

TongTech[®]

东方通数据集成软件

TongDI V2.0

操作手册

2021年07月

东方通 | *TongTech*[®]

北京东方通科技股份有限公司

地址：北京市海淀区中关村南大街2号数码大厦A座22层

客户服务热线：400-650-7088

目 录

1. 产品介绍	1
2. 准备工作	1
2.1 登录系统.....	1
2.2 收藏产品.....	1
3. 项目管理	1
3.1 项目管理.....	1
3.1.1 创建项目.....	1
3.1.2 编辑项目	4
3.1.3 停用/启用项目	4
3.1.4 回收资源.....	4
3.1.5 进入产品页.....	5
3.1.6 项目授权.....	5
3.1.7 申请使用.....	6
3.1.8 删除项目.....	6
3.2 项目申请授权.....	6
3.2.1 申请授权.....	6
3.2.2 申请撤回.....	8
3.2.3 授权查看.....	8
3.2.4 授权冻结/恢复.....	9
3.2.5 授权终止.....	9
3.3 项目申请管理.....	10
3.3.1 申请查看.....	10
3.3.2 申请删除.....	10
4. 分类管理	11
5. 转换管理	11
5.1 转换管理页.....	11
5.2 转换管理-创建转换.....	12
5.2.1 转换-转换设置.....	14
5.2.2 转换-组件视图	18
5.2.3 转换-资源视图	18
5.2.4 转换-转换组件选项	19
5.2.5 转换-转换连接选项	20
5.2.6 转换-转换功能栏.....	20
5.3 转换管理-流程操作.....	20
5.3.1 流程操作-编辑.....	20
5.3.2 流程操作-复制.....	21
5.3.3 流程操作-部署	21
5.3.4 流程操作-回滚.....	22
5.3.5 流程操作-删除.....	23
5.3.6 流程操作-检出.....	23
5.3.7 流程操作-检入.....	23
5.3.8 流程操作-产品版本查看日志.....	24
5.3.9 流程操作-产品版本详情	24
5.3.10 流程批量部署	24
5.3.11 流程导出	24
6. 任务管理	25
6.1 任务管理页.....	25
6.2 任务管理-创建任务.....	25
6.2.1 任务-任务设置.....	27
6.2.2 任务-组件视图.....	30

6.2.3	任务-资源视图	30
6.2.4	任务-任务组件选项	31
6.2.5	任务-任务连接选项	31
6.2.6	任务-任务功能栏	32
6.3	任务管理-流程操作	32
6.4	任务管理-任务监控	32
6.4.1	创建带子转换的任务流程	32
7.	场景服务	34
7.1	异构表-全量同步	35
7.2	异构表-全库备份	37
7.3	异构表-触发器场景	37
7.4	异构表-时间戳(自增序列)场景	38
7.5	异构表-CDC 场景	39
7.6	同构表-全量同步	40
7.7	同构表-全库备份	40
7.8	同构表-触发器场景	40
7.9	同构表-时间戳(自增序列)场景	41
7.10	同构表-CDC 场景	41
7.11	接口数据到库表	41
7.12	接口数据到文件	43
7.13	场景服务列表-操作	44
7.13.1	场景服务流程-库表类型-编辑	44
7.13.2	场景服务流程-接口类型-编辑	44
7.13.3	场景服务流程-部署	45
7.13.4	场景服务流程-删除	45
8.	CDC 任务创建流程	45
8.1	创建抽取	45
8.2	启动任务	46
8.3	停止任务	47
8.4	编辑任务	47
9.	示例资源	47
9.1	典型场景	47
9.1.1	全量同步-任务列表	47
9.1.2	全量同步-转换列表	48
9.1.3	增量同步-任务列表	48
9.1.4	增量同步-转换列表	49
9.2	转换组件示例	49
10.	周期任务开发	49
10.1	选择项目	49
10.2	解决方案	50
10.2.1	创建解决方案	50
10.3	数据开发流程	52
10.3.1	创建周期业务流程	52
10.3.2	新增节点	52
10.3.3	配置节点	53
10.3.4	节点调度配置	58
10.3.5	节点依赖配置	64
10.3.6	配置流程参数	65
10.3.7	提交节点任务	66
10.4	数据处理组件	66
10.4.1	SparkSQL	66
10.4.2	Shell	67
10.4.3	Spark	68

10.4.4 DataFlow	69
11. 资源管理.....	85
11.1 项目表.....	85
11.1.1 创建事实模型	86
11.1.2 创建视图模型	90
11.1.3 创建普通维度模型	92
11.1.4 创建层级维度模型	94
11.1.5 创建枚举维度模型	97
11.2 函数管理.....	99
11.2.1 新增函数	99
11.2.2 查看函数	101
11.2.3 删除函数	102
11.3 自定义算子.....	103
11.3.1 创建资源文件夹.....	103
11.3.2 新增资源	104
11.3.3 查看资源	104
11.3.4 替换资源	105
11.3.5 删除资源	105

1. 产品介绍

东方通数据集成软件具备数据集成和数据开发能力，通过管理控制台，实现产品开发的项目管理、任务管理，任务全生命周期的管理，还有任务转换管理、任务监控、任务转换、日志审计等 Web 管理功能。同时东方通数据集成软件还新增了 Web 版本的集成开发工具，实现了数据平台的可视化开发，具备分布式调度能力。

2. 准备工作

2.1 登录系统

打开系统登录界面，在 Chrome 浏览器中输入登录地址：http://10.10.64.77

备注：登录地址根据实际服务发布的地址进行修改；

进入数据服务化功能模块，首先需要登录数据集成软件 TongDI，如下图所示：



图 2-1

输入账号、密码、验证码点击“登录”。

也可以通过短信验证码方式登录：输入手机号，点击“获取验证码”，输入收到的验证码，点击“登录”。

2.2 收藏产品

输入账号，密码点击“登录”。进入数据集成软件，点击全部产品，展开显示产品信息，选中数据集成下的“TongDI”收藏按钮★，TongDI 系统出现在导航菜单列表。

3. 项目管理

项目管理通过项目形式将平台的一系列业务进行统一管理。当进行数据建模、数据质量管理、ETL 等任务时，需要首先创建项目，在项目中来完成这些任务。对项目有操作、授权、管理权限的一般为系统管理员/项目管理员，其余用户/机构使用项目需要申请，申请需要通过审核，审核通过后，申请的用户/机构才能使用项目。

3.1 项目管理

3.1.1 创建项目

在菜单栏中点击【项目管理】，进入页面后点击“创建项目”，在弹窗中填写项目的相关信息。

点击“下一步”。

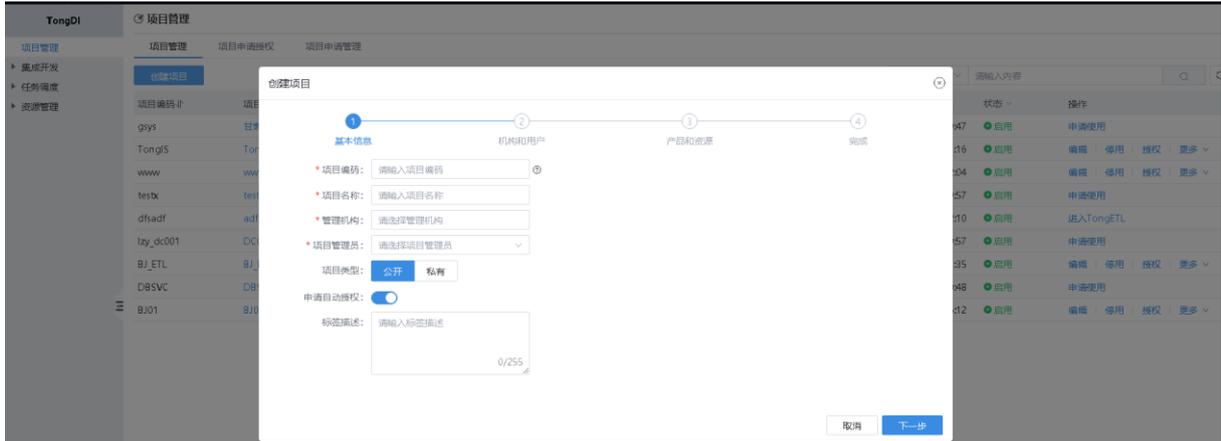


图 3-1

项目编码：全局唯一，不能重复。

项目名称：全局唯一，不能重复。

管理机构：选择正确的管理机构，确保项目的权限和安全。

项目管理员：项目管理员可对项目进行授权和资源分配，要保证正确选择项目管理员，若选择错误人员为项目管理员，可能造成项目无法正常管理使用。

项目类型：公开项目类型是所有人都可以申请使用；私有类项目类型必须由项目管理员指派权限。

申请自动授权：开启后，用户/机构申请项目时，系统将自动评审，自动授权。

注意：如需手动审批项目申请，请关闭“申请自动授权”开关。采用手动批量审批或者手动逐一审批。

点击“增加机构”，可以选择更多有权访问项目的机构。



图 3-2

添加参与人员。

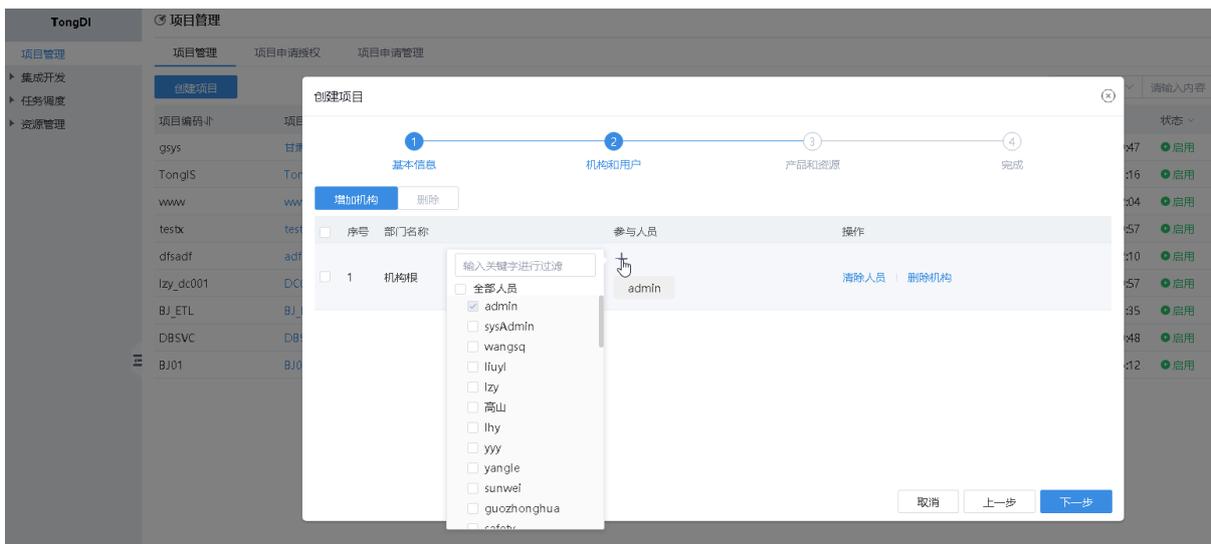


图 3-3

支持删除已添加机构，清除已添加人员。点击“下一步”。

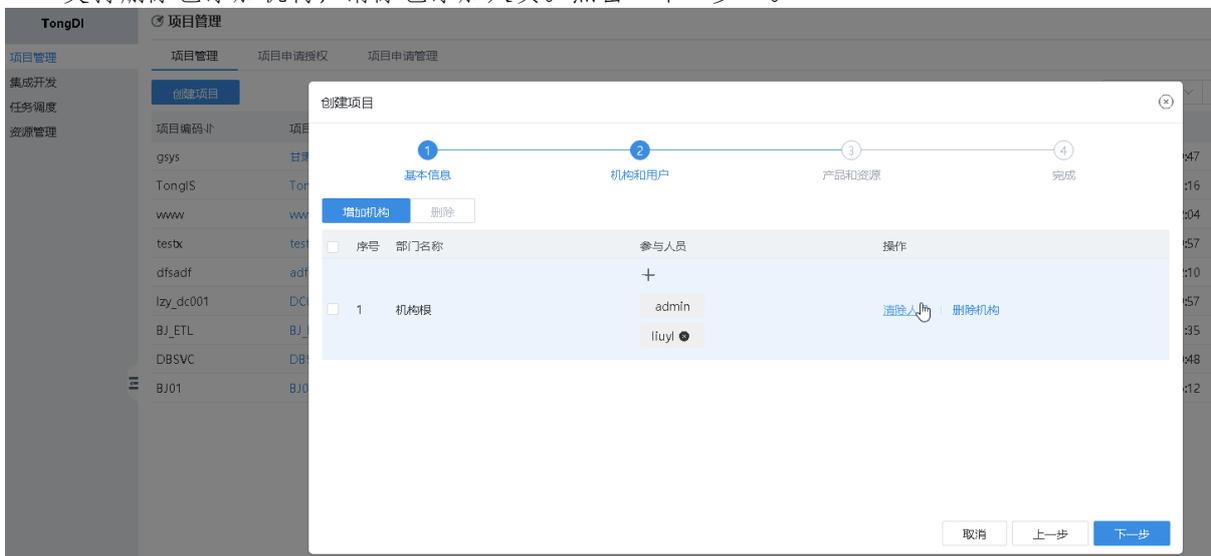


图 3-4

配置产品和资源。选择本项目申请使用的产品、引擎、节点&集群、数据源&表。



图 3-5

点击“提交”，完成项目创建。

3.1.2 编辑项目

需要修改已创建的项目，点击项目列表中的“编辑”。修改项目信息、资源、机构和人员。



图 3-6

3.1.3 停用/启用项目

项目停用后，各产品已有的流程和任务也将被停用，相关信息依然可以查看，但是不能新建流程和任务。

系统管理员/项目管理员点击项目的“停用”操作，设置停用时间和原因说明。点击“确定”，项目将在指定时间停用。

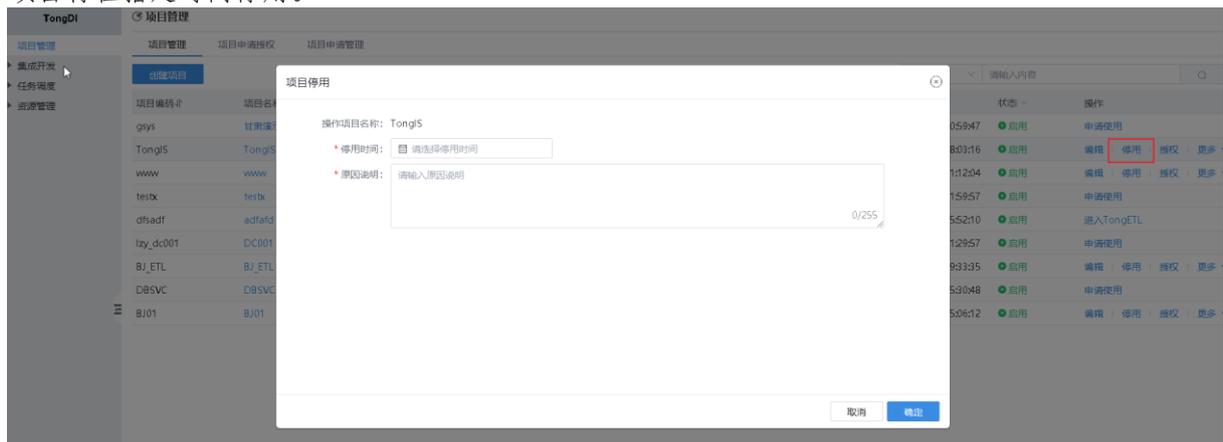


图 3-7

同理，已经停用的项目，再次点击“启用”，启用项目。

3.1.4 回收资源

通过资源回收动作将已部署的任务、流程进行卸载和删除，彻底回收本项目的资源。同时会通过系统消息通知相关的人员。

回收资源前必须先停用项目。然后点击项目操作的“回收资源”按钮。点击“确定”开始回收资源。

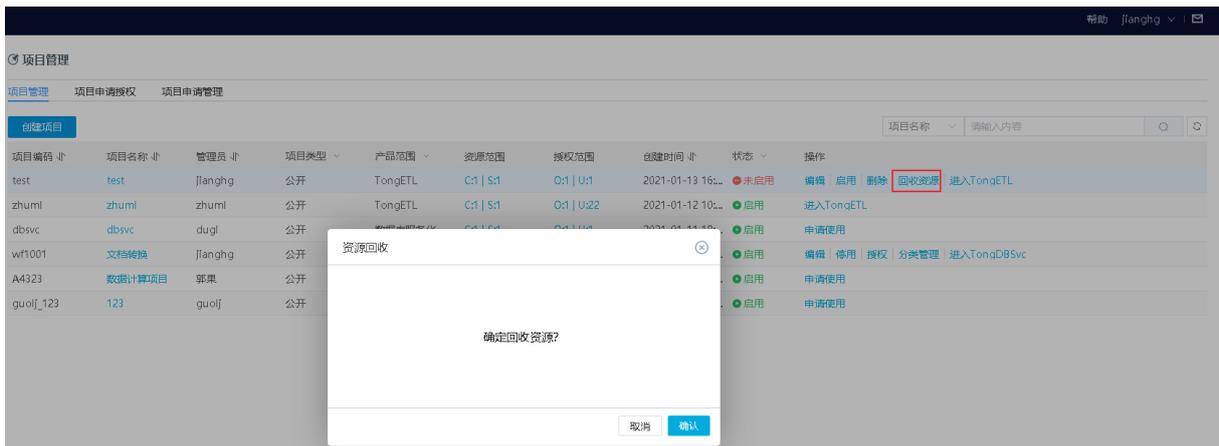


图 3-8

3.1.5 进入产品页

点击项目对应进入产品界面。可以直达项目对应的产品和任务。方便用户联动项目和产品。



图 3-9

3.1.6 项目授权

注意：项目权限与数据角色权限的关系：暂时约定数据角色不对项目内的数据进行权限授权，即在进行数据角色配置时，不允许对 ETL 流程、数据建模、数据计算任务、数据质量任务的相关数据进行数据授权。

注意：项目权限与部门权限的关系：项目权限凌驾于部门权限之上，即当用户对一个项目进行数据操作时，其所能够操作的数据受项目权限控制。项目之外的其他产品数据仍然受部门权限控制。

使用项目管理员用户登录，进入【项目管理】，点击授权按钮，如下图：



图 3-10

项目权限分项目角色管理、项目数据权限、业务数据权限，按照页面的先后顺序进行配置。

3.1.7 申请使用

上面的创建、编辑、启停、回收资源、授权等操作，由项目项目管理员操作，其余的普通用户和成员需要进行申请使用。

可以为本部门或者个人申请，申请需要指定产品和数据权限。

点击要申请的项目操作栏的“申请使用”，弹窗选择适用方，填写申请理由，点击“确定”。



图 3-11

项目申请通过后，“申请使用”变为“进入产品”页面。

3.1.8 删除项目

对于停用的项目，在操作栏会出现“删除”按钮，点击“删除”，删除项目。删除后，项目将从列表中自动剔除，不再展示。已经在使用的产品不能删除。



图 3-12

3.2 项目申请授权

项目申请授权只有项目的管理员能看到，项目申请者本身并不能看到申请的权限和信息。在菜单栏中点击【项目管理】，切换到项目申请授权页签。

3.2.1 申请授权

项目申请授权需要系统管理员/项目管理员进行操作，进入项目申请授权页面，可以看到需要审批的申请。申请授权支持同意、退回和审批操作。



图 3-13

● 同意申请授权

项目管理员选择在审批中的项目，然后点击“同意”，填写审批意见，通过申请授权。

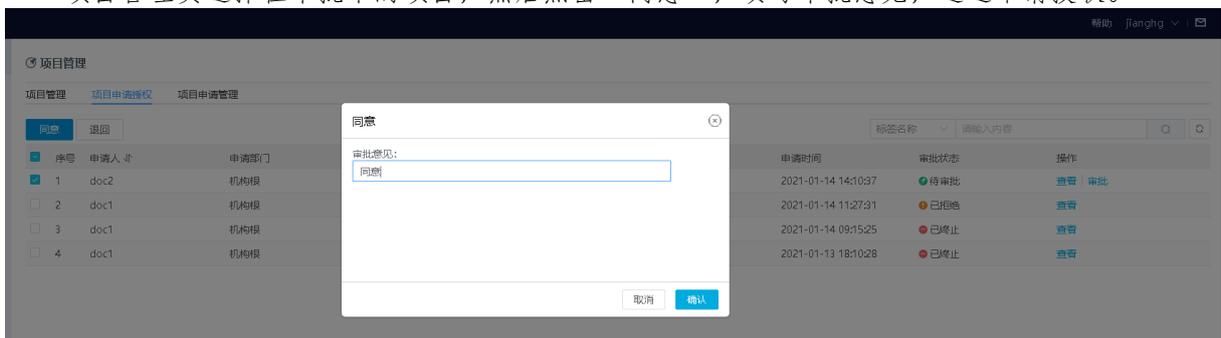


图 3-14

同意后，审批状态是“已授权”，并可以进行查看、冻结、终止操作。



图 3-15

● 退回申请授权

项目管理员选择项目，然后点击“退回”按钮。填写审批意见，点击“确认”。

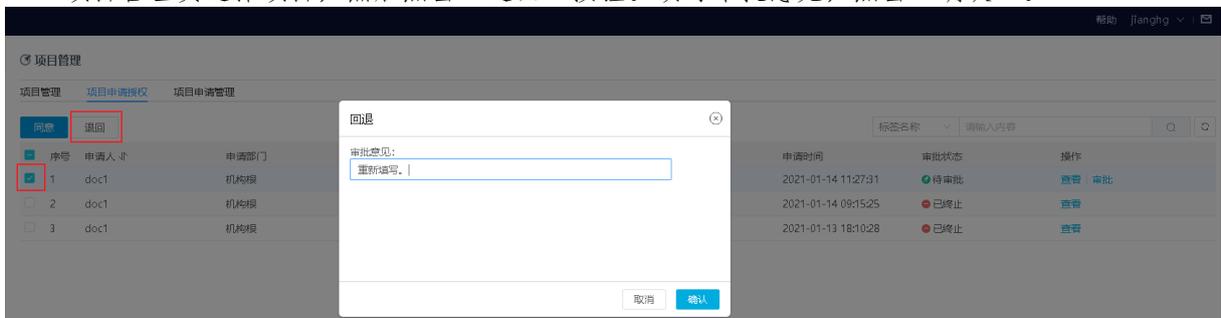


图 3-16

退回的项目申请状态显示“已拒绝”。退回的项目不能再进行申请。



图 3-17

● 审批

审批：使用申请审批包括自动审批和手工审批两类，对于公开类项目，并且在项目创建之初采用自动授权，对于私有项目和未允许自动授权的，采用手动审批方式。

下面介绍自动审批方式：

如下图为公开类项目，采用的是系统自动审批授权，审核状态为“自动授权”。用户发起“申请使用”操作后，系统直接审批通过。



图 3-18

下面介绍手动审批方式：

操作栏点击“审批”，弹窗中填写审批意见，点击“同意”，通过审批。不同意就点击“驳回”。

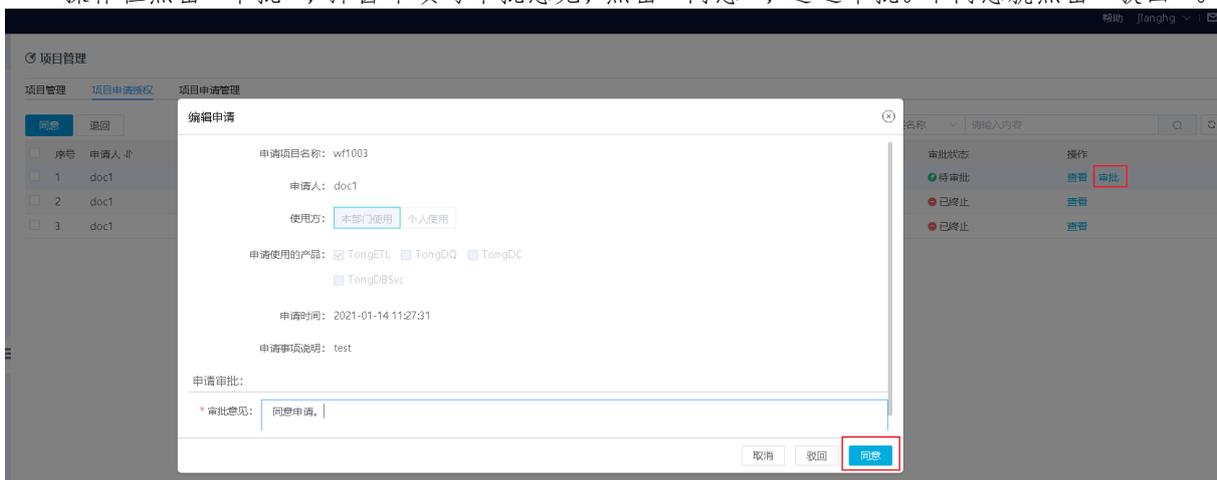


图 3-19

3.2.2 申请撤回

系统管理员/项目管理员对于审批中的使用申请，可以进行撤回操作，已经通过审批或者自动审批的授权不能撤回。

3.2.3 授权查看

项目管理员点击菜单“项目申请授权”，查看项目申请授权记录，包括申请人、项目类型、项

目使用范围、申请时间、审核状态等。点击操作中的“查看”，查看授权详情。

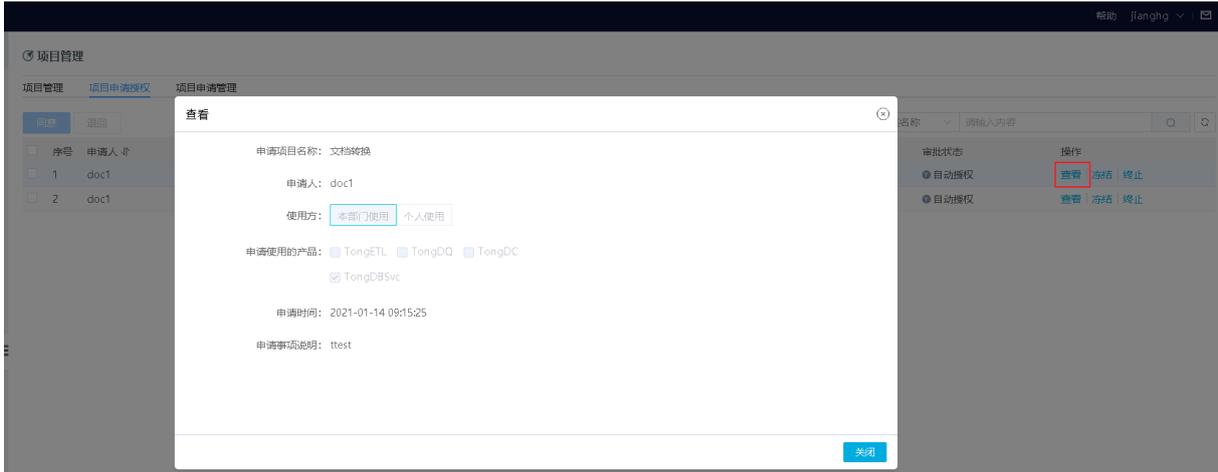


图 3-20

3.2.4 授权冻结/恢复

授权冻结后，用户不能继续使用项目资源，但仍然可以查看自己在项目中操作的信息，其他用户的信息无法查看和操作。

系统管理员/项目管理员在项目申请授权页面，点击操作栏中的“冻结”，并确认，冻结项目。

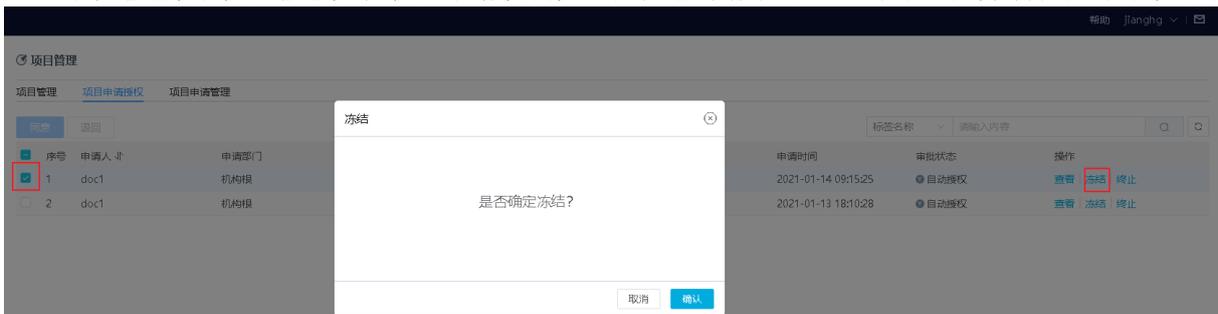


图 3-21

对于已经处于冻结状态的项目，项目管理员点击操作栏中的“恢复”，可以解除冻结。

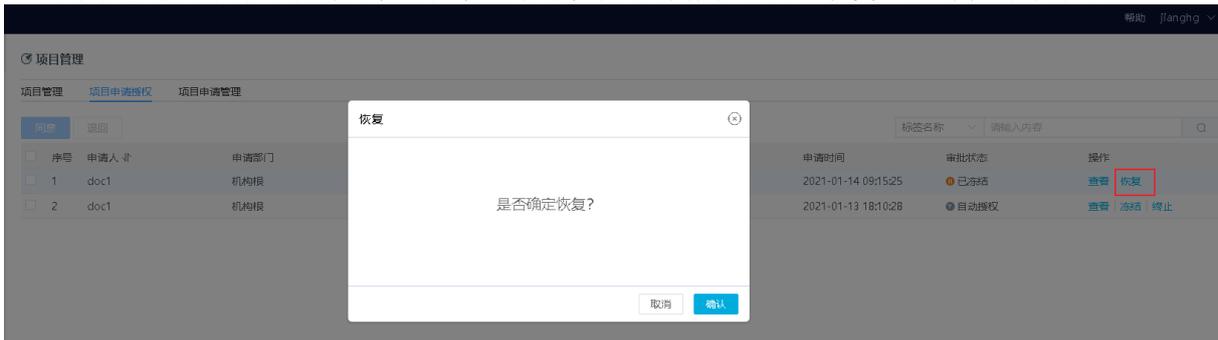


图 3-22

3.2.5 授权终止

终止某个机构/个人的授权后，该机构/个人将失去对项目的使用权限。终止项目权限后，不能进行恢复，如要再次使用项目，需重新申请。

系统管理员/项目管理员在项目申请授权页面，点击操作栏中的“终止”，填写终止原因并确认，项目权限被终止。

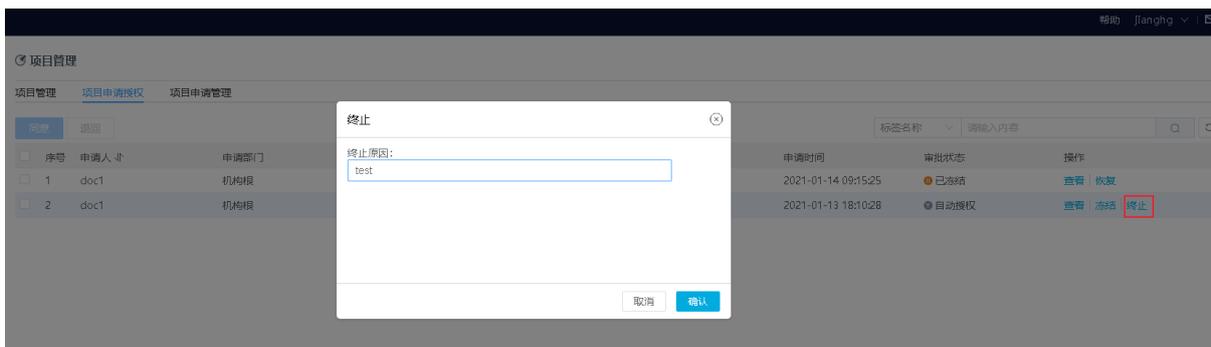


图 3-23

终止后，项目审核状态变为“已终止”。操作只剩下“查看”。



图 3-24

3.3 项目申请管理

在菜单栏中点击【项目管理】，切换到项目申请管理。

3.3.1 申请查看

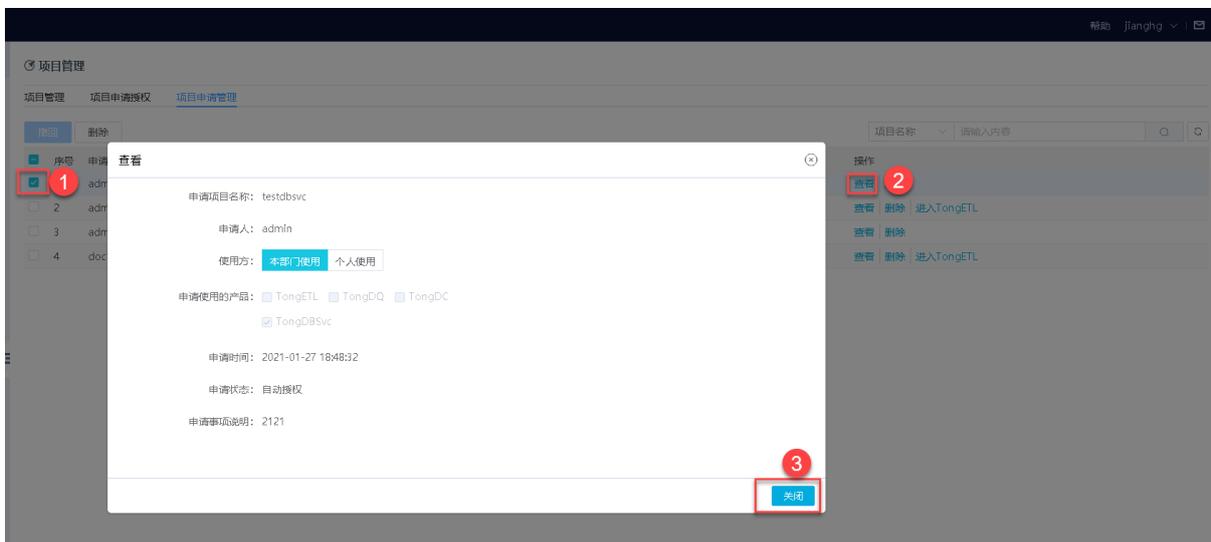


图 3-25

3.3.2 申请删除

选择要删除的申请，点击“删除”按钮，点击确定，如下图所示。

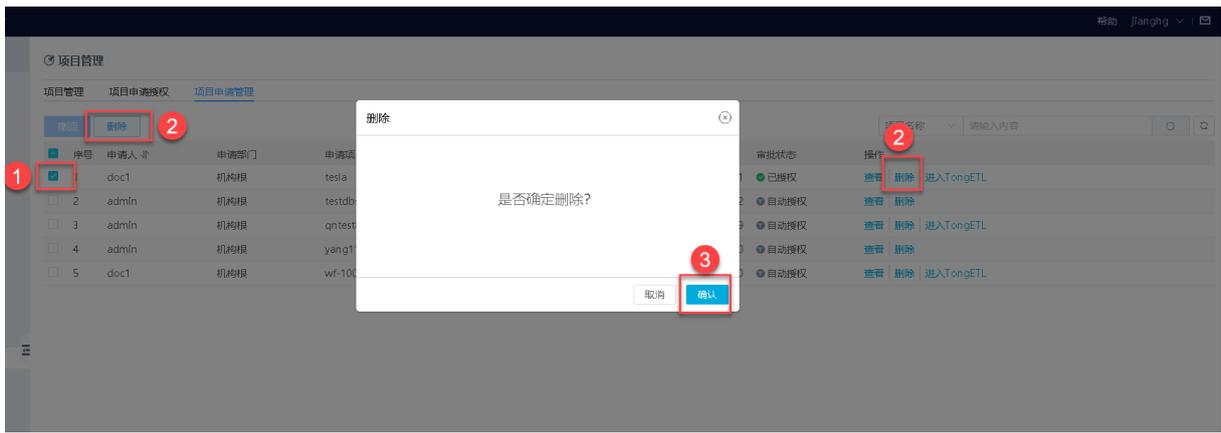


图 3-26

4. 分类管理

已经启用的项目，还可以进行分类管理。点击“分类管理”，在弹窗中设置项目分类。填写分类名称，项目分类维护用于各个项目的统一管理，如下图所示。

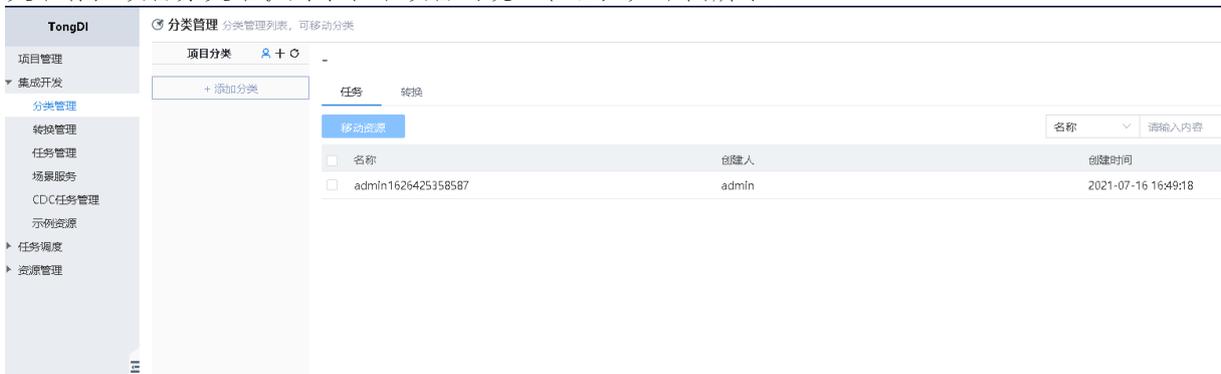


图 4-1

可以添加项目分类同级、子级或者删除。



图 4-2

5. 转换管理

完成数据集成过程中的数据抽取、转换、加载工作，转换流程由多个转换组件编排而成。

5.1 转换管理页

操作步骤：

登录 TongDI 系统，进入“TongETL”->“任务管理”页面，如下图。

图中 1 展示：项目列表，同机构和子机构的用户创建的都展示。

图中 2 展示：转换监控页

图中 3 展示：搜索栏，根据转换名称查找筛选。

图中 4 展示：转换流程列表，同机构和子机构的用户创建的都展示。

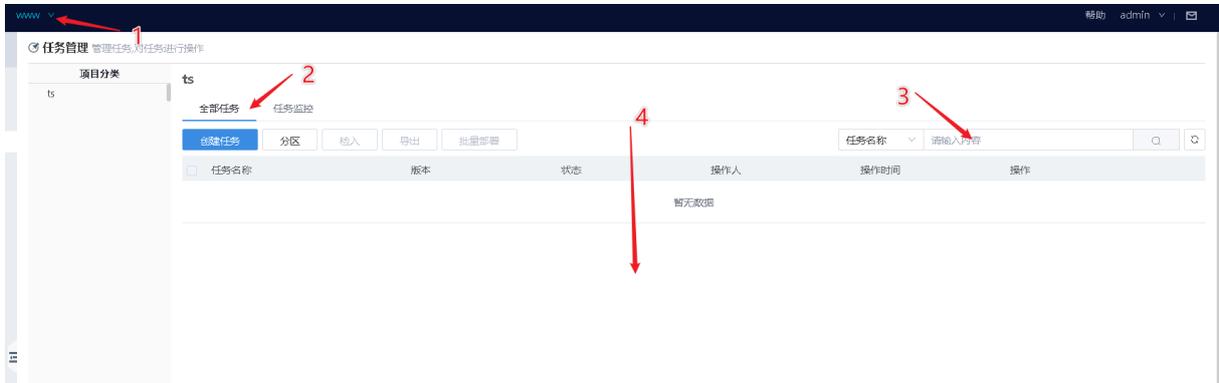


图 5-1

5.2 转换管理-创建转换

操作步骤：

点击【创建转换】，进入转换流程编辑页。如下图。

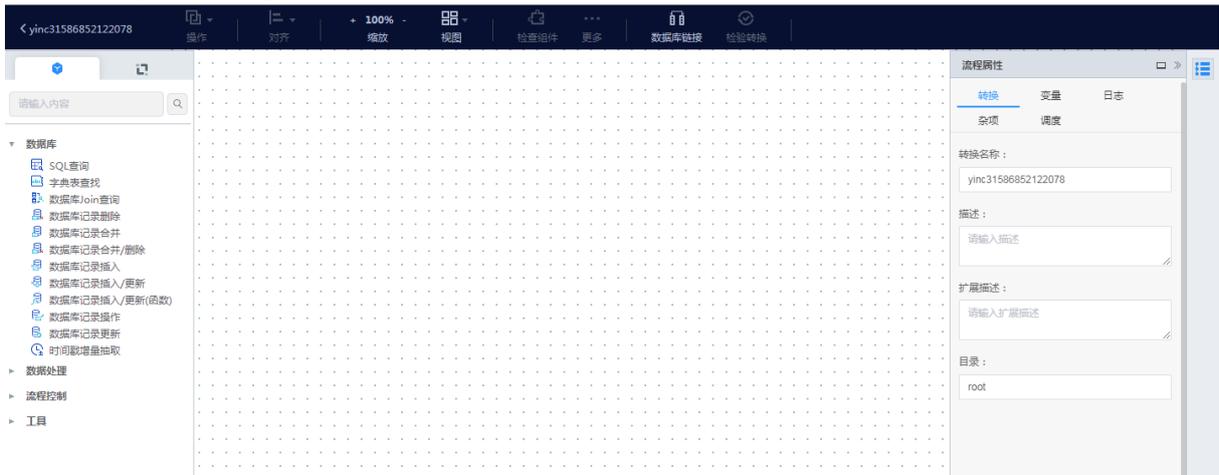


图 5-2

拖拽组件，点击选中组件配置组件属性。

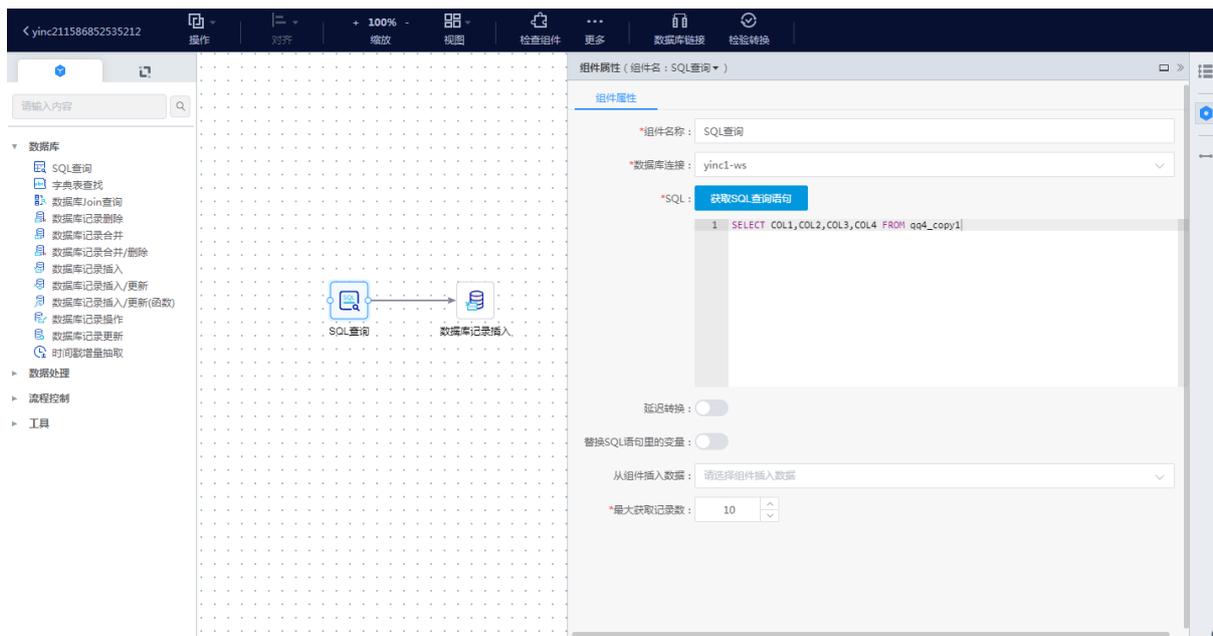


图 5-3

点击画布任意处或点击图中 1 处配置流程属性。

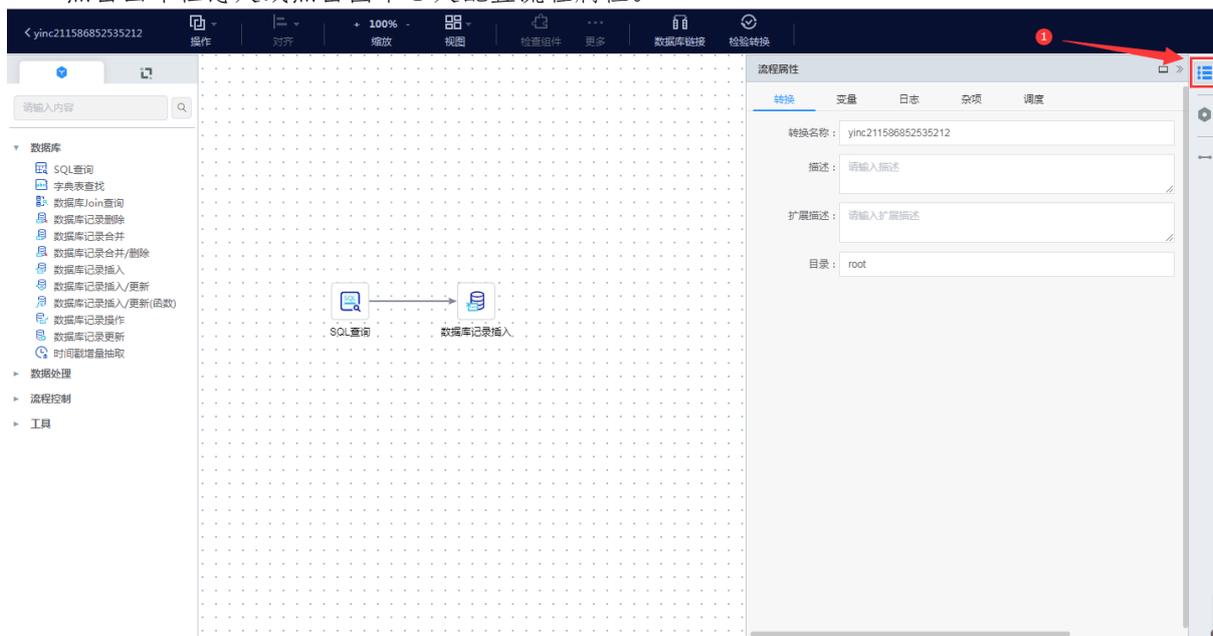


图 5-4

点击左侧顶部位置退出并保存此条流程。

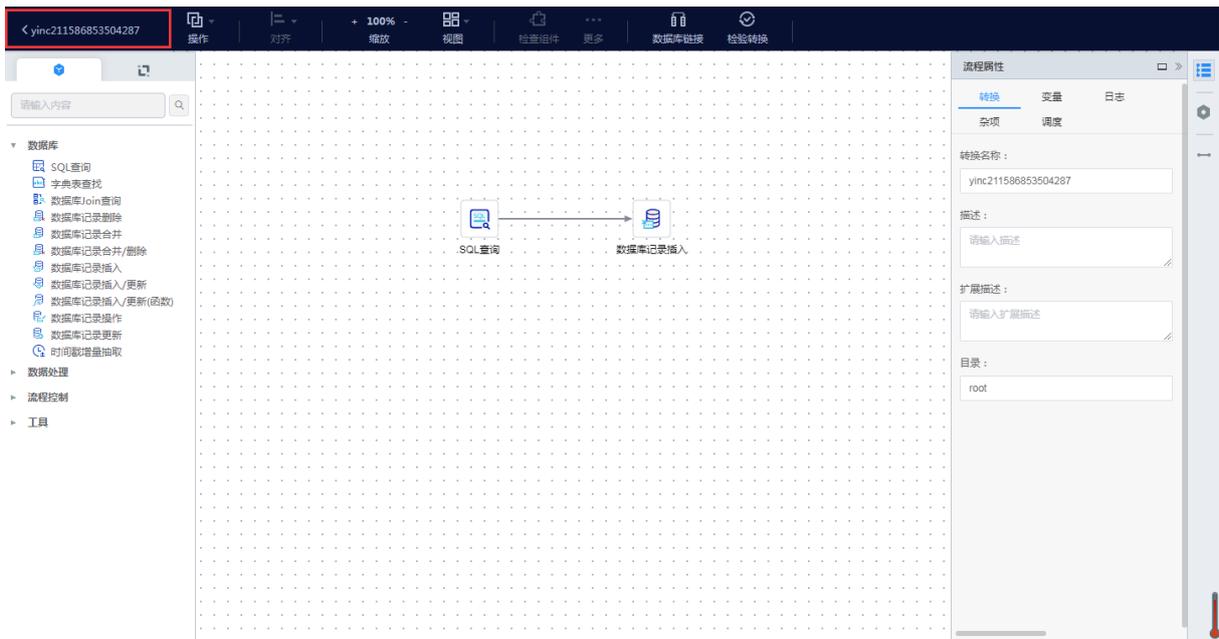


图 5-5

列表页展示刚才配置的流程。



图 5-6

5.2.1 转换-转换设置

进入编辑页面点击画布空白处或点击【流程属性】展示流程属性窗口，转换流程属性包含：转换、变量、日志、杂项、调度，如下图。

转换：配置转换名称、描述、扩展描述、目录。



图 5-7

变量：可新增、删除变量。

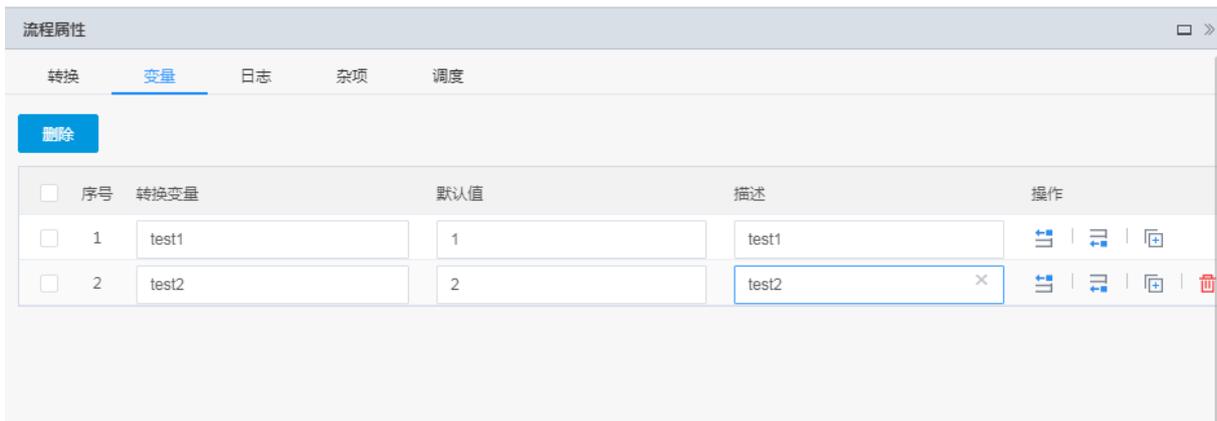


图 5-8

日志-转换：日志内容：从流程整体运行的角度记录流程的运行状态、处理的记录数、运行日志等统计信息。

频次：每次流程开始的时候记录一条流程开始运行的日志，若设置日志记录的时间间隔，会按照时间间隔，每隔一段时间记录一条日志，更新转换日志表中的日志信息，流程结束也会更新转换日志表。在流程的一个运行周期，日志表中只会存在一条日志信息。

功能项	功能说明
日志存储方式	日志的存储方式，目前支持 database、redis、dxn
日志数据库连接	存储为 database 时的连接名称，可使用流程变量，所有流程的日志使用同一个日志数据库
日志表模式	存储为 database 时的模式名称，可使用流程变量
日志表	存储为 database 时的表名称，可使用流程变量
日志写入时间间隔	每间隔多少秒，写入一条日志更新日志表信息

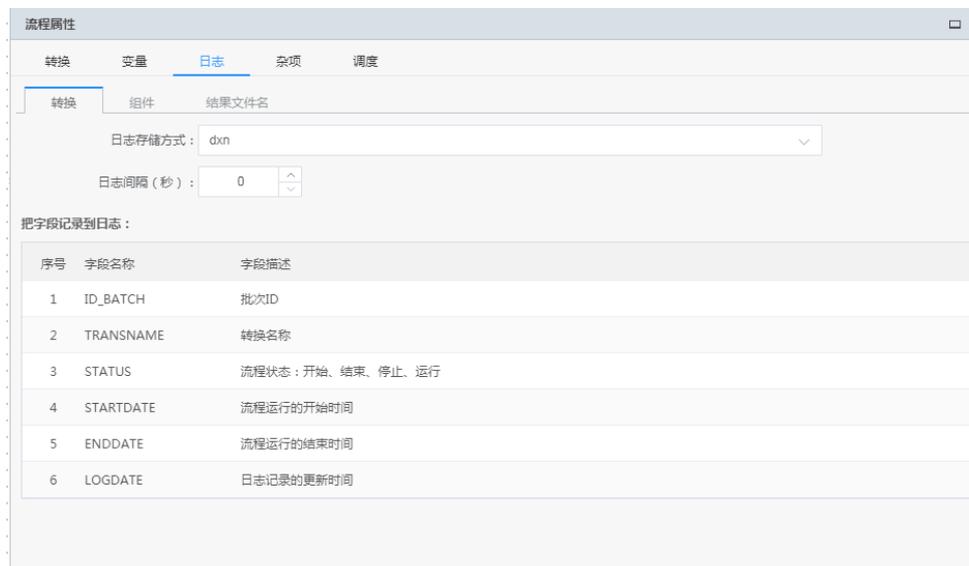


图 5-9

日志-组件：日志内容：从流程运行过程中每个组件的角度记录组件的运行状态、处理的记录数、运行日志、运行时长等统计信息。可以统计流程的历史运行中的处理的记录数等信息。

频次：流程运行结束后，每个组件的每个副本都会记录一条日志到组件日志表中。可配置组件日志的存储方式，记录到日志的字段。

功能项	功能说明
日志存储方式	日志的存储方式，目前支持 database、redis、dxn
日志数据库连接	存储为 database 时的连接名称，可使用流程变量，所有流程的日志使用同一个日志数据库

日志表模式 存储为 database 时的模式名称，可使用流程变量
 日志表 存储为 database 时的表名称，可使用流程变量

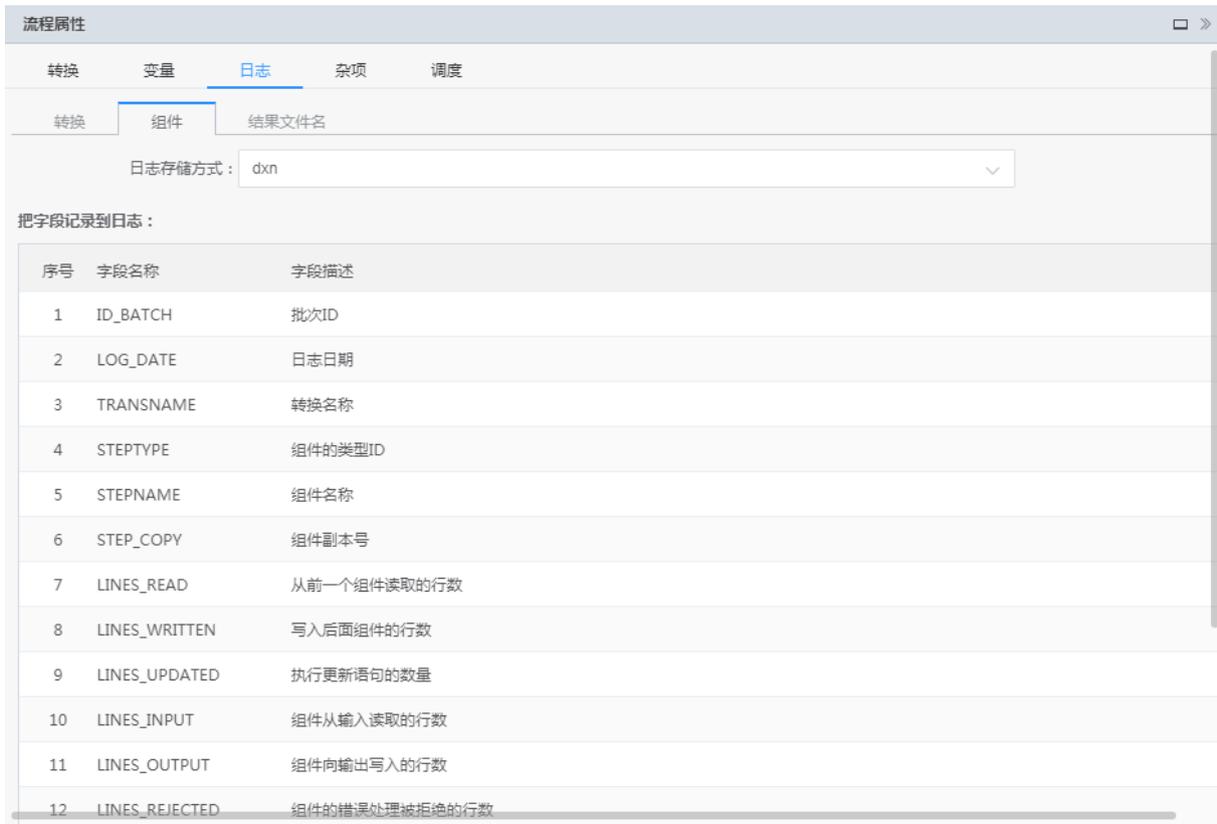


图 5-10

日志-结果文件名：日志内容：转换流程中文件类组件处理过的文件的日志信息。

频次：在流程运行结束后，写入结果文件名日志表。

功能项	功能说明
日志存储方式	日志的存储方式，目前支持 database、redis、dxn
日志数据库连接	存储为 database 时的连接名称，可使用流程变量，所有流程的日志使用同一个日志数据库
日志表模式	存储为 database 时的模式名称，可使用流程变量
日志表	存储为 database 时的表名称，可使用流程变量

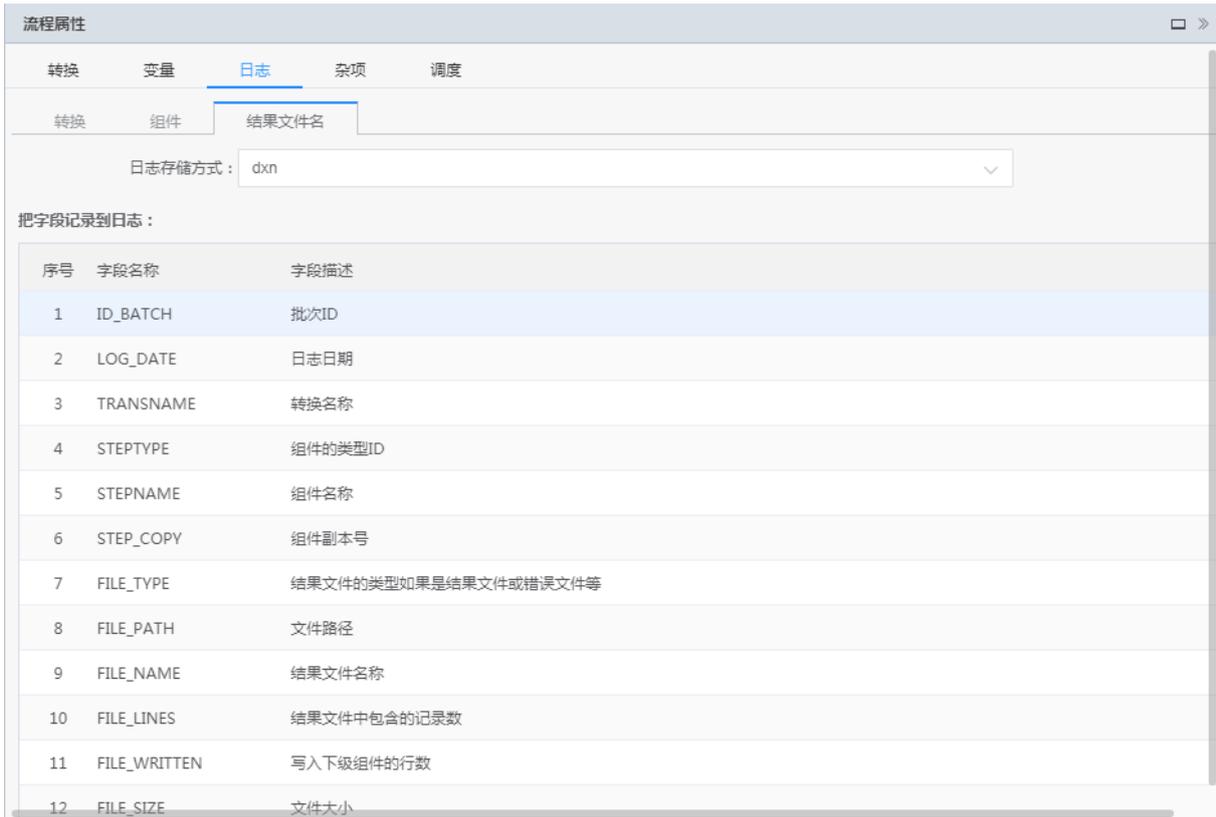


图 5-11

杂项：可配置图中信息。



图 5-12

调度：可设置定时调度方式。

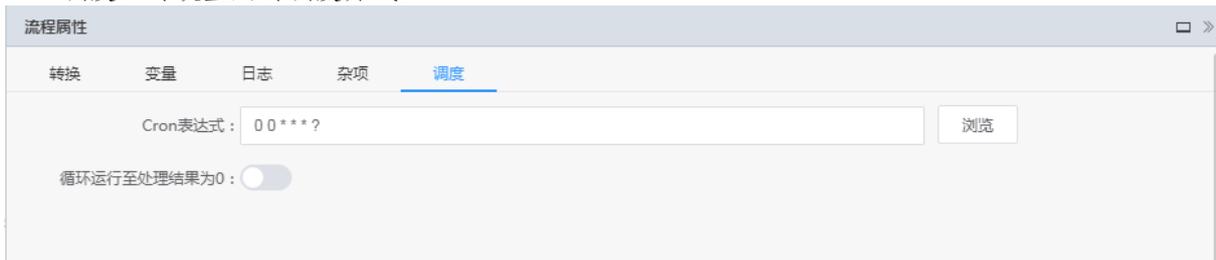


图 5-13

5.2.2 转换-组件视图

转换的最小处理单元，每个转换组件完成一个特殊的数据转换任务（如行转列、数据拆分、数据合并等），一个或多个转换组件编排成一个转换流程。

操作步骤：

进入流程编辑页点击“组件视图”，展示所有组件列表，搜索栏支持模糊搜索组件。



图 5-14

5.2.3 转换-资源视图

操作步骤：

1. 进入流程编辑页点击“资源视图”，展示资源列表。
2. 转换组件中展示流程中的所有组件。
3. 组件连接展示所有组件、之间的连接。
4. 数据库连接展示所有可用的数据源。
5. 分区模式展示此流程拥有分区。



图 5-15

5.2.4 转换-转换组件选项

转换组件选项，是所有转换组件在转换流程编辑内的右键功能菜单（右键上下文菜单），功能列表如下：



图 5-16

功能项	功能说明
连接组件	连接选中的转换组件，需要至少选中 2 个组件。
数据发送	当选中组件后面有多个分支或者有副本时，该功能有用。可选项包括“轮流发送模式”和“复制到下个组件”，当在一个组件上增加多个分支时，会弹出提示选择使用哪种方式。 轮流发送模式： 组件输出数据轮流向后续分支发送数据或者前一个组件以轮流方式向多个副本发送数据； 复制到下个组件： 组件输出数据完整复制到后续每个分支进行处理或者前一个组件将输出数据完整复制到多个副本进行处理。
改变起始的副本数量	设置同时启动多个组件副本同时运行，前一个组件以轮流或复制的方式向同时运行的多个副本发送数据。当副本数量大

复制到剪切板	于 1 时，在组件图标左上角会出现特殊标识，如  。
复制组件	将组件的 XML 描述复制到剪贴板。
分离组件	将在原先的组件低一点的位置创建一个该组件的副本。
显示输入字段	断开此组件与其他组件的连接
显示输出字段	展示所有的输入字段。
对齐/分散	展示所有的输出字段。
检查所选组件	自动排列组件，选中多个组件时可用。
应用分区	检查此组件目前的状态。
取消分区	将配置好的分区应用到此流程。
应用集群	将此流程应用的分区取消。
取消集群	将配置好的集群应用到此流程。
删除	将此流程应用的集群取消
	将从转换中永久的删除组件。

5.2.5 转换-转换连接选项

转换连接，是指在转换流程编辑器中，为转换流程组件之间建立连接线的操作。

转换连接选项，是连接线在转换流程编辑内的右键功能菜单（右键上下文菜单），功能列表如下：

	调换方向	
	使连接失效	
	删除连接	Del
功能项	功能说明	
调换方向	改变选中连接的方向。	
使连接失效	设置选中连接生效或失效。也可以直接单击连接线，在生效和失效状态中切换。	
删除连接	删除选中的连接。	

5.2.6 转换-转换功能栏

转换编辑页的顶部功能栏，列表如下：



返回/保存	保存当前流程并返回任务流程列表
操作	选择组件进行操作
对齐	选择多个组件进行操作
缩放	相当与菜单栏里的视图，放大、缩小工作区
视图	开关右侧功能栏
检查组件	检查组件必填项和数据正确性检查
更多	包含：输入、输出字段，分区，集群，副本数量同转换组件选项功能一致
数据库连接	查看可用的数据库连接
检验转换	检查整个转换的正确性

5.3 转换管理-流程操作

5.3.1 流程操作-编辑

可修改流程名称、描述、组件属性、连接线属性等。

新建流程后，编辑此流程时流程被自动检出，其他人不可操作此流程。
检出流程后编辑流程，保存后自动检入。



图 5-17

5.3.2 流程操作-复制

将指定的流程复制一份，另存为一个新的流程，版本号同新建一致从 0.1 开始。
新建流程后，复制此流程，另建一个新流程，版本号从 0.1 开始。
部署流程后，复制产品版本流程，在草案版本建一个新流程，版本号从 0.1 开始。



图 5-18

5.3.3 流程操作-部署

将任务以指定的调度方式部署到服务运行。每次部署后，新增一个产品版本。
新建流程后，部署此流程时流程被自动检出，部署完成后自动检入。



图 5-19

点击部署后如下图：

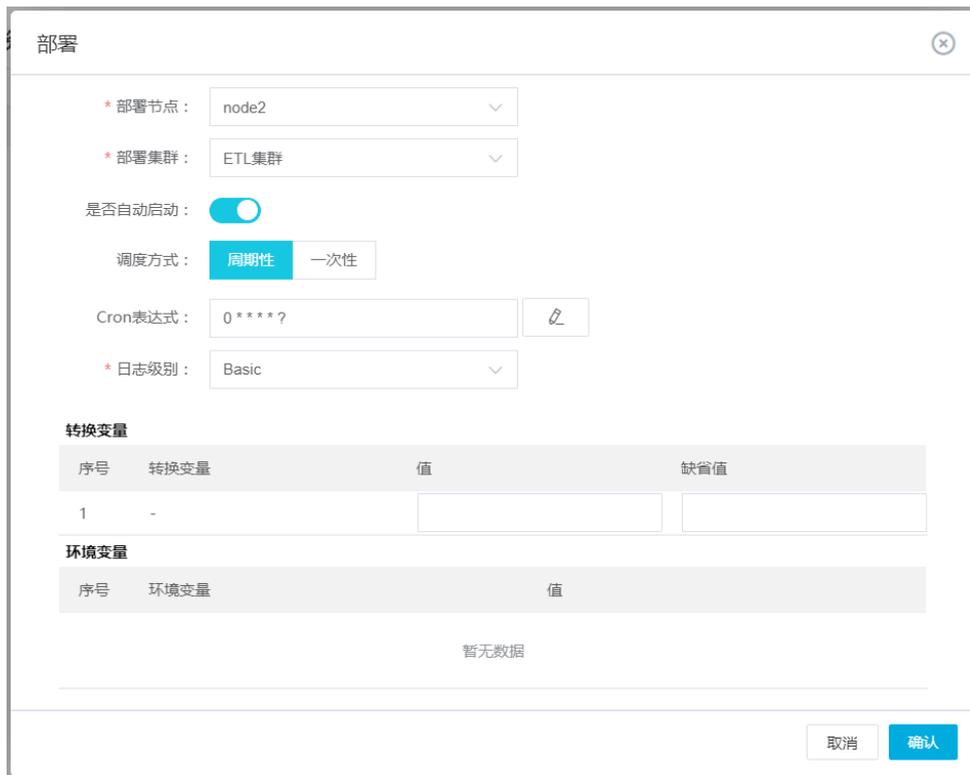


图 5-20

编辑 Cron 表达式如下图:

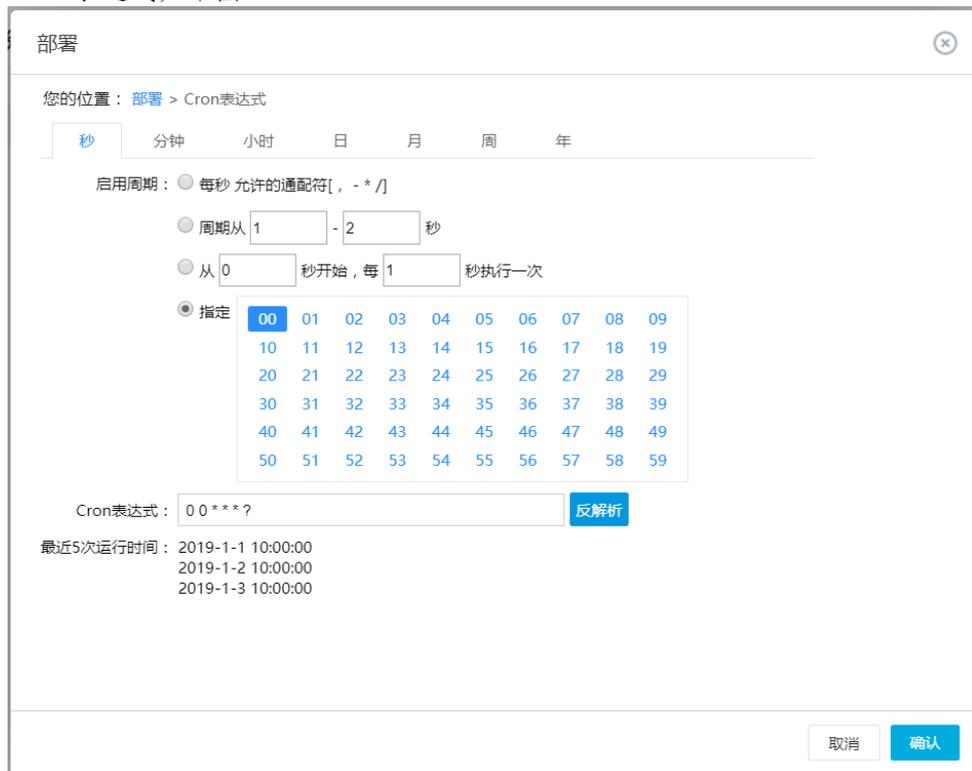


图 5-21

5.3.4 流程操作-回滚

将任务回滚到指定的版本，相当于更新草案版本，回滚后草案版本的版本号的小版本号在原来的基础上加 1。

部署流程后，产品版本操作回滚。



图 5-22

5.3.5 流程操作-删除

删除指定的流程，有操作权限的用户都可删除。

新建流程后，可直接删除草案版本流程。

部署流程后，需卸载流程才可删除产品版本流程。在“转换监控”中先停止流程，然后点击“卸载”流程，最后回到“全部转换”列表中删除流程。



图 5-23

5.3.6 流程操作-检出

当多用户共享的流程，如不想其他人操作此流程，可将流程检出，检出后会锁定此流程。

新建流程后，可将流程检出，检出后仅检出人有操作权限。



图 5-24

5.3.7 流程操作-检入

只有检出后的流程才可做检入操作，不检入无法部署。

检出流程后，可检入流程继续共享此流程。

检出流程后，编辑流程保存后，流程自动检入。



图 5-25

5.3.8 流程操作-产品版本查看日志

可查看流程运行中的日志情况。



图 5-26

5.3.9 流程操作-产品版本详情

查看流程部署时相关的信息。



图 5-27

5.3.10 流程批量部署

勾选多个流程后点击批量部署，多个流程会同时被部署到服务器上。

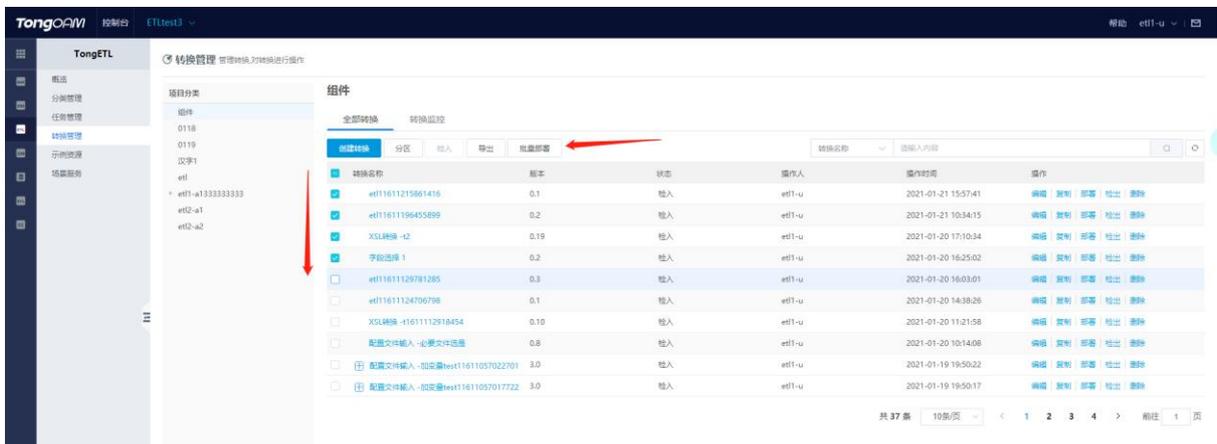


图 5-28

5.3.11 流程导出

选择流程后点击导出，流程会以 xml 格式导出。

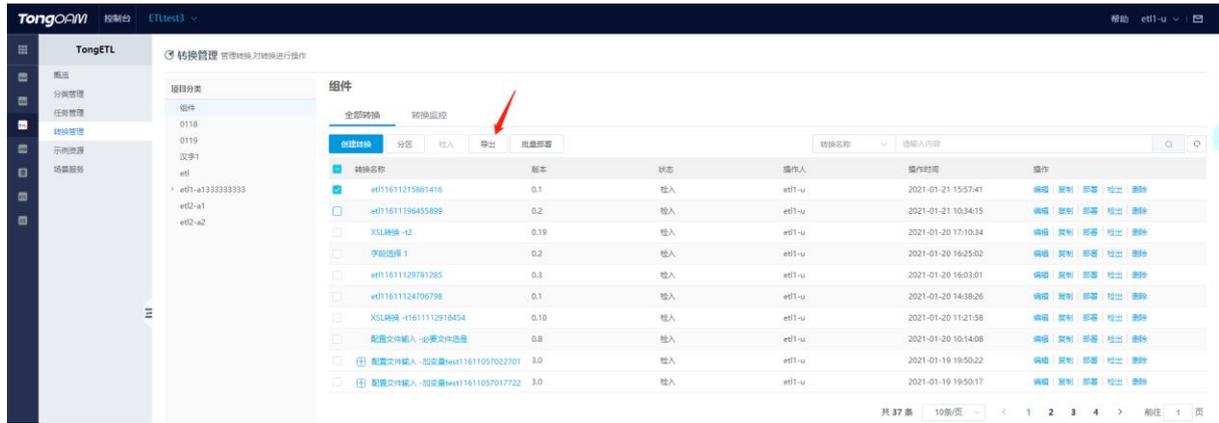


图 5-29

6. 任务管理

提供对转换过程的调度管理，可以在任务流程中实现对多个转换流程和其他任务流程的编排，在出现异常时，提供告警。

任务流程由多个任务组件和转换流程编排而成。

6.1 任务管理页

操作步骤：

登录 TongDI 系统，进入“TongETL”->“任务管理”页面，如下图。

图中 1 展示：项目列表，同机构和子机构的用户创建的都展示，点击项目进行切换。

图中 2 展示：任务监控页。

图中 3 展示：搜索栏，根据任务名称查找筛选。

图中 4 展示：任务流程列表，同机构和子机构的用户创建的都展示。



图 6-1

6.2 任务管理-创建任务

操作步骤：

1. 点击【创建任务】，进入任务流程编辑页。如下图。

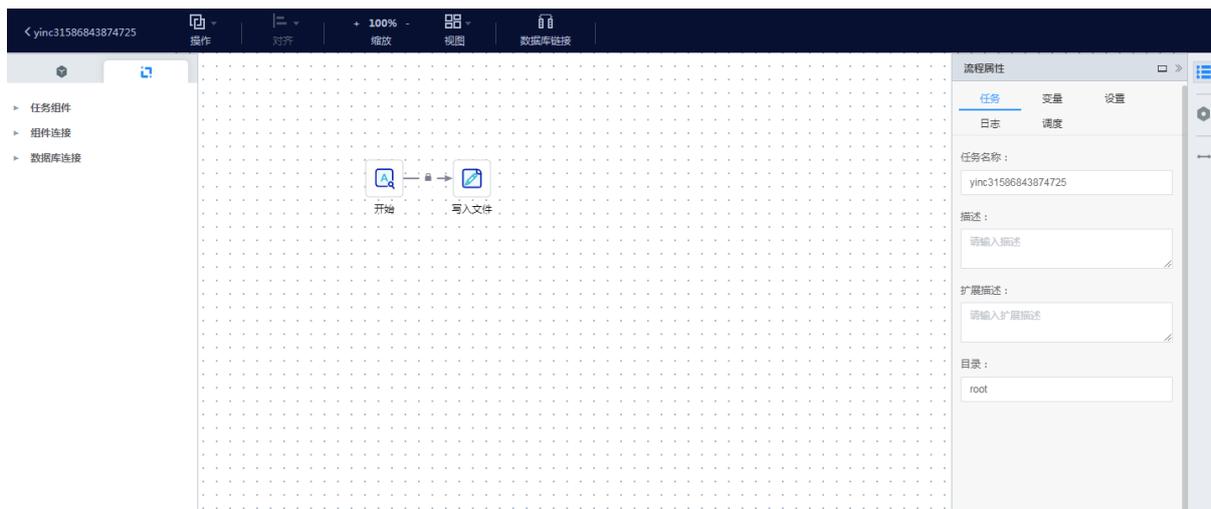


图 6-2

2. 拖拽组件，点击选中组件配置组件属性。

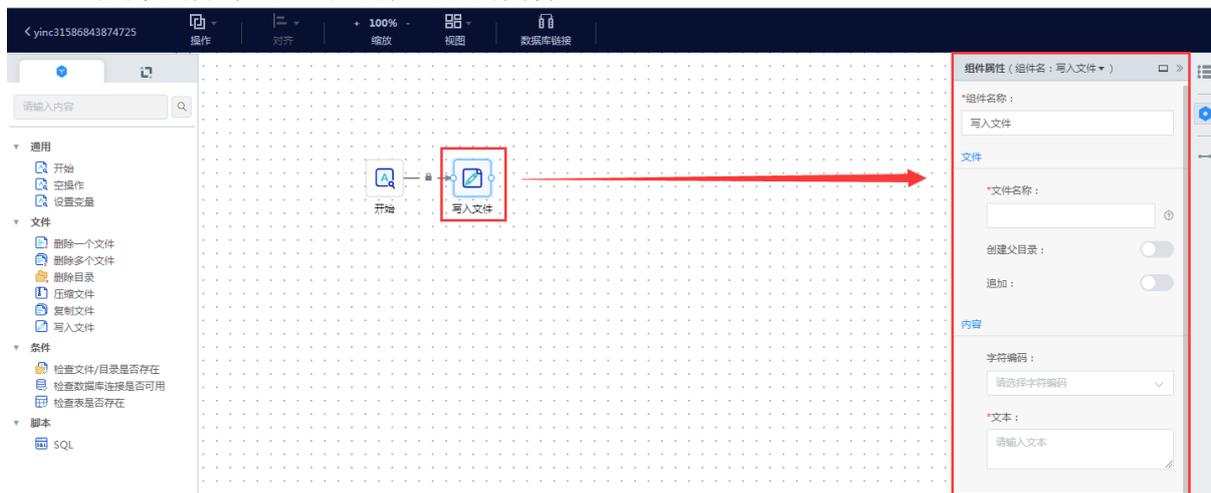


图 6-3

3. 点击画布任意处或点击图中 1 处配置流程属性。



图 6-4

4. 点击左侧顶部位置退出并保存此条流程。

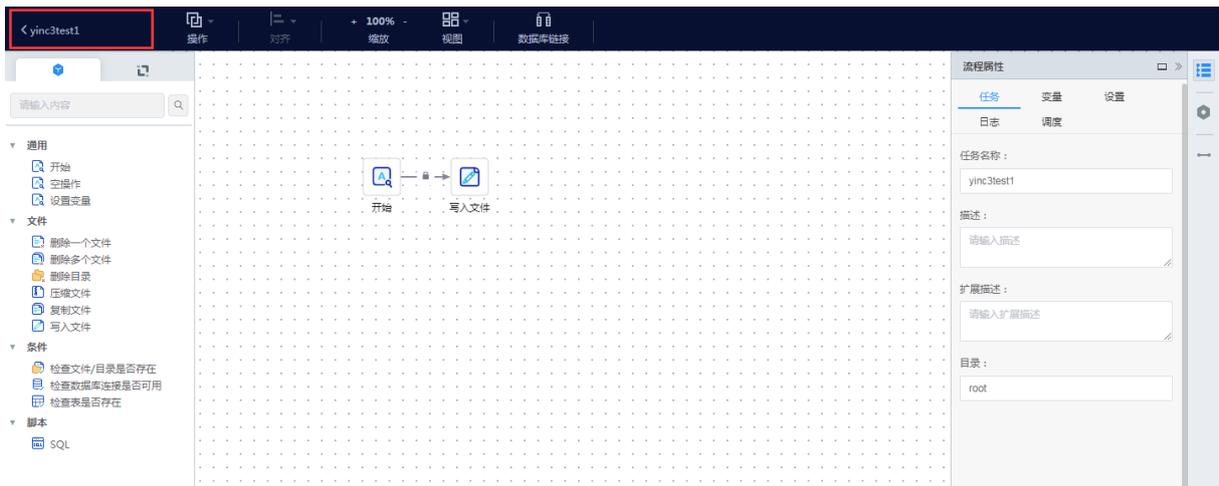


图 6-5

5. 列表页展示刚才配置的流程。



图 6-6

6.2.1 任务-任务设置

操作步骤：

1. 进入编辑页面点击画布空白处或点击【流程属性】展示流程属性窗口，任务流程属性包含：任务、变量、设置、日志、调度，如下图。

任务：可配置任务名称、描述、扩展描述、目录。



图 6-7

2. 变量：可新增、删除变量。



图 6-8

3. 设置：可设置共享文件对象。

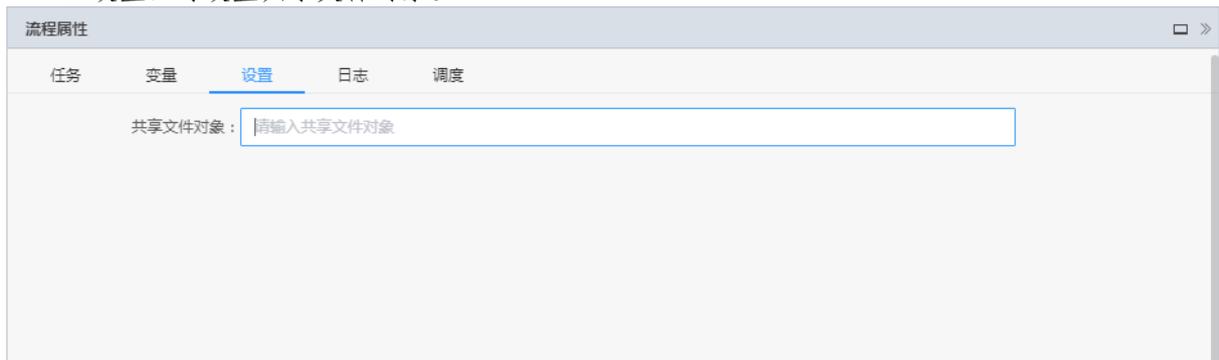


图 6-9

4. 日志-任务：日志内容：从流程整体运行的角度记录流程的运行状态、处理的记录数、运行日志等统计信息。

频次：每次流程开始的时候记录一条流程开始运行的日志，若设置日志记录的时间间隔，会按照时间间隔，每隔一段时间记录一条日志，更新转换日志表中的日志信息，流程结束也会更新转换日志表。在流程的一个运行周期，日志表中只会存在一条日志信息。

功能项	功能说明
日志存储方式	日志的存储方式，目前支持 database、redis、dxn
日志数据库连接	存储为 database 时的连接名称，可使用流程变量，所有流程的日志使用同一个日志数据库
日志表模式	存储为 database 时的模式名称，可使用流程变量
日志表	存储为 database 时的表名称，可使用流程变量
日志写入时间间隔	每间隔多少秒，写入一条日志更新日志表信息

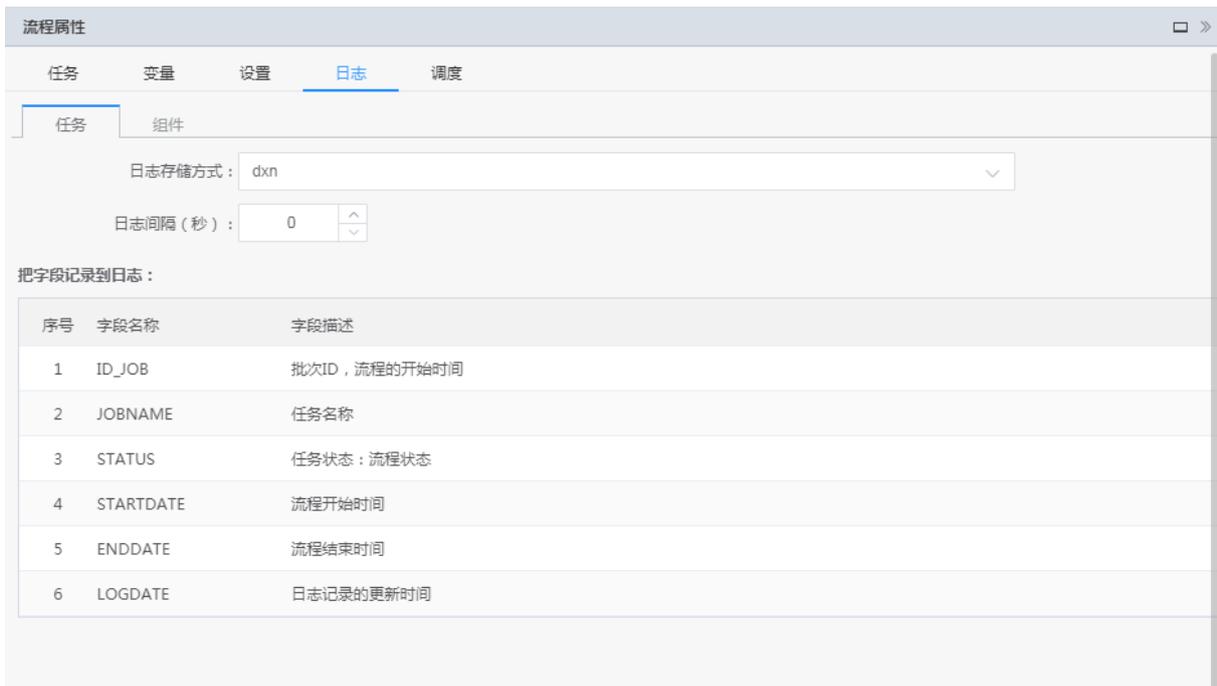


图 6-10

5. 日志-组件：日志内容：从流程运行过程中每个组件的角度记录组件的运行状态、处理的记录数、运行日志、运行时长等统计信息。

频次：流程运行结束后，每个组件的每个副本都会记录一条日志到组件日志表中。

功能项	功能说明
日志存储方式	日志的存储方式，目前支持 database、redis、dxn
日志数据库连接	存储为 database 时的连接名称，可使用流程变量，所有流程的日志使用同一个日志数据库
日志表模式	存储为 database 时的模式名称，可使用流程变量
日志表	存储为 database 时的表名称，可使用流程变量



图 6-11

6. 调度：可设置定时调度方式。

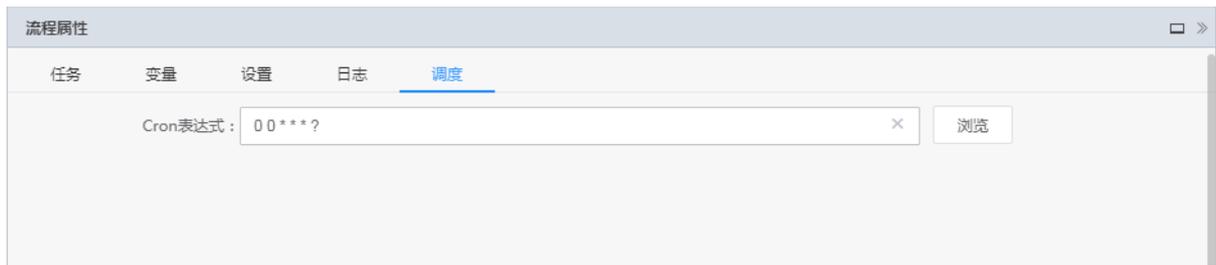


图 6-12

6.2.2 任务-组件视图

任务调度处理的最小单元，一个或多个任务组件可以编排成一个任务流程。

操作步骤：

1. 进入流程编辑页点击“组件视图”，展示所有组件列表，搜索栏支持模糊搜索组件。

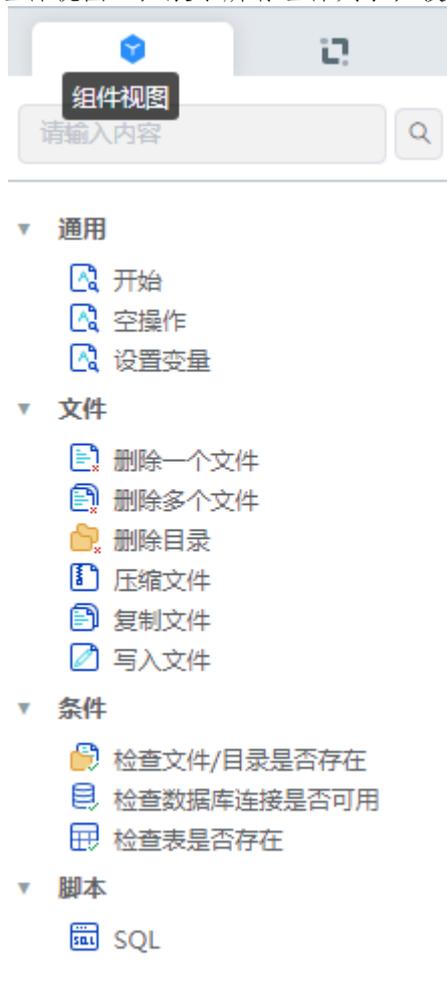


图 6-13

6.2.3 任务-资源视图

操作步骤：

1. 进入流程编辑页点击“资源视图”，展示资源列表：
 - (1) 任务组件中展示流程中的所有组件；
 - (2) 组件连接展示所有组件之间的连接；
 - (3) 数据库连接展示所有可用的数据源。



图 6-14

6.2.4 任务-任务组件选项

任务组件选项，是所有任务组件在转换流程编辑内的右键功能菜单（右键单击图标弹出），功能列表如下：

连接组件	
并行启动下一个任务组件	
复制到剪贴板	Ctrl+C
复制组件	
分离组件	
对齐	▶
删除	Del

图 6-15

功能项	功能说明
连接组件	连接选中的任务组件，需要至少选中 2 个组件
并行启动下一任务组件	如果此任务组件后面同时连接着多个任务组件，则并行启动这些任务组件
复制任务组件	将在原先的任务组件低一点的位置创建一个该任务组件的副本。
复制到剪切板	复制任务组件的 XML 描述到剪贴板。
对齐/分散	自动排列组件，选中多个任务组件时可用
分离任务组件	断开此任务组件与其他任务组件的连接。
删除任务组件	删除所有此任务组件的副本，而不仅仅是这一个！

6.2.5 任务-任务连接选项

任务连接，是指在任务流程编辑器中，为任务流程组件之间建立连接线的操作。

任务连接选项，是连接线在任务流程编辑内的右键功能菜单（右键单击连线弹出），功能列表，见下表：

判断	▶
调换方向	
使连接失效	
删除连接	Del

功能项	功能说明
判断	在创建连接时，指定下一个任务组件将会被执行的条件。

功能项	功能说明
	有以下三个选项，每个选项生成的连接线用不同颜色以示区分： 无条件（黑色） ：当起始组件是不产生结果的组件，比如“开始”组件，能且只能选择该选项，表示始终执行； 当结果为真的时候继续下一步（绿色） ：当起始组件是会产生结果的组件时，比如“检查文件/目录是否存在”组件，指定当结果为真时的任务执行分支。 当结果为假的时候继续下一步（红色） ：当起始组件是会产生结果的组件时，比如“检查文件/目录是否存在”组件，指定当结果为假时的任务执行分支。
调换方向	改变选中连接的方向
使连接失效	设置选中连接生效或失效。也可以直接单击连接线，在生效和失效状态中切换
删除连接	删除选中的连接

6.2.6 任务-任务功能栏

任务编辑页的顶部功能栏，列表如下：



返回/保存	保存当前流程并返回任务流程列表
操作	选择组件进行操作
对齐	选择多个组件进行操作
缩放	相当与菜单栏里的视图，放大、缩小工作区
视图	开关右侧功能栏
数据库连接	查看可用的数据库连接

6.3 任务管理-流程操作

同 [5.3 转换管理-流程操作](#) 一致。

6.4 任务管理-任务监控

6.4.1 创建带子转换的任务流程

1. 进入任务管理，点击【创建任务】，拖拽组件开始、转换，点击图中 1 后，点击图中 2 配置已有的转换流程（在转换管理中创建），保存。

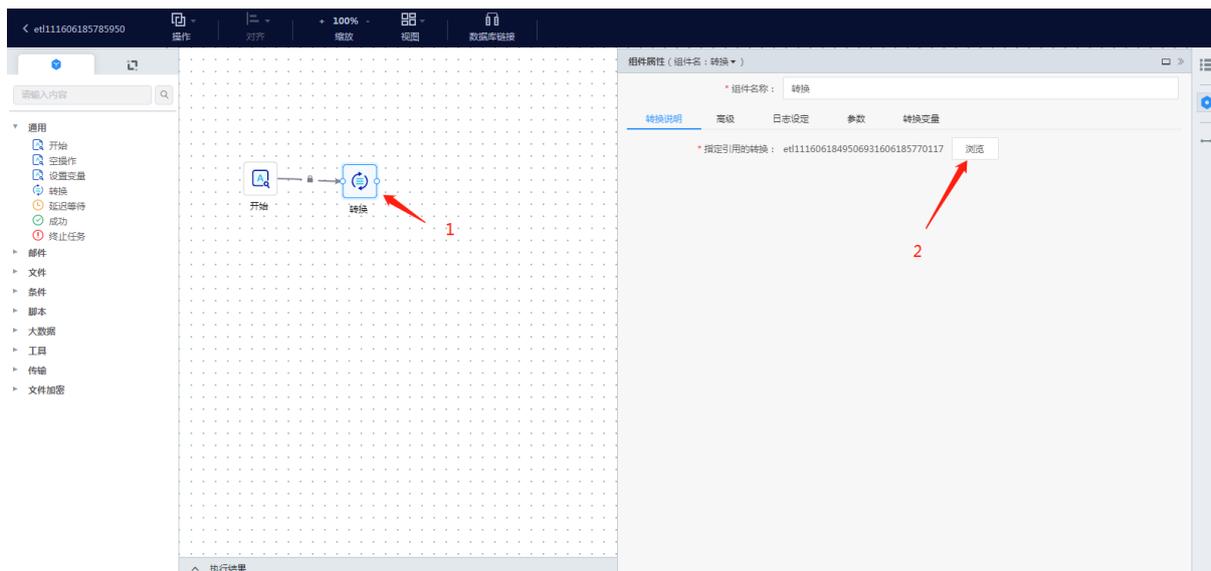


图 6-16

2. 部署流程。



图 6-17

3. 选节点、集群、配置周期性。

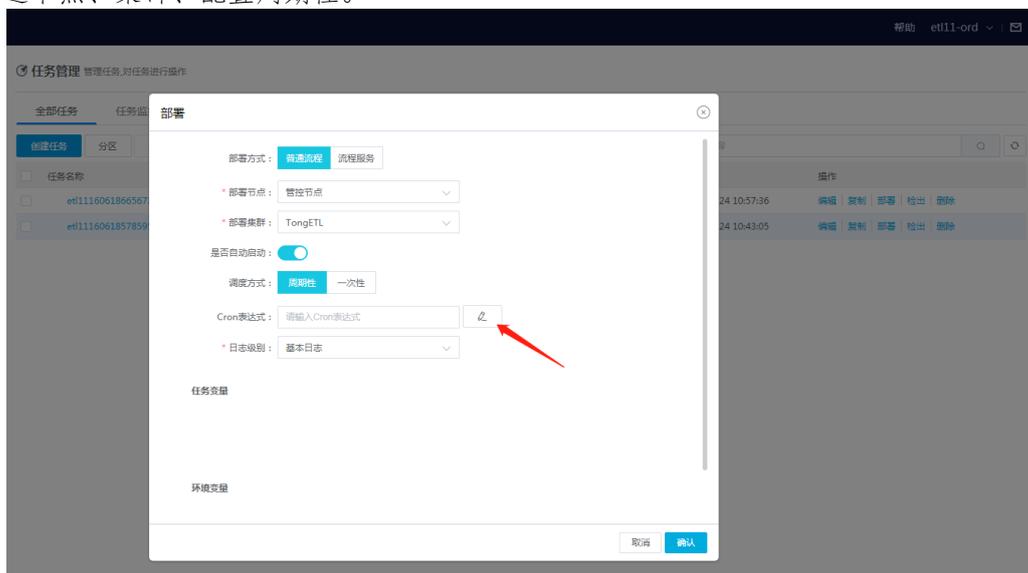


图 6-18

4. 配置 cron 表达式后确定，部署成功。

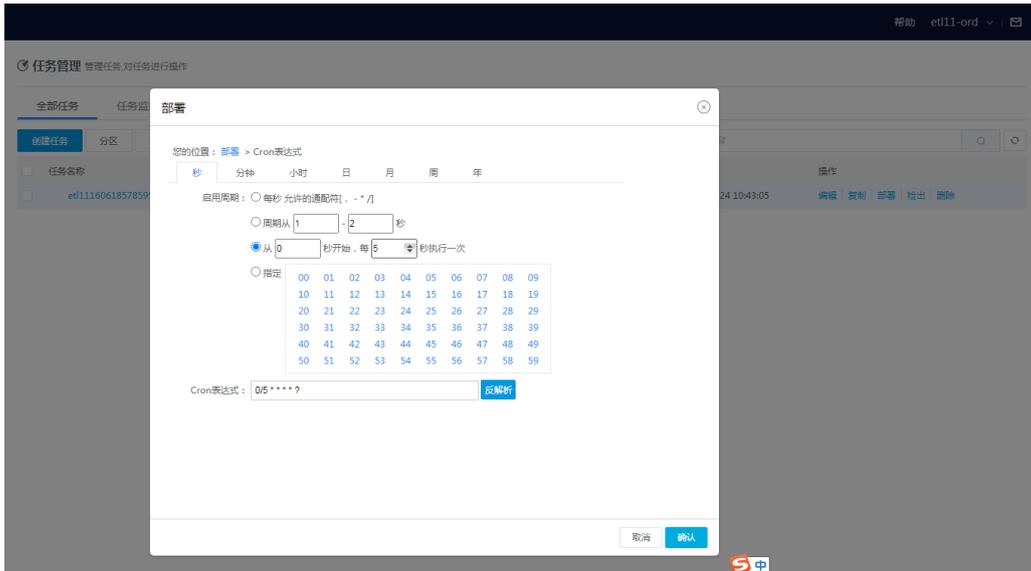


图 6-19

5. 点击任务监控，查看刚才创建的子任务+子转换流程。



图 6-20

6. 点击任务名称进入任务流程页可查看日志。

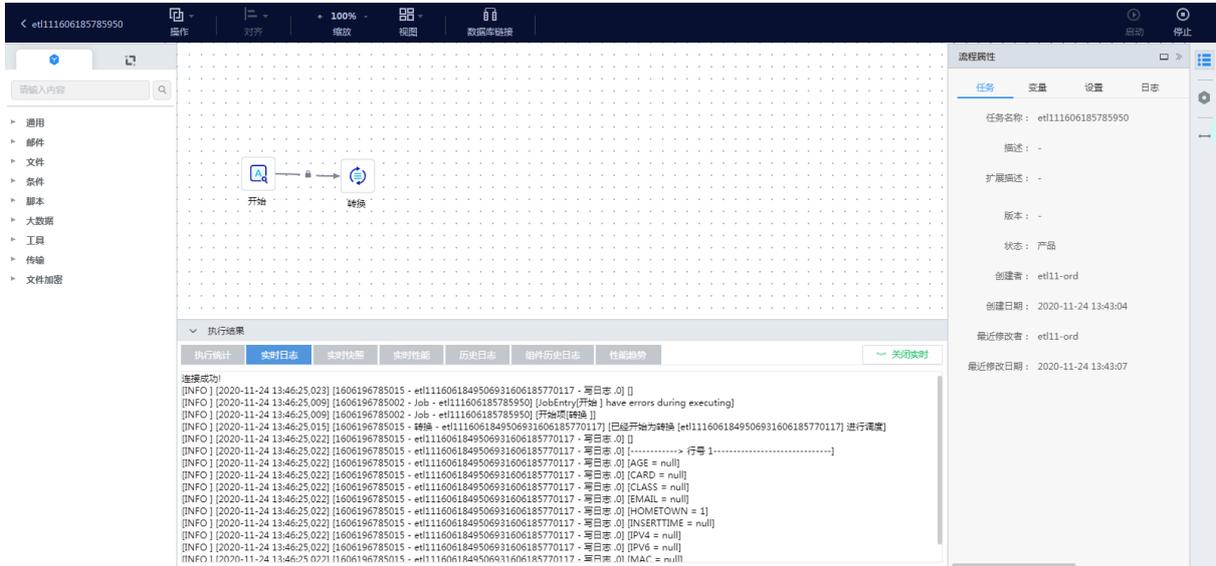


图 6-21

7. 任务监控中流程操作同监控告警-任务监控一致。

7. 场景服务

场景服务通过 ETL 场景化，使用户配置操作简化，无需拖拽功能模块，只需简单配置或者修改

就可以完成复杂的业务逻辑。

7.1 异构表-全量同步

场景服务->点击创建场景，可选对应的场景服务模板。

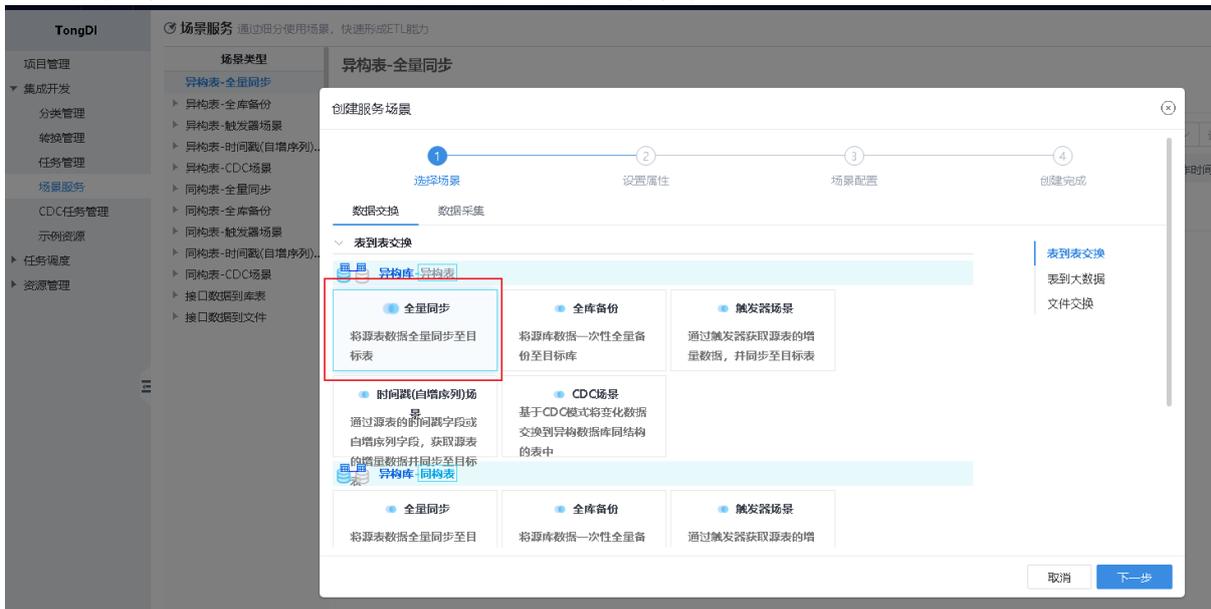


图 7-1

选“全量同步”模板后点击下一步，可配置任务属性：

- 任务名称：根据实际情况填；
- Cron 表达式：点击“？”进行配置，支持秒、分钟、小时、日、月、周、年周期设置；
- 是否清空目标表：勾选后每次运行流程时都会先清空目标表；
- 错误数据存放目录：如在流程运行入库时因主键冲突、数据类型不一致等系列数据问题引发的数据未能入库时，会将未入库的数据抽出来生成.txt 文件放到此目录，此目录可手动配置，如服务器无此目录时会自动创建目录。

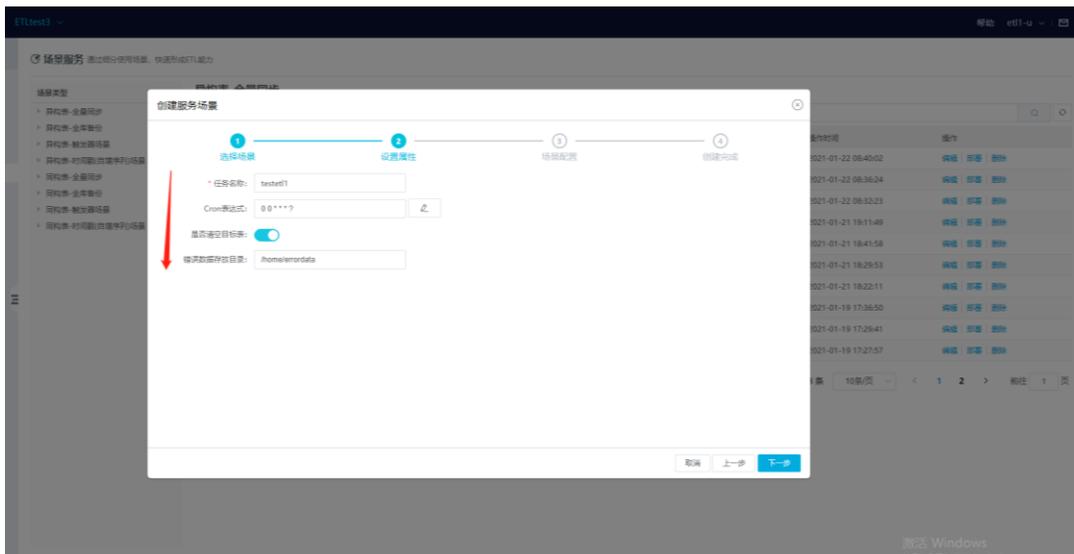


图 7-2

填完以上属性，点击下一步。

点击图 7-3 “+”：可选数据源及表；

可手动连接表与表的对应关系，连接后为红色，点击红色连接线如图 7-4：可手动配置字段映

射关系；

点击图 7-5 中“场景配置”可返回，返回后连接线变绿。

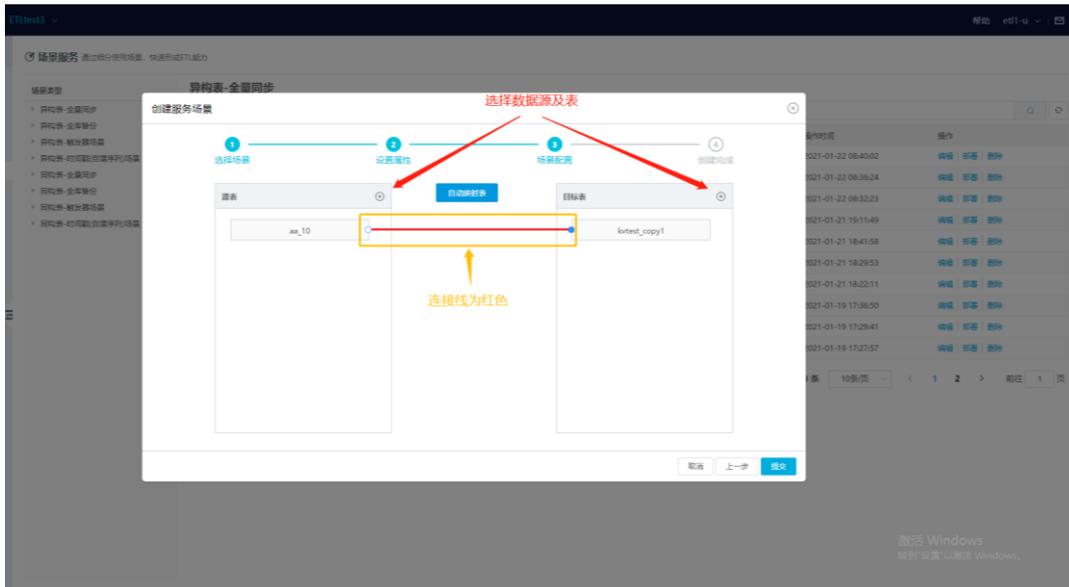


图 7-3

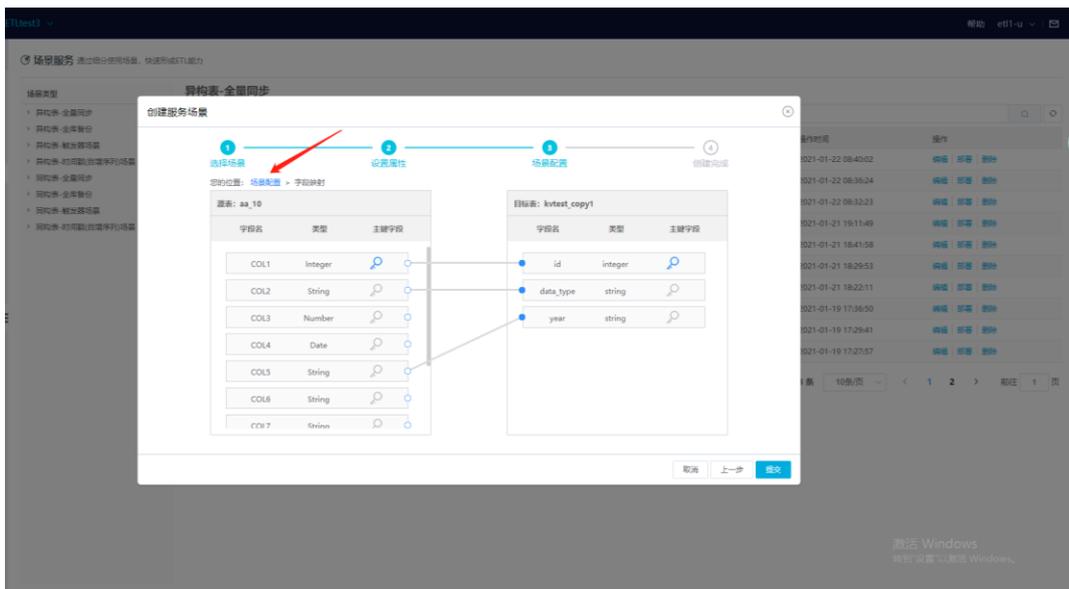


图 7-4

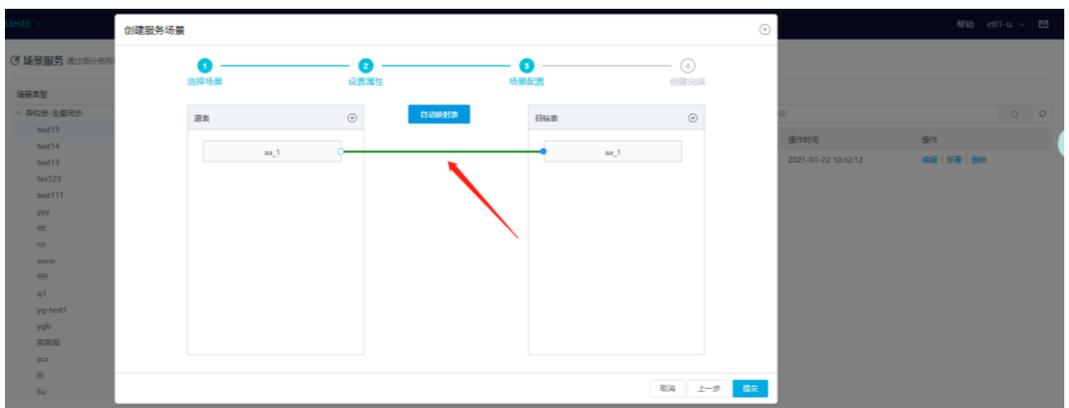


图 7-5

点击提交可保存流程，在流程列表中展示，部署统一见 7.13 场景服务列表-操作。



图 7-6

7.2 异构表-全库备份

前两步选择模板和设置属性同上。

第三步配置场景时只选数据源即可，连接对应的表，配置对应的字段映射。

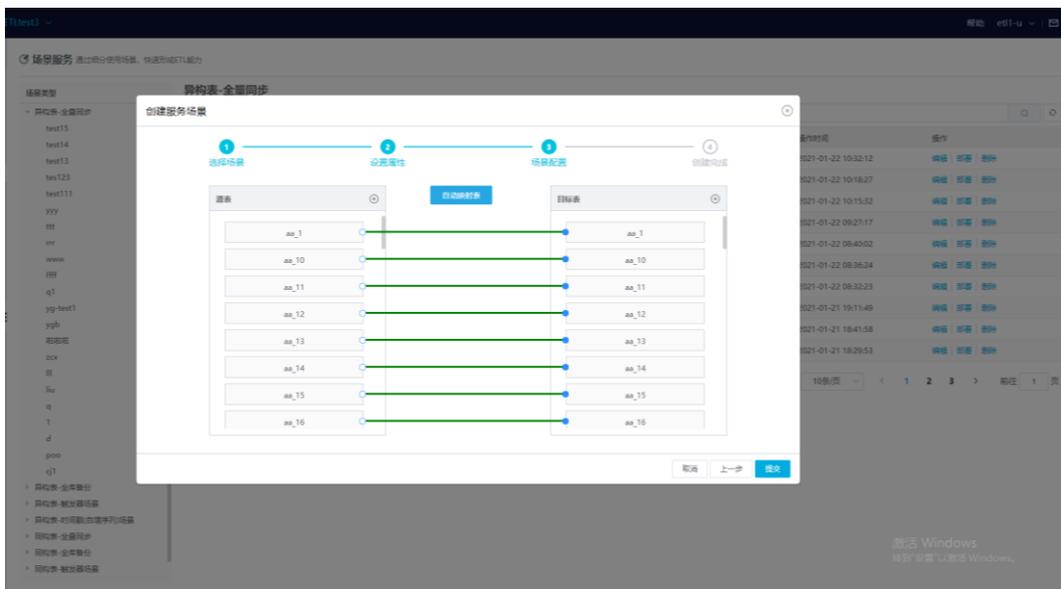


图 7-7

7.3 异构表-触发器场景

点击“创建场景服务”，选择触发器场景，点击下一步。

设置属性页面，配置完后点击下一步。

- 任务名称：根据实际情况填；
- Cron 表达式：点击“”进行配置，支持秒、分钟、小时、日、月、周、年周期设置；
- 最大记录获取数：可根据实际填写；
- 监控插入操作：勾选后原表插入数据会触发交换；
- 监控更新操作：勾选后原表更新数据会触发交换；
- 监控删除操作：勾选后原表删除数据会触发交换；
- 双向同步：暂不支持；
- 是否自动创建触发器：勾选后在部署流程时会创建触发器；
- 入库时先做更新：勾选后流程入库时会先做一次更新操作，然后再入库；

- 主键冲突字符串：如 MySQL 主键冲突字符串：Duplicate entry，配置后如入库时有主键冲突情况会优先更新一次。
- 错误数据存放目录：如在流程运行入库时因主键冲突、数据类型不一致等系列数据问题引发的数据未能入库时，会将未入库的数据抽出来生成.txt 文件放到此目录，此目录可手动配置，如服务器无此目录时会自动创建目录。

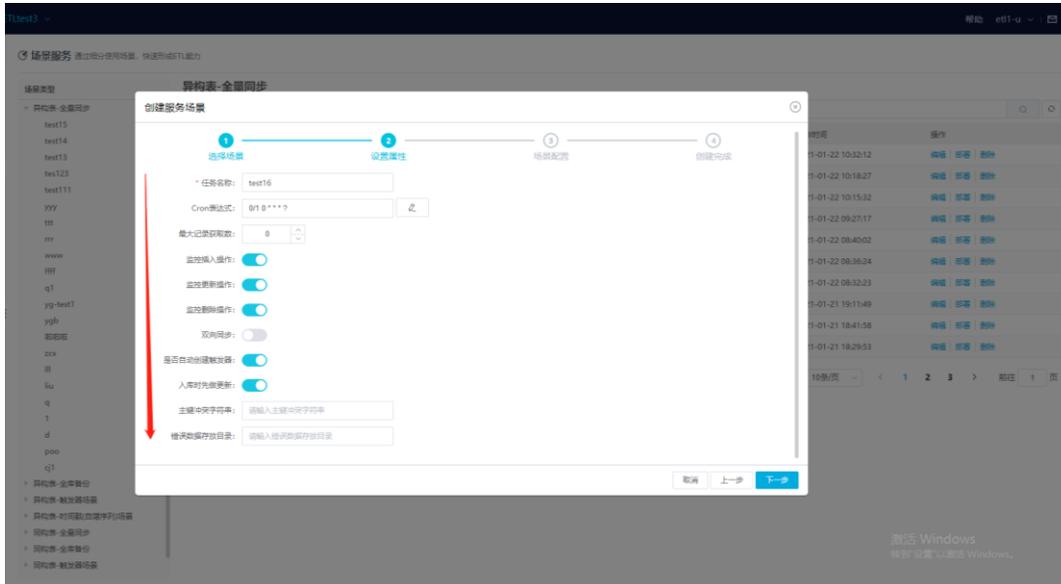


图 7-8

场景配置同上。

7.4 异构表-时间戳(自增序列)场景

- 点击“创建场景服务”，选时间戳场景，点击下一步
- 设置属性页面，配置如下，配置后点击下一步
- 任务名称：根据实际情况填；
- Cron 表达式：点击“？”进行配置，支持秒、分钟、小时、日、月、周、年周期设置；
- 最大记录获取数：根据实际情况填；
- 时间戳字段名称：源表中时间戳字段名；
- 时间戳字段类型：支持 Date,String,Integer；
- 入库先做更新：勾选后流程入库时会先做一次更新操作，然后再入库；
- 主键冲突字符串：如 MySQL 主键冲突字符串：Duplicate entry，配置后如入库时有主键冲突情况会优先更新一次。
- 错误数据存放目录：如在流程运行入库时因主键冲突、数据类型不一致等系列数据问题引发的数据未能入库时，会将未入库的数据抽出来生成.txt 文件放到此目录，此目录可手动配置，如服务器无此目录时会自动创建目录。

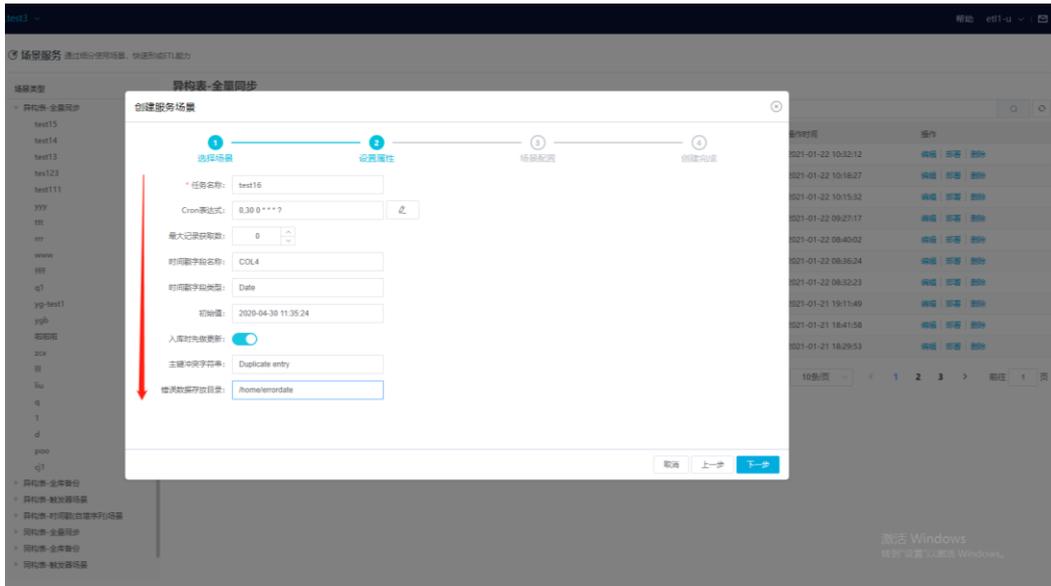


图 7-9

7.5 异构表-CDC 场景

- 点击“创建场景服务”，CDC 场景服务，点击下一步。
- 设置属性页面配置如下：点击下一步。
- 任务名称：根据实际情况填写；
- Cron 表达式：点击“📅”进行配置，支持秒、分钟、小时、日、月、周、年周期设置；
- 最大记录获取数：根据实际情况填写；
- 入库先做更新：勾选后流程入库时会先做一次更新操作，然后再入库；
- 主键冲突字符串：如 MySQL 主键冲突字符串：Duplicate entry，配置后如入库时有主键冲突情况会优先更新一次。
- 错误数据存放目录：如在流程运行入库时因主键冲突、数据类型不一致等系列数据问题引发的数据未能入库时，会将未入库的数据抽出来生成.txt 文件放到此目录，此目录可手动配置，如服务器无此目录时会自动创建目录。

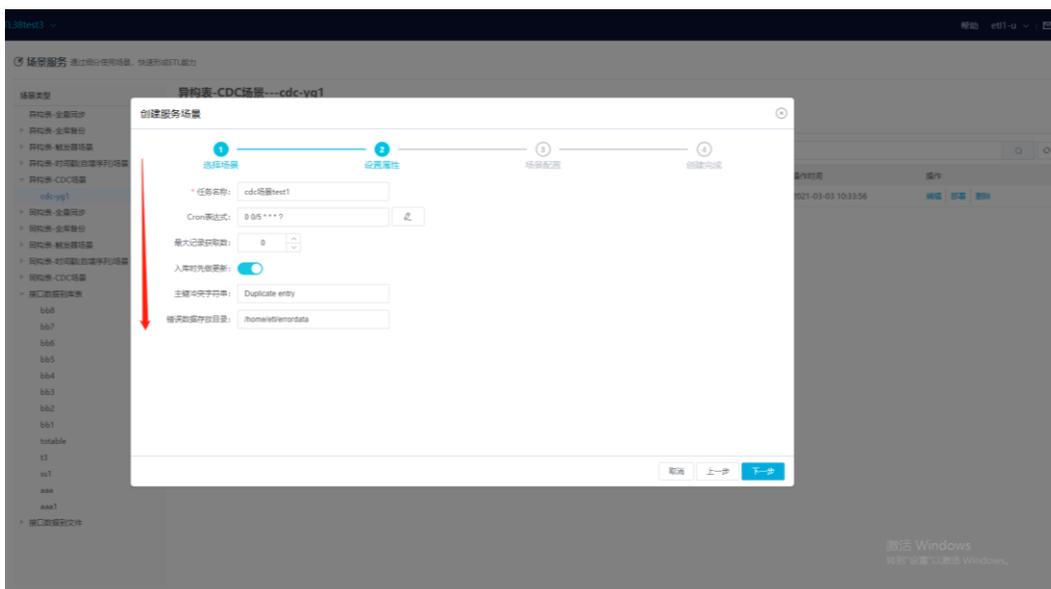


图 7-10

场景配置左侧源表为 CDC 任务中的源表，需要创建 CDC 任务且配置源表此处才能展示（CDC 任务配置及数据库配置详见 8 CDC 任务创建流程），目的表正常配置可操作的不同的源表的异构表即可。

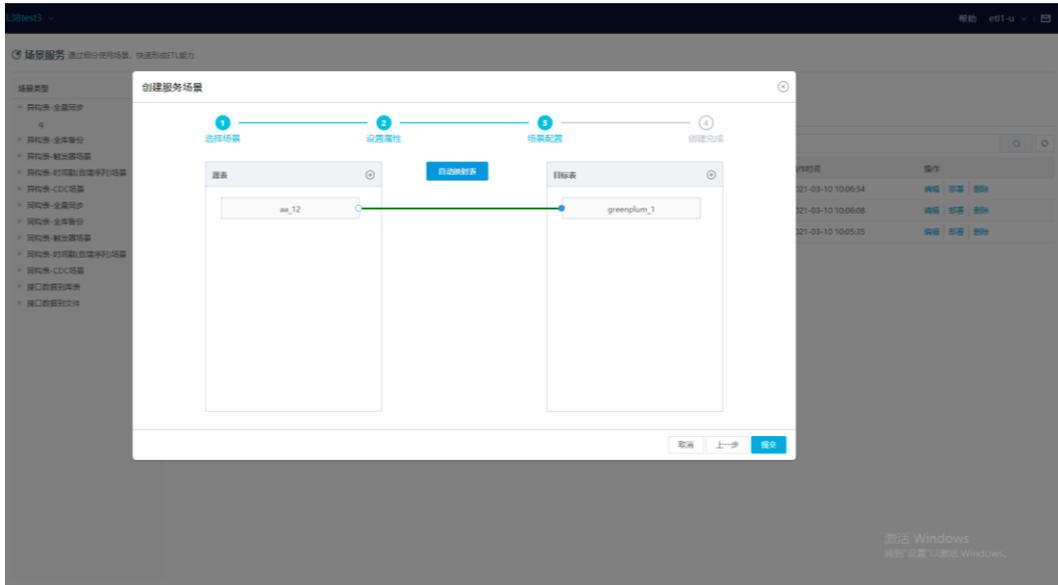


图 7-11

连接表后点击连接线，配置字段映射即可。

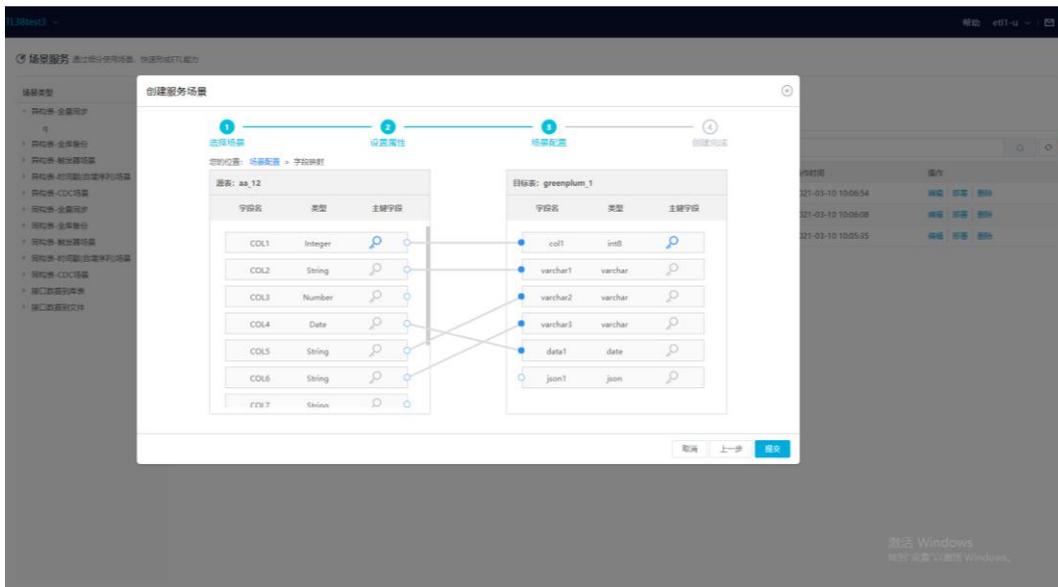


图 7-12

7.6 同构表-全量同步

同异构表——全量同步配置流程一致。

7.7 同构表-全库备份

同异构表——全库备份配置流程一致。

7.8 同构表-触发器场景

同异构表——触发器场景配置流程一致。

7.9 同构表-时间戳(自增序列)场景

同异构表——时间戳（自增序列）场景配置流程一致。

7.10 同构表-CDC 场景

同配异构表——CDC 场景配置流程一致，只是源表和目的表需要是同构表。

7.11 接口数据到库表

点击“创建场景服务”，选数据采集，选接口数据到库表，点击下一步。



图 7-13

设置属性同上，完成后点击下一步。



图 7-14

- 场景配置-输入配置如下：配置完点击下一步。
- 服务名称来自 APIGW：开启此选项可从已发布的网关服务中直接选服务，不开启需要手动填写服务地址。
- 采集服务名称：下拉选 APIGW 服务。

- 服务地址：选择后展示此 APIGW 服务的地址。
- 服务访问方式：选 GET：请求为 GET，选 POST：请求为 POST，请求参数在请求参数/请求头中填写。如果是 APIGW 服务，可自动获取 APIGW 中已填的请求头和体。
- 消息体信息：请求 body，可以传 json 串。
- 请求参数：请求参数配置后会将参数拼接到 url 路径。
- 请求头：请求头。
- 返回值模板：填接口返回的 json 格式串，会根据此格式生成匹配路径，不支持带列表的 json 格式串。

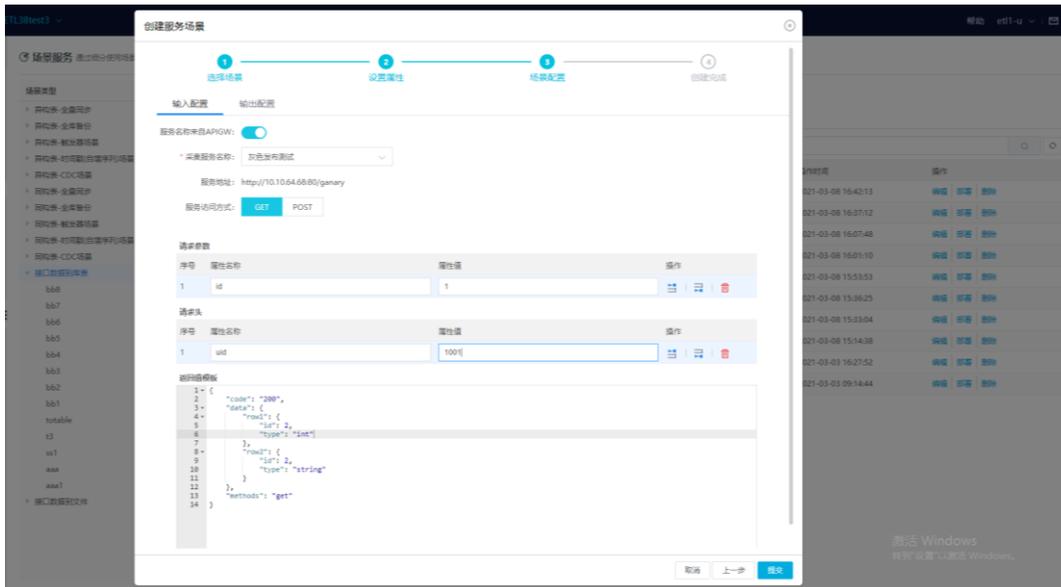


图 7-15

场景配置-输出配置如下：

- 目标数据库：可选；
- 目标模式：可选；
- 目标表：可选；
- 数据节点：根据“输入配置”中“返回值模板”的 json 串匹配出的字段名；
- 连接线：支持手动连接。

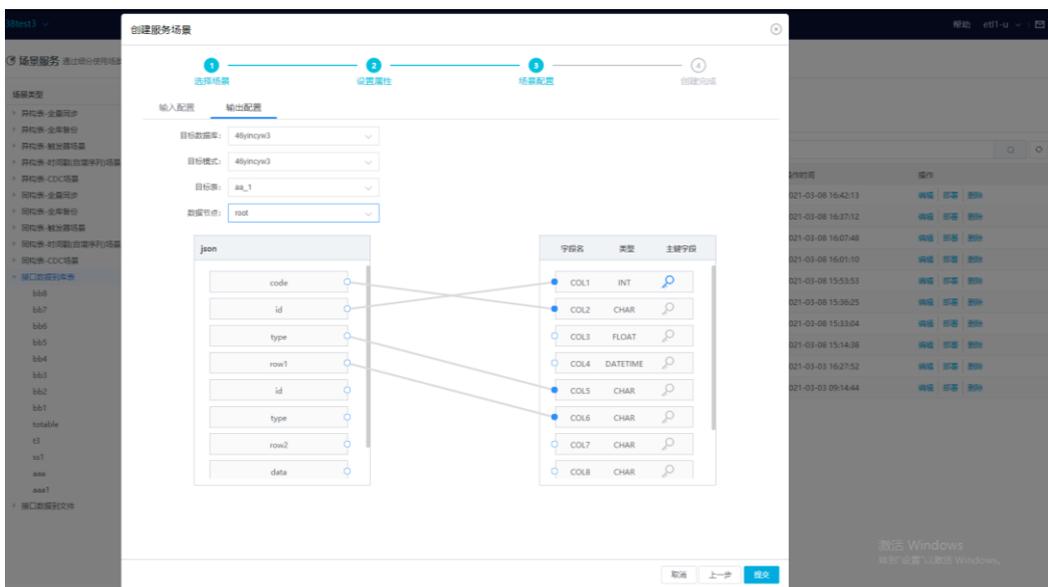


图 7-16

7.12 接口数据到文件

点击“创建场景服务”，选数据采集，选接口数据到文件，点击下一步；



图 7-17

设置任务名称，配置任务启动时间或者重复周期，然后下一步；

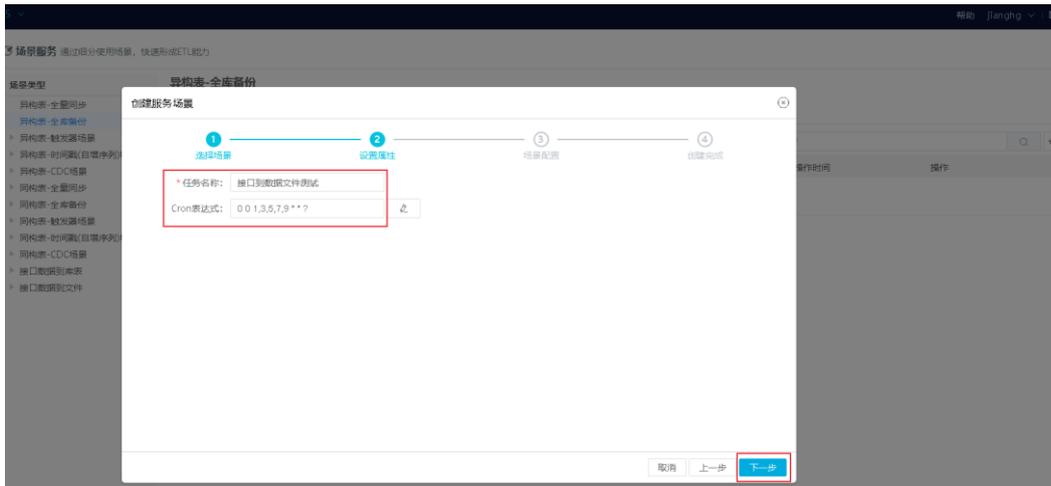


图 7-18

- 选择场景配置：同上；
- 场景配置-输入配置：同上；
- 场景配置-输出配置如下：提交即可；
- 文件：可选 FTP 目录；
- 时间日期格式：可选多种带时间日期的文件；
- 追加：开启追加到文件；
- 数据节点：同上，不同于数据库的是没有目的库，勾选字段指定输出字段。

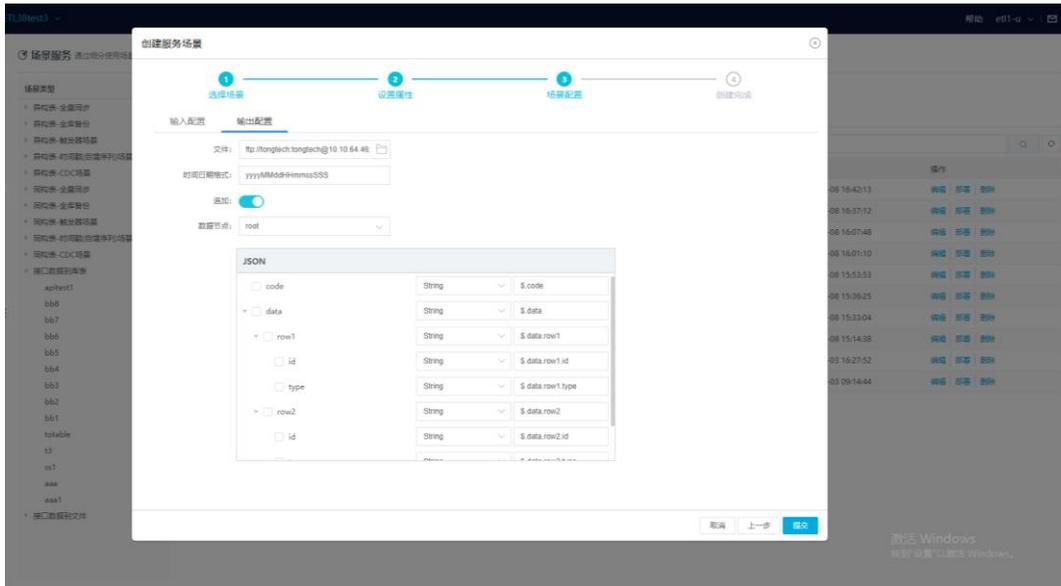


图 7-19

7.13 场景服务列表-操作

7.13.1 场景服务流程-库表类型-编辑

点击“编辑”，可编辑流程场景配置中的异构库表，字段映射关系，同构库表不支持编辑。

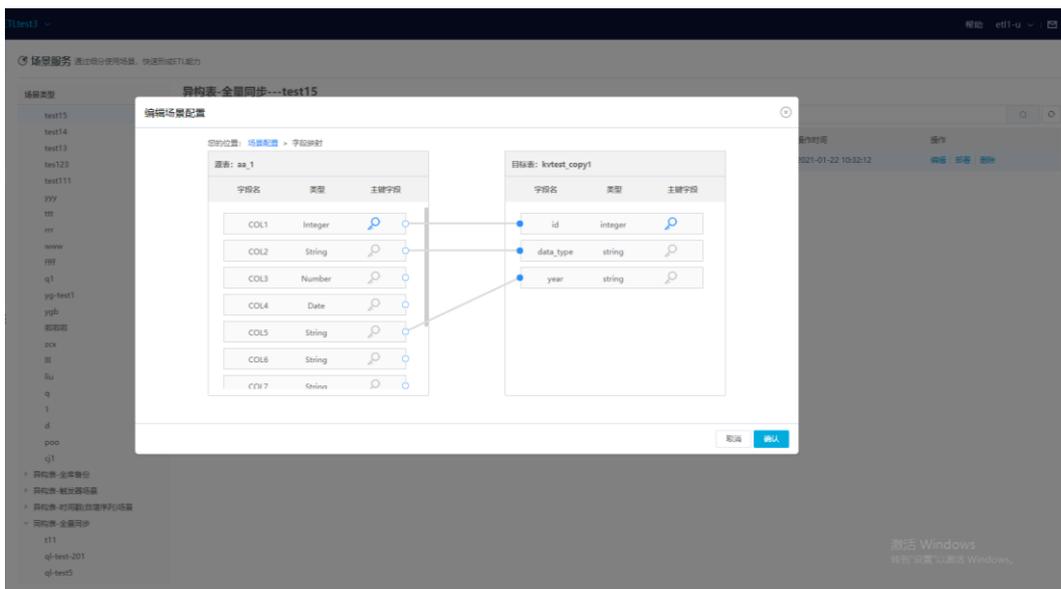


图 7-20

7.13.2 场景服务流程-接口类型-编辑

可编辑输入/输出配置如下：

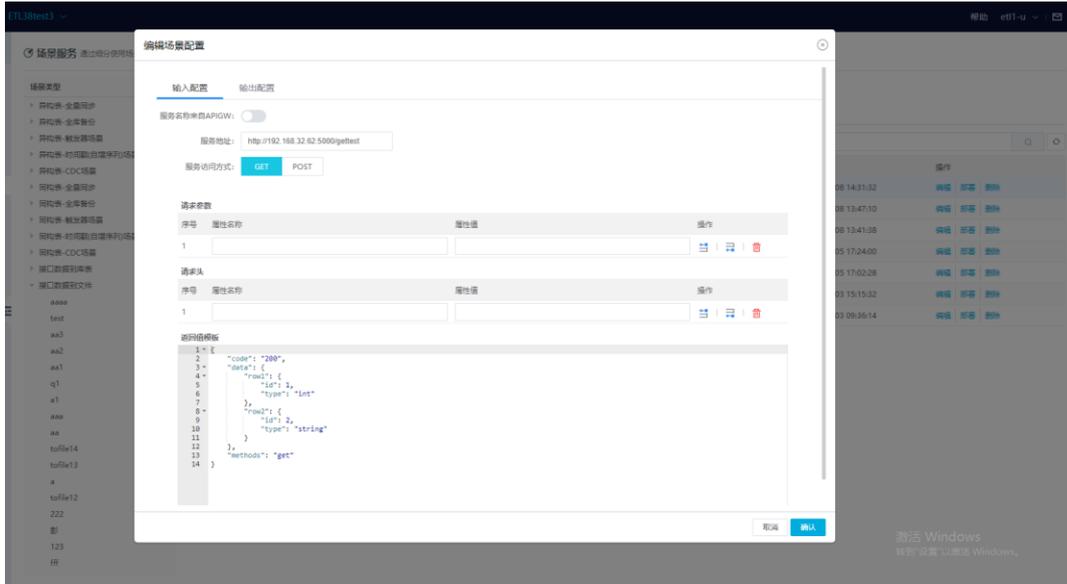


图 7-21

7.13.3 场景服务流程-部署

部署同 5.3.3 流程操作-部署一致。

7.13.4 场景服务流程-删除

点击流程对应的“删除”按钮，删除此流程。

8. CDC 任务创建流程

流程涉及步骤：

- 配置数据源
- 创建抽取任务
- 选择采集的数据源、选择存储的数据源
- 进行任务启动、停止、编辑

8.1 创建抽取

进入 TongCDC 界面，点击创建按钮，打开创建任务的对话框，如下图所示。

创建任务
✕

基本信息

* 任务名称:

* 部署集群:

部署后自动启动:

存储数据源类型:

监听的操作: 插入 更新 删除

抽取配置

* 数据源:

* schema:

* 表:

* 开始位置: [查询](#)

储存配置

* 存储数据源:

默认分片方式: 按日 按月 按年

是否默认分组:

默认分组方式: 按源库

数据保留时间: 天

取消 确定

图 8-1

基本信息:

- 任务名称: 编辑 CDC 任务名称
- 部署集群: 选择 CDC 所在的节点和集群
- 部署后自动启动: 设置自启动, 确定创建之后状态直接变成“部署中”
- 存储数据源类型: 选择【存储数据源配置】中配置的文档型数据源或 Kafka 数据源
- 监听操作: 设置对库表的监听操作, 至少选择一项

抽取配置:

- 数据源: 选择【数据源配置】中配置的关系型数据源
- Schema: 选择上述数据源下的 schema
- 表: 选择上述 schema 下的表, 可选多张表, 监控操作可多选; 可选的表是没被 CDC 任务引用过的表
- 日志文件名: 选择 mysql 数据库下二进制日志文件
- 发布名称: 选择 postgresql 数据库下监控表的发布名称
- 开始位置: 点击查询, 二进制文件中游标位置来设置抽取范围

存储配置:

- 存储源: 选择由存储类型决定的存储数据源
- 默认分片方式: 按日/按月/按年存储抽取数据
- 默认分组方式: 按源库-存储采集数据到一个数据库中
- 数据保留时间: 采集数据在存储源中存储时间

8.2 启动任务

选择刚保存的任务，点击启动的按钮。



图 8-2

8.3 停止任务

选择刚启动的任务，点击停止的按钮。



图 8-3

8.4 编辑任务

选择刚保存的任务，点击编辑的按钮。



图 8-4

9. 示例资源

示例资源可以快速的创建 TongDI 任务资源。提供了典型场景的全量同步、增量同步，以及转换组件示例，如：数据处理组件、文件组件、工具组件和流程控制组件。

9.1 典型场景

9.1.1 全量同步-任务列表

默认提供全量同步的示例资源，切换到任务列表，然后点击“复制”，选择复制的路径，然后点击“确认”，复制一个示例资源。

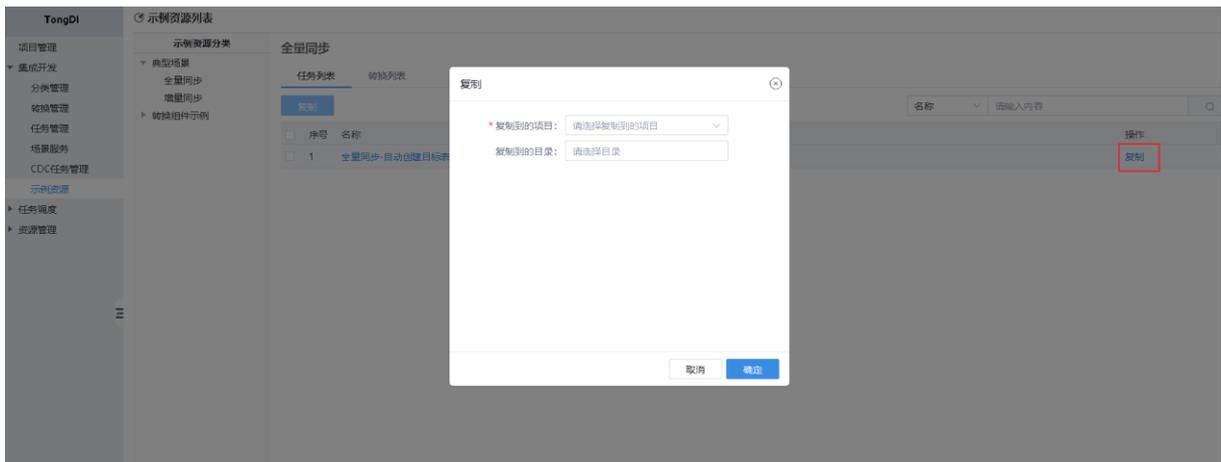


图 9-1

9.1.2 全量同步-转换列表

默认提供全量同步的示例资源，切换到转换列表，点击“复制”，选择复制的路径，然后点击“确认”，复制一个示例资源。

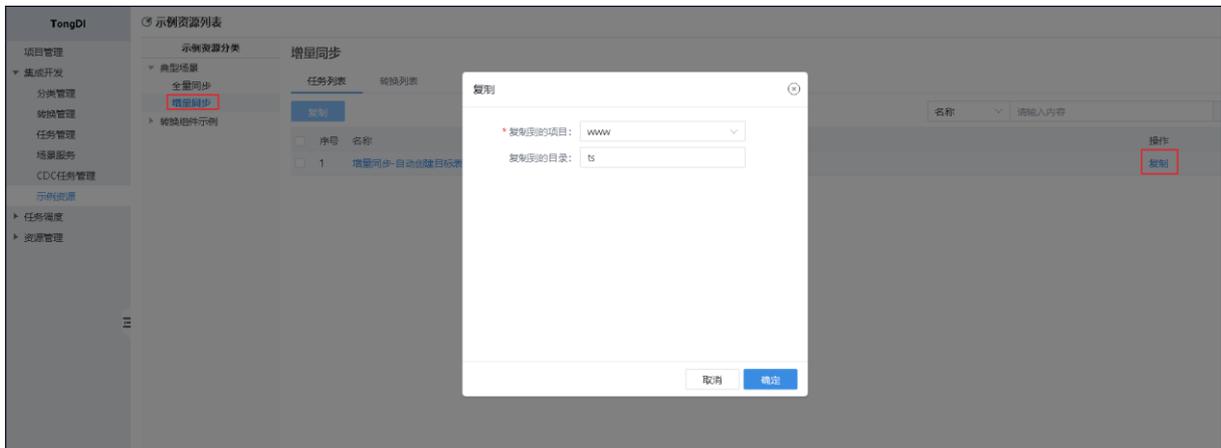


图 9-2

9.1.3 增量同步-任务列表

典型场景中，点击“增量同步”，然后切换到“任务列表”，选择一个场景示例，点击“复制”，选择复制到项目和路径。然后点击“确认”。

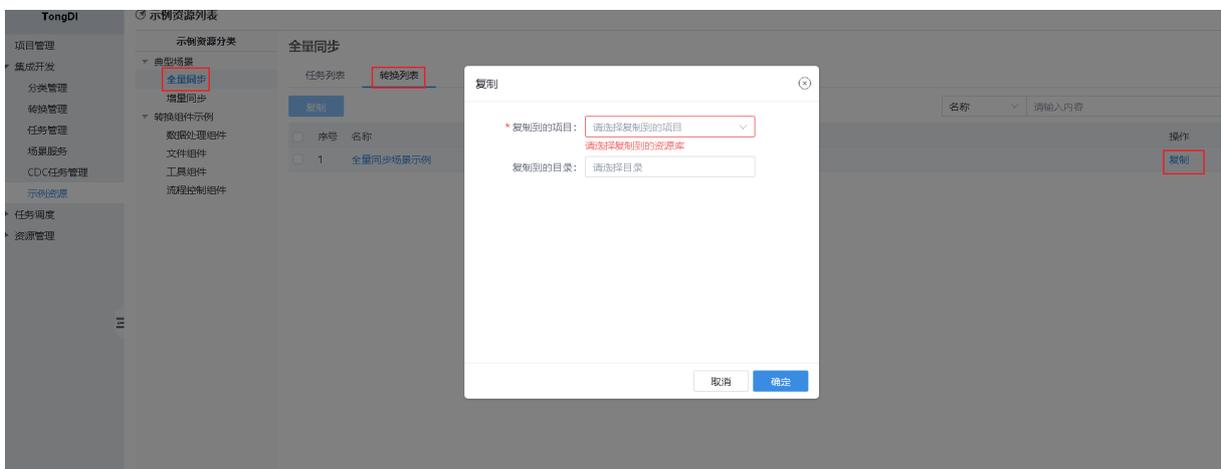


图 9-3

9.1.4 增量同步-转换列表

典型场景中，点击“增量同步”，然后切换到“转换列表”，选择一个场景示例，点击“复制”，选择复制到项目和路径。然后点击“确认”。

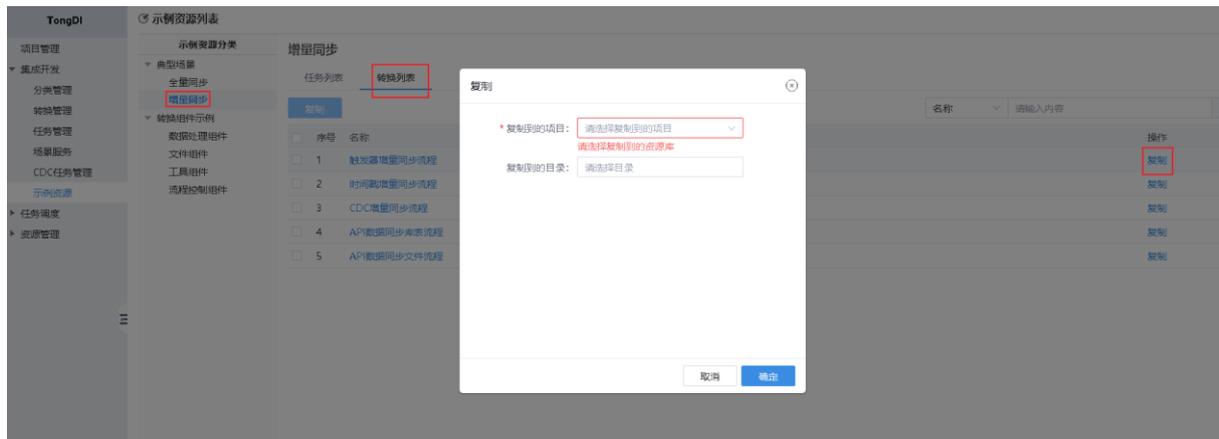


图 9-4

9.2 转换组件示例

转换组件中选择数据处理组件，然后切换的转换列表，选择要使用的数据处理组件，点击复制，最后选择使用的项目和目录，点击“确认”，保存。

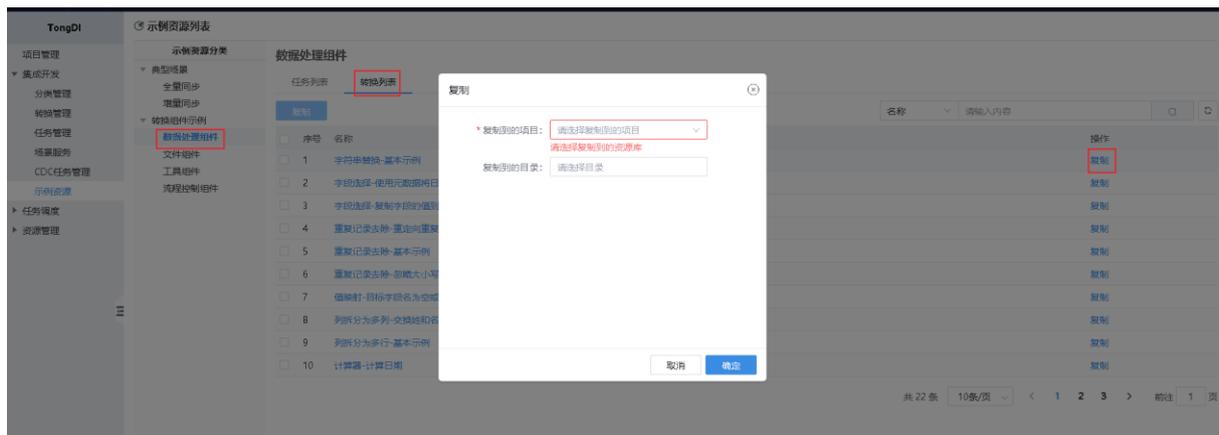


图 9-5

文件组件、工具组件和流程控制组件的使用步骤类似，不做介绍。

10. 周期任务开发

周期业务开发，包括项目->解决方案->业务流程->节点运维任务体系结构，专注业务场景代码开发，提升用户开发体检。

10.1 选择项目

项目空间内的成员代码可以协同开发管理。同部门且具有相同项目权限的数据才能互通，同项目不同部门之间的数据互不可见。

进入离线计算功能模块，点击导航栏，选择已授权的项目，如下图所示：



图 10-1

没有可选择项目时，需要先创建项目，项目创建方法参考 3.1 项目管理。

10.2 解决方案

用户可自定义组合业务流程为一个解决方案。一个解决方案可以包括多个业务流程，解决方案指标可以复用相同的业务流程。

10.2.1 创建解决方案

方式一：根据离线开发模块导航路径【数据开发】->【解决方案】，首次可通过点击“创建解决方案”，设置解决方案名称、业务流程，如下图所示：

说明：此方式只适合该项目没有创建解决方案的情况。

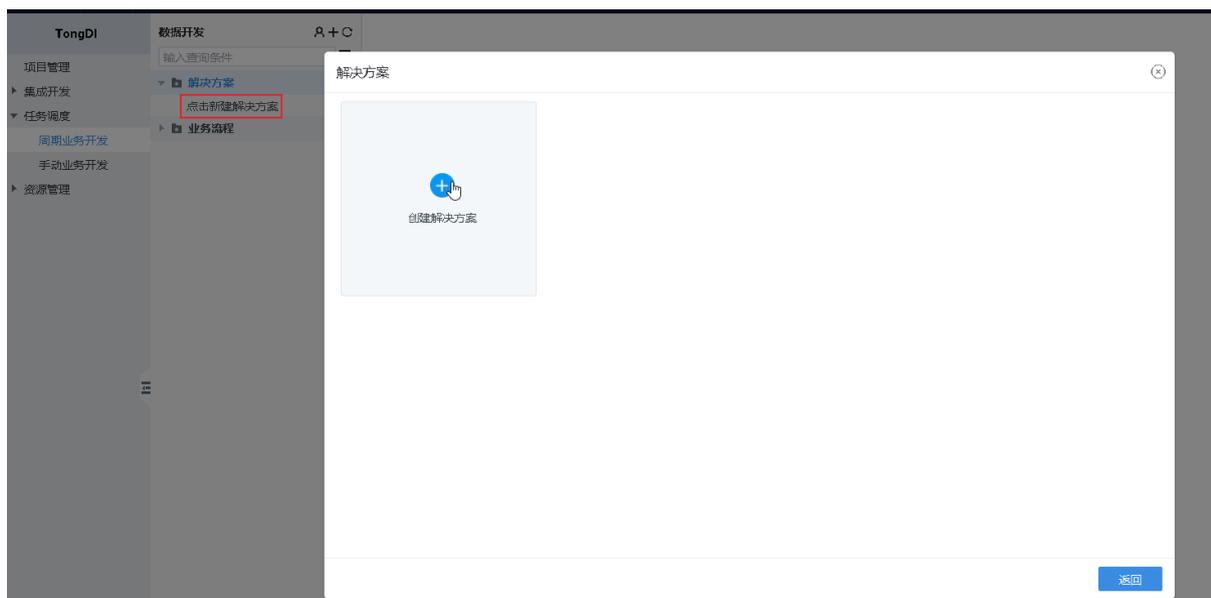


图 10-2

点击确定，可以查看解决方案树下，新增刚才创建的解决方案；

方式二：通过点击图标“+”，点击解决方案，设置解决方案名称、描述、业务流程，如下图所示：



图 10-3

方式三：通过点击图标“☐”，弹出框中点击创建解决方案，设置解决方案名称、描述、业务流程，如下图所示：

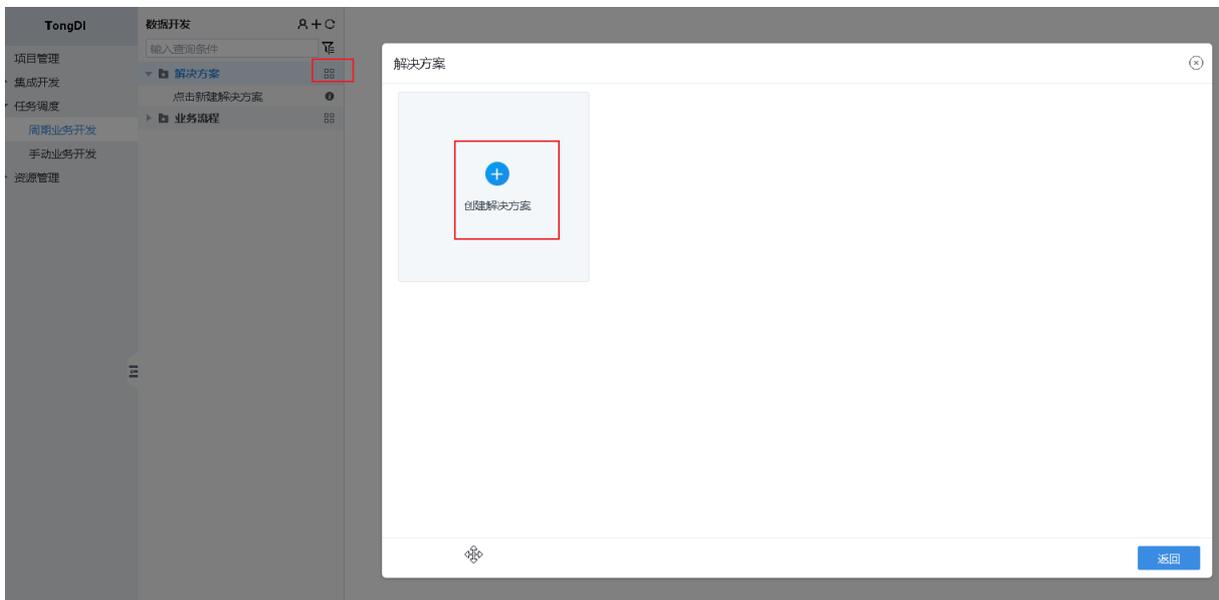


图 10-4



图 10-5

10.3 数据开发流程

数据开发流程能够帮助用户从业务视角组织代码，更加清晰的提供基于任务类型的代码组织式，同时支持用户按照业务流程组织进行发布和运维。业务流程可以被多个解决方案复用。

10.3.1 创建周期业务流程

点击新增图标 ，点击业务流程，设置业务名称、流程类型、部署集群、部署策略、缺省优先级、描述，如下图所示：



图 10-6

10.3.2 新增节点

节点是带有多个输入参数和输出参数的 SQL 代码过程模板，便于用户通过向导方式快速完成开发任务。支持数据集成、数据分析、数据处理等类型。

步骤一：业务流程树中，选择某个业务流程，点击图标 ，点击“新增节点”，在弹出框“新增节点”中选择节点类型、设置节点名称，如下图所示：

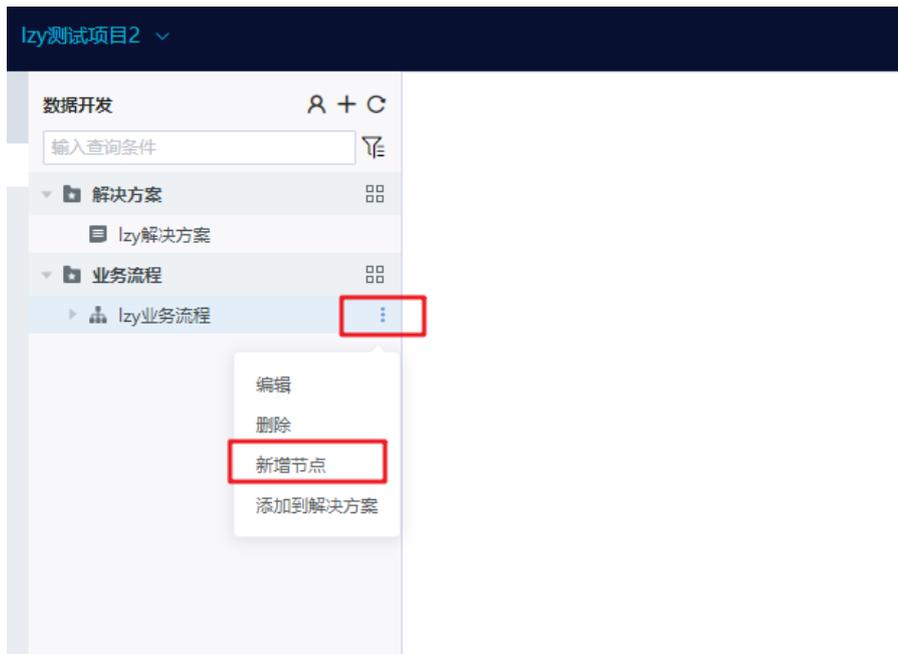


图 10-7

步骤二：点击“确定”，可以查看业务流程树下，刚才新增创建的节点；

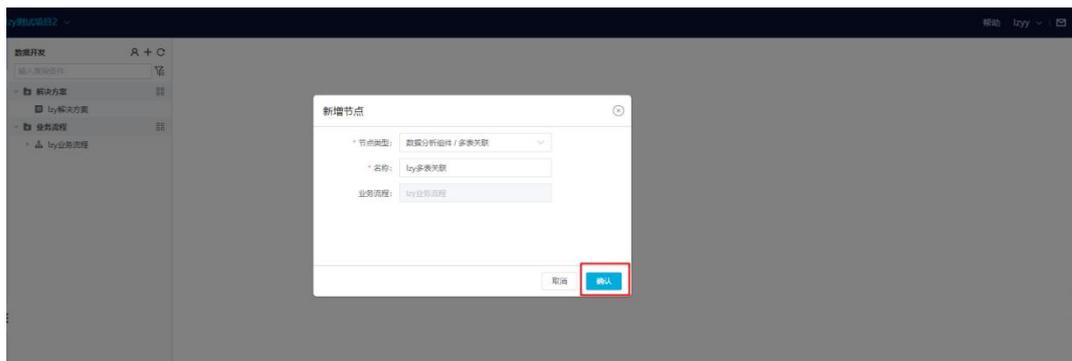


图 10-8

步骤三：业务流程树中，双击该节点，右侧进入节点看板中，配置该节点。

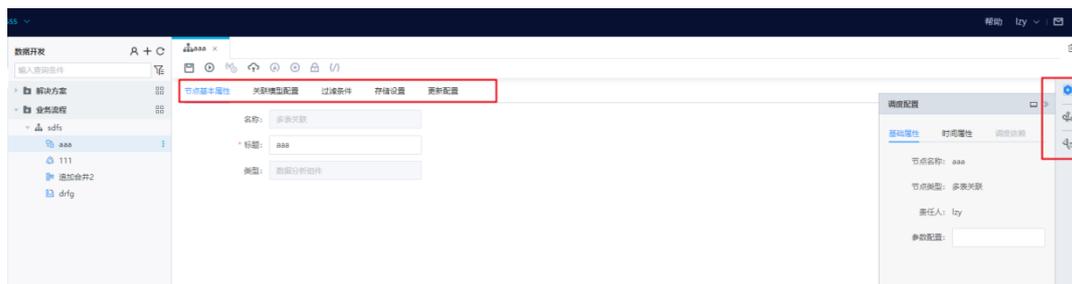


图 10-9

10.3.3 配置节点

这里以多表关联为例配置节点，其他如数据聚合、追加合并和二维转一维操作类似，不再说明。

应用场景：订单数据中一般只存取了用户的 ID，并没有用户的基本信息，如果想要分析商品的用户画像，则需要将订单表和用户基本信息表借助两表共有的用户 id 可汇总成一张新表，对这张新表进行分析即可得到商品的用户画像。数据计算平台支持一表对一表的关联合表，也支持一表对多表（暂不开发）的关联合表。

关联支持三种形式：

- Left join: 左关联，表示以左表为主表进行关联操作。显示左表中的所有记录，不管是否与关联条件相匹配，而右表中的数据只显示与关联条件相匹配的记录。
- Inner join: 等值关联，表示对两表取交集，只返回两个表中联结字段相等的行。
- Full join: 全关联，表示对两表取并集。返回两表中所有数据，不管是否匹配。

10.3.3.1 节点基本属性

业务流程树中，双击多表关联节点，右侧展开该节点面板，基本属性页签，可设置节点名称、标题、类型、执行引擎、目标表表名，如下图所示：

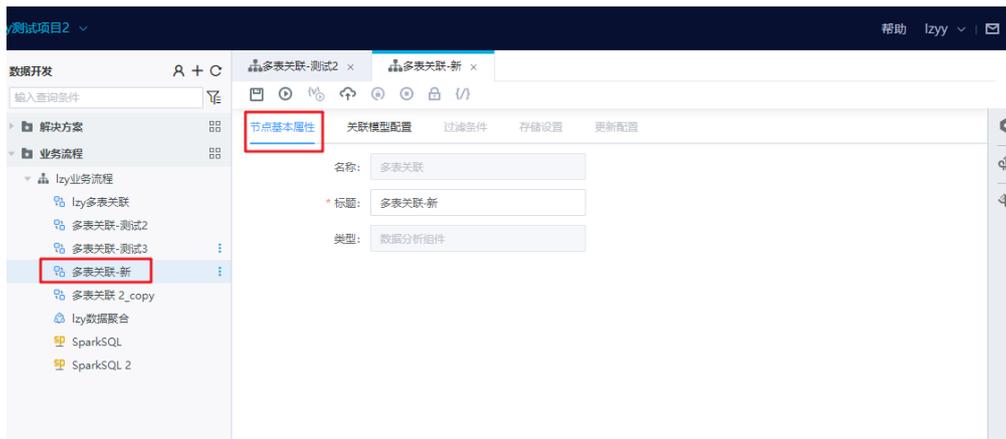


图 10-10

10.3.3.2 关联模型配置

进入关联模型配置页签，从左侧库表树中拖拽 2 张需要关联的表到节点面板中，如下图所示：

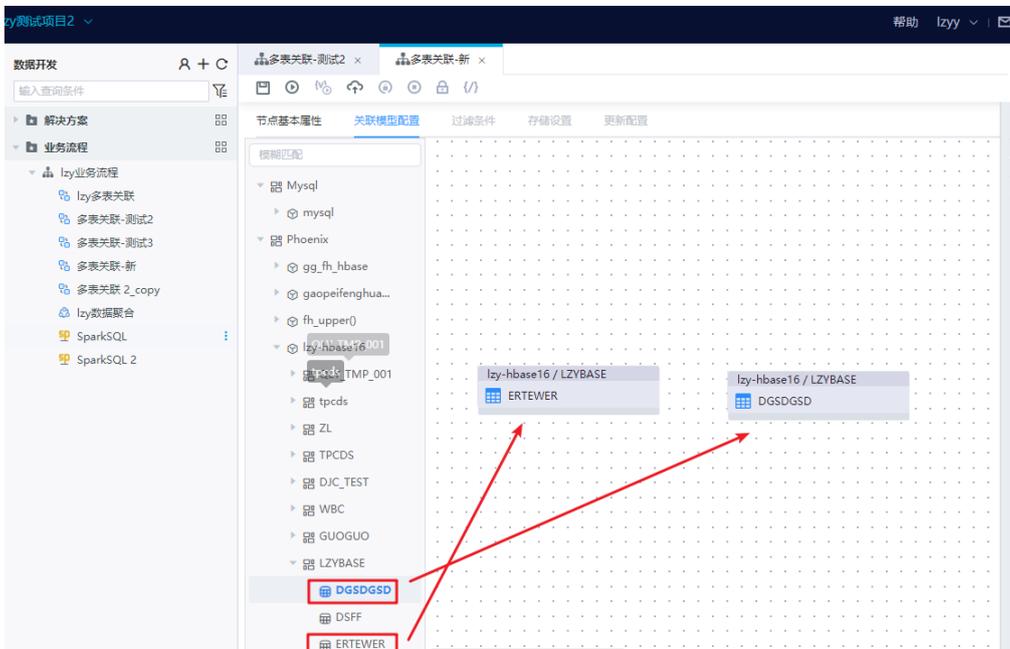


图 10-11

节点面板中，选择表 A 到表 B，进行连线，连线中关联方式默认展示 left join，如下图所示：

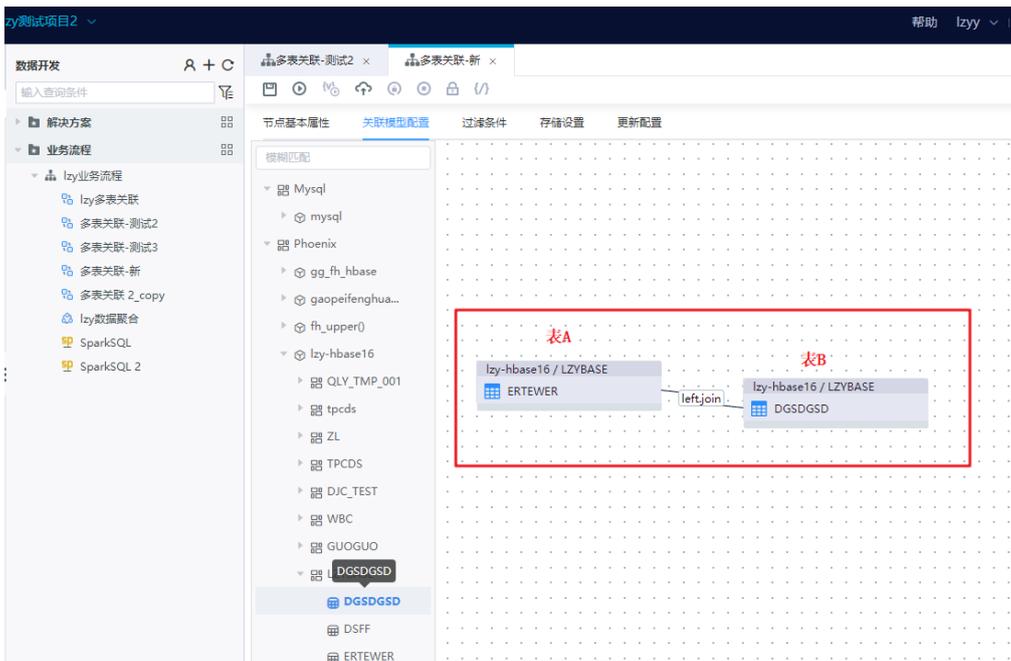


图 10-12

单击关联方式，选择关联方式、配置关联字段，其中“”表示在上方插入字段，“”表示在下方插入字段，“”删除当前字段，如下图所示：

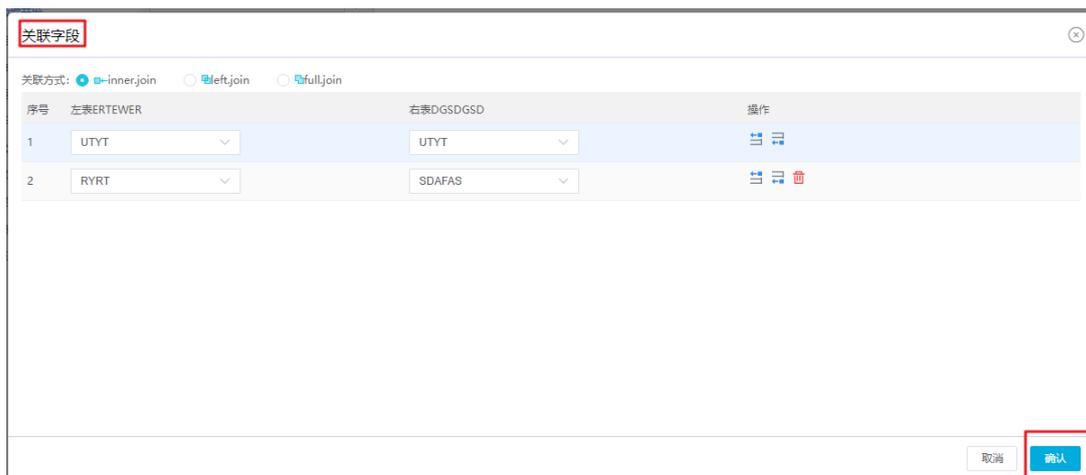


图 10-13

右键单击关联方式，可以删除关联。

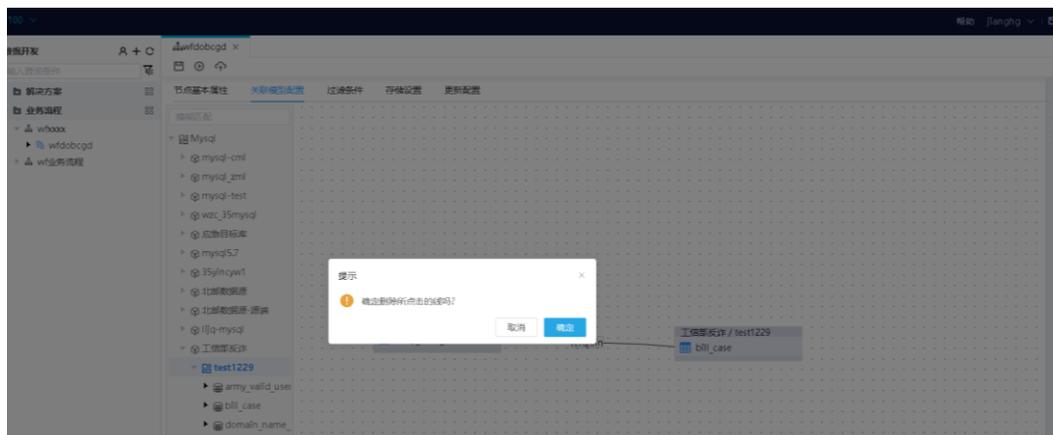


图 10-14

点击确定，节点面板中，更新连线中的关联方式，如下图所示：

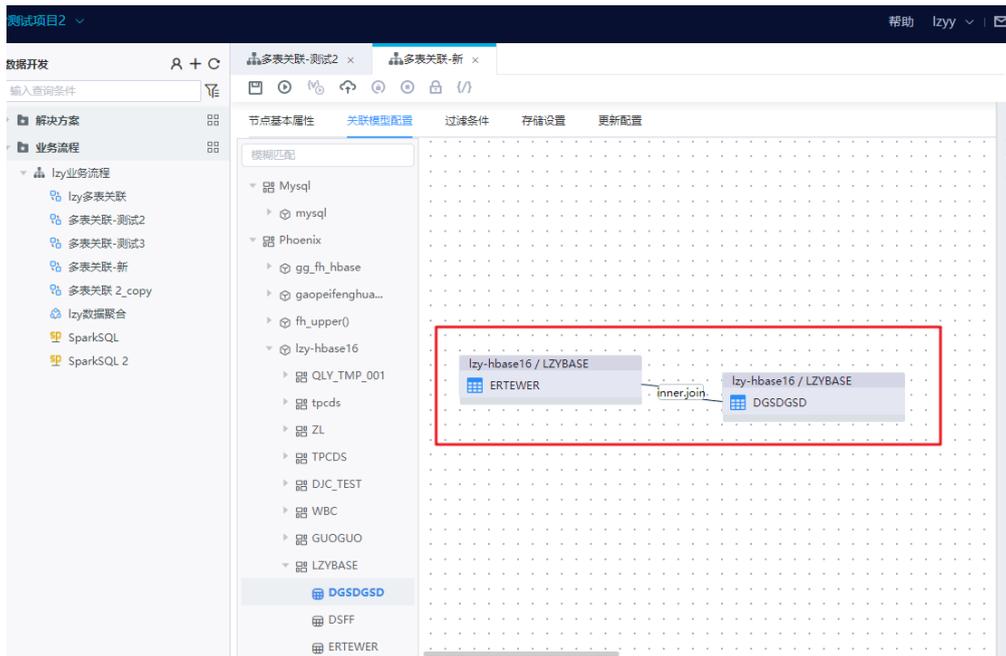


图 10-15

10.3.3.3 过滤条件

设置过滤条件可以过滤掉不需要的数据或是关联后为空的数据，关联为空的数据可以在异常报告中查看到。

步骤 1：进入过滤条件页签，可选择筛选方式、满足条件、过滤条件，如下图所示：

说明：过滤条件支持新增、删除功能。

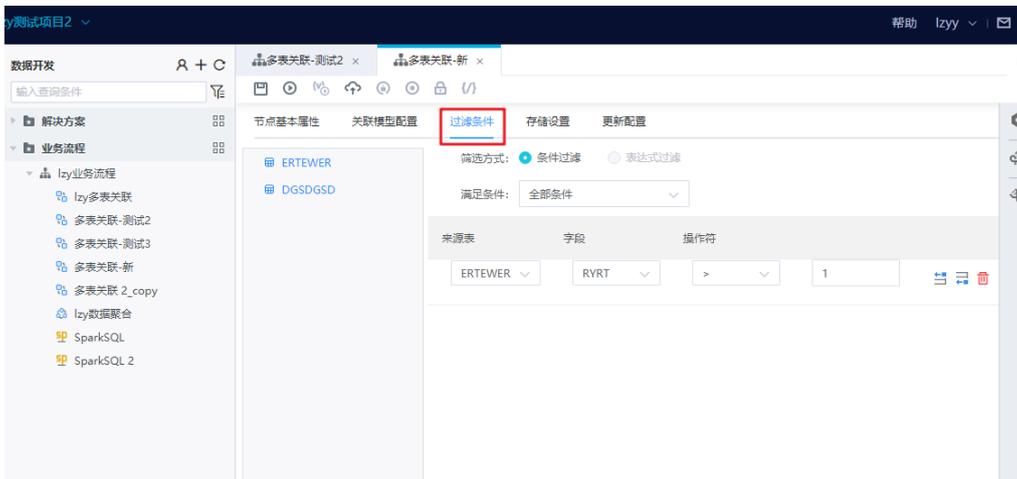


图 10-16

异常报告，如图所示：

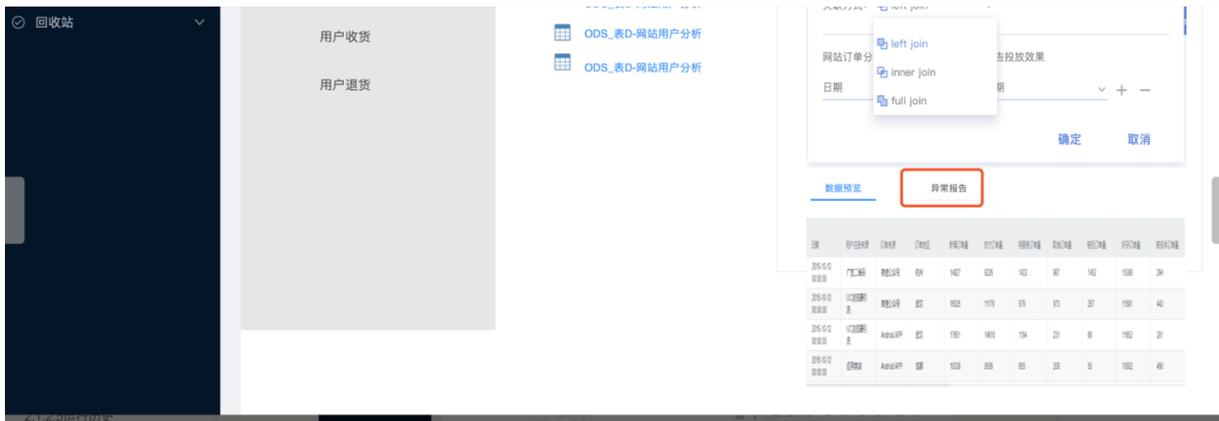


图 10-17

10.3.3.4 存储设置

提交节点时会根据字段设置创建新表，作为多表关联输入的入库表。

步骤 1：进入存储设置页签，可设置数据源、schema、表名、所属类目、生命周期、描述、字段信息，如下图所示：

说明：字段信息中，默认展示关联表的所有字段（约束条件除外），支持删除字段、新增关联表的字段、移位。约束条件（可空、主键）根据不同数据库特性设置。多表关联创建新表时需自定义主键字段。

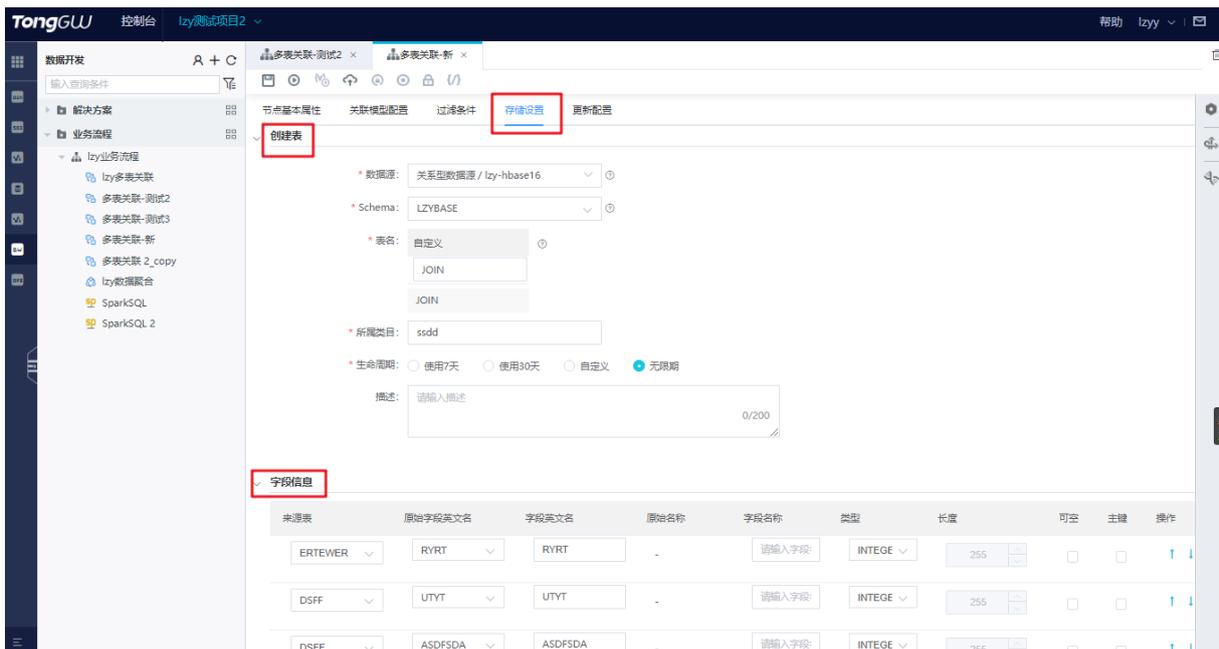


图 10-18

说明：若该节点提交后，会根据该存储设置中配置的表信息，直接下发到物理库生成物理表，无需审批。同时，在【数据建模】【数据地图】模块中生成该项目表 V1 版本记录，影响如下页面功能：

- 【模型设计】中该记录类型为项目表，其表字段不可编辑、发版申请；
- 【审批管理】->【发版已审批】【上线已审批】该记录为系统审批，并且审批通过；
- 【生命周期管理】中新增该模型 V1 版本记录；
- 【上线管理】中该记录为启用中，并且不支持停用；
- 【数据检索】中生成该物理表记录，类型为项目表，该记录不支持申请授权；

【所属类目】中对应类目下新增该模型记录；

【数据资产】【模型监控】中统计该记录。

10.3.3.5 更新配置

定义数据的更新机制。增量标识字段可以选择多张表。

步骤 1：进入更新配置页签，可设置更新方式、表、增量标识字段、统计周期，如下图所示：

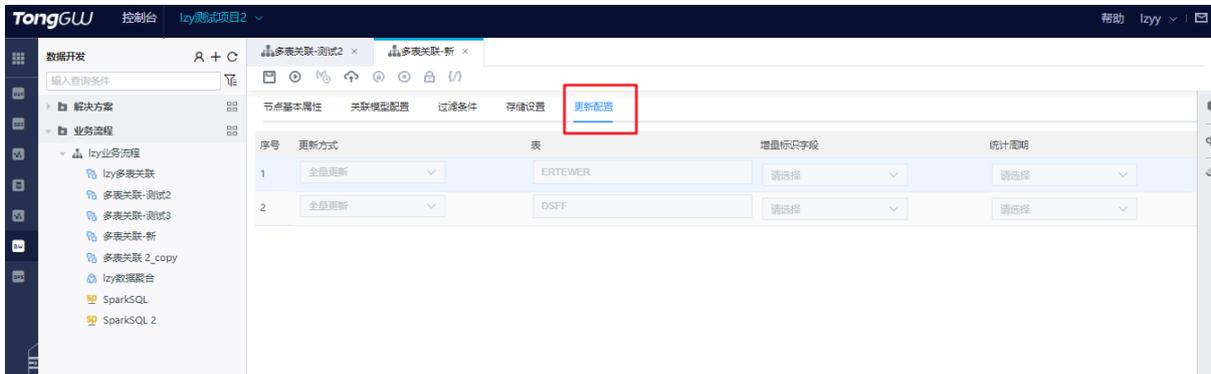


图 10-19

10.3.4 节点调度配置

说明：若是手动业务流程中的节点，则不支持此步骤。

说明：若为手动业务流程节点，提交节点后在【运维中心】->【手动运维任务】查看该节点的任务和实例；若为周期业务流程节点，提交节点后在【运维中心】->【周期运维任务】查看该节点的任务和实例。

10.3.4.1 调度配置

任务调度配置分为：基础属性、时间属性、依赖属性等。

根据导航路径【离线计算】->【数据开发】，业务流程树中，选择某个节点；该节点面板右侧栏，点击图标“”，弹出“调度配置”，进行调度配置，如下图所示：

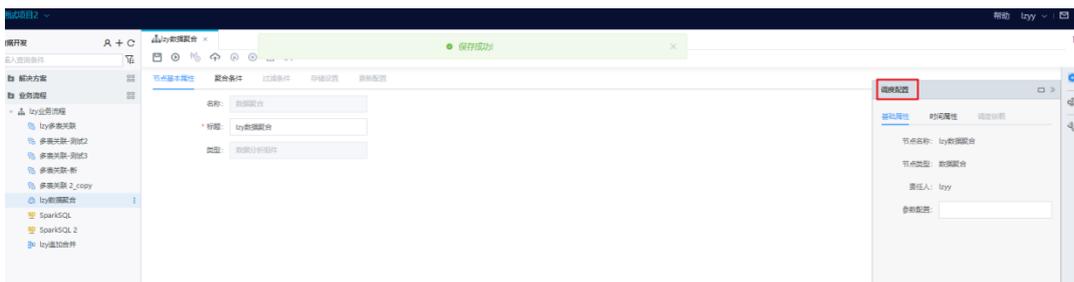


图 10-20

弹框“调度配置”中，点击图标“”，可将调度配置弹出框放大，如下图所示：

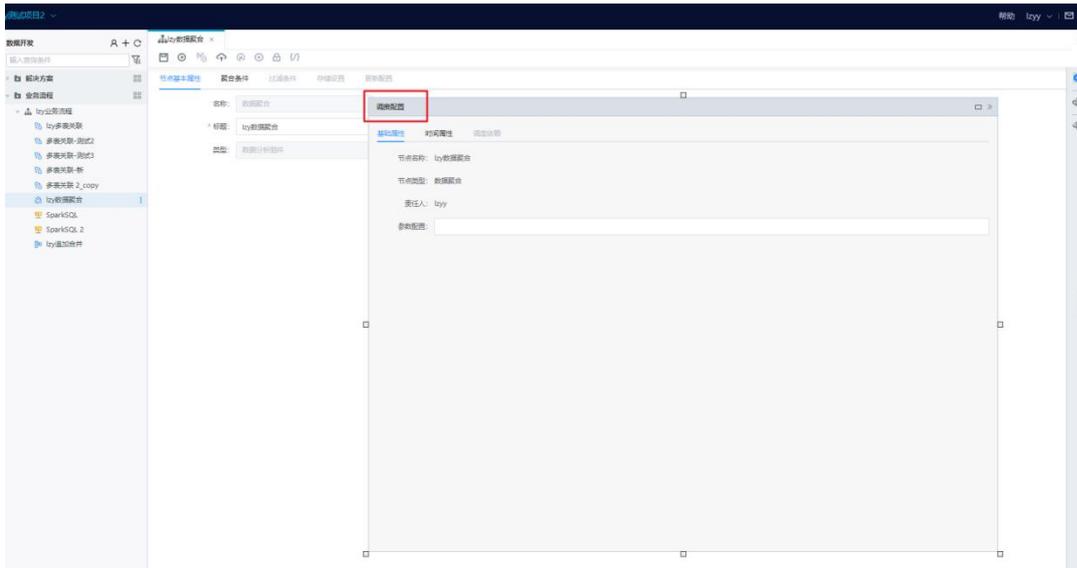


图 10-21

弹框“调度配置”中，点击图标“>”，可将调度配置弹出框缩小，如下图所示：

备注：弹框“调度配置”已缩小时，点击图标“>”，会隐藏该弹框。

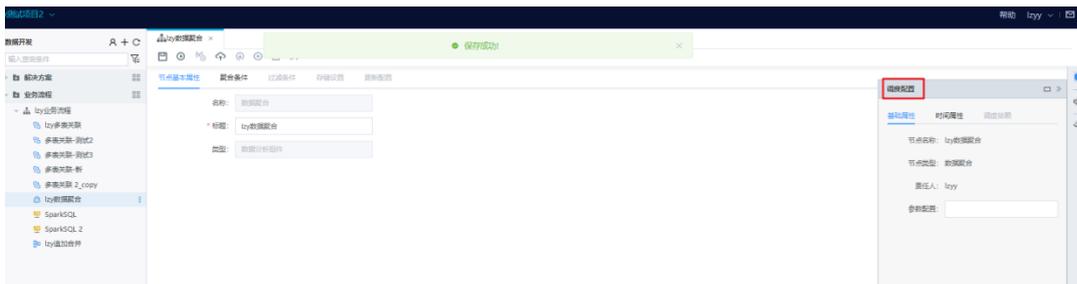


图 10-22

10.3.4.2 基础属性

进入基础属性页签，设置节点名称、节点类型、责任人、描述，如下图所示：

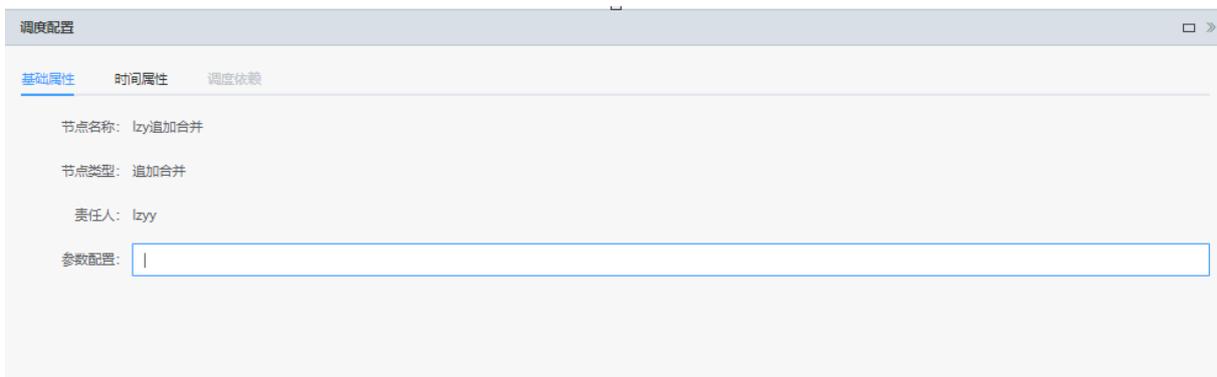


图 10-23

节点名	新建节点时输入的节点名，可以在目录树右键单击节点，选择重命名进行修改。
节点类型	新建节点时选择的节点类型，不可以修改。
责任人	新建的节点责任人默认是当前登录的用户，用户可以根据自身需求修改责任人。只能选择当前工作空间下的成员为责任人。
描述	通常用于描述节点业务、用途。

参数

任务调度时，给代码中的变量赋值。

说明：常见的数据研发场景中，各类型节点的代码需要根据调度周期动态赋值(例如日期时间)，替换变量后再进行计算。调度参数功能就是为了满足此类业务场景。需要注意，如果在业务流程中配置了参数，优先级是大于节点参数的。

10.3.4.3 时间属性

进入时间属性页签，可设置生成实例方式、出错重跑机制、生效时间、调度时间等内容，如下图所示：

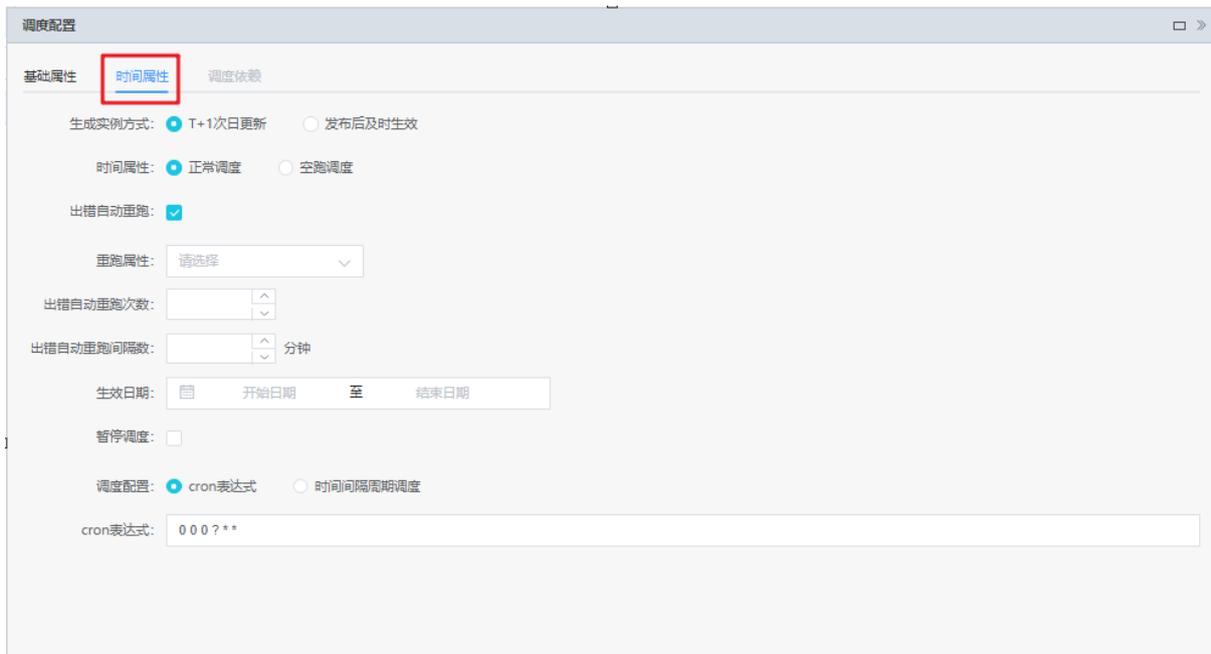


图 10-24

时间属性	说明
实例生成方式	节点的实例生成方式包括 1+1 次日生成和发布后即时生成： 1、T+1 次日生成：全量转实例。 每天 23:30 前提交发布的任务，第二天实例生效。 每天 23:30 后提交发布的任务，第三天实例生效。 2、发布后及时生效。
时间属性	1、正常调度：按照调度周期的时间配置调度，并正常执行，通常任务默认选中该项。 2、空跑调度：按照调度周期的时间配置调度，但都是空跑执行，即一调度到该任务便直接返回成功，没有真正的执行任务。
出错自动重跑	只有勾选后，才支持设置重跑属性、出错自动重跑次数、出错自动重跑间隔数。
重跑属性	可选运行成功或失败后皆可重跑、运行成功后不可重跑、运行失败后可以重跑、运行成功或失败后皆不可重跑，用户根据自身需求进行选择。 如果未选择重跑属性，则不可以提交节点。 只有选择运行成功或失败后皆可重跑、运行失败后可以重跑，支持用户可以配置任务出错自动重跑的次数和间隔。
出错自动重跑次数	默认重跑 3 次，用户可以进行修改。出错自动重跑的次数最少设置为 1 次，最多设置为 10 次。
出错自动重跑间隔数	默认每次重跑的间隔为 2 分钟，用户可以进行修改。出错自动重跑间隔最少设置为 1 分钟，最多设置为 30 分钟。
生效日期	调度在有效日期内生效并自动调度，超过有效期的任务将不会自动调度。
暂停调度	暂停调度后，会根据调度周期的时间配置调度，但是一旦调度到该任务会直接返回失

时间属性	说明
	败，不会执行。通常用于某个任务暂时不用执行，但后面还会继续使用的场景。
调度配置	选择 corn 表达式时，才可设置 corn 表达式。 选择时间间隔周期调度时，才可设置调度周期、定时调度。
调度周期	支持选择分钟、小时、天、周、月。
定时调度	支持选择具体调度时间。

当一个任务被成功提交后，底层的调度系统从第二天开始，将会每天按照该任务的时间属性生成实例，在调度配置的调度时间属性决定某个时间点到达时，任务实例不会马上运行，而是先检查上游运行结果，并根据上游依赖的实例运行结果和时间点运行，运行情况如下：

上游依赖的实例没有全部运行成功，并且已到定时运行时间，则实例仍为未运行状态。

上游依赖的实例全部运行成功，并且未到定时运行时间，则实例进入等待时间状态。

上游依赖的实例全部运行成功，并且已到定时运行时间，则实例进入等待资源状态准备运行。

- 重跑属性可选项，如下图所示：



图 10-25

- cron 表达式

cron 表达式是 Quartz 格式的。生成对应表达式，可根据生成器核对 <https://cron.qqe2.com/>



图 10-26

- 调度周期可选项，如下图所示：



图 10-27

1、分钟调度：即每天指定的时间段内，调度任务按 N*指定分钟的时间间隔运行一次。

2、小时调度：即每天指定的时间段内，调度任务按指定的时间间隔运行一次。例如，每天 01:00~04:00 的时间段内，每 1 小时运行一次。

时间周期根据左闭右闭原则计算。例如，配置调度任务在 0 点 3 点的时间段内，每隔 1 个小时运行一次。表示时间区间为 [00:00, 03:00]，间隔为 1 小时，调度系统每天将生成 4 个实例，分别在 0 点、1 点、2 点和 3 点运行。

3、日调度：即调度任务每天自动运行一次。新建周期任务时，默认的时间周期为每天 0 点运行一次。用户可以根据需要自行指定运行时间点。调度任务需要满足上游任务执行成功，并且已到定时时间两个条件，任务才能成功执行。任何一个条件没有满足都无法执行，两个条件没有先后顺序。

4、周调度：即调度任务每周的特定几天，在特定时间点自动运行一次。如果在没有被指定的日期时，为保证下游实例正常运行，系统会生成实例后直接设置为运行成功，而不会真正执行任何逻辑，也不会占用资源。

5、月调度：即调度任务在每月的特定几天，在特定时间点自动运行一次。如果在没有被指定的日期时，为保证下游实例正常运行，系统会每天生成实例后直接设置为运行成功，而不会真正执行任何逻辑，也不会占用资源。

10.3.4.4 调度依赖

在数据开发过程中，经常需要配置作业的上下游依赖任务，这样作业之间便会组成一个有向无环图（DAG），同时会配置作业的开始调度时间。

设置正确的调度依赖关系是构建有序业务流程的根本，可以保障业务数据有效、适时地产出，形成规范化的数据研发场景。规范的依赖关系配置后，无需关心当前节点依赖表由哪个节点产出，通过自动解析一键设置任务依赖。

依赖关系分为：自依赖、上游依赖、互斥依赖、~~下游依赖~~。只支持任务流中配置。

可以通过以下三种方式设置节点的依赖关系：

- 在业务流程面板连线形成依赖上下游关系。
- 自动解析配置形成依赖关系。
- 自定义依赖。

进入调度依赖页签，可设置自依赖、上游依赖、互斥依赖等，如下图所示：

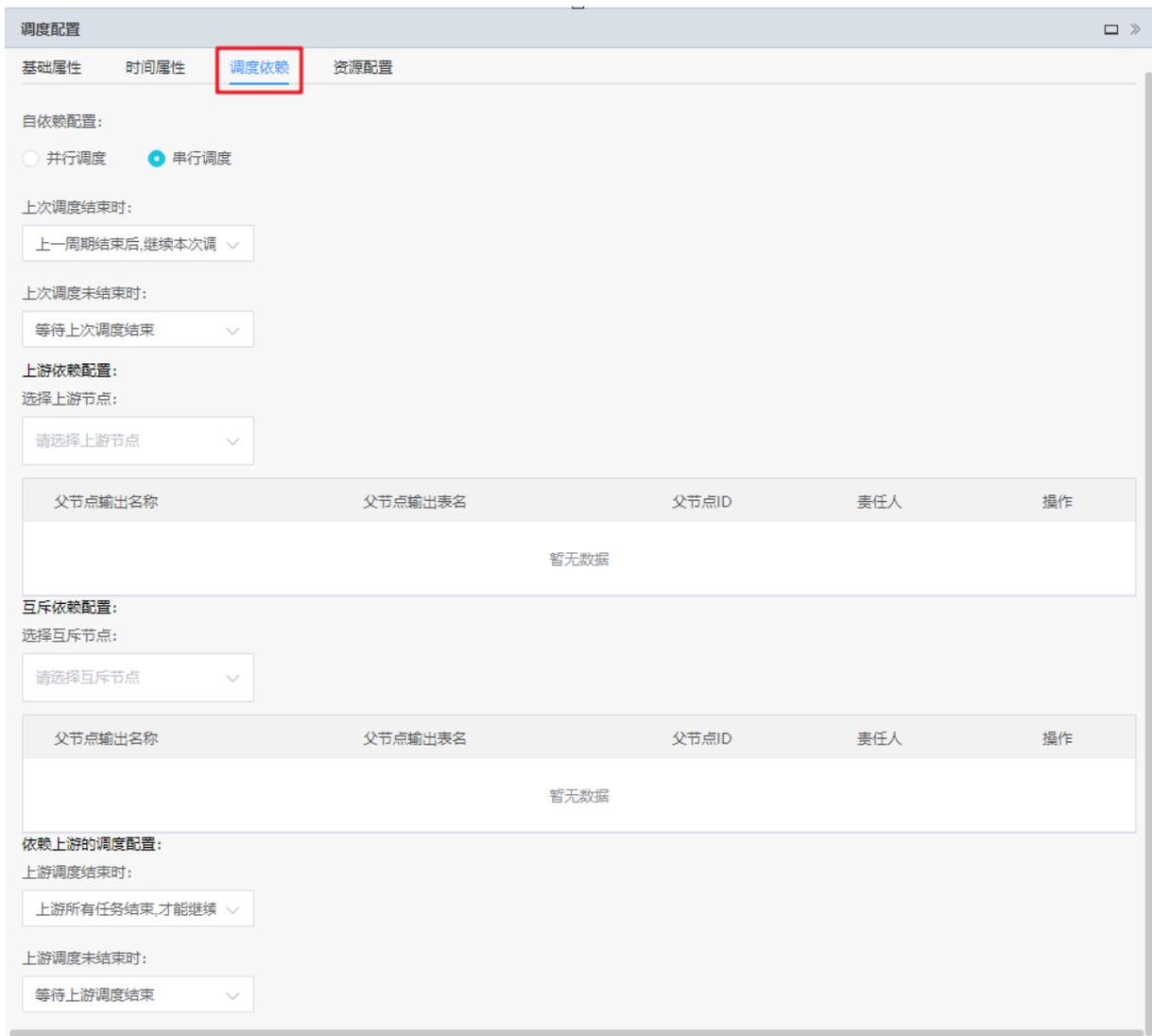


图 10-28

- 自依赖是根据自身调度周期，结合自依赖配置进行任务调度：

并行调度：节点任务根据调度周期，按周期运行

串联调度：节点任务根据调度周期运行时，会结合自上次调度的实例情况运行。共分以下情况：

- 上游依赖是强依赖上游节点的周期运行情况，结合上游依赖配置进行任务调度；
- 互斥依赖是强依赖互斥节点的运行状态，进行任务调度。互斥节点不能同时运行（即节点 A 和节点 B 互斥，则 A 运行中，B 取消运行本次调度，反之亦然。

备注：

- 1) 不支持同时配置相同节点为上游依赖和互斥依赖，即节点 A 不能同时成为节点 B 的上游节点和互斥节点。
- 2) 若节点同时存在自依赖、上游依赖、互斥依赖关系，则运行优先级为互斥依赖>自依赖>上游依赖。

10.3.4.1 资源配置

支持用户自定义节点在 spark 环境中计算时可用资源配置。

进入资源配置页签，设置对列名、内存大小、执行器数量、每个执行器内存、每个执行器核数，如下图所示：

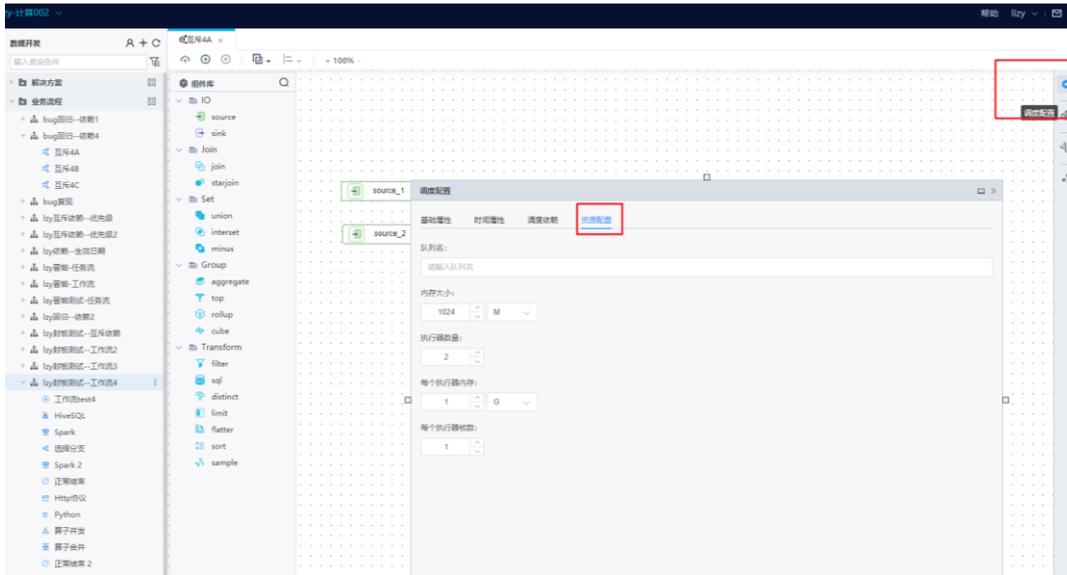


图 10-29

说明：只有在 spark 环境中计算的组件才需要配置，无需再 spark 环境中计算的组件配置不会生效。

10.3.5 节点依赖配置

步骤一：业务流程树中，选择某个业务流程并双击，进入业务流程看板，如下图所示：

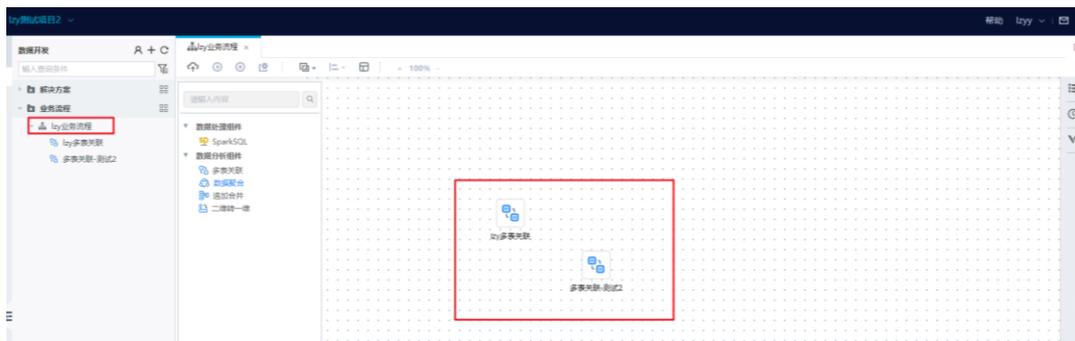


图 10-30

步骤二：业务流程面板中，选择节点 1 到节点 2，进行连线，如下图所示：

注意：节点间连线存在顺序关系，连线时将光标放在图标右侧的连接点上，光标变为十字形。



图 10-31

步骤三：查看节点血缘关系。

方式一：进入步骤二中，节点 1 的看板，检点看板右侧，点击血缘关系图标“🔗”，查看依赖视图，如下图所示：

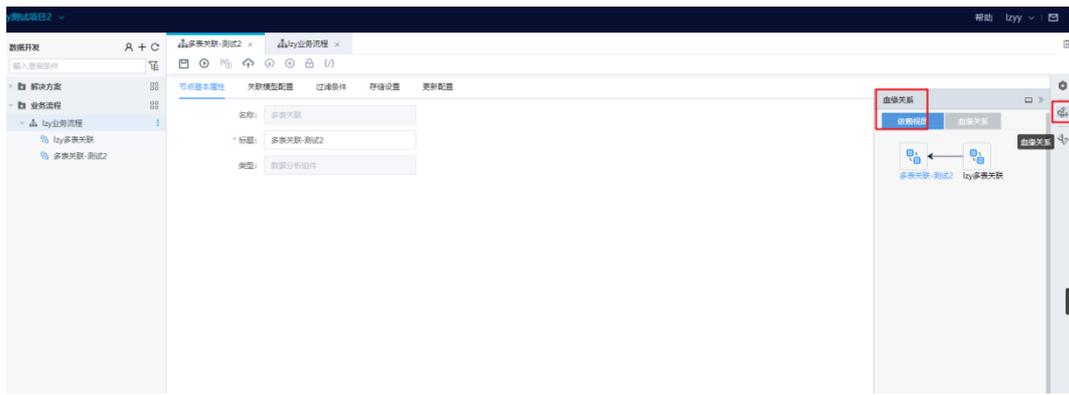


图 10-32

方式二：业务流程面板中，选择节点1，鼠标右键点击查看节点血缘关系，查看该节点血缘关系，会弹出“血缘关系”，如下图所示：

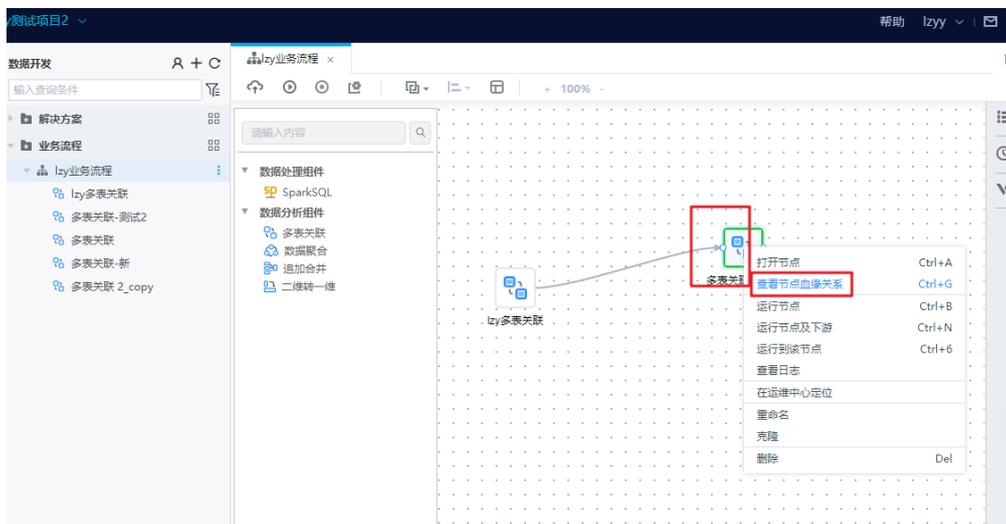


图 10-33

10.3.6 配置流程参数

当整个业务流程需要对同一个变量进行统一的赋值或替换参数值时，使用流程参数功能。

配置步骤：在流程树中，选择某个业务流程，业务流程看板右侧，点击业务流程图标“”，弹出框“流程参数”中，设置参数名称、参数值或表达式，如下截图所示：

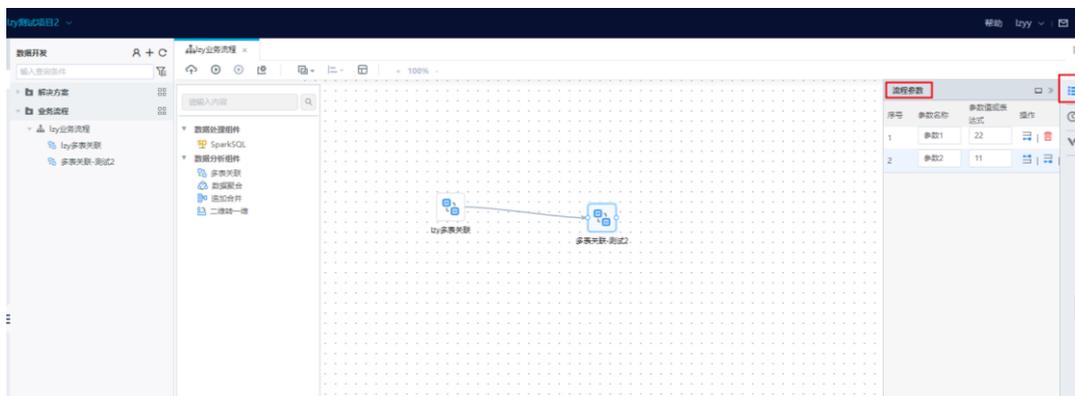


图 10-34

说明：如果业务流程配置了参数、业务流程中的节点也配置了参数，以业务流程参数为主。具体配置操作参考章节错误！未找到引用源。错误！未找到引用源。。

10.3.7 提交节点任务

支持提交单一节点任务和多节点任务。

方式一：提交单节点任务。配置完节点后，点击提交图标“”，如下图所示：

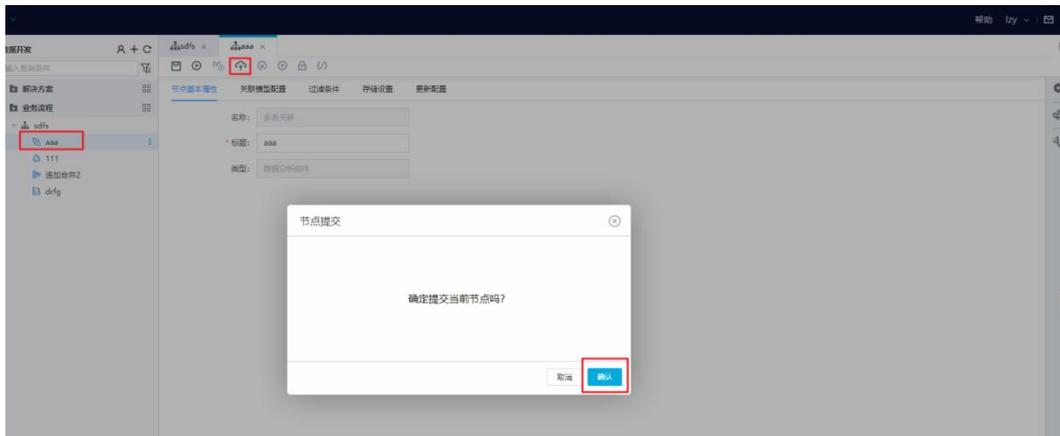


图 10-35

点击确定，进入【运维中心】->【手动周期任务】查看节点调度任务和实例。

方式二：批量提交节点任务。配置业务流程的节点后，进入业务流程看板，点击提交图标“”，如下图所示：

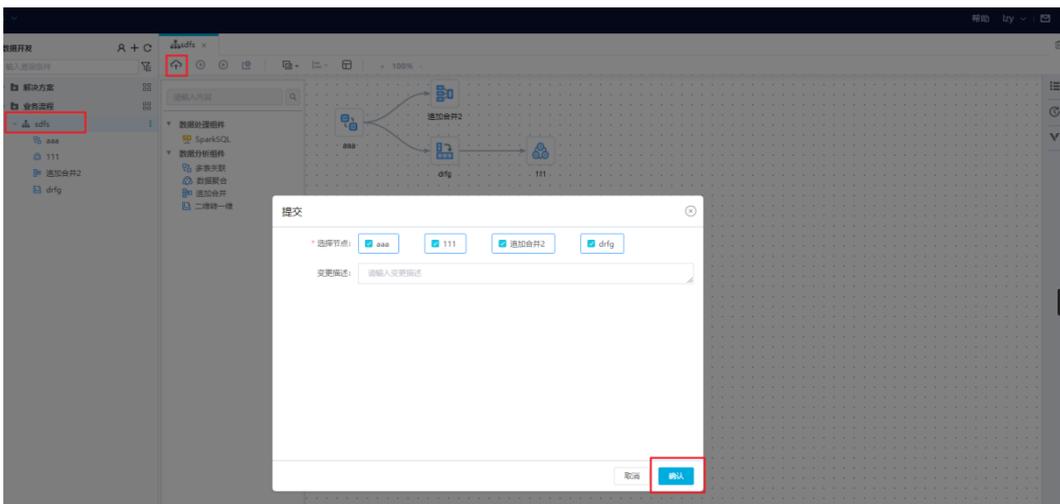


图 10-36

勾选需要提交的节点，点击确定，进入运维中心查看节点调度任务和实例。

说明：若该节点进行了存储设置，则提交节点后会根据存储设置中配置的表信息，直接下发到物理库生成物理表，无需审批。

10.4 数据处理组件

10.4.1 SparkSQL

10.4.1.1 节点基本属性

业务流程树中，新增 SparkSQL 组件，可设置节点名称，如下图所示：

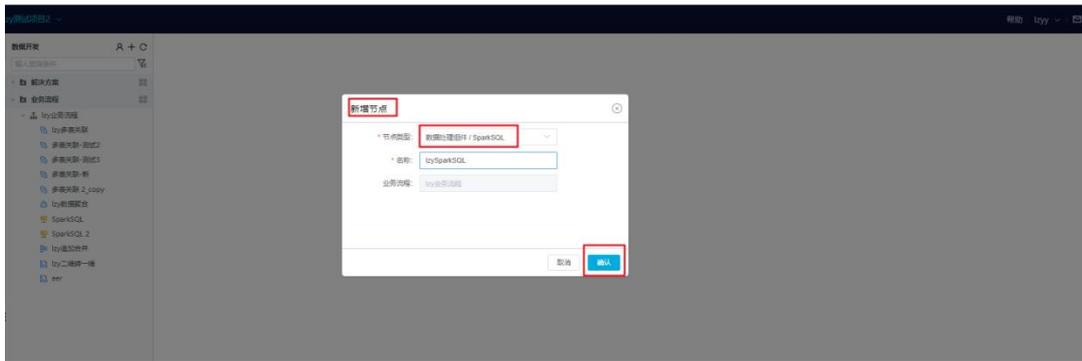


图 10-37

10.4.1.2 节点详情

业务流程树中，双击 SparkSQL 节点，进入该节点详情，如下图所示：

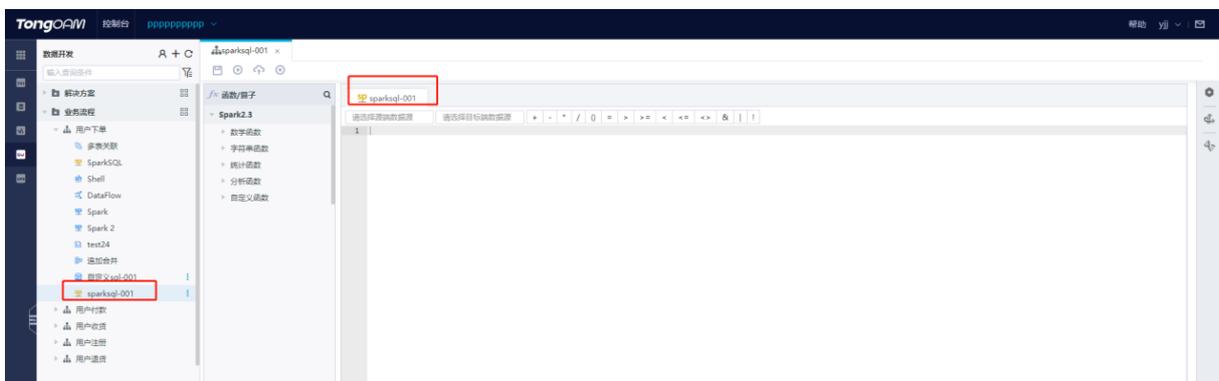


图 10-38

选择数据源，包源端以及目标端，进行 SQL 语句编写。

注意：SparkSql 的源表和目标表的字段名称不一致的话要通过 as 修改名称让其一致，这样才能建立字段的映射关系。

10.4.2 Shell

10.4.2.1 节点基本属性

业务流程树中，新增 Shell 组件，可设置节点名称，如下图所示：

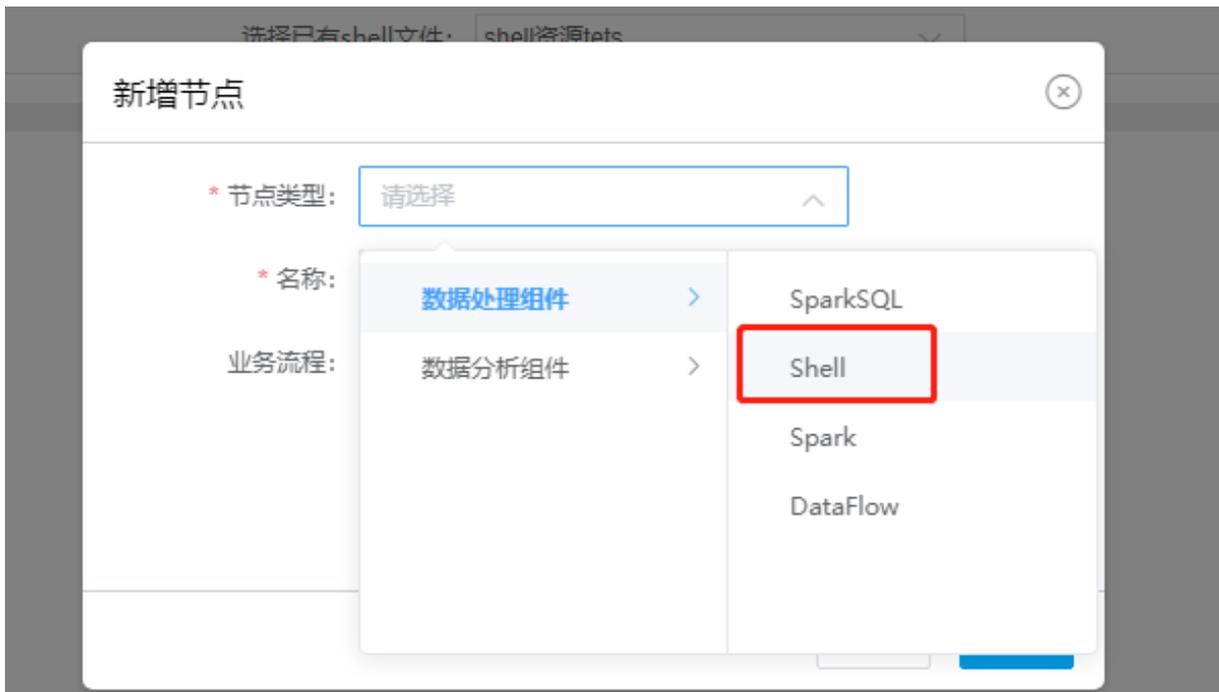


图 10-39

10.4.2.2 节点详情

业务流程树中，双击 Shell 节点，进入该节点详情，可以选择手动输入或者选择已有的 shell 如下图所示：

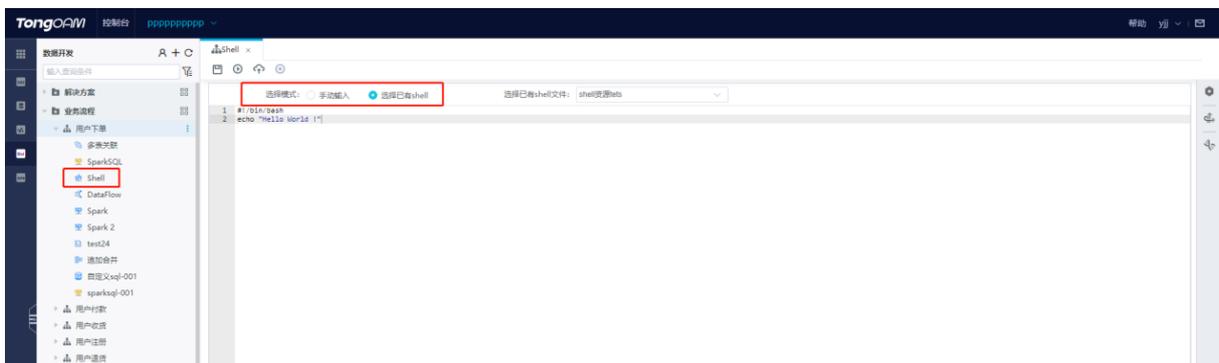


图 10-40

【手动输入】：可以按照实际需求输入 shell 脚本。

【选择已有 shell】：需要先将已经写好的 shell 脚本在【资源管理】进行上传，之后在此处可以选择，选择后操作台会显示已选择的脚本内容。

10.4.3 Spark

10.4.3.1 节点基本属性

业务流程树中，新增 Spark 组件，可设置节点名称，如下图所示：



图 10-41

10.4.3.2 节点详情

业务流程树中，双击 Spark 节点，进入该节点详情，可以设置：选择 spark 文件、选择 MainClass 文件、配置参数，如下图所示：

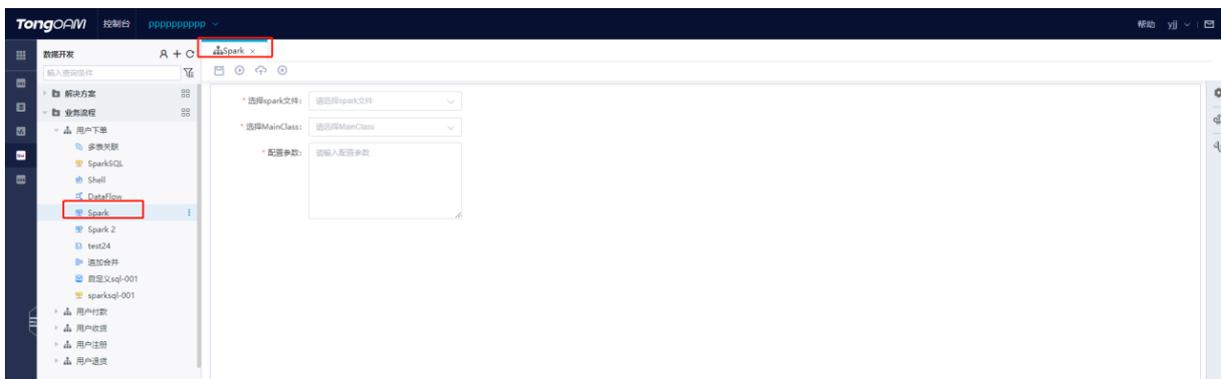


图 10-42

【选择 spark 文件】：需要先将已经写好的 spark 脚本在【资源管理】进行上传，之后在此处可以选择。

【选择 MainClass】：上传 spark 脚本之后，在此处可以选择对应的类。

【配置参数】：根据不同的类去填写。

10.4.4 DataFlow

10.4.4.1 节点基本属性

业务流程树中，新增 DataFlow 组件，可设置节点名称，如下图所示：



图 10-43

新建完成后，双击节点，可以进入选择组件进行操作，组件库的“”可以进行组件的查询，如下图：

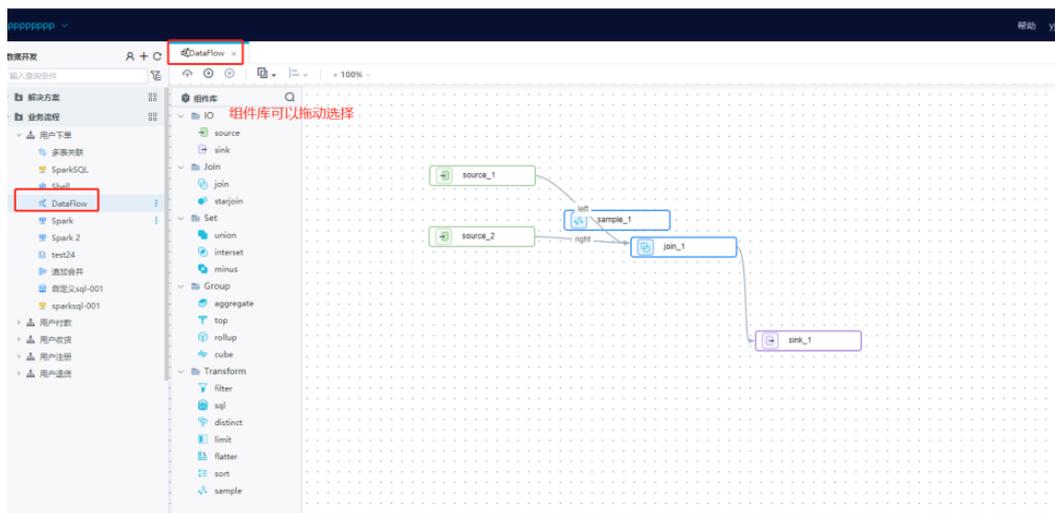


图 10-44

10.4.4.2 组件 source

Source：源库表，dataflow 的必要组件，也是开始组件。在操作台双击 source 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，参数（数据源，增量表示，时间基准，时间周期），输出（待选字段，已选字段）。具体如下图：



图 10-45

【节点 ID】：节点生成时自动生成的，不可修改。

【节点名称】：节点名称，可以自行修改，修改后操作台的名称也随之修改，如下图：

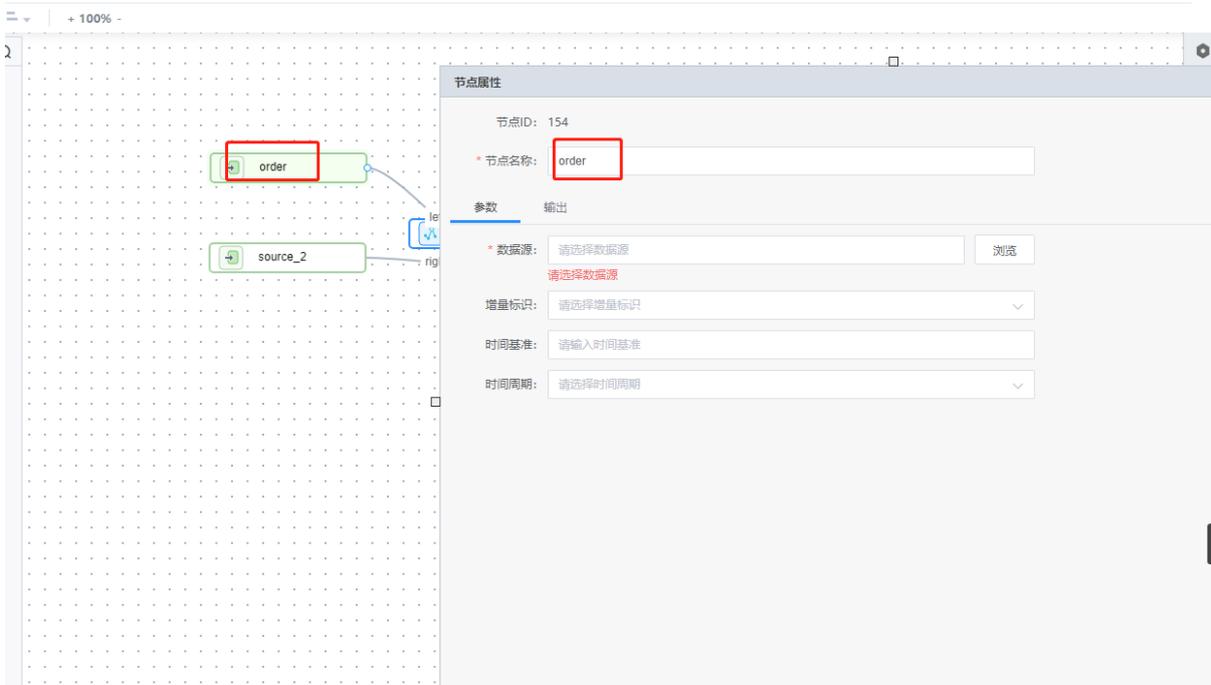


图 10-46

【数据源】：点击浏览可以进入选择数据源页面，有刷新树以及折叠树按钮，同时还可以输入关键字进行过滤，如下图：

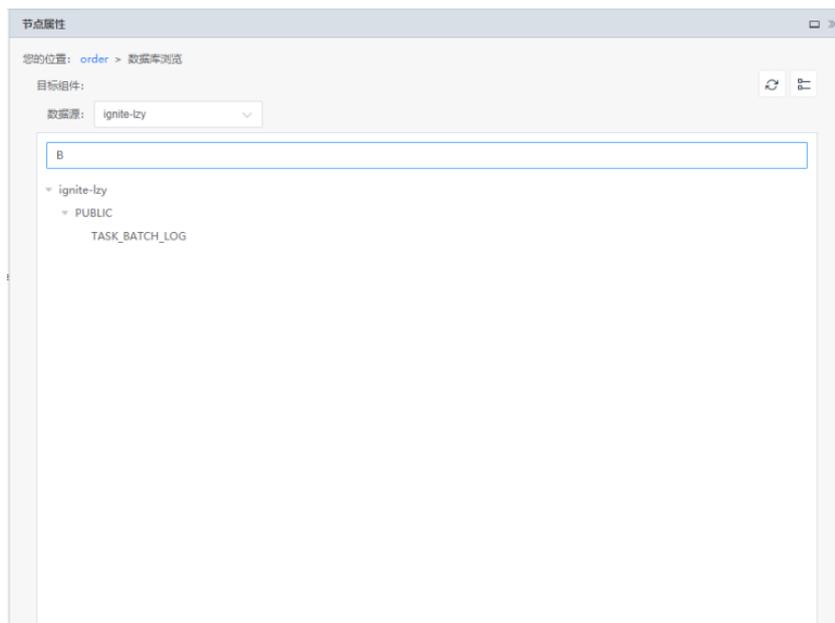


图 10-47

在鼠标放置对应的源表上时，会出现预览前 100 行及更多按钮，可以根据需要进行预览查看，如下图：

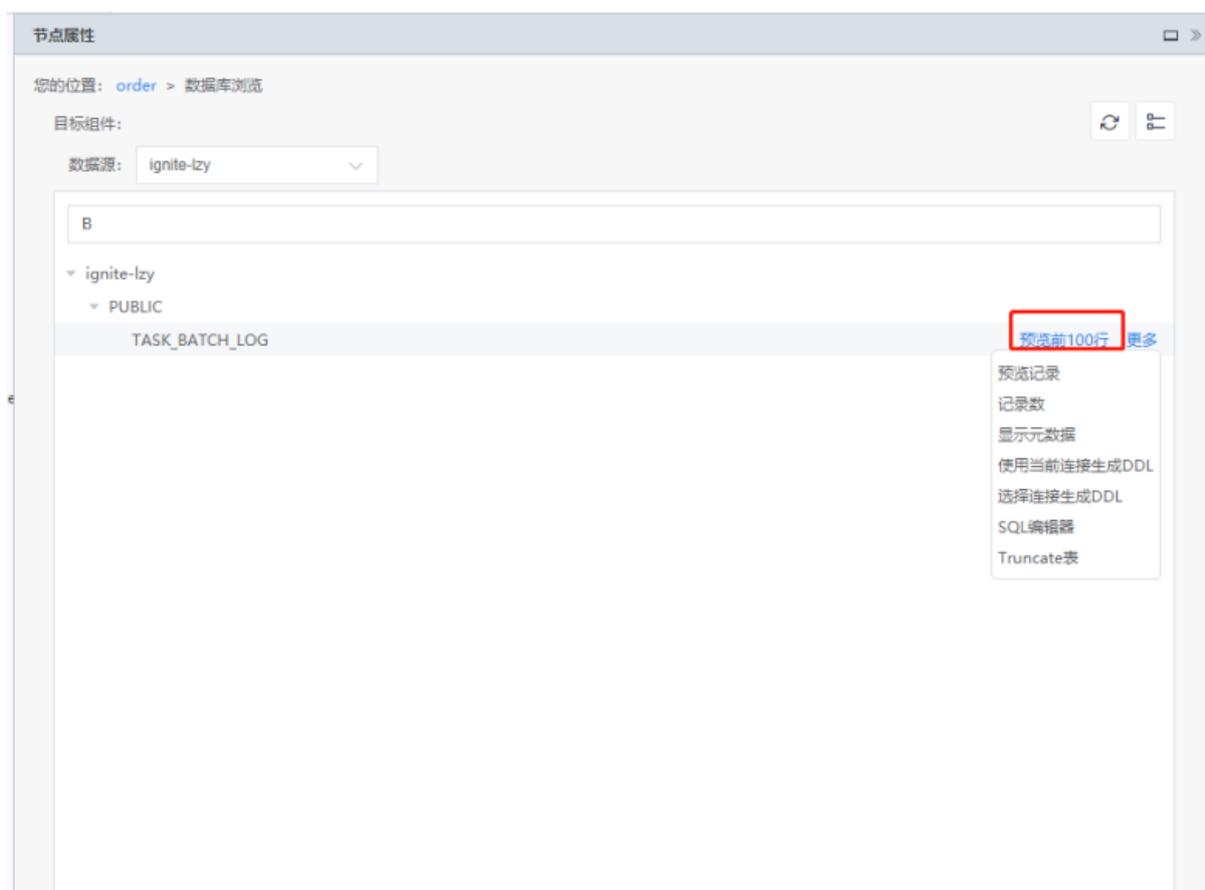


图 10-48

选择完成数据源表之后，在组件属性页面可以看到数据源字段显示选中的表。

【增量标识】：

【时间基准】：

【时间周期】：选择时间周期

【待选字段】：源表待选字段，可以选中单个或多个，移动到已选字段，也可以全部移动

【已选字段】：在待选字段选择移动后，可以展示在已选字段中，具体如下图：

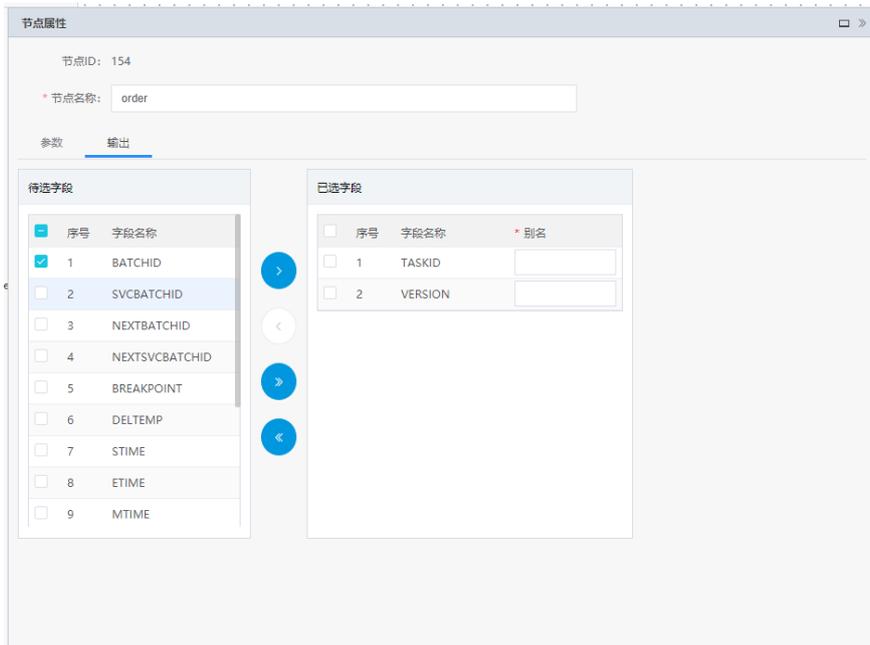


图 10-49

10.4.4.3 组件 sink

Sink：目标库表，dataflow 的必要组件，也是结束组件。在操作台双击 sink 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，参数（数据源，加载方式），输出（待选字段，已选字段），如下图：

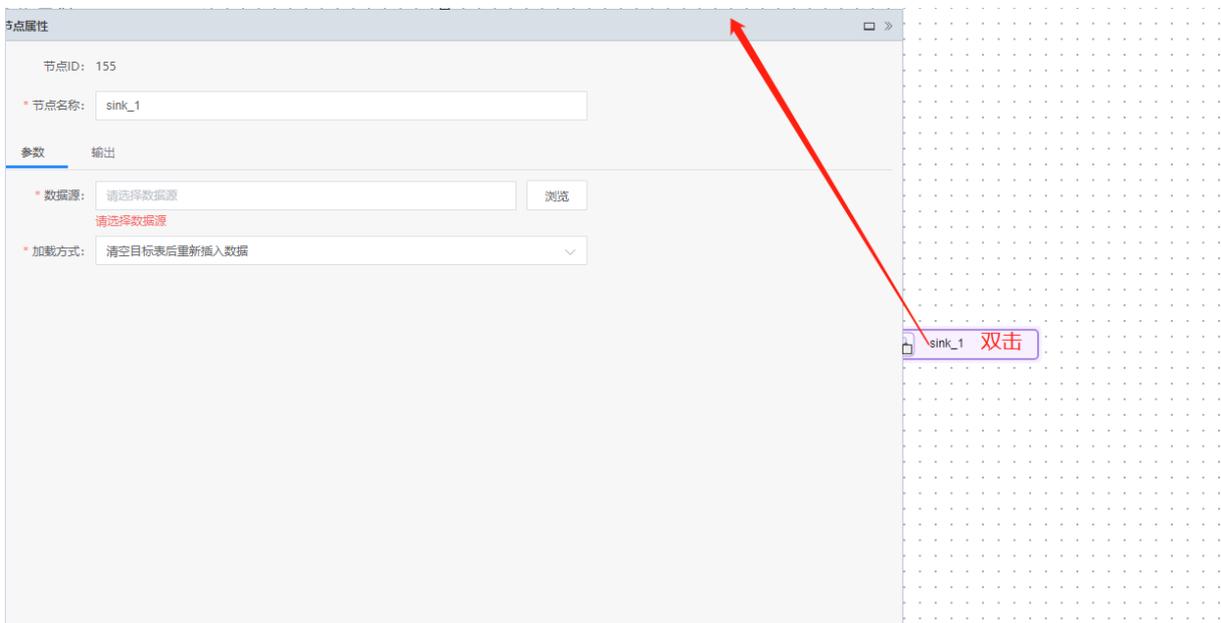


图 10-50

【节点 ID】：节点生成时自动产生的，不可以修改

【节点名称】：设置节点名称

【数据源】：选择目标数据源，与源端数据源选择相同

【加载方式】：三种方式：清空目标表后重新插入数据，根据业务主键数据插入或更新，数据

全部执行插入操作

【待选字段】：同 source 组件的待选字段

【已选字段】：同 source 组件的已选字段

10.4.4.4 组件 join

Join 组件：两表多字段等值关联，在操作台双击 join 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，left(source)，right(source)，关联方式，关联条件，待选字段，已选字段，如下图：

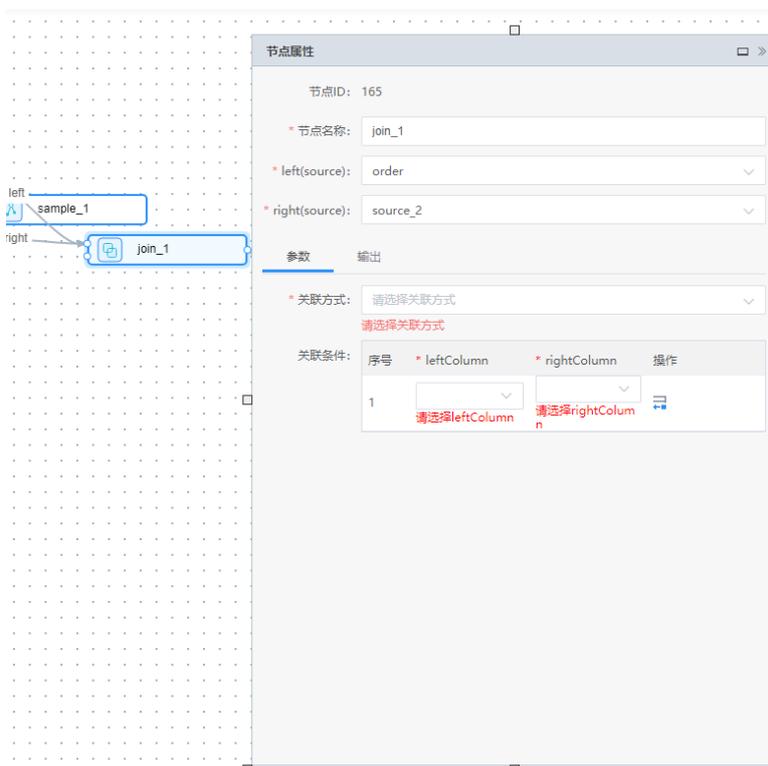
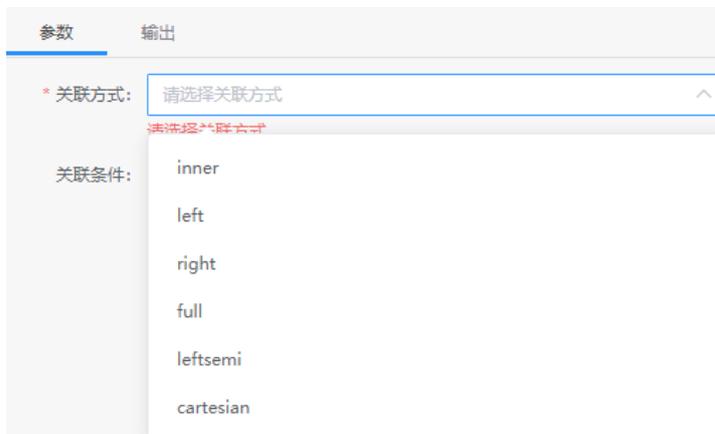


图 10-51

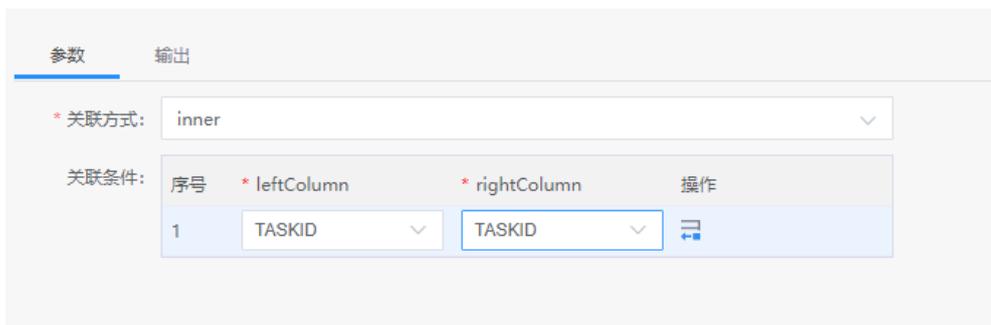
【left(source)】：源端左连接的 source 组件，可以更换

【right(source)】：源端右连接的 source 组件，可以更换

【关联方式】：inner, left, right, full, leftsemi, cartesian



【关联条件】：可配置关联字段，如下图：



10.4.4.5 组件 sql

Sql 组件：自定义 sparksql，仅支持 select，将 sql 组件从组件库拖到操作台之后，在操作台双击 sql 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，sql，待选字段，已选字段，如下图：

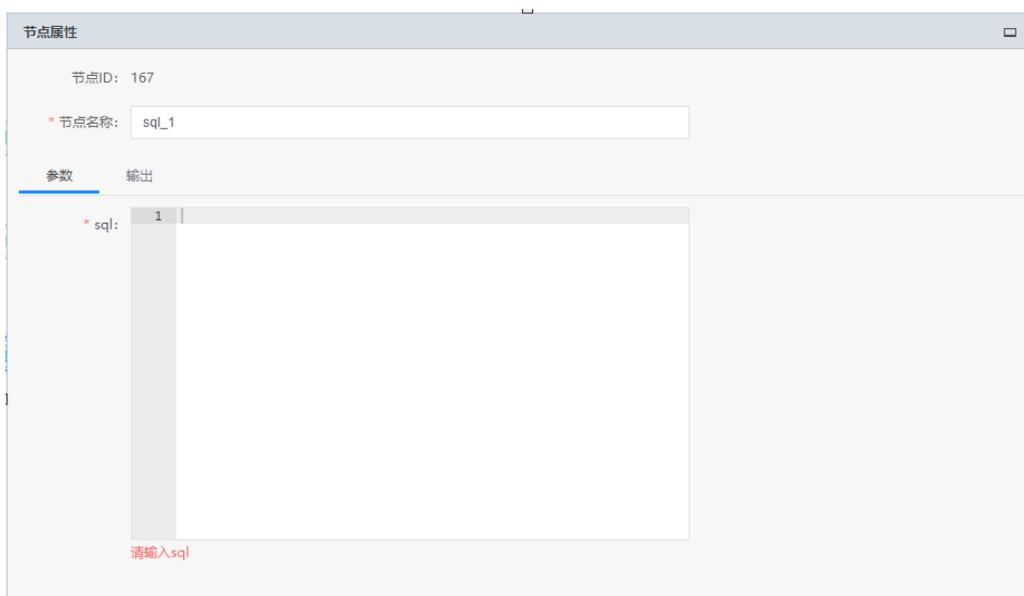
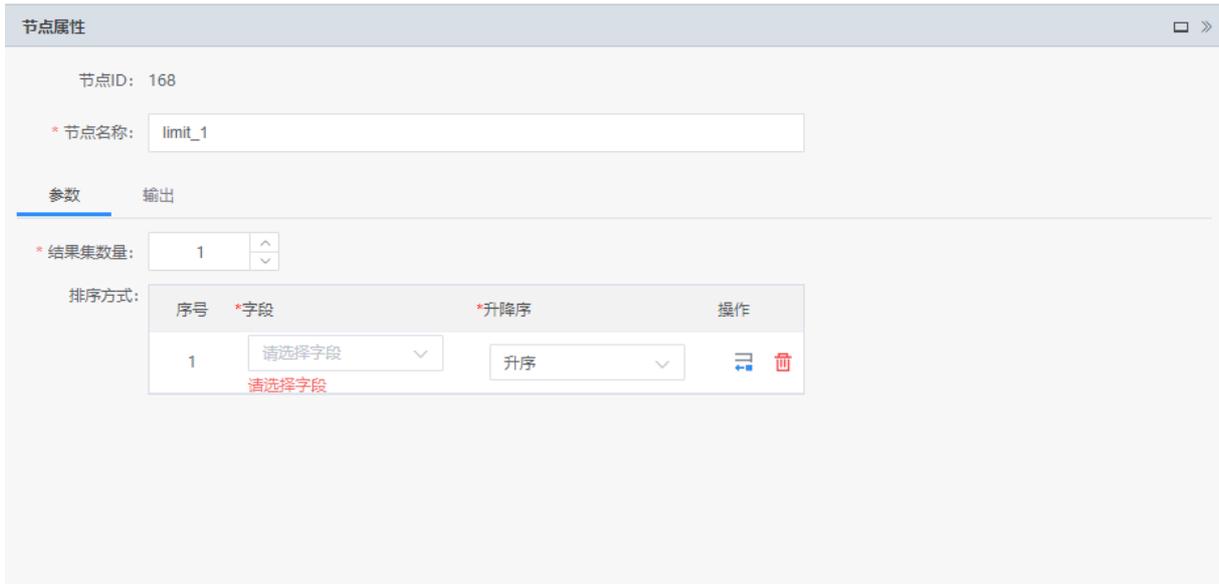


图 10-52

【sql】：根据实际需要填写 sql，目前仅支持 select

10.4.4.6 组件 limit

limit 组件：排序数据集取前若干行数据，将 limit 组件从组件库拖到操作台之后，在操作台双击 limit 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，结果集数量，排序方式，待选字段，已选字段，如下图：



【结果集数量】：取结果集前多少行的数据，例如 1，就是前 1 的数据，即仅取一行

【排序方式】：将结果集按照某一个字段进行排序

10.4.4.7 组件 starjoin

starjoin 组件：将事实表与多张维度表之间进行 join，将 starjoin 组件从组件库拖到操作台之后，在操作台双击 starjoin 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，事实表，关联条件，待选字段，已选字段，如下图：



图 10-53

【事实表】：选择需要关联的事实表

【关联条件】选择维表以及维表字段，与事实表的字段进行关联

10.4.4.8 组件 aggregate

aggregate 组件：聚合运算，将 aggregate 组件从组件库拖到操作台之后，在操作台双击 aggregate 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，分组条件，关联条件，如下图：

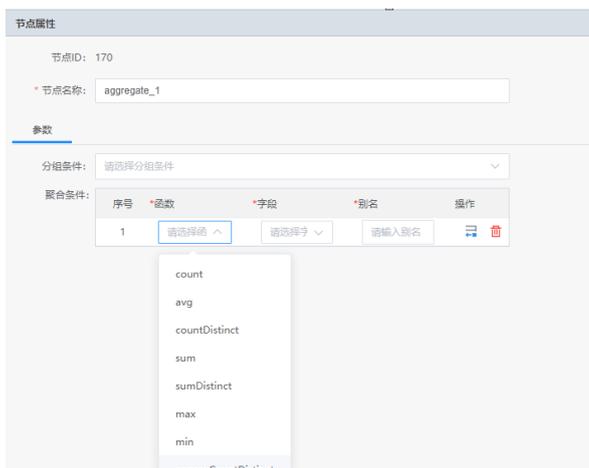


图 10-54

【分组条件】：

【聚合条件】：有 count, avg, countDistinct, sum, sunDistinct, max, min, approxDistinct 函数，选择函数后选择字段以及别名即可

10.4.4.9 组件 union

union 组件：求集合的并集，将 union 组件从组件库拖到操作台之后，在操作台双击 union 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，输入，待选字段，已选字段，如下图：

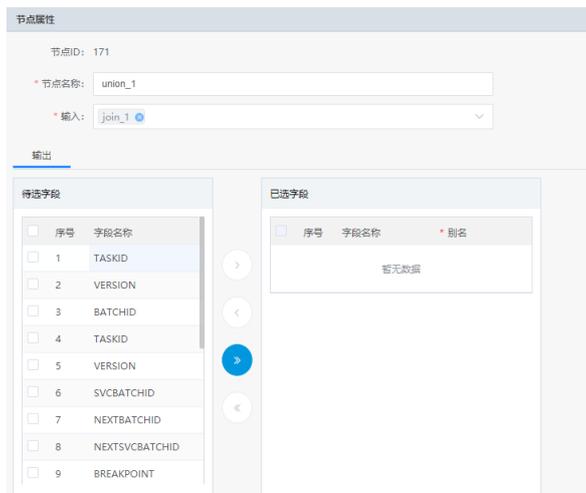


图 10-55

【输入】：选择需要求并集的组件

10.4.4.10 组件 filter

filter 组件：过滤数据集，将 filter 组件从组件库拖到操作台之后，在操作台双击 filter 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，表达式，待选字段，已选字段，如下图：

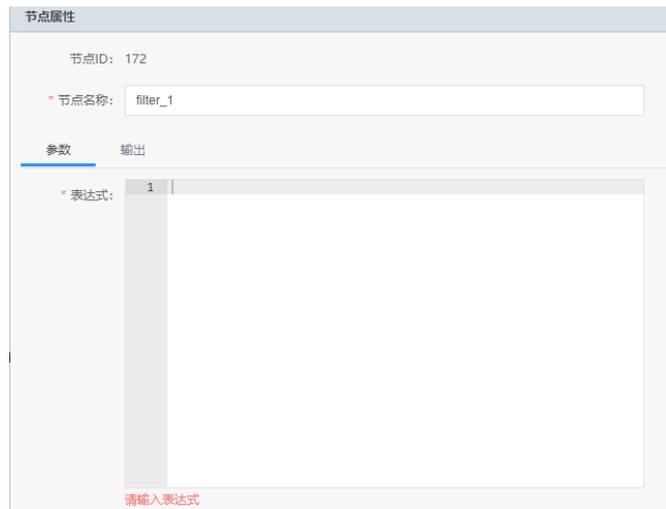


图 10-56

【表达式】：

10.4.4.11 组件 top

top 组件：基于分组内排序数据集，并取每组前若干行数据，将 top 组件从组件库拖到操作台之后，在操作台双击 top 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，抽取方式，抽取比例，分组条件，排序方式，待选字段，已选字段，如下图：



图 10-57

【抽取方式】 rowcount, rate

【抽球比例】：按照比例抽取

【分组条件】：根据分组情况，每组抽取数据

【排序方式】：根据设置的字段以及排序规则进行排序

10.4.4.12 组件 interset

interset 组件：基于两个集合的交集，类似 union，区别是取交集 $A \cap B$ ，将 interset 组件从组件库拖到操作台之后，在操作台双击 interset 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，输出（待选字段，已选字段），如下图：

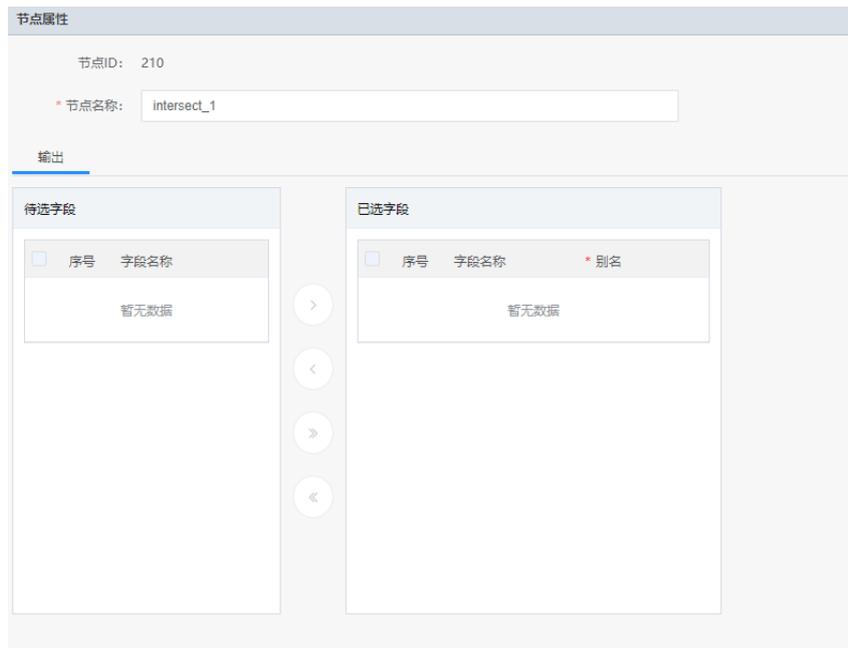


图 10-58

10.4.4.13 组件 minus

minus 组件：基于两个集合的交集，类似 union，区别是取差集 A-B。将 minus 组件从组件库拖到操作台之后，在操作台双击 minus 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，输出（待选字段，已选字段），如下图：



图 10-59

10.4.4.14 组件 rollup

rollup 组件：分组分析类算子，向上卷积，例如： $rollup(a,b)=group\ by(a,b)+group\ by(a)+group\ by(all)$ 。将 rollup 组件从组件库拖到操作台之后，在操作台双击 rollup 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，参数（分组条件，聚合条件），如下图：



图 10-60

【分组条件】根据数据选择需要分组的字段

【聚合条件】：可以设置单个也可以设置多个，可选的函数有：avg,sum,max,min,mean. 字段就是需要进行聚合的字段，别名可以根据需要进行设置。这里的函数可以单个多个按照需要进行聚合计算。

10.4.4.15 组件 cube

cube 组件：分组分析类算子，多维度卷积，例如： $cube(a,b)=rollup(a,b)+rollup(b,a)$ 。将 cube 组件从组件库拖到操作台之后，在操作台双击 cube 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，参数（分组条件，聚合条件），如下图：



图 10-61

【分组条件】根据数据选择需要分组的字段，可以选择多维度

【聚合条件】：可以设置单个也可以设置多个，可选的函数有：avg,sum,max,min,mean. 字段就是需要进行聚合的字段，别名可以根据需要进行设置。这里的函数可以单个多个按照需要进行聚合计算。

10.4.4.16 组件 distinct

distinct 组件：分组分析类算子，数据集去重，将 distinct 组件从组件库拖到操作台之后，在操作台双击 distinct 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，输出（待

选字段，已选字段），如下图：

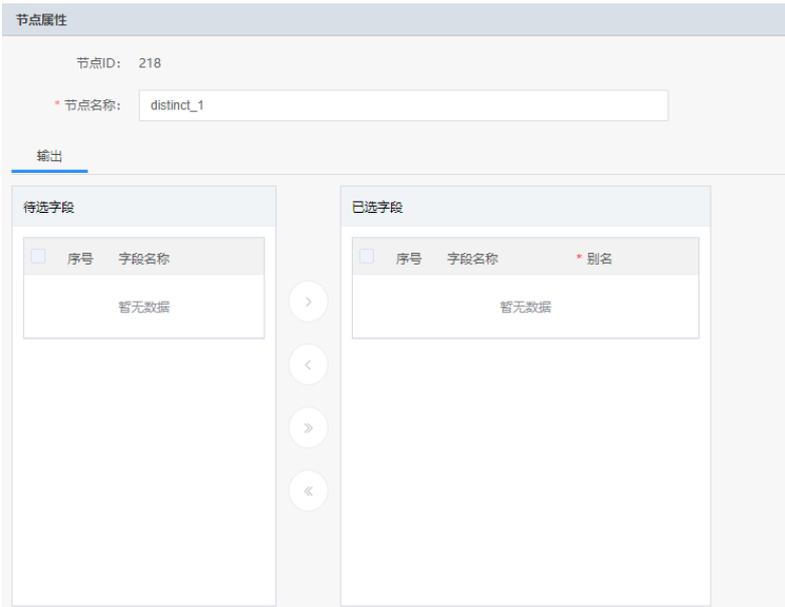


图 10-62

注意：这里的去重是根据 source 传递过来的数据字段进行去重，即如果按照 source 传递过来的字段为两个，那么在源表中的这两个字段完全一致的，才算重复的。

10.4.4.17 组件 flutter

flutter 组件：分组分析类算子，列转行（二维转一维），flutter 组件从组件库拖到操作台之后，在操作台双击 flutter 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，参数（待选字段，已选字段），输出（待选字段，已选字段），新增字段，如下图：



图 10-63

【参数】根据转换的需要，选择转换字段设置



图 10-64

【输出】根据实际的需要，选择输出字段设置

【新增字段设置】根据转换的需要，输入转换字段名称

10.4.4.18 组件 sort

sort 组件：分组分析类算子，按照选择的字段及排序规则对数据集进行排序，sort 组件从组件库拖到操作台之后，在操作台双击 sort 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，参数（排序方式），输出（待选字段，已选字段），新增字段，如下图：



图 10-65

【参数】根据排序的需要，选择排序的字段以及排序的规则（升降序）



图 10-66

【输出】根据实际的需要，选择输出字段设置

10.4.4.19 组件 sample

sample 组件：分组分析类算子，按照选择的字段及排序规则对数据集进行排序，sample 组件从组件库拖到操作台之后，在操作台双击 sample 组件，可以打开组件属性窗口，可以设置：节点 ID，节点名称，参数（采样比例，是否启用替换采样），输出（待选字段，已选字段），新增字段，如下图：

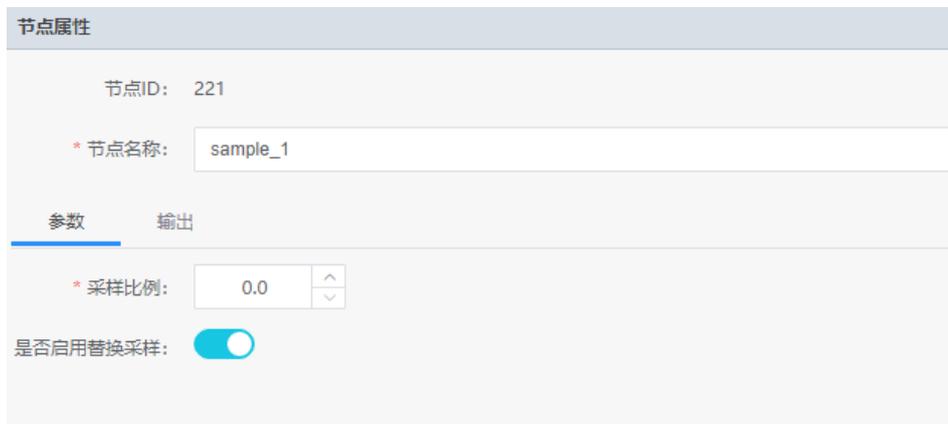


图 10-67

【参数】根据采样的需要，输入采样比例，是否启用替换采样（是：抽样后放回去再采，否：无放回）



图 10-68

10.4.4.20 组件关联

业务流程树中，双击 DataFlow 节点，进入该节点详情，从左侧组件库中拖拽对应的组件到节点面板中，如下图所示：

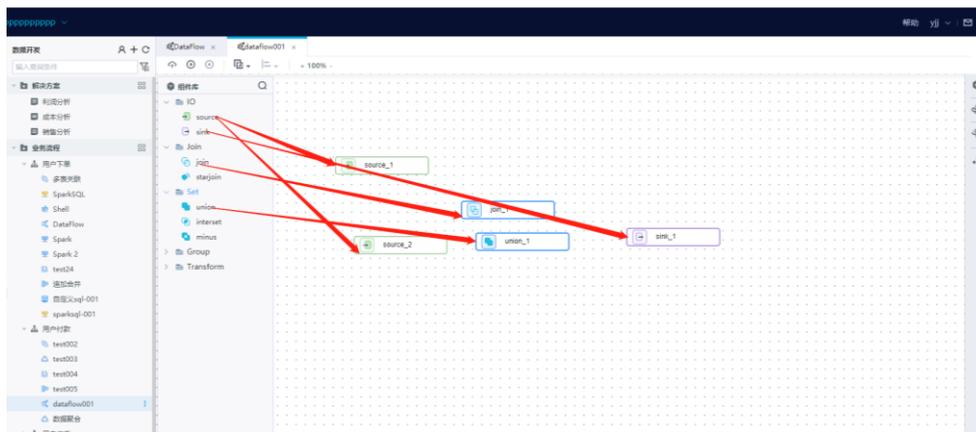


图 10-69

节点面板中，选择对应的组件，进行连线，如下图所示：

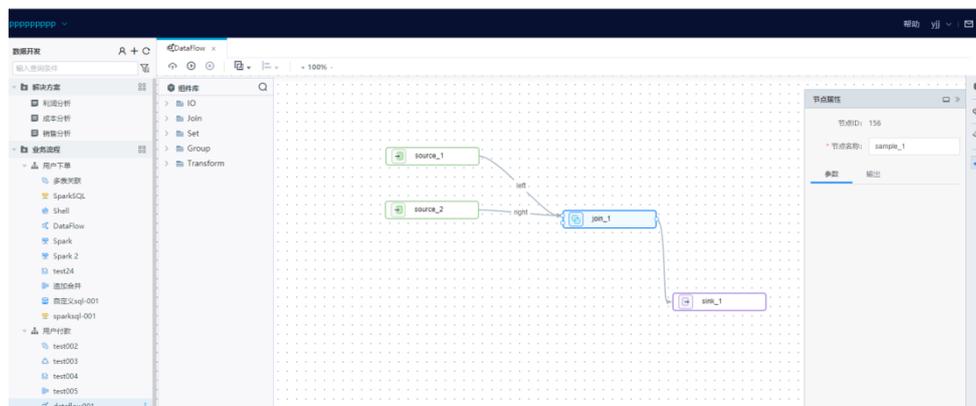


图 10-70

双击组件可以展示组件的基本属性，展示组件属性，下图以 join 组件为例：

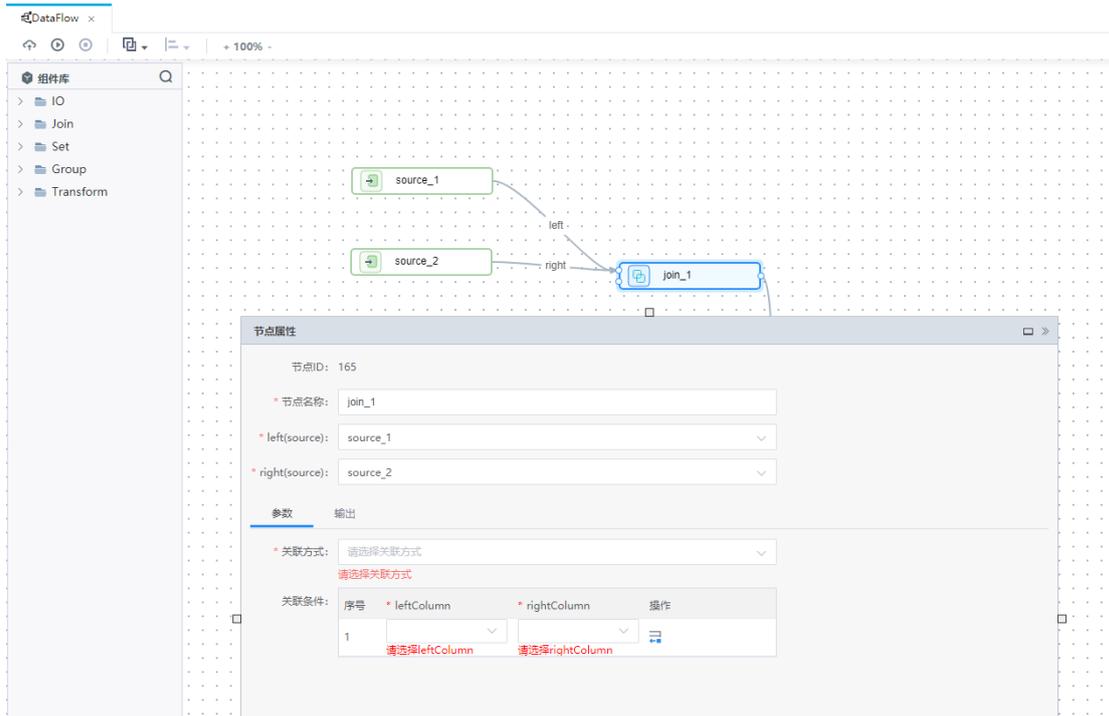


图 10-71

11. 资源管理

11.1 项目表

创建模型主要包含创建事实模型、视图模型、维度模型三种模型，事实模型相当于创建一个物理表。视图模型相当于创建一个视图。维度模型相当于创建一个与其他维度之间存在关联的物理表。上线后的模型会在已选择的数据源物理库中创建相应的物理表/视图。

维度模型主要储存公共数据，作为分析的对象；事实模型关联维度相关的明细，通过事实模型多关联维度模型，达到多维分析的目的，事实模型会存储关联维度表的主键，分析的时候，维度作为对象，去过滤事实表的数据。

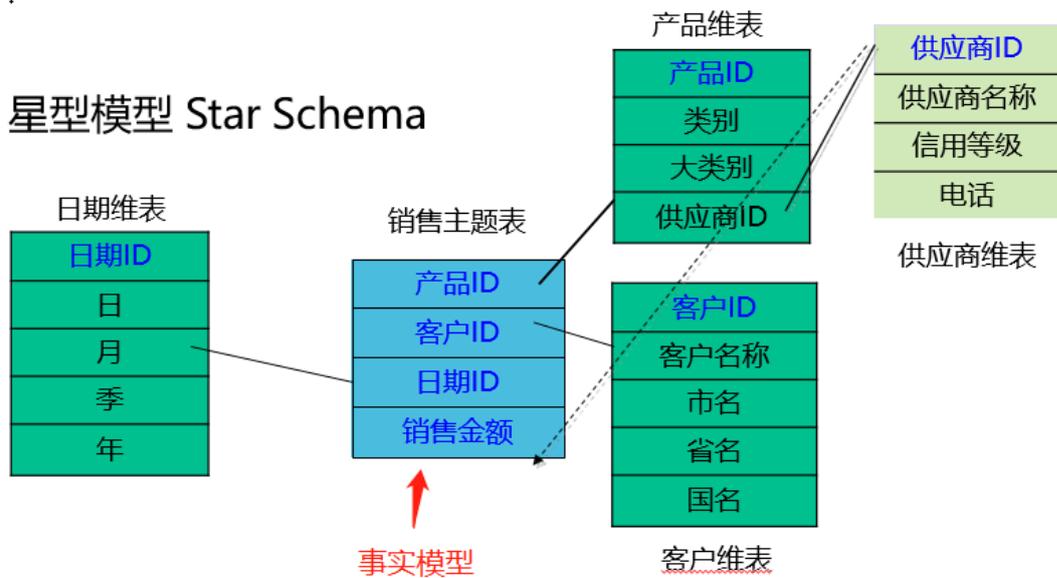


图 11-1

模型定义为两种类型，分别为表和视图。表和视图的模型设计理念和流程基本一致，在创建的过程中会有各自属性定义的差异。比如，视图模型设计更关注多表的关联查询，而表模型设计更关注自身的结构。

字段释义：

表名：用户在进行模型创建的时候，需按照配置中心的定义规则进行表命名；

表描述：用户在进行表设计时，可输入表的描述字段，最长 200 字符；

数据源名称：用户在进行表设计时，需选择模型所属的数据源名称；

数据库类型：用户在进行表设计时，需选择模型所属的数据库类型；

模型层级：用户在进行模型创建的时候，需按照配置中心的定义模型层级进行表的层级归属选择；

主题域：用户在进行模型创建的时候，需按照配置中心的定义主题域进行表的主题域归属选择。

生命周期：定义表的生命周期，可手动选择常用的生命周期(365、999)，也可自定义表的生命周期，超过定义的生命周期后，模型会被自动删除。

负责人：模型创建人；

最后修改时间：模型最后被修改的日期，需要显示到秒级；

最新版本：最新经过上线审批的模型版本；

线上版本：线上正在启用的版本；（暂定线上只能存在一个有效版本）

状态：此处获许模型的最新线上状态，如有线上正在使用的版本，状态为启用，反之为停用。

11.1.1 创建事实模型

创建步骤分两步，第一步模型基本信息填写：模型设计页面中，点击“创建”选择“创建事实模型”，添加基本信息如图，点击下一步即可。



图 11-2

第二步字段信息设置：主要作用为模型选取字段、设置索引、列簇设置。字段选取可通过数据库表、手工录入、数据仓库表、DDL 导入、原子指标、衍生指标、视图、模板导入的方式，导入演示如下图。

字段信息添加完成点击“提交”即可创建成功如图数据库表、数据仓库表可直接选择库中字段导入。



图 11-5

可手动输入 sql 语句创建字段。



图 11-6

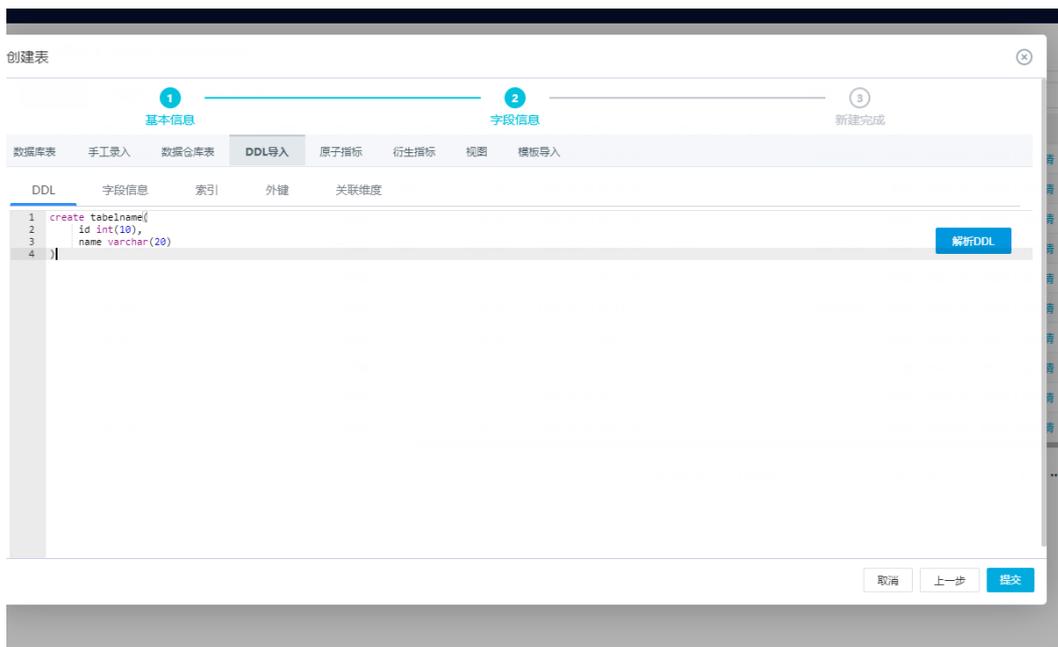


图 11-7

11.1.2 创建视图模型

创建步骤分两步，第一步模型基本信息填写：模型设计页面中，点击“创建”选择“创建事实模型”。



图 11-8

基本信息页中添加视图信息如数据源、schema、所属类目、生命周期等，点击下一步即可。

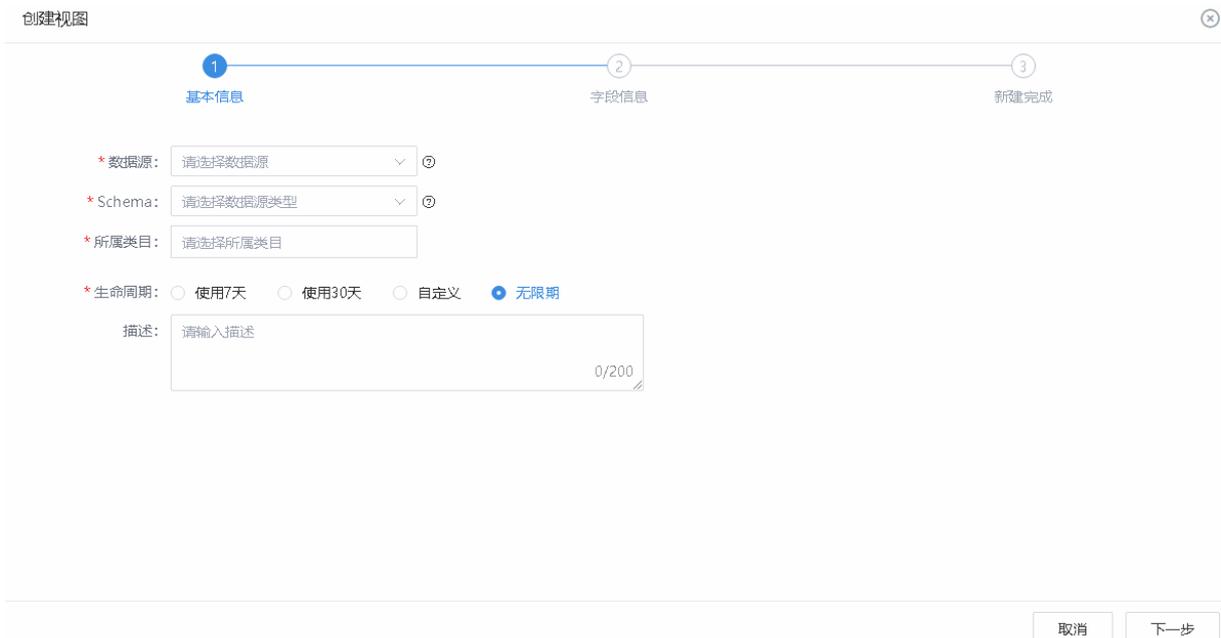


图 11-9

第二步为视图模型选择关联表：表选择可以选择从数据库表、数据仓库表两种方式获取，选择要关联的表字段即可。

注：数据库代表物理库中所有的表，数据仓库表显示所有数据仓库已上线的模型，依赖元数据的自动采集，元数据那边一小时更新一次，也可以手动立即更新。

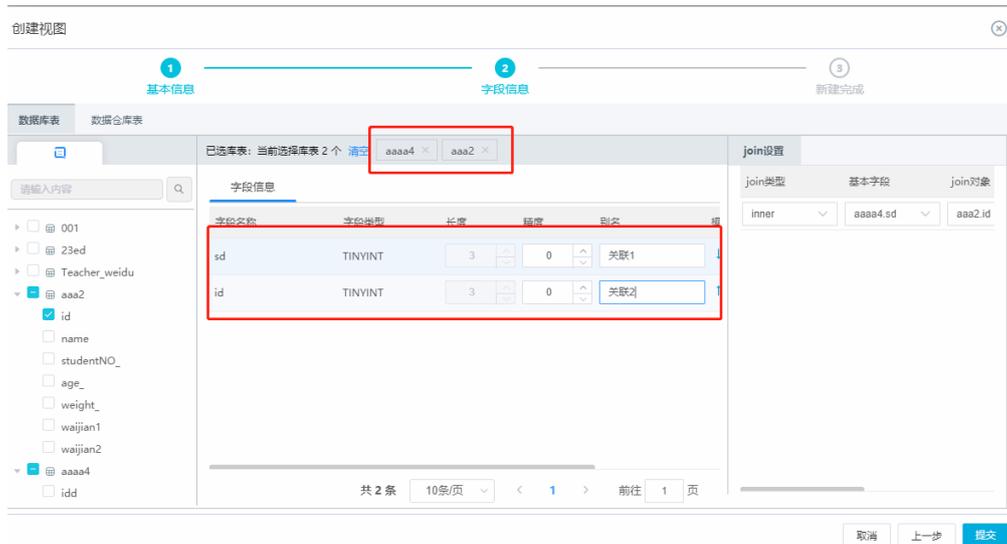


图 11-10

第三步为视图模型添加 join 设置：join 设置支持四种 join 类型，inner、left outer、right outer、full outer 选取不同表 join 方式，添加关联字段选取要 join 的对象如下图，点击提交即可完成视图模型创建。

以下图为例 test_a 与 test_b 进行 inner join，对应 sql 为，

```
SELECT test_a.aid,test_c.aid FROM test_a INNER JOIN test_c ON test_a.aid=test_c.cid
```

注：mysql 数据源暂不支持 full outer、ignite 数据源不支持 join

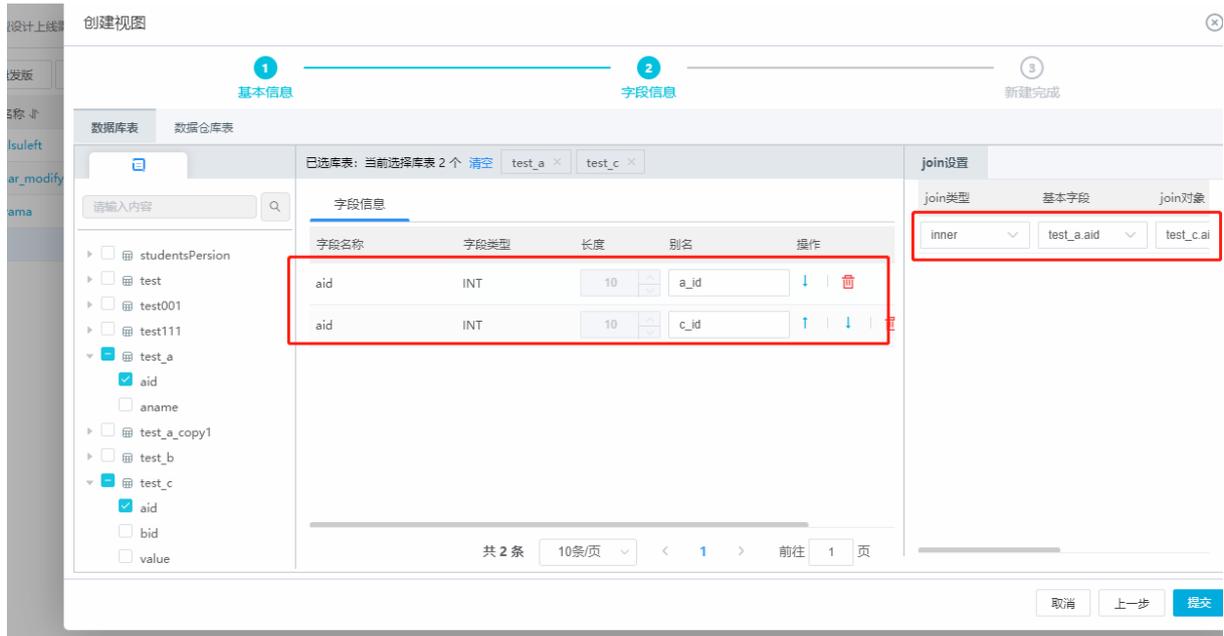


图 11-11

11.1.3 创建普通维度模型

普通维度模型创建主要分四步：

1、维度模型基本信息填写：模型设计页“创建”按钮中选择维度模型如图，进入维度模型创建流程页后填写维度模型基本信息。



图 11-12

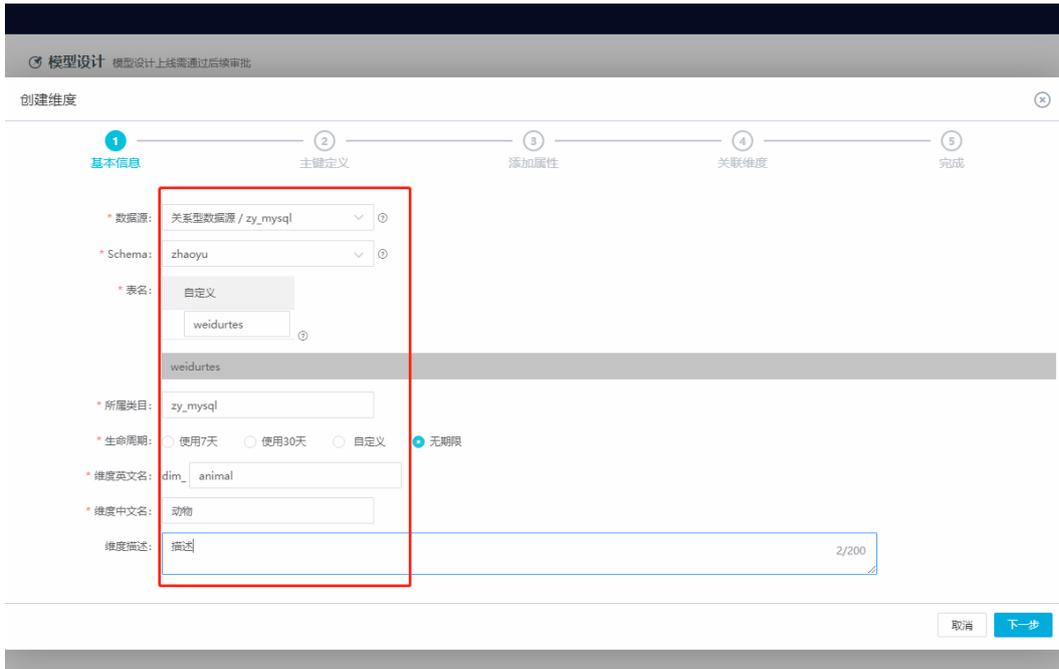


图 11-13

2、添加维度主键：主键可用作其他模型的关联建，添加主键信息即可如下图。

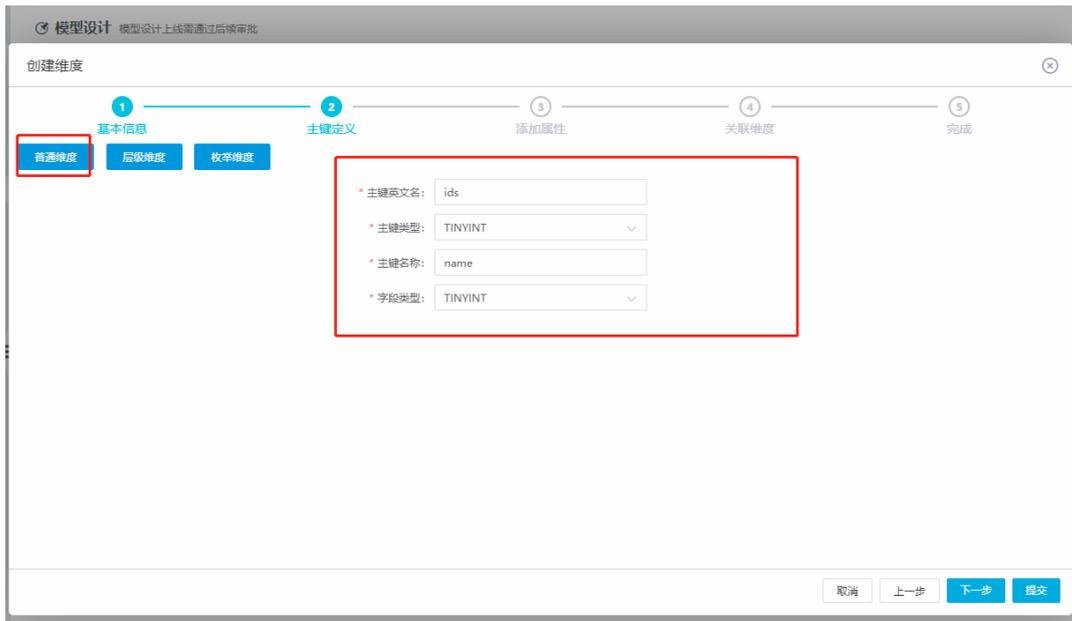


图 11-14

3、添加维度属性：为维度模型添加字段信息。

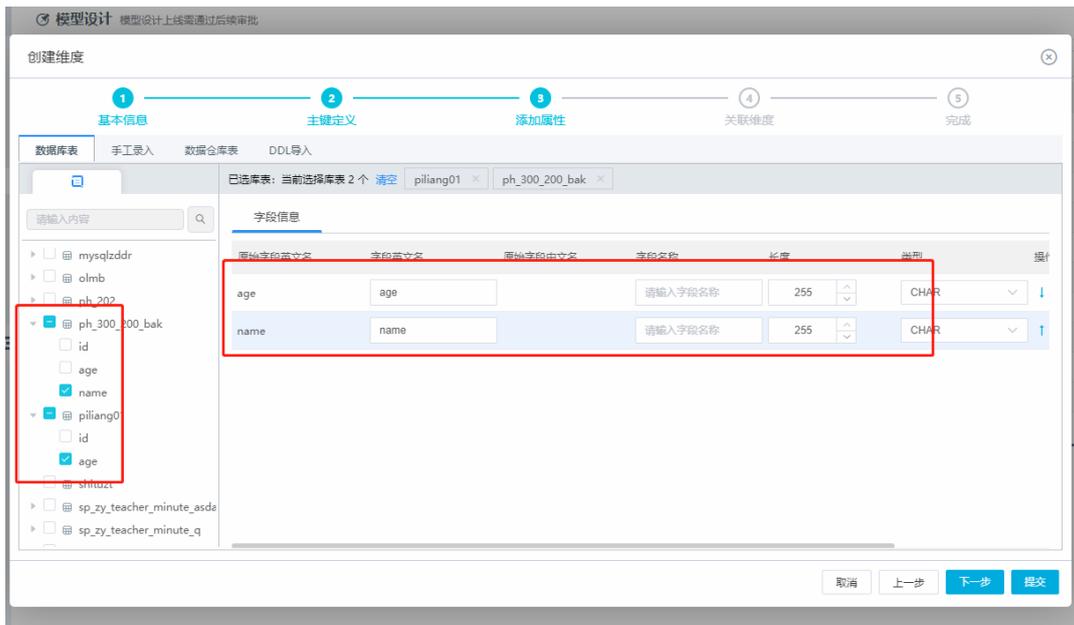


图 11-15

4、关联维度：为当前模型添加维度关联，关联维度表中的表指当前数据源下已上线的维度模型。选择要关联的维度模型，添加字段英文名、字段名。注：字段英文名可以理解为当前表的外键，原始字段英文名可理解为关联表的主键如下图，关联信息填写后点击“提交”，维度模型创建完毕。



图 11-16

11.1.4 创建层级维度模型

层级维度模型中定义了字段间的层级关系。目前系统中还不能使用层级关系，仅可配置层级间的逻辑关联信息。

层级维度模型创建主要分六步：

1、维度模型基本信息填写：模型设计页“创建”按钮中选择维度模型如图，进入维度模型创建流程页后填写维度模型基本信息。

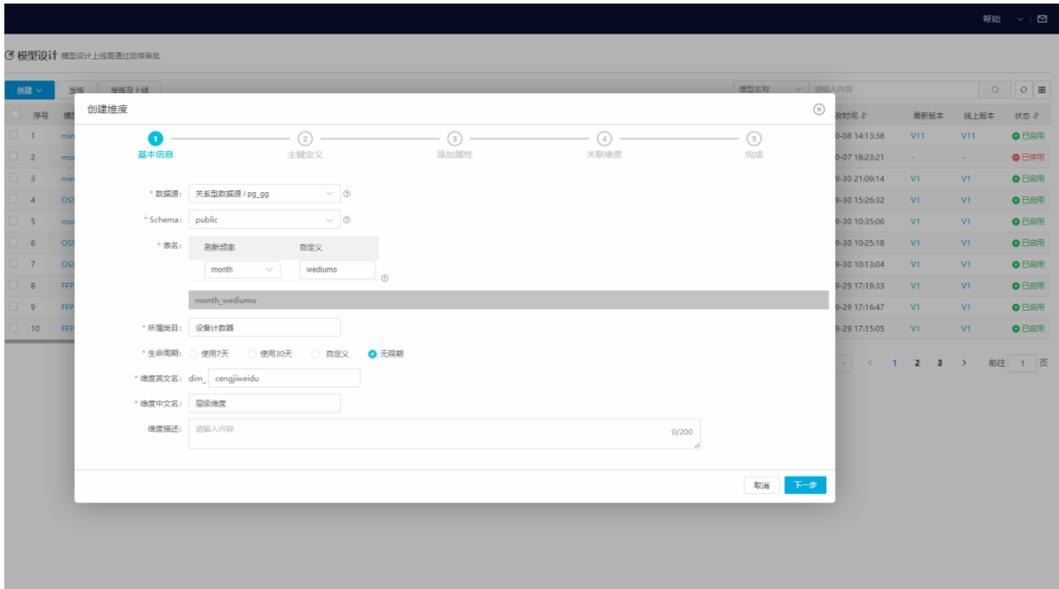


图 11-17

2、添加维度主键：主键可用作其他模型的关联建，添加主键信息即可如下图。

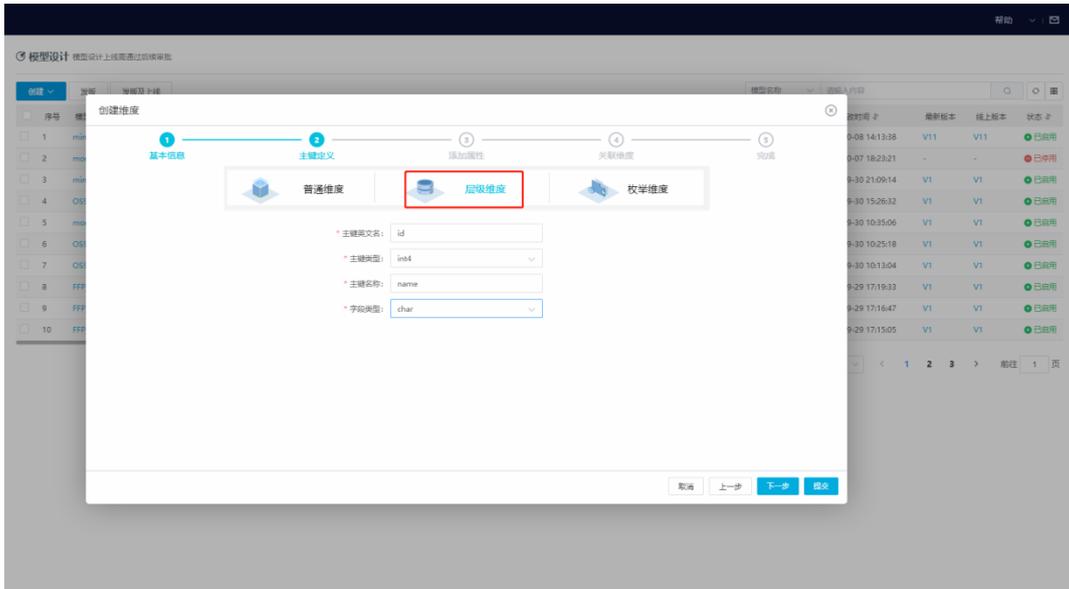


图 11-18

3、添加维度属性：为维度模型添加字段信息。

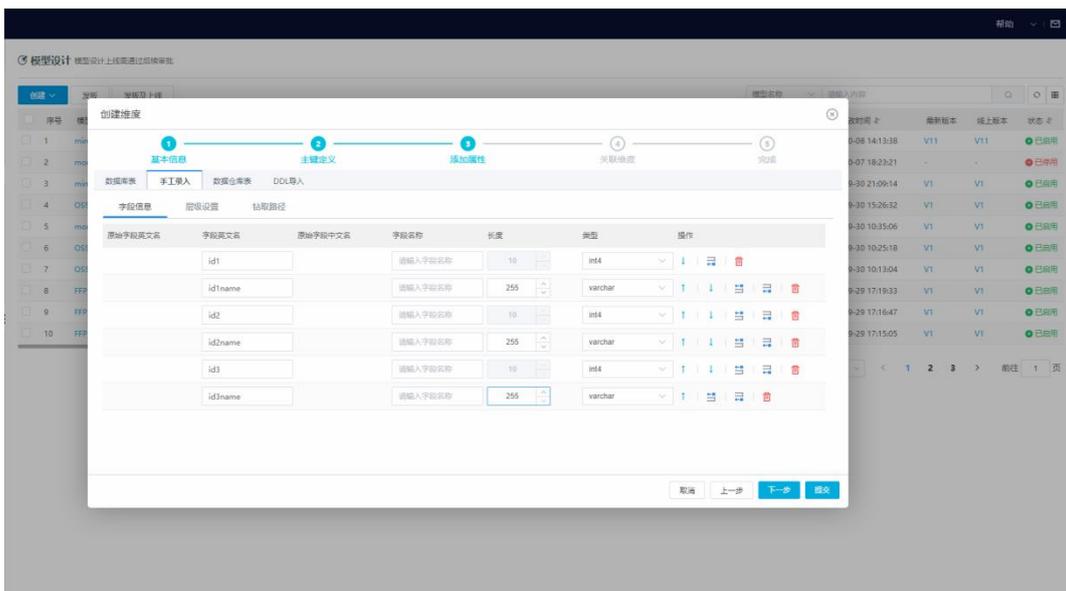


图 11-19

4、层级设置：为层级维度模型定义字段所属层级。

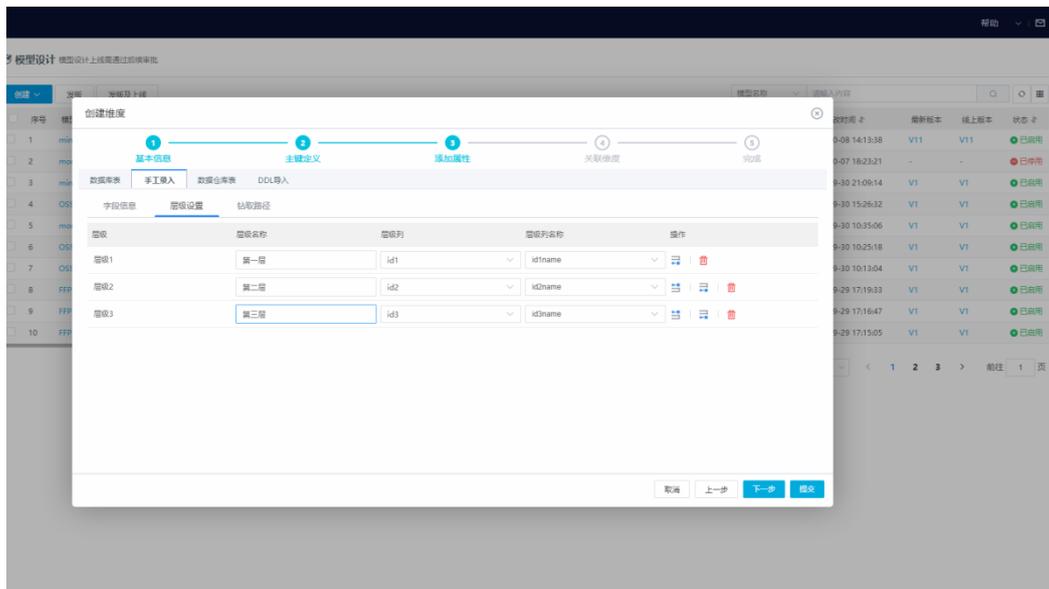


图 11-20

5、钻取路径：为层级维度模型定义钻取路径。

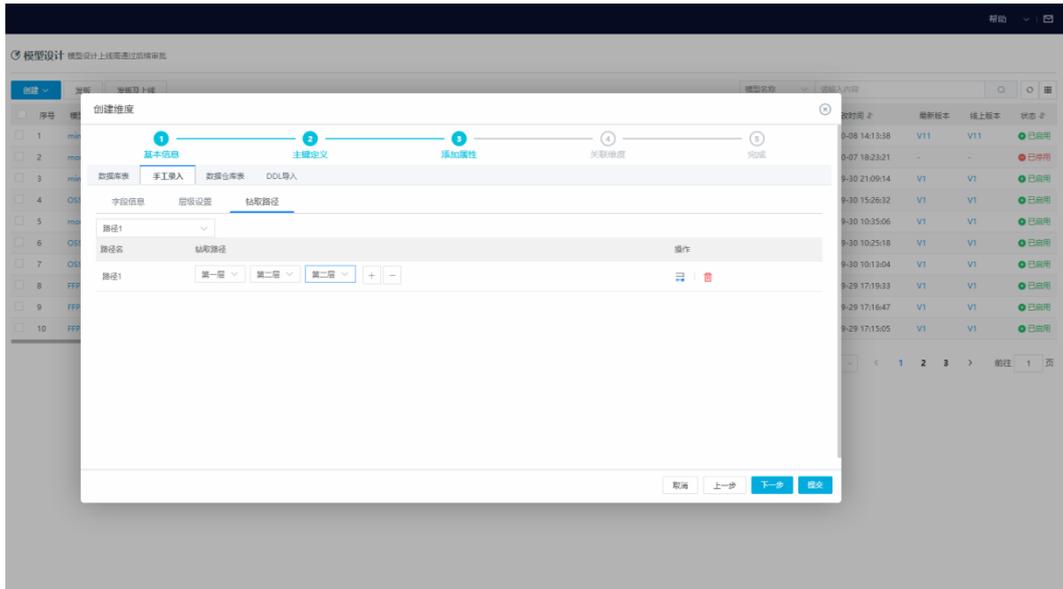


图 11-21

6、关联维度：为当前模型添加维度关联，关联维度表中的表指当前数据源下已上线的维度模型。选择要关联的维度模型，添加字段英文名、字段名。注：字段英文名可以理解为当前表的外键，原始字段英文名可理解为关联表的主键如下图，关联信息填写后点击“提交”，层级维度模型创建完毕。

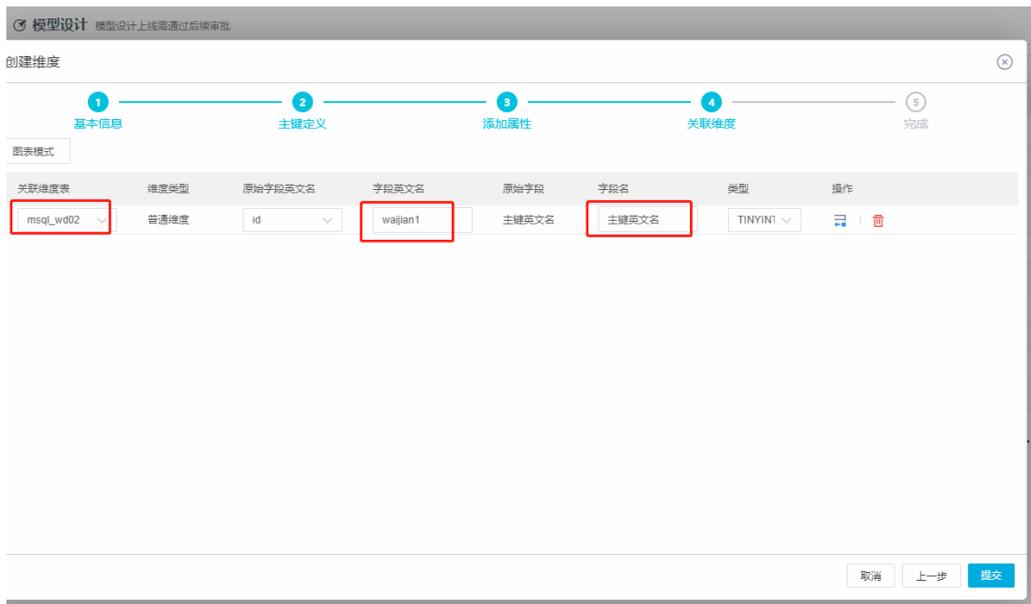


图 11-22

11.1.5 创建枚举维度模型

枚举维度模型相当于字典表，将 code 对应 value 存储到物理数据库中。

枚举维度模型不可关联其他维度模型，也不可被其他维度模型关联。

层级维度模型创建主要分三步：

1、维度模型基本信息填写：模型设计页“创建”按钮中选择维度模型如图，进入维度模型创建流程页后填写维度模型基本信息。

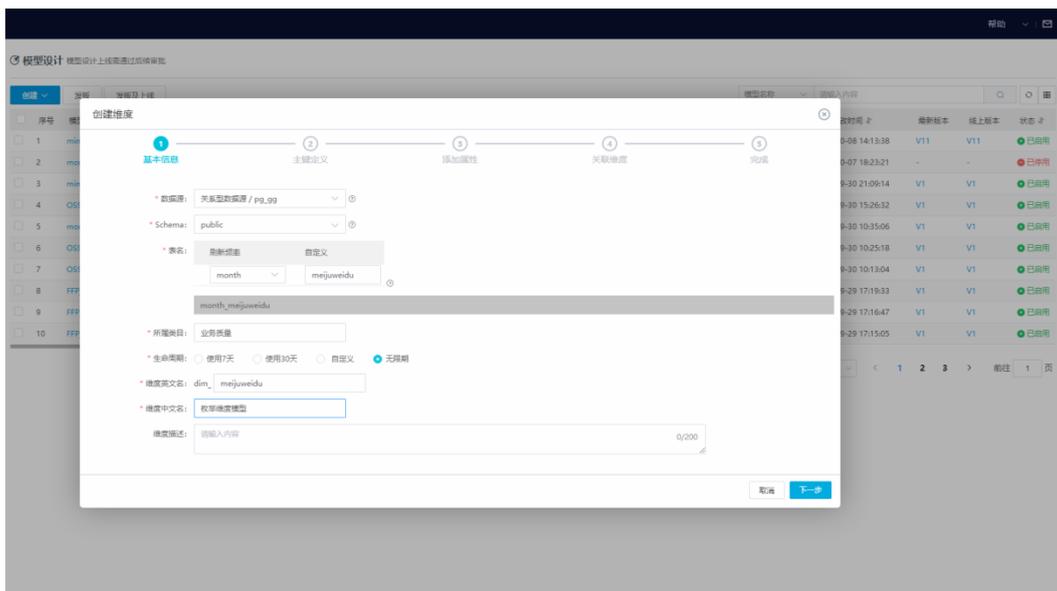


图 11-23

2、添加维度主键：主键可用作其他模型的关联建，添加主键信息即可如下图。

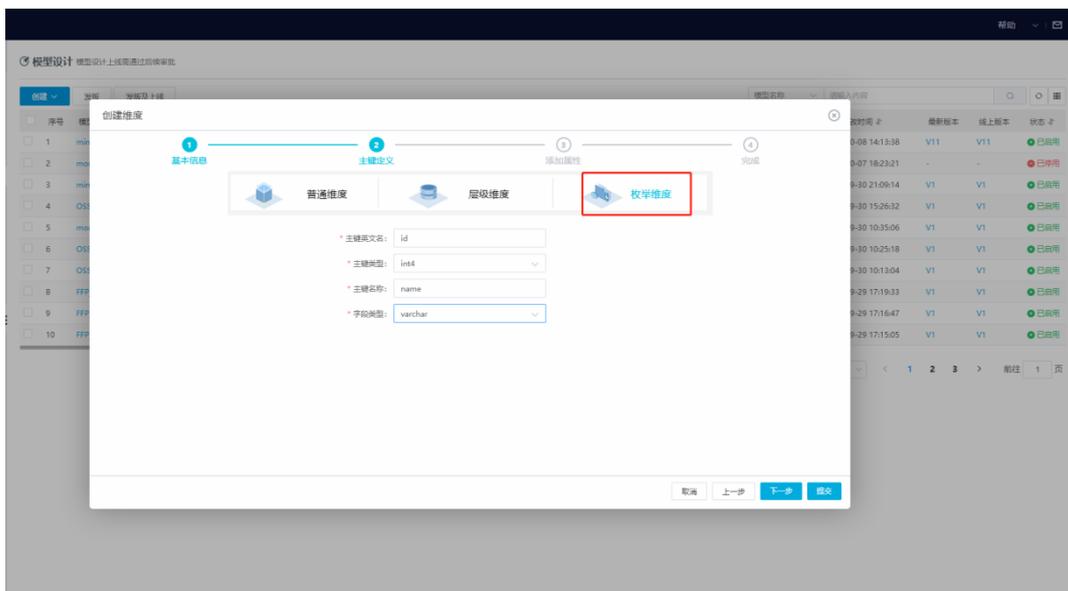


图 11-24

3、添加真实数据：为维度模型添加 code 对应 value 的字典值，点击提交即完成枚举维度模型的创建。

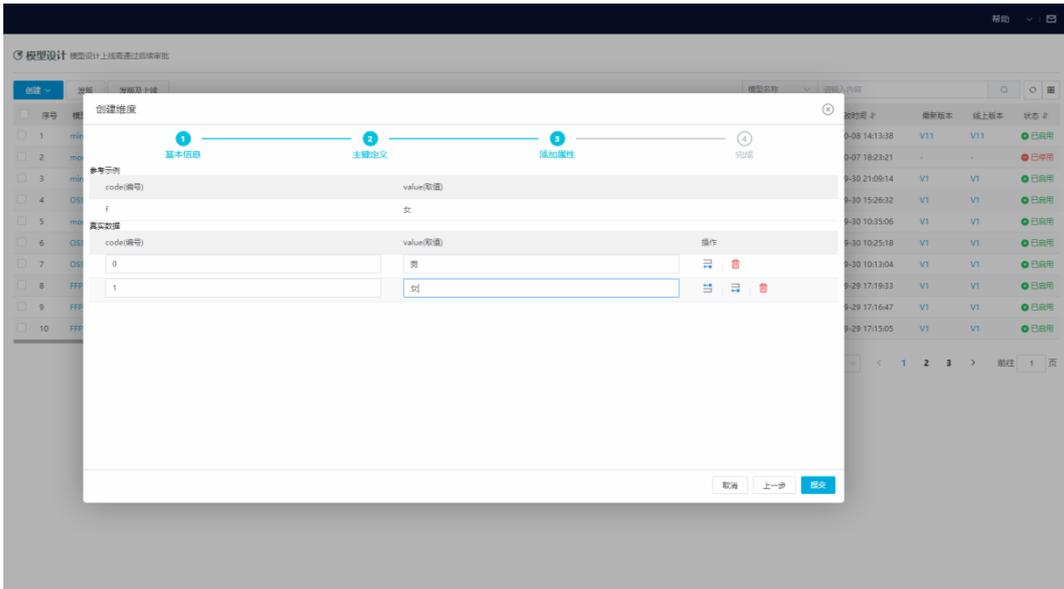


图 11-25

11.2 函数管理

数据处理工作对象大部分为 SQL 类型的脚本和任务。在编辑 SQL 类型的脚本和任务的代码时，常需要使用各种函数对数据做标准化处理。函数管理，是平台提供的专用于对 SQL 编辑时需要的系统函数和自定义函数进行管理的功能，在此页面可以进行新建目录、新建函数的操作。函数管理模块下显示的全部函数，无论是系统默认的还是自定义函数，仅用于 SQL 类型的任务和脚本。

函数的应用场景如下图所示：

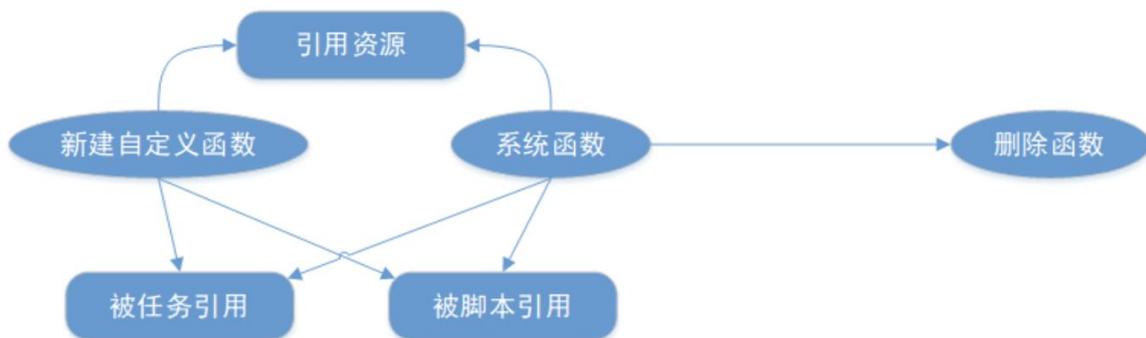


图 11-26

11.2.1 新增函数

点击【函数管理】，函数管理树中选择某数据库类型查看函数，选择自定义函数，点击图标，点击新增，弹出“创建函数”框，如下图所示：



图 11-27



图 11-28

设置函数名称、资源、类名、参数说明、存储位置，点击确定，函数管理树中新增该函数。

11.2.2 查看函数

在函数目录树中，选择某个函数，点击“”，点击“查看”，查看函数的详情。如下两图所示。



图 11-29

查看的函数详情如下。

函数说明



函数名称

GROUP

用途

GROUPING 用于区分数据中 NULL 值 和由 GROUP BY 类函数 (ROLLUP、CUBE、GROUPING SETS) 返回的 NULL 值。作为 ROLLUP、CUBE 或 GROUPING SETS 操作结果返回的 NULL 是 NULL 的特殊应用。它在结果集内作为列的占位符,表示全体。GROUPING 表示 GROUP BY 列表中的表达式是否参与分组,返回 1 表示不参与分组,返回 0 表示参与分组。对普通 GROUP BY 表达式, GROUPING 返回 0。

参数说明

```
GBase> create table t1(i int,v varchar(10)); GBase> insert into t1 values (2,'a'),(2,'b'); GBase> select * from t1; -> ' | i | v | | 2 | a | | 2 | b | '
```

创建

系统于2020-09-22 17:43:27创建

最后修改

系统于2020-09-22 17:43:27编辑

返回

图 11-30

11.2.3 删除函数

点击要删除的自定义函数的“”，点击“删除”。

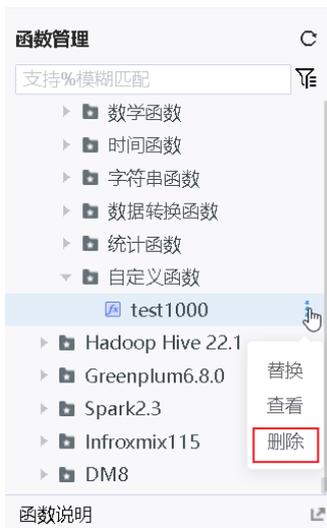


图 11-31

“确认”删除，删除后函数不能再被找回。

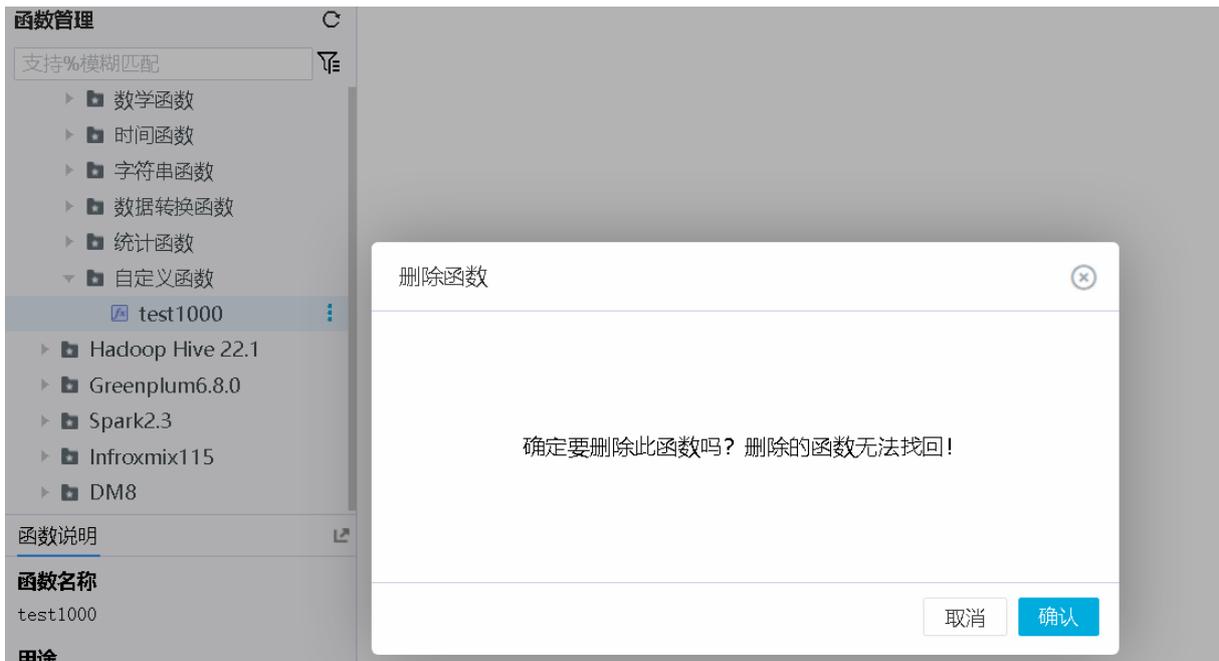


图 11-32

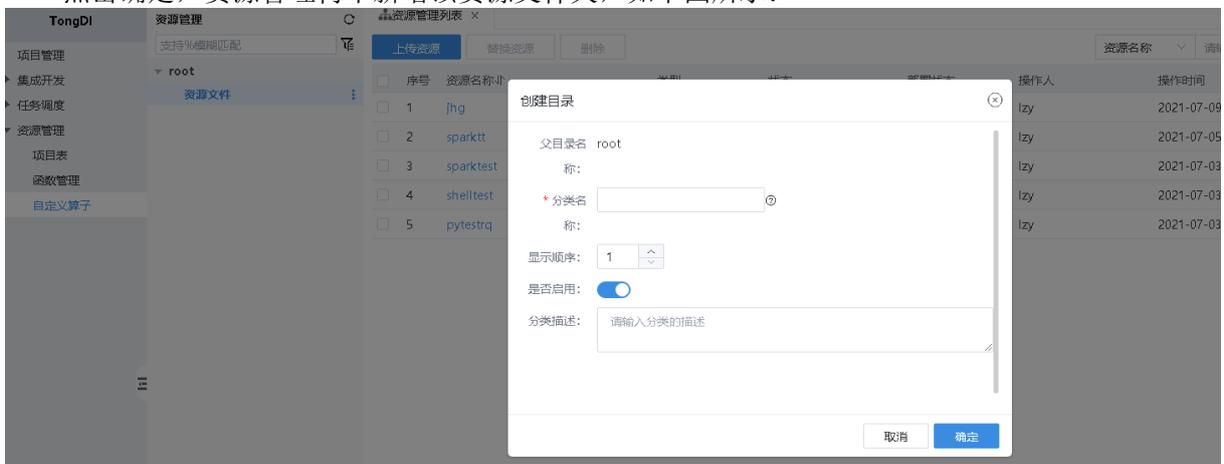
11.3 自定义算子

11.3.1 创建资源文件夹

资源管理树中，根目录资源管理后点击图标“⋮”，点击“创建同级”，弹出“新建文件夹”，可设置目录名称、选择目录位置，如下图所示：



点击确定，资源管理树中新增该资源文件夹，如下图所示：

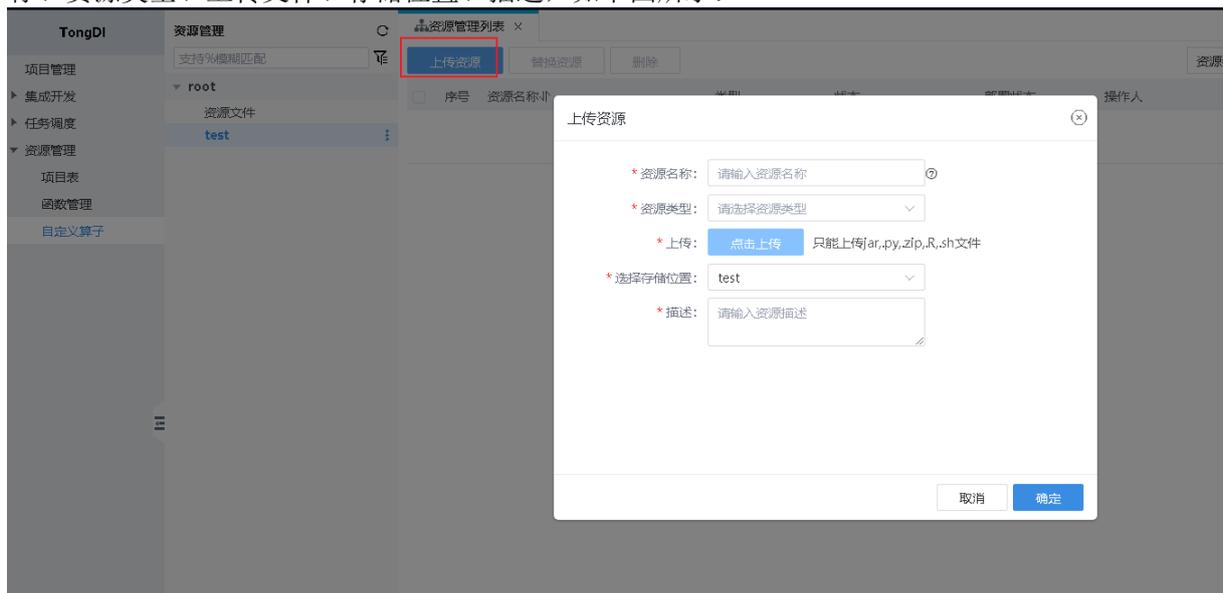


资源管理树中，该资源文件夹提供新增资源、编辑资源文件夹、删除资源文件夹功能，如下图所示：



11.3.2 新增资源

资源管理树中，选择该资源文件夹，点击图标“上传资源”，弹出“上传资源”，设置资源名称、资源类型、上传文件、存储位置、描述，如下图所示：



上传成功后，点击确定，资源管理树中，该资源文件夹下新增该资源文件。

11.3.3 查看资源

点击资源的名称，可以查看资源详情。



11.3.4 替换资源

点击替换，弹出“替换资源”，如下图所示：



可修改资源类型、资源文件、描述，点击确定，替换成功。

11.3.5 删除资源

首先停用资源，然后点击删除，如下图所示：

