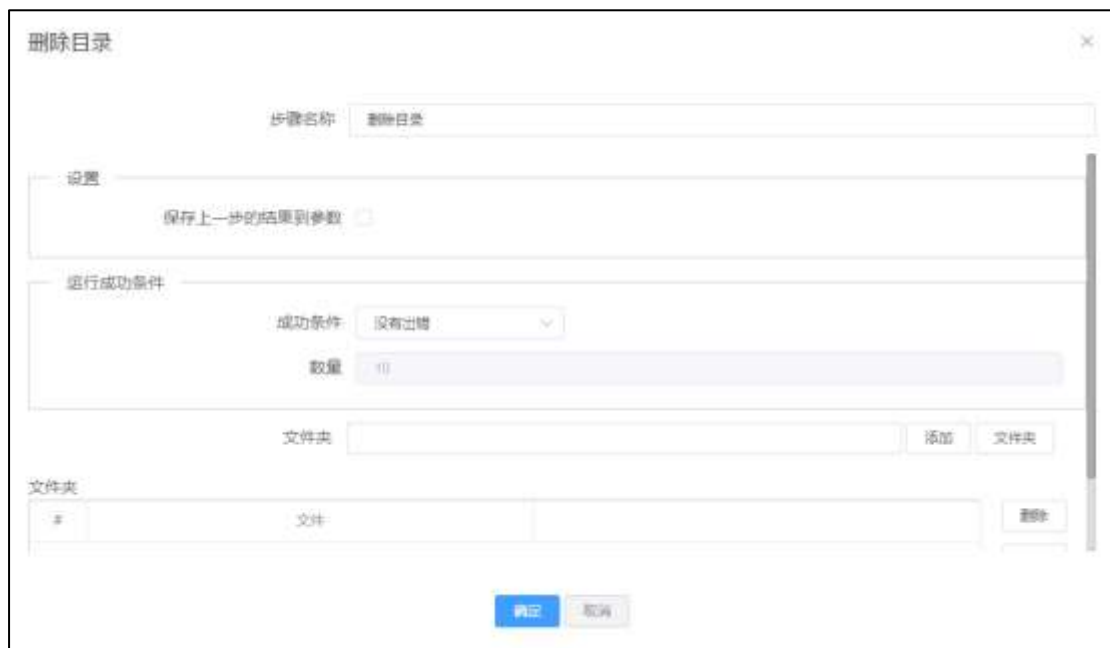


图 49

### 3.2.3.5 赋值/移动结果文件

图标: 

作用：复制或移动结果文件

参数：

步骤名称：当前步骤的名称

动作：进行的操作类型

目标文件夹：移动的终点文件夹

创建目标文件夹：是否创建一个新文件夹

替换已经存在的文件夹：是否替换现有的文件夹

从结果中移除源文件名：是否从结果中移除源文件的文件名

添加日期到文件名：是否将日期添加到文件名

添加时间到文件名：是否将时间添加到文件名

指定日期时间格式：是否指定日期时间格式

日期时间格式：日期时间显示的格式

在扩展名前添加日期：是否在扩展名前添加日期

动作范围：进行操作的动作范围

通配符（正则表达式）：匹配的正则表达式

排除通配符（正则表达式）：反向匹配的正则表达式

成功条件：成功的条件

数值：允许错误的最大数目

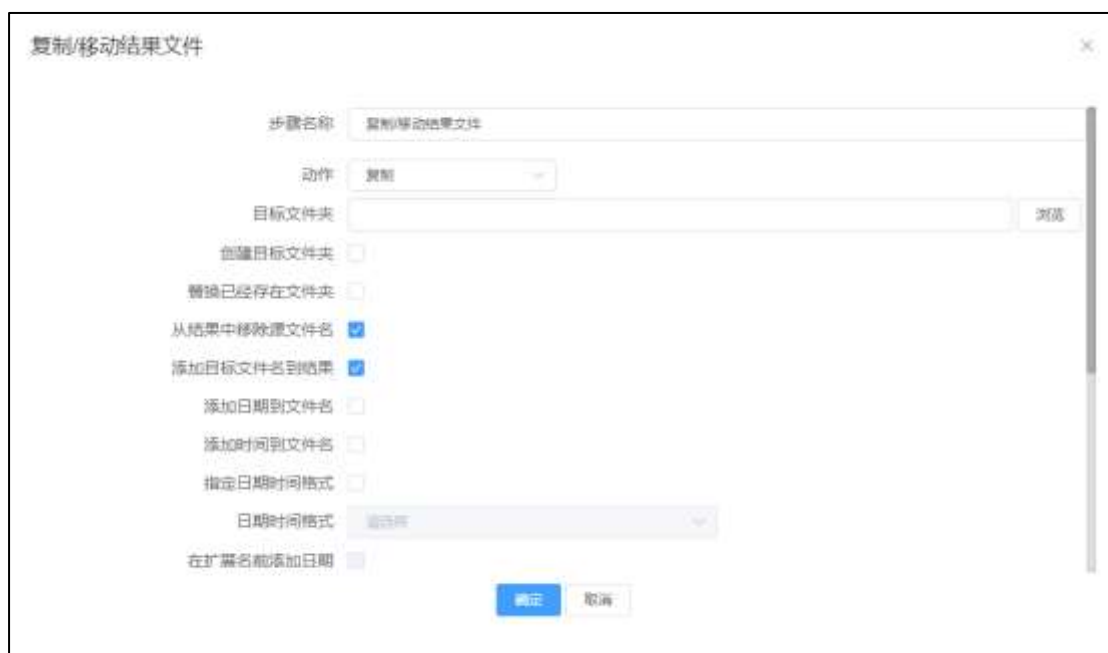


图 50

### 3.2.3.6 复制文件

图标：

作用：复制一个文件

参数：

步骤名称：当前步骤的名称

文件/目录：进行复制的文件及目标目录

包括子目录：复制过程是否复制子目录

目标是文件：复制的是否为单个文件

复制空目录：是否复制空目录



创建目标目录：在目录不存在的情况下是否创建目标目录

替换已存在的目录：是否替换已存在的目录

移除源文件：是否移除源文件

复制上一个作业结果为参数：是否复制上一个作业结果作为参数

增加文件：是否添加文件

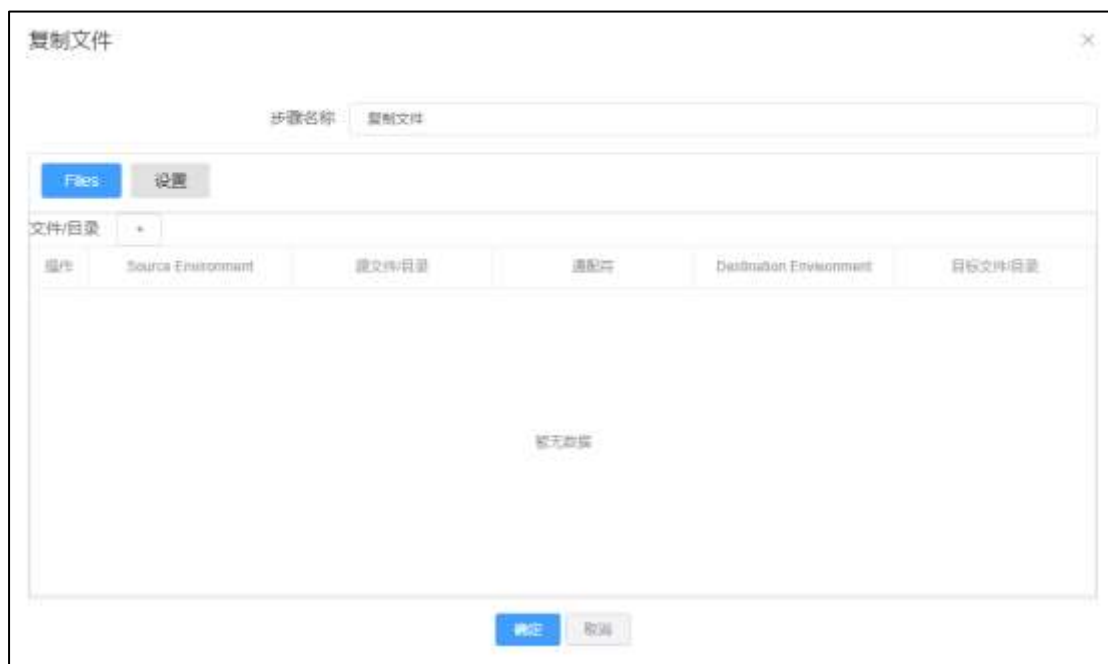


图 51

### 3.2.4 条件

#### 3.2.4.1 检查数据库连接

图标：



作用：检查一个数据库连接是否正常

参数：

步骤名称：当前步骤的名称



图 52

### 3.2.4.2 检查检查表是否存在

图标:

作用: 检查某个连接下的某张表是否存在

参数:

步骤名称: 当前步骤的名称

数据库连接名称: 数据连接的名称, 点击新建按钮可新建

表名: 需要检查的表



图 53

## 3.2.1 文件传输

### 3.2.1.1 检查数据库连接



作用: 通过 SFTP 的方式上传文件

参数:

步骤名称: 当前步骤的名称

SFTP 服务器名称/IP: 上传目标服务器的 IP

端口号: 目标服务器的上传端口号

用户名: 目标服务器用户名

密码: 目标服务器用户名对应的密码

使用私钥文件: 是否使用私钥文件

私钥文件: 私钥文件文件所在的地址

密钥: 密钥

代理类型: 代理的类型

代理主机: 代理的主机 IP

代理端口: 代理主机的端口

代理用户名: 代理主机的用户名

代理密码: 代理主机对应用户名的密码

压缩: 压缩的方式

将上一作业项的结果作为参数: 是否将上移作业项的结果作为参数

Copy previous result files to: 复制以前的结果至指定位置

本地目录: 上传文件所在的目录

通配符 (正则表达式): 上传文件的匹配正则表达式

当本地没有文件时运行成功: 当本地没有文件时运行是否成功

SFTP 上传后: 上传成功后的动作

目标文件夹：选择上传成功后移动的目标文件

创建目标文件夹：是否创建目标文件夹

添加文件到结果文件列表：是否将文件添加到结果文件列表

远程目录：上传至主机的目录

创建文件夹：是否创建文件夹



图 54

### 3.2.1.2 SFTP 下载

图标：



作用：下载 SFTP 服务器中的某个文件

参数：

步骤名称：当前步骤的名称

步骤名称：当前步骤的名称

SFTP 服务器名称/IP：上传目标服务器的 IP

端口号：目标服务器的上传端口号

用户名：目标服务器用户名

密码：目标服务器用户名对应的密码

使用私钥文件：是否使用私钥文件

私钥文件：私钥文件文件所在的地址

密钥：密钥

代理类型：代理的类型

代理主机：代理的主机 IP

代理端口：代理主机的端口

代理用户名：代理主机的用户名

代理密码：代理主机对应用户名的密码

压缩：压缩的方式

将上一作业项的结果作为参数：是否将上移作业项的结果作为参数

远程目录：SFTP 服务器中的文件目录

通配符（正则表达式）：下载文件匹配的正则表达式

获取后删除服务器文件：是否在成功下载后删除服务上的文件

目标目录：复制至本地的目标目录

创建目标文件：文件不存在是否创建

添加文件名到结果：是否添加文件名到结果



图 55

## 4 用户权限管理

用户权限管理仅限超级管理员使用，对用户使用权限进行管理

### 4.1 组织架构

#### 4.1.1 新增组织架构

如图 56 所示，点击【新增】按钮，则会弹出如图 57 所示的对话框。

**组织架构名称：**必填项。

**描述：**必填项。

**上级组织架构：**选填项。

**排序优先级：**必填项。

**状态：**组织架构是否可用。

The screenshot shows the 'Organization Management' (组织架构管理) page. It includes a search bar for organization names and creation times, and a table listing existing organizations. The table has columns for ID, Organization Name, Description, Parent Organization, Status, Creation Time, Creator, and Actions. Three organizations are listed: '市场部' (Marketing Dept), '研发部' (R&D Dept), and '销售部' (Sales Dept). The status for all is '启用' (Enabled).

ID	组织架构名称	描述	上级组织架构	状态	创建时间	创建人	操作
1	市场部	市场部		启用	2017/11/28	admin	编辑 删除
2	研发部	研发部		启用	2017/9/11	admin	编辑 删除
3	销售部	销售部(销售部ON产品)		启用	2017/2/21	admin	编辑 删除

显示 1 到 3 条记录，总共 3 条记录

图 56

**添加组织架构**

\*组织架构名称:

\*描述:

上级组织架构:

- 销售部门
  - 创新部
  - 产品部

\*排序优先级:

状态:  可用  不可用

图 57

#### 4.1.2 删除组织架构

在组织架构表的第一列进行勾选，再点击【删除】，即可将勾选的组织架构进行删除。如图 58 所示。

	组织架构名称	描述	上级组织架构	状态	创建时间	创建人	操作
<input checked="" type="checkbox"/>	产品部	产品部		启用	2017/11/20	admin	<input type="button" value="编辑"/> <input type="button" value="删除"/>
<input type="checkbox"/>	创新部	创新部		启用	2017/5/11	admin	<input type="button" value="编辑"/> <input type="button" value="删除"/>
<input type="checkbox"/>	销售部门	销售部门销售OA产品		启用	2017/2/21	admin	<input type="button" value="编辑"/> <input type="button" value="删除"/>

显示第 1 到第 3 条记录，总共 3 条记录

图 58

#### 4.1.3 编辑组织架构

点击【编辑按钮】，可对已有的组织架构进行编辑，可编辑的内容与新建组件架构相同，如图 59 所示。

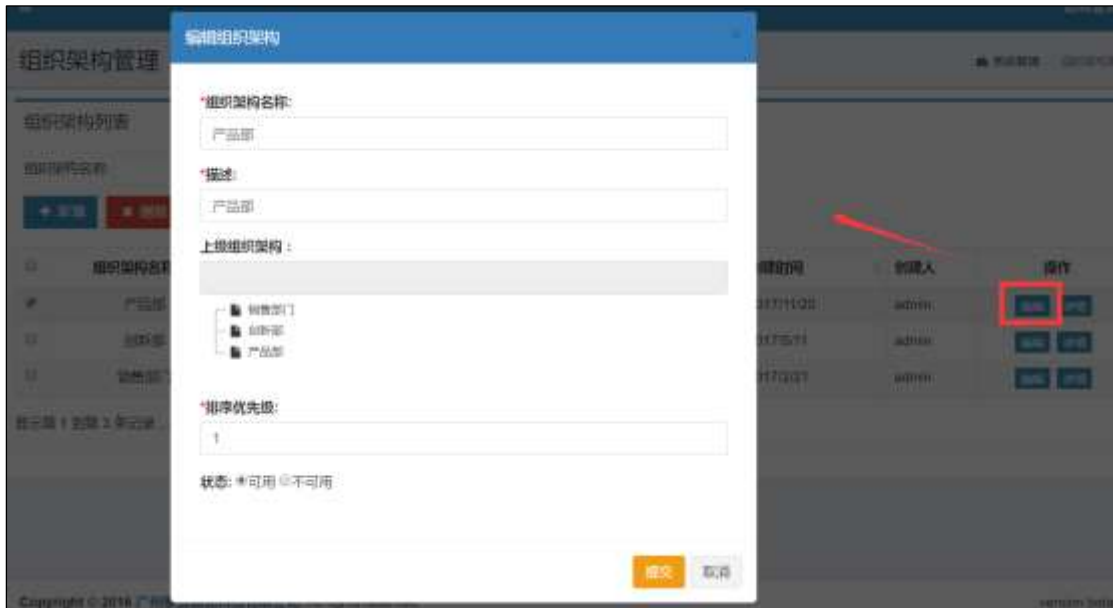


图 59

#### 4.1.4 查看组织架构信息

点击【详情】可查看组织架构的详细信息，包括：组织架构名称、描述、上级组织架构、排序优先级、状态、创建者、创建时间，如图 60 所示。



图 60



## 4.2 角色管理

### 4.2.1 新增角色

点击【新增】可新增角色，如图 61 所示。新增角色需要填写的信息如图 62 所示。

**角色名：**必填项，建议使用英文字母。

**描述：**必填项，对角色进行描述。

角色列表

角色名:  创建时间:

<input type="checkbox"/>	角色名	描述	是否系统默认角色	创建时间	创建人	操作
<input type="checkbox"/>	guest	来宾	是	2015/10/28	admin	<input type="button" value="编辑"/> <input type="button" value="详情"/> <input type="button" value="设为默认角色"/> <input type="button" value="删除角色"/> <input type="button" value="角色可搜索用户"/>
<input type="checkbox"/>	performancetest		否	2017/4/12	admin	<input type="button" value="编辑"/> <input type="button" value="详情"/> <input type="button" value="设为默认角色"/> <input type="button" value="删除角色"/> <input type="button" value="角色可搜索用户"/>
<input type="checkbox"/>	admin	超管	否	2015/10/19	admin	<input type="button" value="编辑"/> <input type="button" value="详情"/> <input type="button" value="设为默认角色"/> <input type="button" value="删除角色"/> <input type="button" value="角色可搜索用户"/>

显示第 1 到第 3 条记录，总共 3 条记录

图 61

添加角色

\*角色名:

\*描述:

图 62

### 4.2.2 删除角色

通过勾选角色，点击【删除】，可以删除已经存在的角色。如图 63 所示。



图 63

### 4.2.3 编辑角色

点击【编辑】按钮可以对角色的描述及状态进行编辑，如图 64 所示。

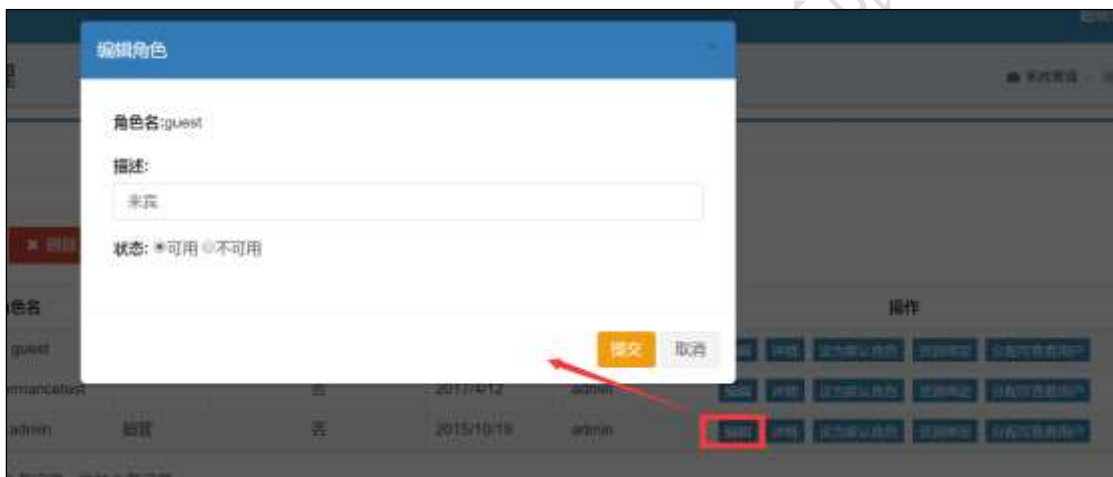


图 64

### 4.2.4 资源绑定

可给不同的角色绑定不同的资源。如图 67 所示。



图 65

#### 4.2.5 查看角色详情

可查看角色的详情，包括：用户名、描述、角色资源、状态、创建者、创建时间，如图 66 所示。



图 66

## 4.2.6 设为默认角色

如下图 67 所示。



图 67

## 4.1 用户管理

### 4.1.1 新增用户

如图 68 所示，点击【新增】按钮可新增用户，则会弹出如图 69 所示的对话框，填写相应信息即可新增用户。

**用户名：**必填项。

**邮箱：**必填项。

**密码：**必填项。

**别名：**必填项。

**组织架构：**可选项。



图 68

添加用户

\*用户名:

\*邮箱: admin

\*密码: .....

\*别名:

组织架构(按住Ctrl键可多选):

- 销售部门
- 生产部
- 研发部

提交 取消

图 69

#### 4.1.2 编辑用户

如图 70 所示。

编辑用户

用户名: yanger

邮箱: weruzz@tpdm.com

别名: yanger

状态:  锁定  未锁定

组织架构(按住Ctrl键可多选):

- 销售部门
- 生产部
- 研发部

提交

操作: 删除, 重置, 忘记密码, 编辑, 详情, 冻结/解冻, 锁定/解锁, 禁用/启用

图 70

#### 4.1.3 查询用户

如图 71 所示。

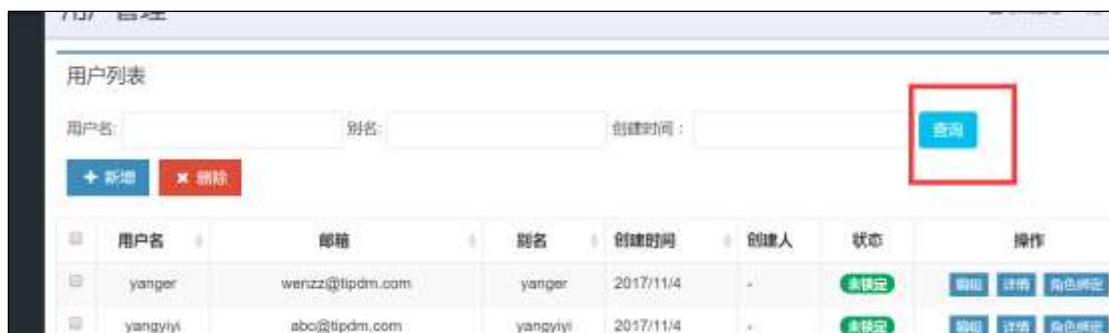


图 71

#### 4.1.4 角色绑定

如图 72 所示。

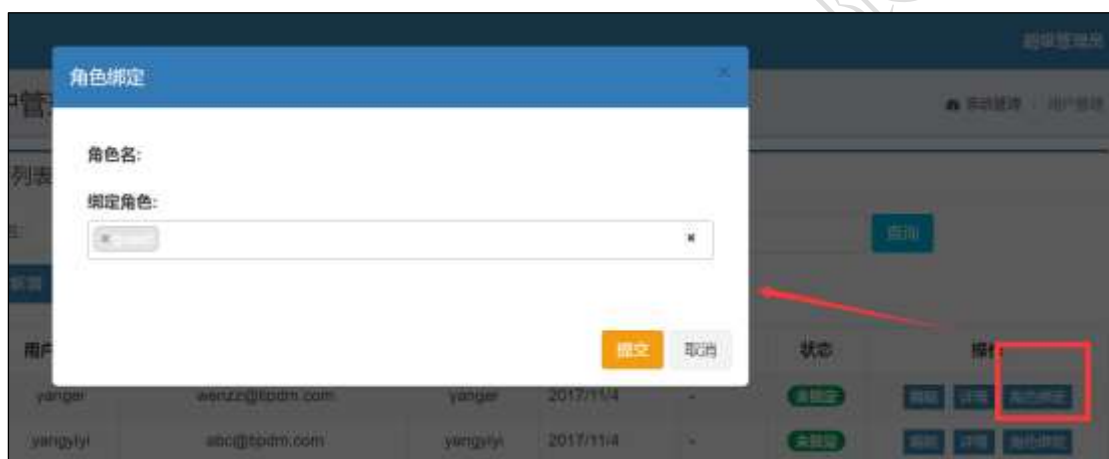


图 72

### 4.2 资源管理

#### 4.2.1 资源绑定

资源绑定包括基础权限资源及数据分析平台资源。

基础权限资源：主要针对用户管理系统部分的资源进行绑定。

数据分析平台权限资源：主要针对挖掘分析平台部分资源进行绑定，具体见 4.2.1。

如图 73、图 74 所示。

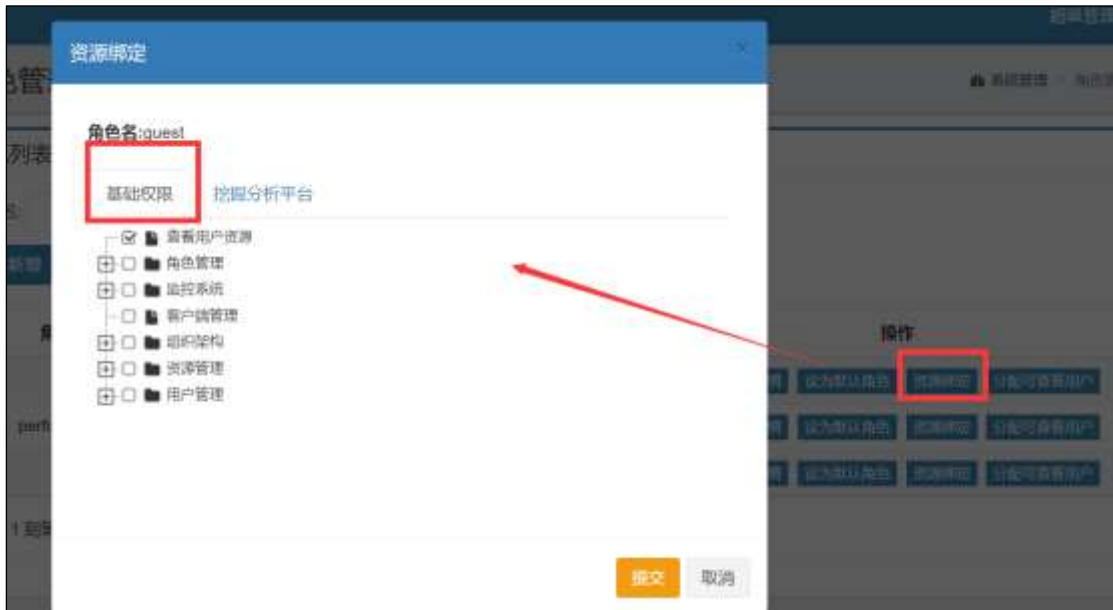


图 73



图 74

#### 4.2.1 资源管理

资源管理包括：基础权限资源、挖掘分析平台资源。

基础资源具体包括：

序号	一级菜单	二级菜单
----	------	------

1	角色管理	绑定角色
2		创建角色
3		删除角色
4		编辑角色
5		查询角色
6		设为系统默认角色
7	监控系统	监控系统
8	客户端管理	客户端管理
9	组织架构	创建组织架构
10		编辑组织架构
11		删除组织架构
12		查询组织架构
13	资源管理	创建资源域
14		创建资源
15		编辑资源
16		查询资源
17		删除资源
18	用户管理	绑定角色
19		创建用户
20		删除用户
21		编辑用户
22		查询用户
23	查看用户资源	查看用户资源

挖掘建模平台资源具体包括：

序号	一级菜单	二级菜单	序号	一级菜单	二级菜单
1	模型管理	检验模型	53	工程管理	生成 URL
2		分享模型	54		删除历史版本
3		删除分类/模型	55		另存为模板
4		获取个人模型列表	56		恢复到指定版本
5		分享的模型列表	57		创建历史版本
6		新增模型分类	58		停止流程调度
7		保存模型	59		执行该节点
8		导入模型	60		从此节点开始执行
9		模型导出	61		执行到此处
10		卸载模型	62		查看运行结果
11		获取个人的模型列表	63		查看运行日志
12		获取分享模型列表	64		输出节点数据可视化
13	横幅链接配置	社区	65	查看算法源码	
14		帮助文档	66	模板克隆	



15		过滤任务数据	67		获取子级节点
16		开始定时任务	68		获取个人模型评估列表
17		添加任务数据	69		新生成模型-重新部署
18	任务管理	更新任务数据	70	模型评估	部署模型
19		删除任务数据	71		新生成模型-部署模型
20		获取任务数据	72		获取个人模型列表
21		介绍定时任务	73		模型预测 (CSV)
22		获取表的结构信息	74		模型预测 (RDBMS)
23		从 HDFS 路径新建 HIVE 表	75		模型预测
24	Hive 表管理	新建 HIVE 表	76	用户信息管理	用户信息管理
25		过滤表	77		获取表的结构信息
26		测试数据库连接	78		检查表是否存在
27		新建数据库连接信息	79		删除表
28		同步表数据	80		创建表
29	文件管理	检查文件是否存在	81	数据源管理	数据预览
30		上传数据文件	82		测试数据库连接
31		上传算法包文件	83		预览数据
32	模板管理	模板列表	84		预览分享数据
33		模板删除	85		新建数据库连接
34	HDFS 文件管理系统	查看内容的 URL	86		同步表数据
35		判断 URL	87		过滤表
36		删除 URL	88		搜索分享列表
37		获取创建的 URL	89		搜索数据
38		匹配相应字符的路径	90		分享数据源
39	系统设置	修改文件内容	91	组件管理	上移
40		获取内容	92		获取算法集合
41		获取配置文件	93		删除组件参数
42	工程管理	执行流程	94		修改类别名
43		删除工程	95		删除组件分类
44		获取工程	96		获取自分类的算法
45		创建工程	97		删除组件
46		新增工程分类	98		更新组件数据
47		同步工程	99		获取组件数据
48		另存为工程	100		添加组件
49		工程排序上移	101		添加组件分类
50		查看数据	102		添加基础组件类型
51		查看工程的历史版本	103		获取基础组件集合
52	删除工程分类	104	系统组件		

### 4.3 客户端管理

如下图 75 所示。



图 75

### 4.4 监控系统

如下图 76 所示。



图 76