

阿里云 智能数据构建与管理 Dataphin 研发

文档版本：20200217

法律声明

阿里云提醒您 在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云文档中所有内容，包括但不限于图片、架构设计、页面布局、文字描述，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

| 格式 | 说明 | 样例 |
|--|------------------------------------|--|
|  | 该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。 |  禁止： 重置操作将丢失用户配置数据。 |
|  | 该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。 |  警告： 重启操作将导致业务中断，恢复业务时间约十分钟。 |
|  | 用于警示信息、补充说明等，是用户必须了解的内容。 |  注意： 权重设置为0，该服务器不会再接受新请求。 |
|  | 用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。 |  说明： 您也可以通过按Ctrl + A选中全部文件。 |
| > | 多级菜单递进。 | 单击设置 > 网络 > 设置网络类型。 |
| 粗体 | 表示按键、菜单、页面名称等UI元素。 | 在结果确认页面，单击确定。 |
| Courier字体 | 命令。 | 执行cd /d C:/window命令，进入Windows系统文件夹。 |
| ## | 表示参数、变量。 | bae log list --instanceid Instance_ID |
| []或者[a b] | 表示可选项，至多选择一个。 | ipconfig [-all -t] |
| { }或者{a b} | 表示必选项，至多选择一个。 | switch {active stand} |

目录

| | |
|-----------------------|-----|
| 法律声明..... | I |
| 通用约定..... | I |
| 1 研发页面概览..... | 1 |
| 2 数据引入..... | 6 |
| 2.1 数据引入概述..... | 6 |
| 2.2 管理同步任务文件夹..... | 7 |
| 2.3 创建同步目标表..... | 11 |
| 2.4 创建同步任务..... | 14 |
| 2.5 配置同步任务..... | 16 |
| 2.6 调度配置..... | 24 |
| 2.7 执行与管理同步任务..... | 34 |
| 3 规范建模..... | 40 |
| 3.1 数据建模研发概述..... | 40 |
| 3.2 规范定义-维度..... | 43 |
| 3.2.1 新建维度..... | 43 |
| 3.2.2 修改维度..... | 52 |
| 3.2.3 下线或删除维度..... | 53 |
| 3.3 规范定义-业务过程..... | 54 |
| 3.3.1 新建业务过程..... | 54 |
| 3.3.2 管理业务过程..... | 58 |
| 3.4 逻辑表-维度逻辑表..... | 65 |
| 3.4.1 编辑模型信息和表信息..... | 65 |
| 3.4.2 主表编辑..... | 69 |
| 3.4.3 物理化配置..... | 70 |
| 3.4.4 调度配置..... | 71 |
| 3.4.5 下线或删除维度逻辑表..... | 72 |
| 3.5 逻辑表-事实逻辑表..... | 72 |
| 3.5.1 新建与修改事实逻辑表..... | 72 |
| 3.5.2 下线或删除事实逻辑表..... | 77 |
| 3.6 规范定义-原子指标..... | 78 |
| 3.6.1 新建原子指标..... | 78 |
| 3.6.2 管理原子指标..... | 85 |
| 3.7 规范定义-业务限定..... | 103 |
| 3.7.1 新建业务限定..... | 103 |
| 3.7.2 修改业务限定..... | 103 |
| 3.7.3 下线或删除业务限定..... | 104 |
| 3.8 规范定义-派生指标..... | 106 |
| 3.8.1 新建派生指标..... | 106 |
| 3.8.2 修改派生指标..... | 107 |
| 3.8.3 下线或删除派生指标..... | 108 |

| | |
|---------------------------|------------|
| 3.8.4 查看相关..... | 108 |
| 3.9 逻辑表-汇总逻辑表..... | 109 |
| 3.9.1 新建汇总逻辑表..... | 109 |
| 3.9.2 物理化配置..... | 109 |
| 3.9.3 查看表详情..... | 110 |
| 3.9.4 新增自定义指标与派生指标管理..... | 111 |
| 3.9.5 删除汇总逻辑表..... | 112 |
| 3.10 规范建模-建模引擎..... | 112 |
| 4 数据处理..... | 116 |
| 4.1 数据处理概述..... | 116 |
| 4.2 代码任务..... | 119 |
| 4.2.1 新建代码任务..... | 119 |
| 4.2.2 修改代码任务..... | 122 |
| 4.2.3 下线或删除代码任务..... | 126 |
| 4.2.4 查看版本信息..... | 127 |
| 4.3 同步任务..... | 128 |
| 4.4 函数管理..... | 133 |
| 4.4.1 新建用户自定义函数..... | 133 |
| 4.4.2 管理用户自定义函数..... | 135 |
| 4.5 资源管理..... | 137 |
| 4.5.1 新建资源..... | 137 |
| 4.5.2 修改资源..... | 139 |
| 4.5.3 删除资源..... | 141 |
| 5 即席查询..... | 142 |
| 5.1 即席查询概述..... | 142 |
| 5.2 新建即席查询..... | 143 |
| 5.3 管理即席查询..... | 144 |
| 6 发布管理..... | 148 |
| 7 数据萃取..... | 153 |
| 7.1 数据萃取概述..... | 153 |
| 7.2 行为中心..... | 155 |
| 7.2.1 行为中心概述..... | 155 |
| 7.2.2 管理行为域..... | 156 |
| 7.2.3 管理业务线..... | 164 |
| 7.2.4 管理动作..... | 172 |
| 7.2.5 管理对象..... | 181 |
| 7.2.6 管理对象属性..... | 191 |
| 7.2.7 新建行为规则..... | 199 |
| 7.2.8 管理行为规则..... | 212 |
| 7.2.9 查看行为数据..... | 238 |
| 7.3 标签中心..... | 241 |
| 7.3.1 新建工厂标签..... | 241 |
| 7.3.2 管理工厂标签..... | 250 |
| 7.3.3 标签逻辑表..... | 278 |

| | |
|---------------------|------------|
| 7.3.4 管理标签类目..... | 283 |
| 8 运维调度..... | 294 |
| 8.1 运维调度概述..... | 294 |
| 8.2 全局运维..... | 297 |
| 8.2.1 周期任务..... | 298 |
| 8.2.2 手动任务..... | 301 |
| 8.2.3 周期实例..... | 304 |
| 8.2.4 手动实例..... | 307 |
| 8.2.5 补数据实例..... | 312 |
| 8.3 逻辑表运维..... | 314 |
| 8.3.1 逻辑表任务..... | 314 |
| 8.3.2 逻辑表实例..... | 319 |
| 8.4 萃取运维..... | 322 |
| 8.4.1 行为规则任务..... | 323 |
| 8.4.2 标签任务..... | 328 |
| 8.4.3 行为规则实例..... | 335 |
| 8.4.4 标签实例..... | 338 |
| 8.4.5 补数据实例..... | 343 |
| 9 监控报警..... | 348 |
| 9.1 任务监控配置..... | 348 |
| 9.1.1 新建任务监控..... | 348 |
| 9.1.2 任务监控配置管理..... | 350 |
| 9.2 任务报警记录..... | 352 |

1 研发页面概览

本文为您介绍如何进入研发页面、研发页面简介，以及各功能区和按钮的使用说明。

进入研发页面

在Dataphin产品首页，单击顶部菜单栏的研发，或者通过中间的研发功能快速入口，均可进入研发页面。



研发页面简介

研发页面主要包括开发、萃取、发布、运维、权限模块。

| 模块 | 说明 |
|----|---|
| 开发 | <ul style="list-style-type: none"> 规范建模：构建指标模型。 数据处理：构建同步任务、代码任务、函数和资源文件等。 即席查询：任务发布前，进行数据查询、验证。 <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p> 说明： 即席查询功能仅开发环境可见。</p> </div> |
| 萃取 | <ul style="list-style-type: none"> 行为中心：行为中心提供明细行为的配置及行为数据统计信息。 标签中心：标签中心提供标签可视化配置工作台，可自动化生成标签，并对不同开发状态的标签任务进行管理。 |

| 模块 | 说明 |
|----|--|
| 发布 | <ul style="list-style-type: none"> 查看从开发环境提交的任务，并将其发布到生产环境进行数据生产。 查询历史发布记录。 <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;">  说明： 发布功能仅开发环境可见。 </div> |
| 运维 | <ul style="list-style-type: none"> 运维中心：对周期任务/实例、手动任务/实例、逻辑表任务/实例、补数据实例等进行查看运维。 监控报警：设置和查看生产环境任务的监控报警信息。监控报警功能仅生产环境可见。 |
| 权限 | <p>为确保数据资源使用安全可控，对数据表的使用权限进行治理，即先申请，且需要精确到字段粒度，得到审批后方可使用。</p> |

使用说明

· 研发首页

进入研发页面后，首先可以看到整个研发过程的操作指引图。用户可以单击指引图中每个模块下方的  图标，快速新建对应的任务文件。



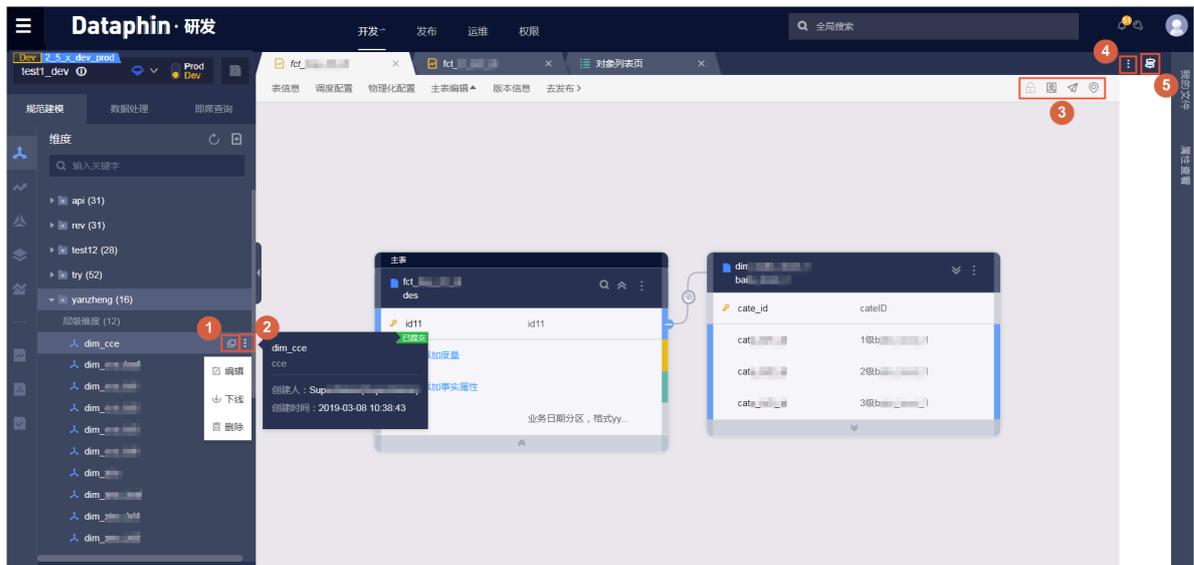
· 研发页面功能区



| 功能区编号 | 说明 |
|-------|--|
| 1 | <p>环境类型标识按钮，单击可以在开发环境和生产环境之间进行切换。Prod表示生产环境，Dev表示开发环境。</p> <p> 说明: 若当前打开的是生产环境中的项目A，单击环境标识按钮切换到开发环境后，编号②的项目也切换为A的孪生项目A_dev。</p> |
| 2 | <p>展示当前打开的项目名称，以及项目的环境类型（Prod或Dev）。单击右侧的下拉箭头可以切换到开发环境或生产环境中的其他项目。</p> <p> 说明: 如果当前打开的是开发环境中的项目，切换到生产环境中的项目后，编号①的环境类型标识也随之切换为生产环境，反之亦然。</p> |
| 3 | 新建任务/文件按钮，单击可以快速新建规范定义、逻辑表、代码任务。 |
| 4 | 刷新、新建文件按钮，单击可以刷新下方的任务列表、新建任务。 |
| 5 | 搜索框，输入关键字可以搜索任务。 |
| 6 | 展示已创建的任务/对象列表。 |
| 7 | 几种不同的规范定义类型，单击选择后可以进行维度、业务过程、原子指标、业务限定、派生指标的查看、新建和下线等操作。 |

| 功能区编号 | 说明 |
|-------|---|
| 8 | <p>三种逻辑表类型（事实逻辑表、维度逻辑表、汇总逻辑表），单击选择后可以查看、新建、删除、下线逻辑表等。</p> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p> 说明： 维度逻辑表不可手动创建，维度创建成功后，系统会自动创建对应的维度逻辑表。</p> </div> |
| 9 | 单击可以查看已创建的规范定义对象列表。 |
| 10 | 研发功能包含的四大模块，单击不同的页签切换到对应的功能页面。 |
| 11 | 全局搜索框，可以搜索物理表、逻辑表、规范定义、函数、资源等，支持选择过滤条件以快速搜索目标对象。 |
| 12 | 我的文件按钮，单击会展示用户最近打开的文件列表，选择列表中的文件，会显示文件编辑页面。 |
| 13 | 属性查看按钮，单击会展示用户最近打开的对象列表，选择列表中的对象，页面刷新显示该对象的属性信息及可操作项（如编辑、引用等）。 |

· 研发页面按钮



| 按钮编号 | 说明 |
|------|--|
| 1 | 鼠标放置在对象/任务上，右侧会显示该按钮，单击可复制对象/任务的名称。 |
| 2 | 鼠标放置在对象/任务上，右侧会显示该按钮，单击会展示更多操作项，包括编辑、下线、删除等。 |

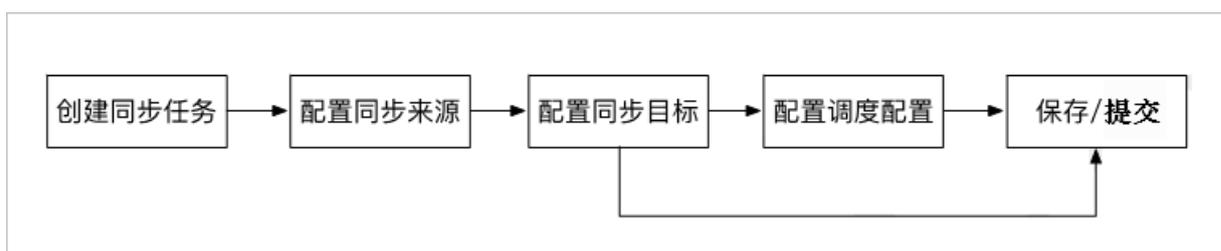
| 按钮编号 | 说明 |
|------|--|
| 3 | <p>从左到右依次为抢锁（或锁定）、保存、提交、定位按钮。</p> <ul style="list-style-type: none"> - 抢锁（或锁定）：任务或对象默认被创建者锁定。在有权限的前提下，其他用户若要编辑任务或对象，需要先抢锁，才可以操作。 - 保存：保存当前页面的任务/对象信息。 - 提交：提交任务。任务提交后，可以在发布管理页面进行发布管理。 - 定位：单击可以快速定位当前任务/对象在左侧功能区列表中的位置。 |
| 4 | <p>单击可以快速切换或关闭已打开的页面。</p> |
| 5 | <p>单击可以切换到操作指引图页面。</p>  |

2 数据引入

2.1 数据引入概述

数据引入是通过创建数据同步任务，将源数据库的表导入至目标数据库的过程。例如，将MySQL数据库中A表的数据导入至PostgreSQL数据库的B表。

数据引入的整体流程如下。



具体步骤如下：

1. [管理同步任务文件夹](#)。
2. [创建同步目标表](#)。
3. [创建同步任务](#)。
4. 配置同步来源、同步目标、参数信息等，详情请参见[配置同步任务](#)。



说明：

配置同步目标时，需要用到步骤2中创建的同步目标表。系统默认为同步任务分配0.5核CPU、1GB内存，暂不支持扩大资源量。

5. 配置同步任务的调度参数，并保存、提交，详情请参见[调度配置](#)。



说明：

- 如果创建的是手动任务，直接保存、提交即可。
- 如果创建的是周期性任务，则需要进行调度配置，再进行保存、提交。

2.2 管理同步任务文件夹

数据引入，即是通过数据同步任务，将建模研发所需的源数据同步至目标数据库。为了分类管理同步任务，我们需要创建文件夹，用于存放同步任务文件。你可以按需对文件夹进行移动、重命名和删除等操作。

新建文件夹

1. 登录Dataphin。
2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。
3. 选择开发 > 数据处理 > 同步任务。



4. 新建文件夹。

单击左侧功能区的新建文件夹按钮，输入文件夹名称，回车后即可新建一个文件夹。



5. 新建子文件夹。

将鼠标悬停于上述步骤中创建的文件夹上，单击右侧的新建文件夹按钮。输入子文件夹名称，回车后即可新建一个子文件夹。



移动文件夹

1. 将鼠标悬停于上述步骤中创建的子文件夹上，单击右侧的  按钮，选择移动。



2. 在移动文件夹对话框中，选择目录，单击确定。



重命名或删除文件夹

将鼠标悬停于某个文件夹上，单击右侧的  按钮，选择重命名、删除，即可重命名或删除该文件夹。



说明:

当文件夹非空（例如，包含子文件夹）时，不允许被删除。

2.3 创建同步目标表

本文为您介绍如何创建数据同步的目标表。

1. 登录Dataphin。
2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。

3. 选择开发 > 即席查询，单击左侧功能区的新建文件按钮。



4. 在新建文件对话框中，填写文件名称、描述，选择目录，单击确定。

新建文件

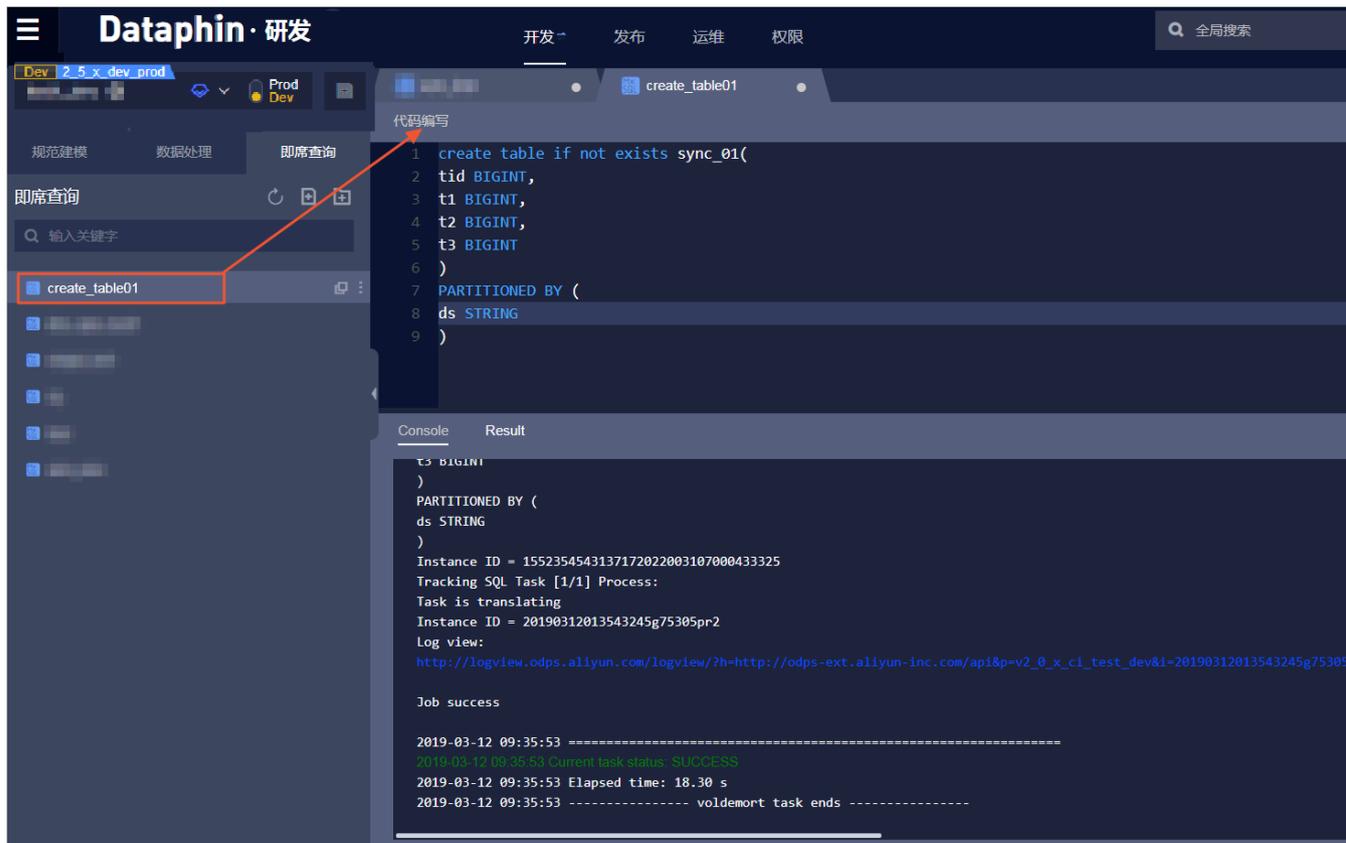
* 名称: create_table01

描述: 创建同步目标表

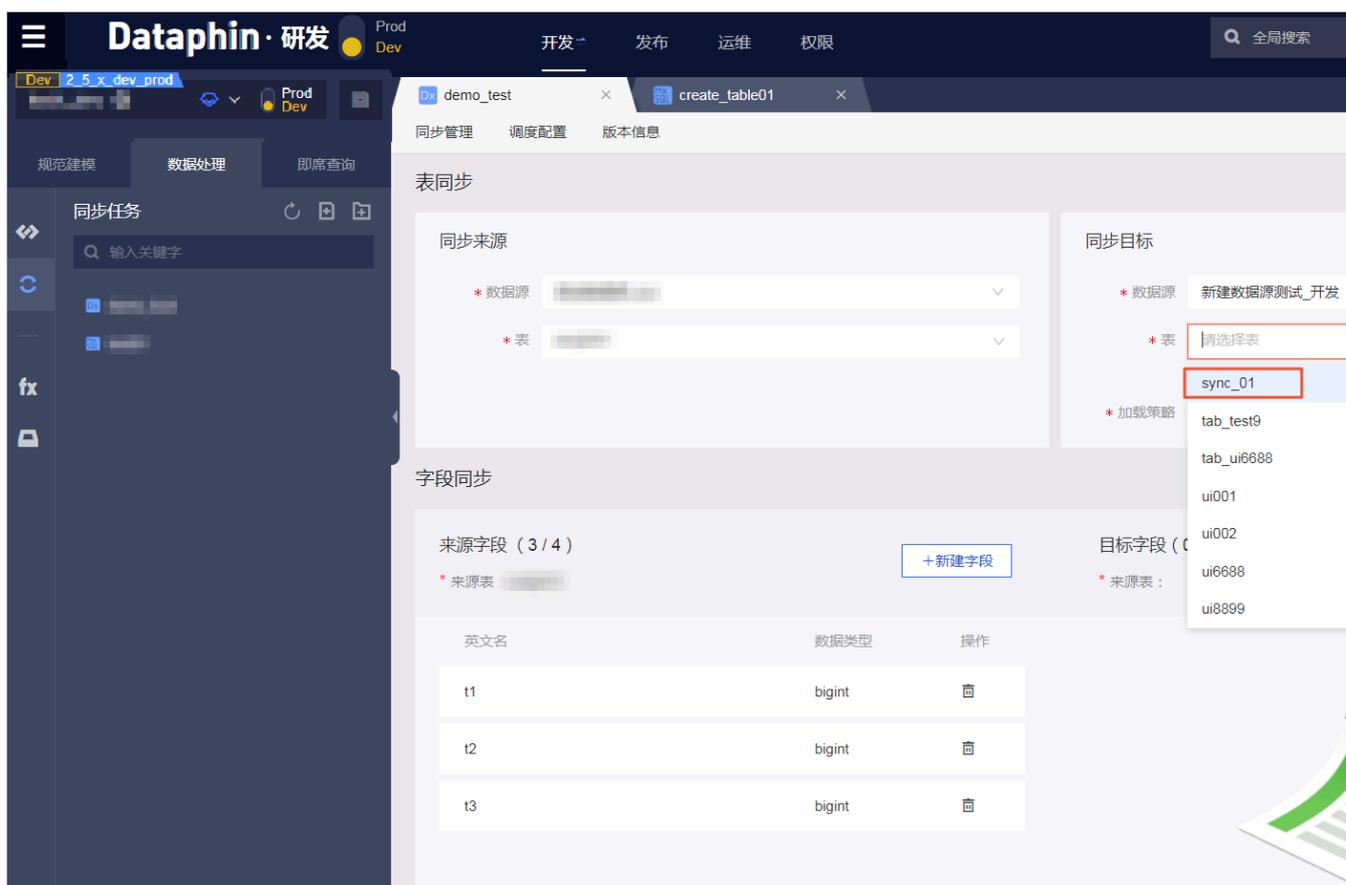
选择目录: 临时代码

取消 确定

- 单击上述步骤中创建的文件，进入代码编写页面。您可以根据实际需求，编写创建同步目标表的SQL语句（即编写建表语句），然后单击页面右上方的执行。



6. 建表语句执行成功后，您可以在配置同步任务时使用刚创建的同步目标表。



2.4 创建同步任务

本文为您介绍如何创建同步任务。

操作步骤

1. 登录Dataphin。
2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。

3. 选择开发 > 数据处理 > 同步任务，单击左侧功能区的新建文件按钮。



4. 在新建文件对话框中，填写同步任务的名称、调度类型、描述，并选择目录，单击确定。

新建文件 ×

* 名称

* 调度类型 周期性节点 手动节点

描述

选择目录



说明:

调度类型可以选择周期性节点或者手动节点。

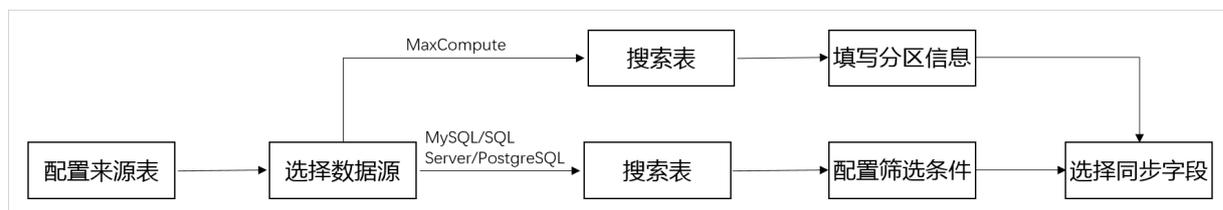
- 周期性节点：指需要定期执行的任务。
- 手动节点：指没有依赖关系，需要手动触发的任务。

2.5 配置同步任务

本文为您介绍如何配置同步任务。

配置流程

同步任务的配置流程如下图。



来源表和目标表均支持多种数据源。按照数据源类型，将支持的数据源分类如下：

- 关系型数据库：MySQL、Vertica、Oracle、SQL Server、PostgreSQL、RDS、DRDS。
- 阿里云大数据数仓存储：MaxCompute。
- 开源大数据数仓存储：Hive。
- 非结构化数据存储：FTP、HDFS、ElasticSearch。
- NoSQL数据存储：MongoDB。



说明：

HBase作为数据源或者目标，支持HBase V1.1.x、HBase V0.94.x两类版本。

同时，支持以MySQL、PostgreSQL、Oracle、SQL Server、Vertica为基础存储的同库分表作为来源数据的数据同步形式。

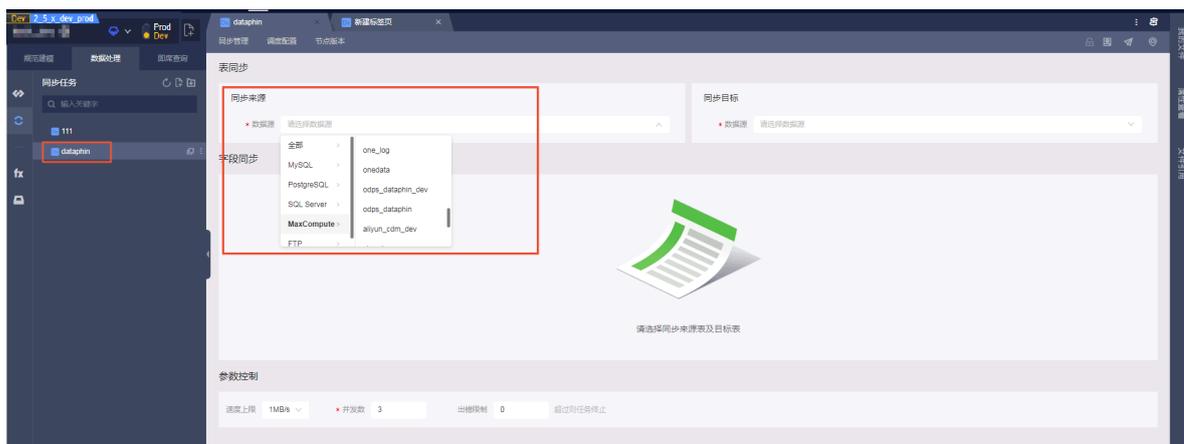
操作步骤

1. 登录Dataphin。在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。

2. 选择开发 > 数据处理 > 同步任务。



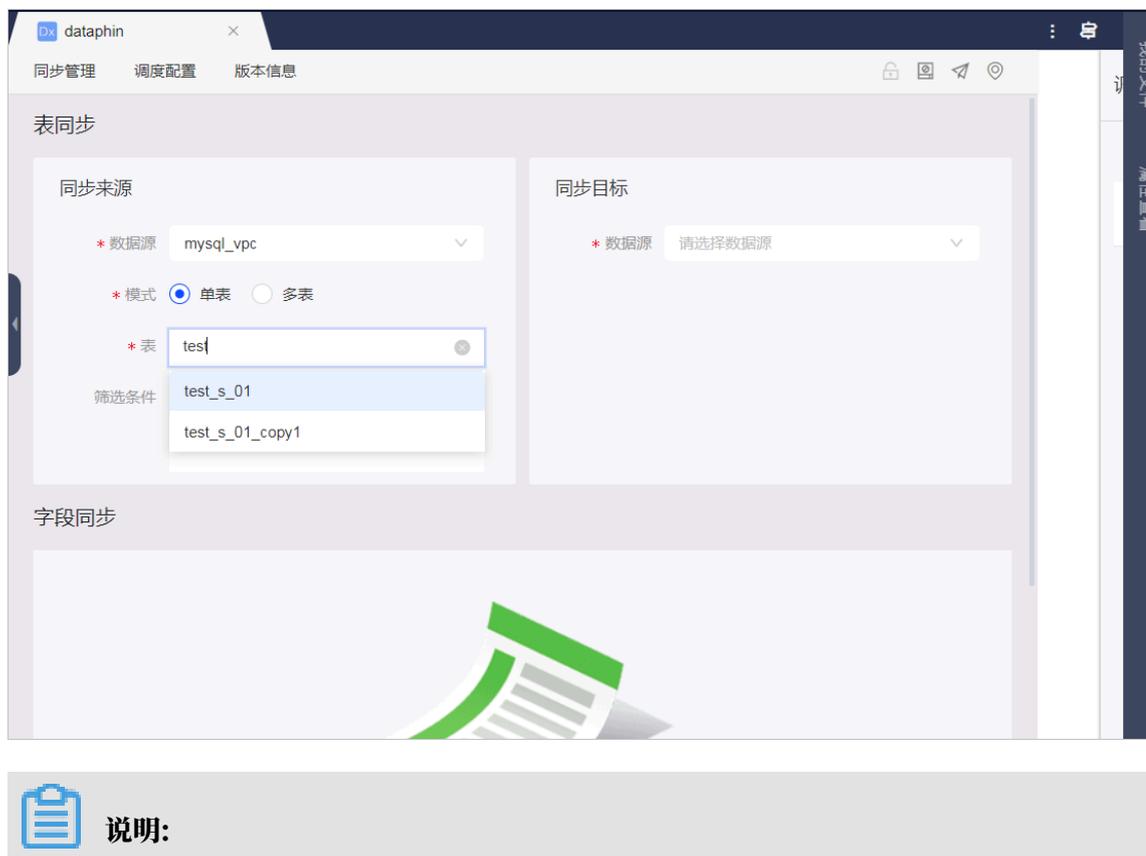
3. 单击选择已创建的同步任务，选择同步来源的数据源。



4. 选择同步来源的表。如果数据源为MySQL、PostgreSQL、SQL Server、Oracle，则需要根据需求选择单表或多表模式。

- 单表模式

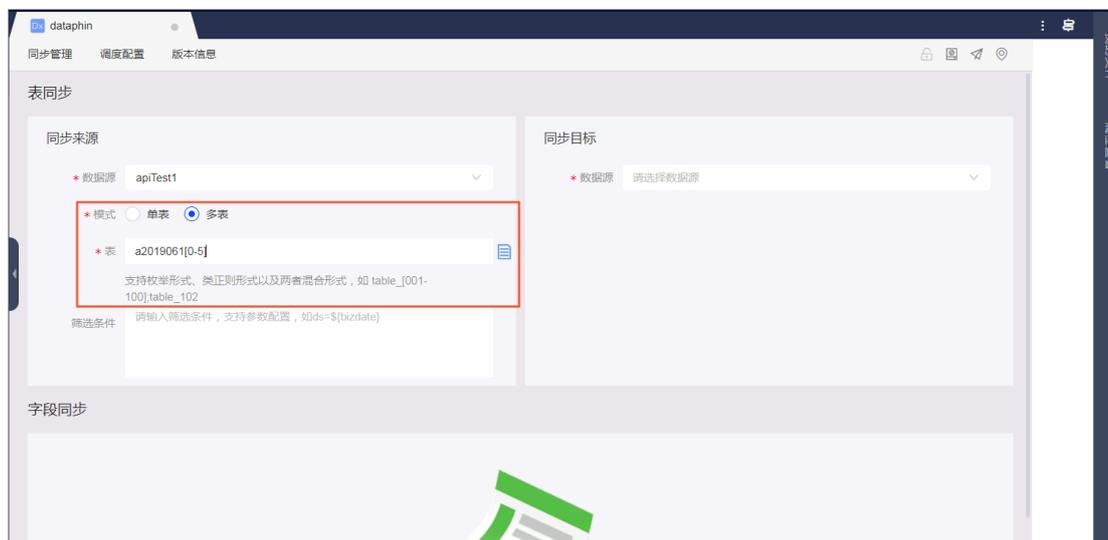
单表模式下，可以搜索表名，支持前缀匹配，在匹配到的下拉列表中选择需要的来源表。



数据源仅支持选择物理数据源。

- 多表模式

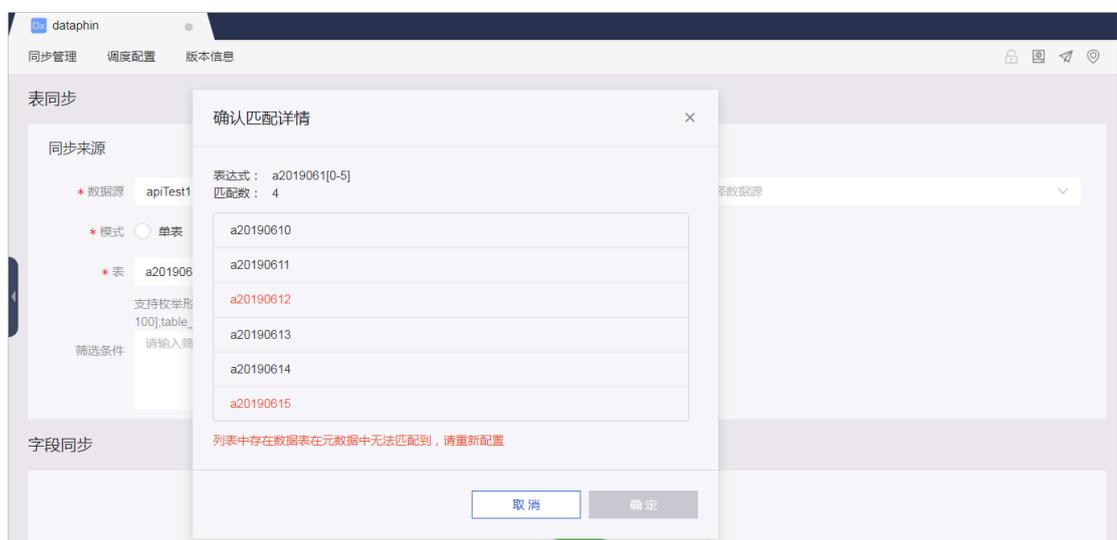
- a. 多表模式下，支持按照输入的表达式规则，匹配多张相同结构的数据表。例如，输入 `a2019061[0-5]`。



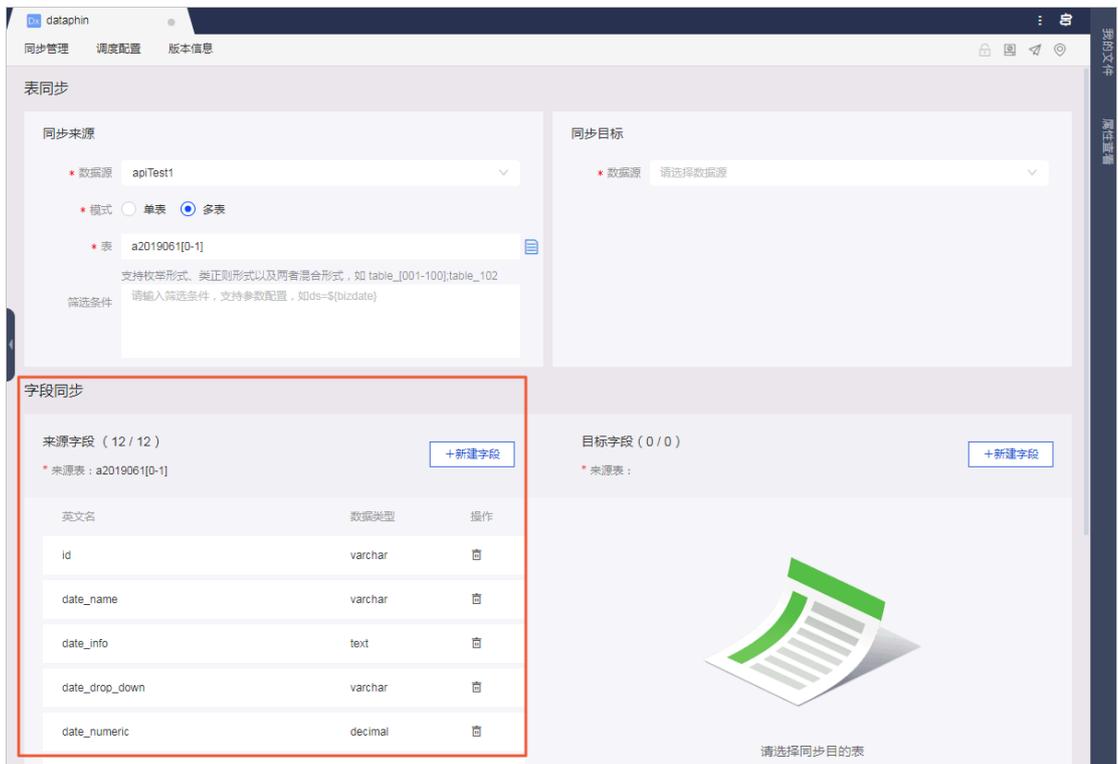
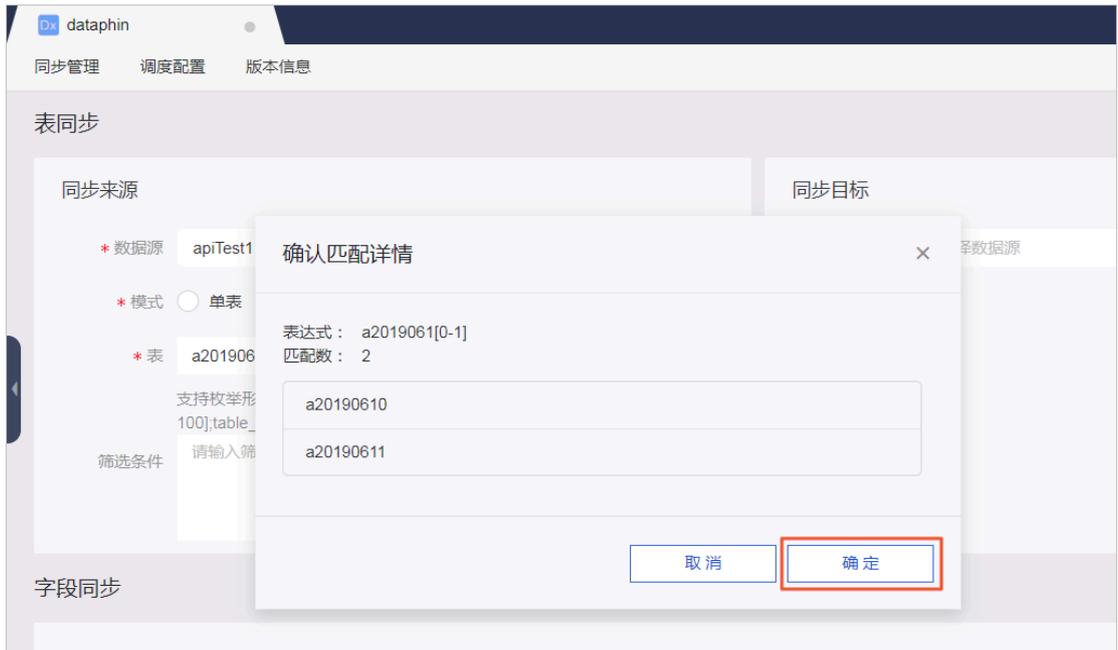
说明:

数据源仅支持选择物理数据源。

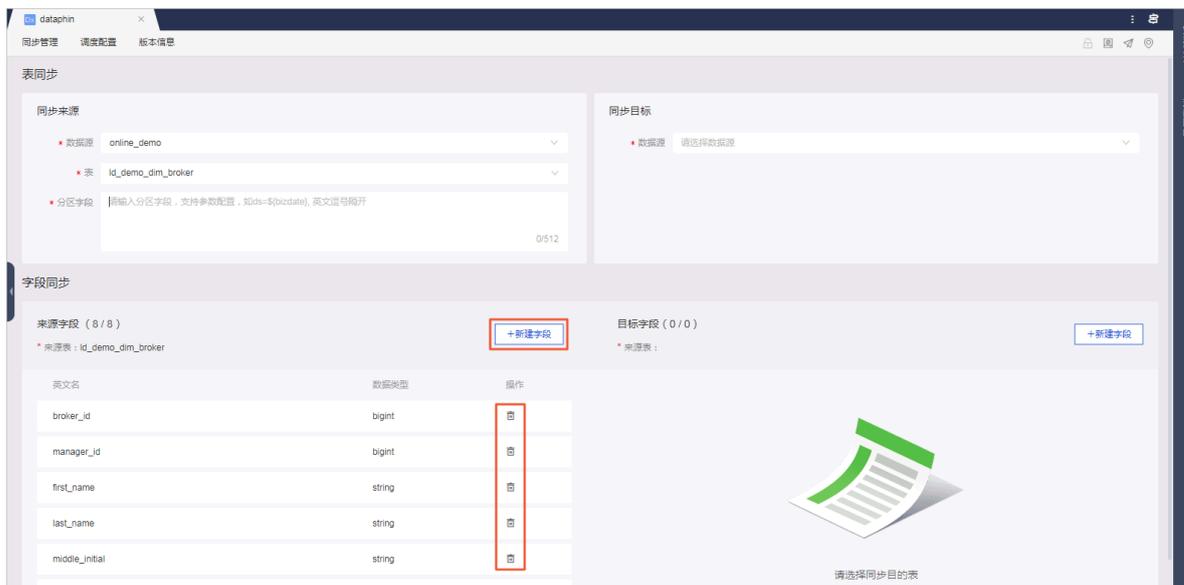
- b. 输入表达式后，单击右侧的确认匹配详情按钮，可以识别到a20190610、a20190611、a20190613、a20190614四张数据表，且这四张表会与源库的元数据进行匹配。



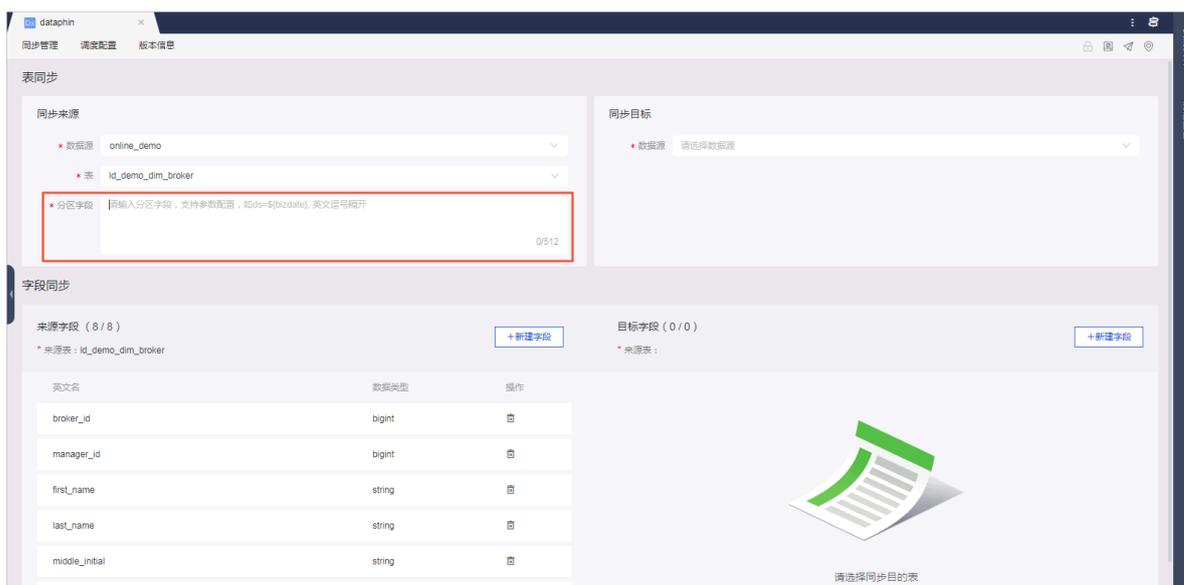
- c. 当数据表不在源库中时，会高亮显示无法匹配到的表名称。此时确定按钮是置灰的、无法操作，则不会加载字段元数据信息。
- d. 当数据表在源库中都存在时，单击确定后，会以首个被匹配到的表为基础，加载字段元数据信息。



5. 选择完来源表之后，表的字段会展示出来。可以删除展示出来的字段，删除字段后，如果需要重新添加，则单击新建字段。



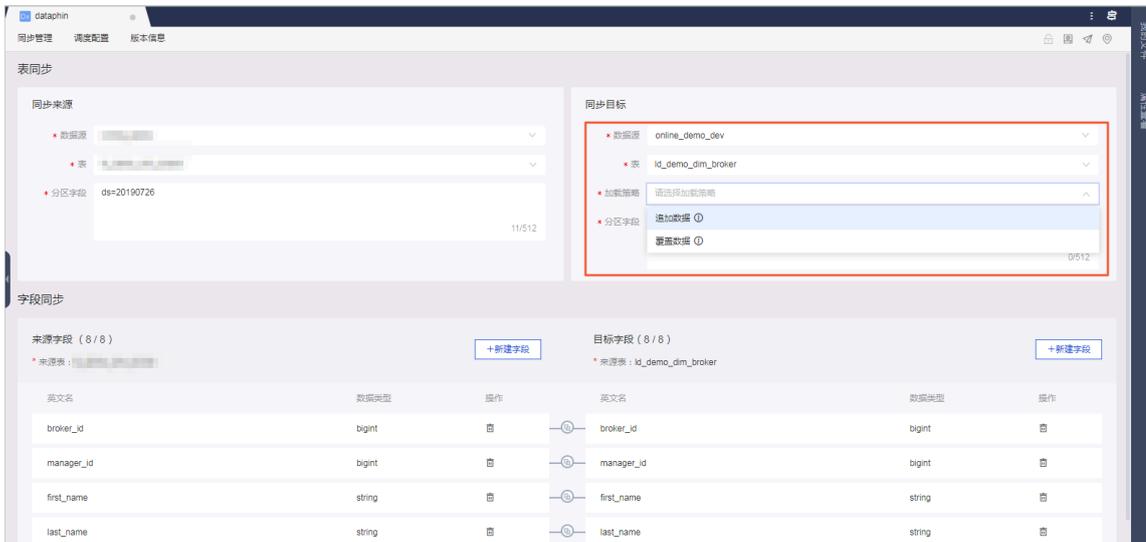
6. 填写表的分区字段。



- 如果是分区表，则必须填写分区字段。分区字段的格式为：支持Linux Shell通配符，*表示0个或多个字符，?表示任意一个字符。例如，分区表的分区字段配置格式为：month=201701,ds=20170101,/*query*/ month>=201701，动态分区字段的格式为：ds=\${bizdate}。
- 当MySQL、SQL Server、PostgreSQL作为来源数据源时，没有分区信息，但需要配置筛选条件。配置筛选条件时，要填写where条件，但不包含where字段。例如，需要导入的表只包含id大于2并且name为dataphin的记录，则应该配置筛选条件为：id>2 and name="dataphin"。

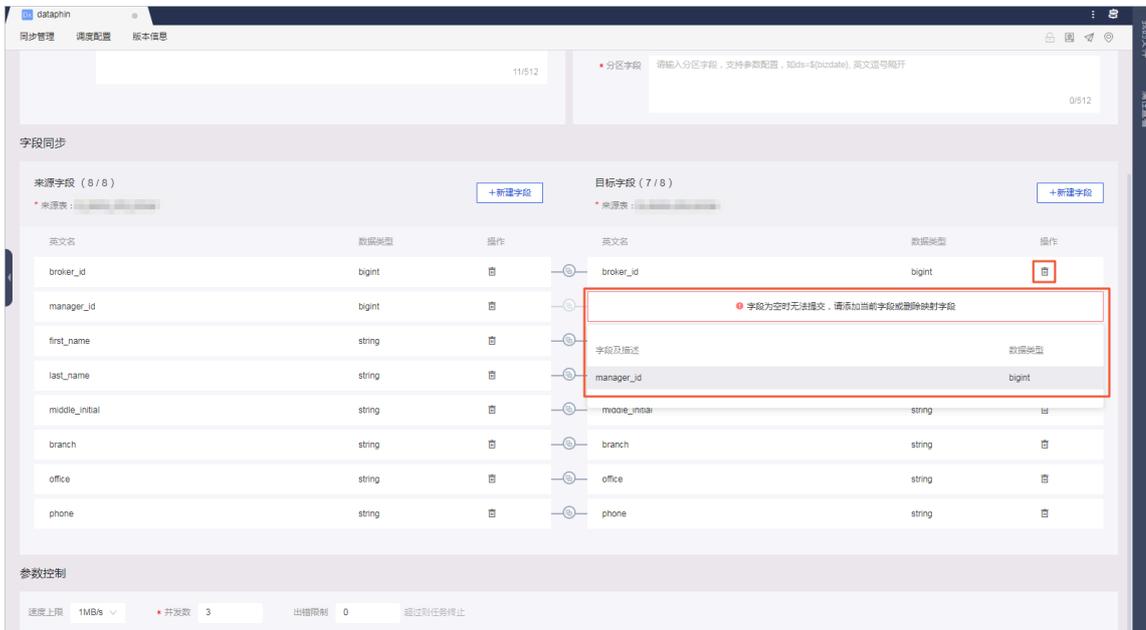
7. 配置同步目标。

a. 选择数据源、表、加载策略（包括两种：追加数据和覆盖数据）及分区字段。

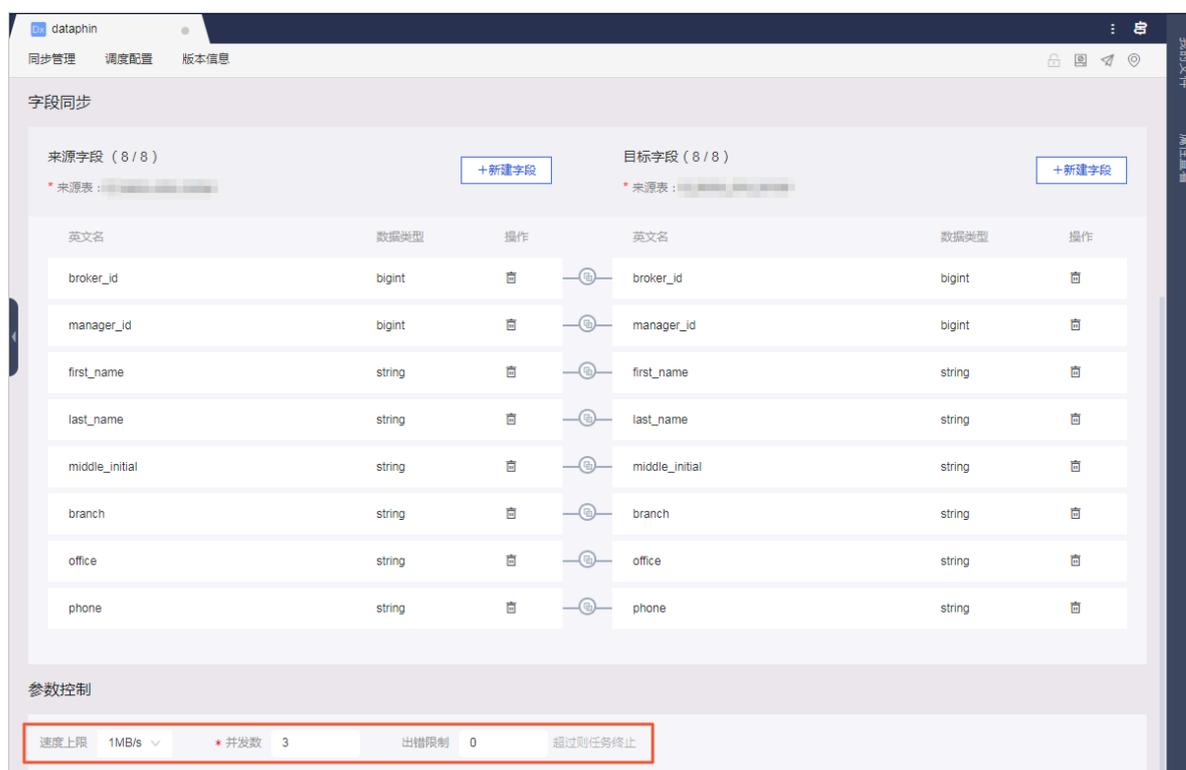


 **说明:**
数据源仅支持选择物理数据源。

b. 调整字段。同步来源和同步目标两边的字段名如果是一致的，则会默认匹配上。如果字段名不一致，则需要手动调整。可以先删除不匹配的字段，然后将鼠标悬停于字段列表的空白处，选择对应字段即可。



8. 配置参数控制。配置参数控制是为了调整同步任务的并发数和容错率等。通常情况下，用户不需要配置，保持默认值即可。



| 参数控制 | 说明 |
|------|--|
| 速度上限 | 数据同步的传输速率，默认为1MB/s。系统数据同步的传输速率会尽可能地达到速度上限，但是不会超过速度上限。速度上限会影响整个调度的资源情况，速度上限设置的越高，运行任务的调度资源量会越大。 |
| 并发数 | 启动并发抽取数据的任务并发数。 |
| 出错限制 | 指数据同步过程中，出错的记录数超过多少条则任务中止。默认为0条，即不允许出错。 |

9. 配置同步任务的调度参数，并保存、提交，详情请参见[调度配置](#)。

2.6 调度配置

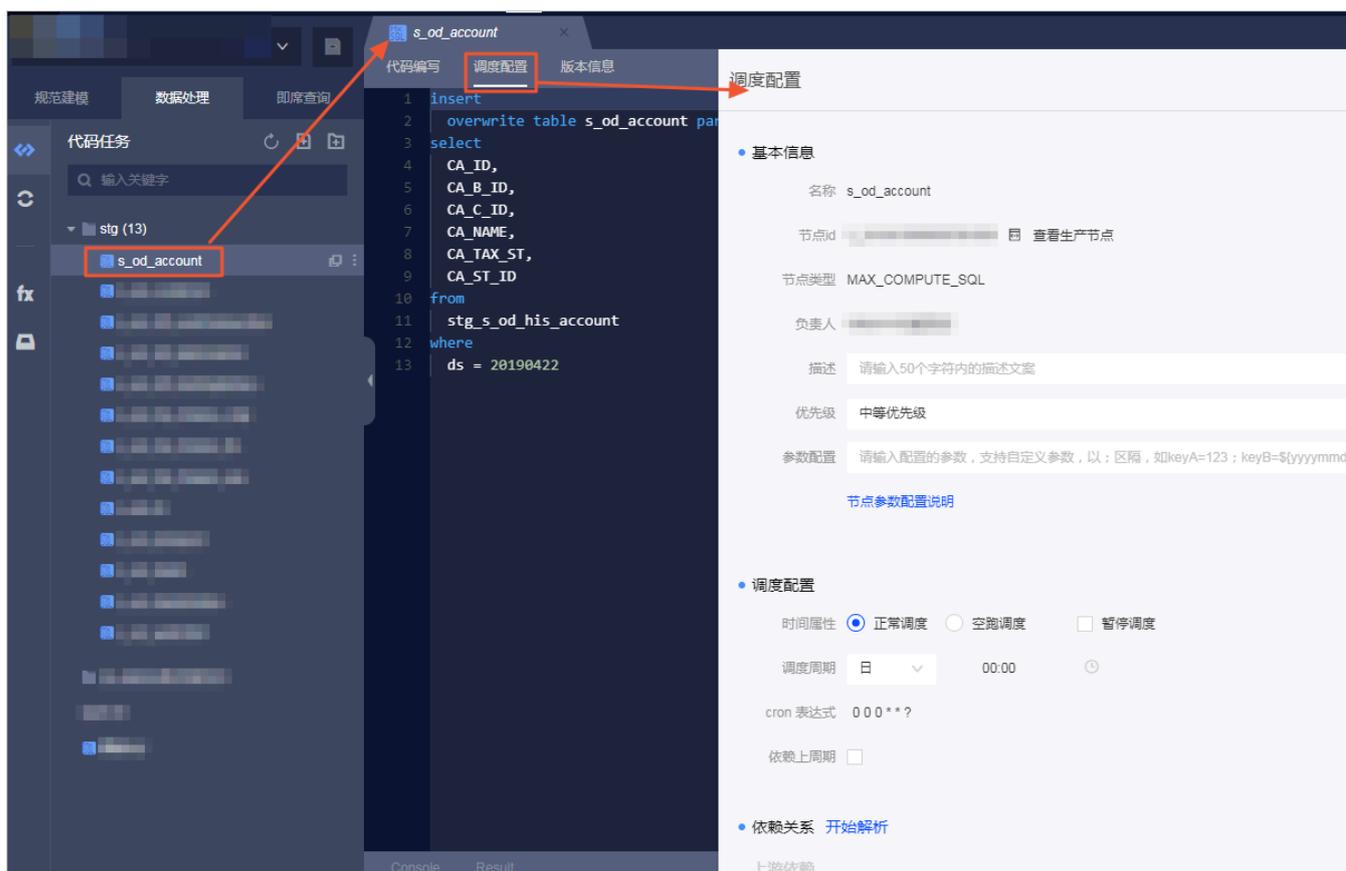
调度配置是整个代码任务和同步任务中的通用功能。参与调度的任务都需要进行调度配置。本文以代码任务为例，为您介绍如何进行调度配置。

1. 登录Dataphin。
2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。

3. 选择开发 > 数据处理 > 代码任务。



4. 在左侧功能区，单击列表中的任意一个代码任务。单击顶部的调度配置，打开调度配置页面。



5. 填写参数配置。单击下方的节点参数配置说明，会弹出参数配置的详细说明。您可以根据实际需求和填写说明，填写参数配置。

The screenshot displays the configuration interface for a node named 's_od_account'. On the left, the code editor shows the following SQL query:

```
1 insert
2 overwrite table s_od_account par
3 select
4 CA_ID,
5 CA_B_ID,
6 CA_C_ID,
7 CA_NAME,
8 CA_TAX_ST,
9 CA_ST_ID
10 from
11 stg_s_od_his_account
12 where
13 ds = 20190422
```

The right-hand side of the interface is the '调度配置' (Scheduling Configuration) panel, which includes the following sections:

- 基本信息 (Basic Information):** Name: s_od_account, Node ID, Node Type: MAX_COMPUTE_SQL, Assignee, Description (placeholder: 请输入50个字符内的描述文案), Priority: 中等优先级 (Medium Priority).
- 参数配置 (Parameter Configuration):** A text input field with the placeholder: 请输入配置的参数，支持自定义参数，以；分隔，如keyA=123；keyB=\${yyyyymmdd-7}；. A button labeled '节点参数配置说明' (Node Parameter Configuration Instructions) is located below this field.
- 调度配置 (Scheduling Configuration):** Time Property: 正常调度 (Normal Scheduling) selected, 空跑调度 (Empty Run Scheduling) and 暂停调度 (Pause Scheduling) are unselected. Scheduling Cycle: 日 (Daily), 00:00. Cron Expression: 0 0 0 * * ?. Depend on Previous Cycle: unchecked.
- 依赖关系 (Dependency Relationship):** 开始解析 (Start Parsing), 上游依赖 (Upstream Dependency), and a button labeled '新建上游依赖' (New Upstream Dependency).

6. 填写调度配置信息。

调度配置

参数配置 请输入配置的参数，支持自定义参数，以；分隔，如keyA=123；keyB=\${yyyyymmdd-?}；

[节点参数配置说明](#)

- 调度配置**
 - 时间属性 正常调度 空跑调度 暂停调度
 - 调度周期 日 00:00
 - cron 表达式 0 0 0 * * ?
 - 依赖上周期
- 依赖关系** [开始解析](#)
 - 上游依赖 [新建上游依赖](#)

| 输出名称 | 节点名 | 节点ID (实例ID) | 负责人 | 操作 |
|--|--|-------------|-----|----|
| virtual_root_node_3 3202474892053708 8 | virtual_root_node_3 3202474892053708 8 | | | |
 - 当前节点 [新增](#)

| 输出名称 | 节点名 | 节点ID (实例ID) | 负责人 | 操作 |
|--------------------------|--------------|-------------|-----|----|
| demo_devs_od_acc ount | s_od_account | | | |

[取消](#) [确定](#)

时间属性

- **正常调度**：按照配置的调度周期正常调度执行。
- **空跑调度**：按照配置的调度周期正常调度执行，但都是空跑。调度到该任务时，会直接返回成功，不会执行实际的任务代码。
- **暂停调度**：暂停调度任务，按照配置的调度周期正常调度执行。调度到该任务时，会直接返回失败，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。

调度周期

调度周期包括日、周、月、小时和分钟。

- **日调度**：每日当上游节点运行成功后，会按照设置的时间运行。即当上游节点全部执行成功且定时时间到达，才会执行该任务。默认运行时间是00:00:00。

• 调度配置

时间属性 正常调度 空跑调度 暂停调度

调度周期 日

依赖上周期

- **周调度**：指定每周几（可多选）调度。未被选定的时间段，会执行空跑调度。

• 调度配置

时间属性 正常调度 空跑调度 暂停调度

调度周期 周

依赖上周期

- **月调度**：指定每月的具体日期（可多选）调度。未被选定的时间，会执行空跑调度。

• 调度配置

时间属性 正常调度 空跑调度 暂停调度

调度周期 月

依赖上周期

- **小时调度**：每日时间里每N个小时调度一次。

• 调度配置

时间属性 正常调度 空跑调度 暂停调度

调度周期 小时 时间段 时间点

开始 00:00 间隔 10 小时 结束 00:00

- **分钟调度**：每日时间里每N分钟调度一次。

• 调度配置

时间属性 正常调度 空跑调度 暂停调度

调度周期 分钟 v

开始 00:00 ⌚ 间隔 20 v 分钟 结束 00:00 ⌚

依赖上周期

勾选依赖上周期，表示当前节点的执行时间依赖其他节点（也可以选当前节点）的上一个周期。

• 调度配置

时间属性 正常调度 空跑调度 暂停调度

调度周期 分钟 v

开始 00:00 ⌚ 间隔 20 v 分钟 结束 00:00 ⌚

该任务配有针对未完成的监控报警，更改调度周期将导致对应监控规则失效，请谨慎修改！

依赖上周期 自定义 v n_1446476114 x

7. 填写依赖关系。

调度配置

描述 请输入50个字符内的描述文案

参数配置 请输入配置的参数，支持自定义参数，以英文分号分隔，如keyA=123;keyB=\${yyyyymmdd-7}

调度配置

时间属性 正常调度 空跑调度 暂停调度

调度周期 日 00:00

依赖上周期

依赖关系

上游依赖 新建上游依赖

| 输出名称 | 节点名 | 节点ID | 负责人 | 操作 |
|---------------------------|---------------------------|-----------|----------------------------------|----|
| virtual_root_node_1060008 | virtual_root_node_1060008 | n_1060009 | 18228972862627182115289738813070 | |

< 1 >

当前节点 新增

| 输出名称 | 节点名 | 节点ID | 负责人 | 操作 |
|------|-------------|-----------|-----|----|
| node | dataphin_01 | n_1180000 | - | |

< 1 >

取消 确定

a. 设置上游依赖。设置当前节点的上游依赖节点，要依赖的节点必须是已经存在的。



- 当任务的节点类型是SQL类时，您可以单击开始解析，系统会将SQL中依赖的表自动解析为当前节点的上游依赖。
- 您也可以单击新建上游依赖，输入节点输出名称搜索，找到需要的上游依赖，单击确定新增，完成上游依赖的设置。



说明:

- 上游依赖节点的输出名称是全局唯一的，不同节点的输出名称不能相同。通常情况下，SQL节点的输出名称是表名。

- 每个租户（企业）在初始化时都会有一个根节点，是以virtual开头的虚拟节点。

b. 设置当前节点的输出名称。单击新增，填写节点输出名称，单击确定新增。

当前节点

1

| 输出名称 | 节点名 |
|----------------------|----------------------|
| <input type="text"/> | <input type="text"/> |

2

3

新增当前节点输出

仅支持英文大小写、数字、点(.)、下划线(_)、中划线(-)!

当前节点

| 输出名称 | 节点名 | 节点ID | 负责人 | 操作 |
|------------------------------------|-----------------------------|-----------|-----|----|
| tmp_xsf_import_sqlserver_data_dest | xsf_import10_odps2sqlserver | n_1147885 | | |



说明:

- 新建上游依赖时，搜索上游节点就是根据节点的输出名称来搜索的。节点的输出名称是节点的产出信息。
- 当前节点的输出名称可以设置为多个，是供其他节点依赖使用的。输出名称的命名规则请尽量统一，便于其他用户创建节点时容易找到上游节点。

8. 调度配置完成后，单击确定。



9. 保存和提交同步任务。



2.7 执行与管理同步任务

本文为您介绍同步任务的执行和管理操作。

执行同步任务

请确认您创建的同步任务已提交，且已成功发布到生产环境，发布操作请参见[发布管理](#)。

如果您创建的同步任务的调度类型为周期任务，系统会每天自动生成调度任务。如果您创建的同步任务的调度类型为手动任务，需您手动触发调度任务。手动触发的操作步骤如下：

1. 单击研发 > 运维，进入运维中心。



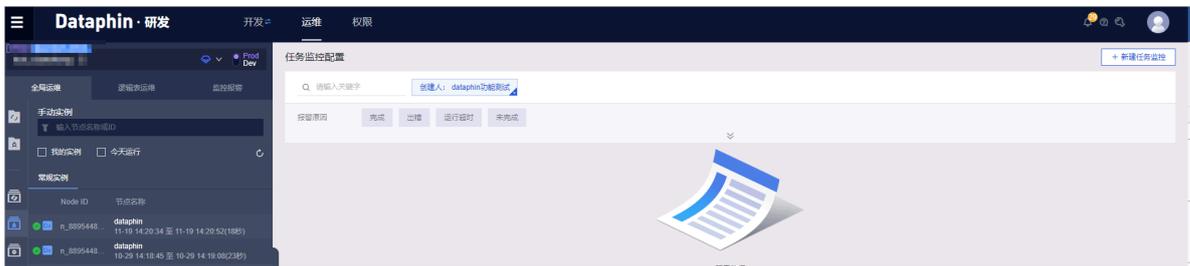
2. 单击切换到生产环境，单击左侧导航栏中的，进入手动任务页面。



3. 单击页面右上角的运行，即可触发调度类型为手动任务的同步任务。

验证同步任务执行结果

1. 在生产环境中的运维页面，单击左侧导航栏中的  或 ，进入手动实例或周期实例页面，查看实例是否运行成功。



2. 运行成功的实例，可以通过即席查询验证同步任务。单击  到开发环境，选择开发 > 即席查询，进入即席查询页面。通过编写 `select * from 目标表名称 where ds='任务调度日期'` SQL语句查询同步任务执行的结果，详细内容请参见[即席查询](#)。

管理同步任务

· 编辑、重命名、移动

1. 登录Dataphin。
2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。
3. 选择开发 > 数据处理 > 同步任务。



4. 将鼠标悬停于某个同步任务上，单击后面浮现的  按钮，选择编辑、重命名、移动，可以对该同步任务进行编辑、重命名、移动等操作。



| 操作项 | 说明 |
|-----|--|
| 编辑 | 可以修改同步任务的同步来源、同步目标、字段同步、参数控制、调度配置等信息，详情请参见 配置同步任务 和 调度配置 。 |
| 重命名 | 可以修改同步任务的名称。 |
| 移动 | 可以将同步任务移动至指定的文件目录下。 |

- 上线、下线、偷锁、删除

同步任务文件默认是由创建人锁定的。在有操作权限的情况下，对其他用户创建的同步任务进行编辑、上线（提交/发布）、下线、删除等操作时，系统会提示不可操作，需要先偷锁。



- 定位

在页面右侧，单击各个任务的页签切换时，左侧功能区的任务列表不会随之切换。此时，可以单击右上方的定位按钮，快速找到该任务在左侧任务列表中的位置。



· 搜索、刷新

- 您还可以单击刷新按钮来刷新同步任务当前的状态等信息，如下图中编号1所示。
- 在同步任务页面，支持用户输入任务名称的关键字来搜索同步任务，如下图中编号2所示。
- 将鼠标悬停于某个同步任务上，可以查看该任务的状态、创建人、创建时间，如下图中编号3所示。



3 规范建模

3.1 数据建模研发概述

数据建模研发，是基于数据仓库理论的工具化，提供了体系化、系统化建模及研发的功能。您可以通过数据建模研发功能，自顶向下快速构建企业维度、业务过程，并进一步细化开发维度表、事实表、汇总表、应用层，沉淀出企业数据公共层，便于业务分层的数据应用，同时优化计算和存储。

进入数据建模研发页面

通过Dataphin首页的研发，或者研发工作台的快速入口，都可以进入到数据建模研发页面。



数据研发模式分为Basic模式和Dev-Prod模式。进入数据建模研发界面时，需要选择项目空间（目前有Basic模式项目和Dev-Prod模式项目），在对应的项目空间内进行数据规范建模作业。两种模式的项目介绍看参见[#unique_16](#)。

进入项目空间之后，可以根据默认的操作流程指引图（请参见[使用说明](#)）了解规范定义的维度、业务过程、业务限定、原子指标、派生指标、建模研发的维度逻辑表、事实逻辑表、汇总逻辑表，以及其他开发工作内容及相互关系。

在左侧功能区，鼠标悬停于任务之上，会提示任务当前的发布状态等信息。



进行数据建模研发之前，首先需要将外部数据引入到项目绑定的数据源中。关于如何引入数据，请参见数据引入文档[创建同步目标表](#)。

Dataphin公测期间，支持华东2（上海）、华南1（深圳）、华北2（北京）区域的智能研发版、基础研发版和数据萃取、数据服务、资产治理增值功能包，详情请参见[#unique_17](#)。

数据建模前提

- 仅绑定业务板块的项目可以通过规范建模功能在项目空间进行开发任务。如绑定错误或未绑定业务板块，则规范建模功能不可用。关于项目和业务板块的绑定，请参见[#unique_16](#)。
- Basic模式下的Basic项目不支持执行规范建模操作，仅可以使用数据处理和即席查询功能。
- Dev-Prod模式下，开发环境的项目可以直接创建、编辑、删除任务操作。新建完成并提交到待发布列表中，需要执行发布操作。发布成功后可以在对应生产环境下进行自动调度，并可查看。

页面介绍

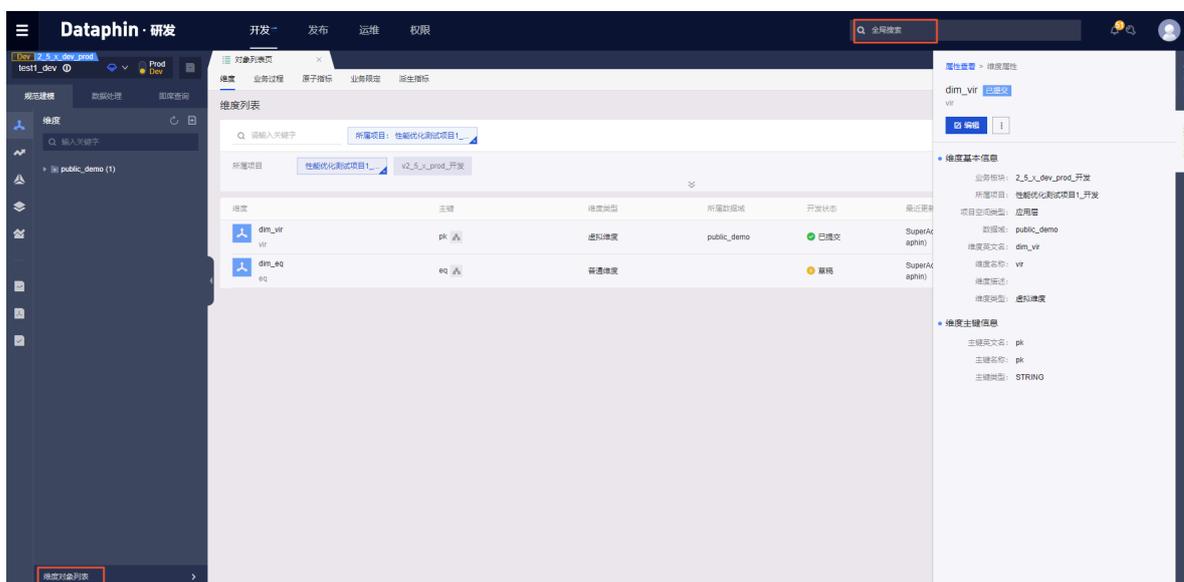
1. 登录Dataphin。

2. 在Dataphin首页，单击顶部菜单栏的研发，或者通过研发工作台的快速入口进入研发页面。



3. 在研发页面，单击左侧导航栏规范建模中下方的维度对象列表，进入所有对象的列表查看管理页面。

- 通过签换页签，查看维度、业务过程、原子指标、业务限定、派生指标信息。
- 您也可通过在搜索框中输入关键字进行快速搜索。
- 通过单击某个维度、业务过程、原子指标、业务限定、派生指标，在右侧弹框中查看基本信息。
- 您还可以通过界面右上方的全局搜索框搜索和查看任务。在搜索框内输入任务关键词，可以通过任务类型、归属范围进行快速筛选。其中，仅购买了数据萃取功能的用户可以在全局搜索框查询萃取对象。



3.2 规范定义-维度

3.2.1 新建维度

维度是人们观察事物的角度，是指一种视角，是确定事物的多方位、多角度、多层次的条件和概念。本文为您介绍如何新建维度。

背景信息

维度即进行数据统计的对象。通常情况下，维度是实际存在、不因事件发生而存在的实体，例如时间维度、地区维度、产品维度等。创建维度，即从顶层规范业务中实体（或称主数据）的存在性及唯一性。

操作步骤

1. 登录Dataphin。

2. 在Dataphin首页，单击顶部菜单栏的研发，或者通过研发工作台的快速入口，进入研发页面。



3. 选择开发 > 规范建模 > 维度。



4. Basic模式下的Basic项目和Dev-Prod模式下的Dev项目可以直接新建维度。在开发环境新建完成并提交后，再执行发布操作，生产环境中对应的项目才会新增维度。您可以通过以下任何一种方式进入新建维度页面：

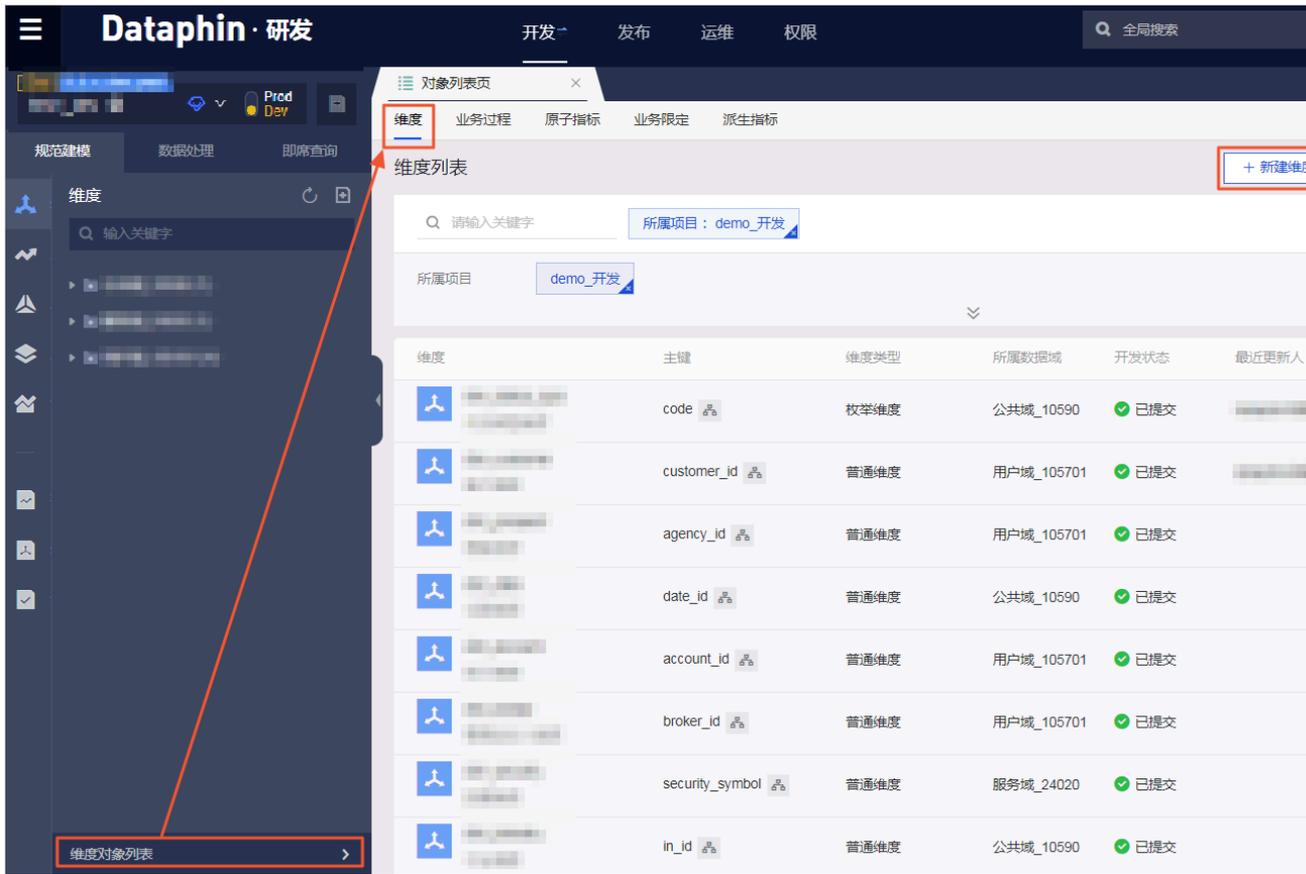
- 单击项目名称右侧的  按钮，选择规范定义 > 维度。



- 单击左侧功能区右上方的新建维度按钮。



- 单击左侧功能区下方的维度对象列表，切换至维度页签，然后单击页面右侧的新建维度。



- 在右侧的操作指引图上，单击维度下方的  按钮。



5. 在新建维度页面，根据页面提示配置维度基本信息。

新建维度

维度基本信息

业务板块 [模糊] 所属项目 [模糊]

项目空间类型 应用层 * 数据域 请选择数据域

* 维度英文名 dim_ 请输入维度英文名 * 维度名称 请输入维度名称

维度描述 请输入维度描述

| 参数 | 说明 |
|-------|---|
| 维度英文名 | 填写维度英文名。维度英文名由字母、数字、下划线组合组成，且默认以dim_作为前缀。 |
| 维度名称 | 填写维度名称。维度名称由汉字、字母、数字、下划线或-组合组成。 |
| 维度描述 | 对维度简单的描述。 |
| 数据域 | <p>选择该维度所属的数据域。如果您还没有创建该维度所属的数据域，请参见#unique_20/unique_20_Connect_42_section_bmc_31g_2hb创建数据域。</p> <p> 说明： 数据域的划分请参见#unique_21。</p> |

6. 配置维度逻辑信息。

维度逻辑信息用于定义维度对象范围和对象特征，确保维度真实且唯一性存在。Dataphin为您提供四种维度类型，以满足不同维度对象建设需求。



| 参数 | 说明 |
|-----------|--|
| 普通维度 | 需定义主键及维度主从关系。普通维度创建完成并已成功发布后自动生成对应的维度逻辑表。 |
| 普通维度 (层级) | 需定义主键、来源逻辑及层级，支持使用多个维度进行分析，不同维度之间存在稳定的层级关系。维度创建完成并已成功发布后自动生成一组有层级关联关系的维度逻辑表。 |
| 枚举维度 | 需定义主键，支持列举及标准化枚举值以标识枚举维度对象。 |
| 虚拟维度 | 需定义主键，支持没有具体业务实体承载、没有可固化数据范围逻辑定义的维度对象，也可以规范化业务建模。 |

- 如果您选择普通维度，根据页面提示完成参数配置。



| 参数 | 说明 |
|-------|-------------------------------|
| 主键英文名 | 填写维度的主键英文名。英文名由字母、数字或下划线组合组成。 |

| 参数 | 说明 |
|--------|--|
| 主键名称 | 填写维度的主键名称。主键名称由汉字、数字、字母、下划线或-组合组成。 |
| 主键类型 | 选择主键类型。主键类型包括：STRING、BIGINT、DOUBLE、DATETIME。 |
| 主键计算逻辑 | 填写维度主键计算逻辑，代码中需限制时间分区条件，否则可能出现运维报错或数据错误： a. 单击主键计算逻辑后面的参考示例，可以查看SQL编写示例，例如。 <pre>select province from dataphin_test where ds = '\${bizdate}'</pre> b. SQL编写完成后，单击主键计算逻辑后面的规范性校验，可以校验您编写的SQL是否符合规范。 |
| 所属主维度 | 可选择是否，如果选择是，同时请您在右侧下拉框中选择所属的主维度。 |

· 如果您选择普通维度（层级）。

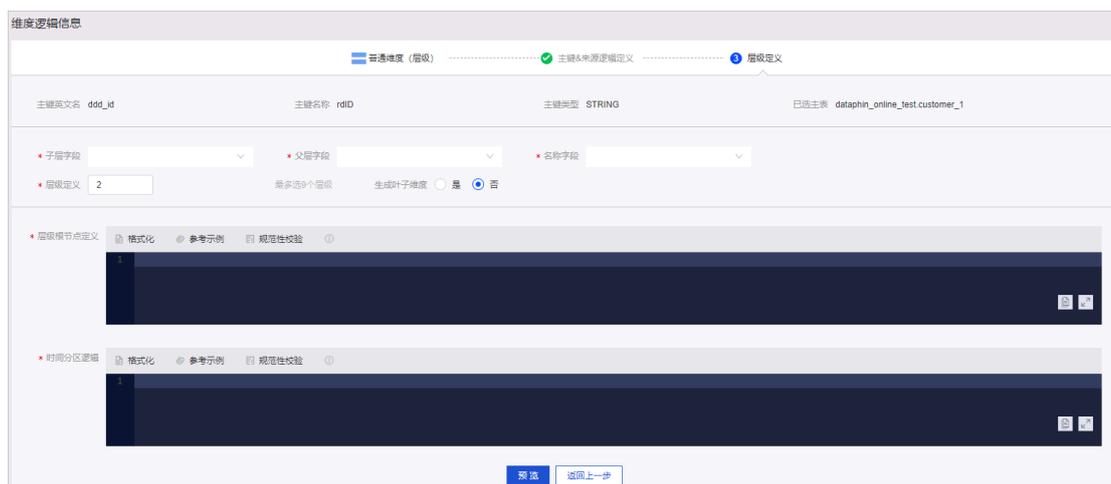
a. 在主键&来源逻辑定义，根据页面提示完成参数配置，并单击下一步。



| 参数 | 说明 |
|-------|--|
| 主键英文名 | 填写维度主键英文名。英文名由字母、数字或下划线组合组成。 |
| 主键名称 | 填写主键名称。维度名称由汉字、字母、数字、下划线或-组合组成。 |
| 主键类型 | 选择主键类型。主键类型包括：STRING、BIGINT、DOUBLE、DATETIME。 |

| 参数 | 说明 |
|------|---|
| 主表表名 | 选择主表表名，即同步任务中的目标表。  说明： 建议您选择生产环境中的表。 |

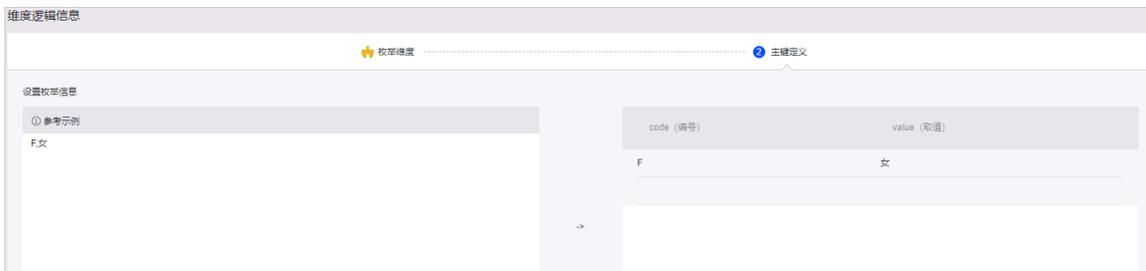
b. 在层级定义，根据页面提示完成参数配置。



| 参数 | 说明 |
|---------|--|
| 子层字段 | 选择层级字段中的子层字段。 |
| 父层字段 | 选择层级字段中的父层字段。 |
| 名称字段 | 选择名称所在字段。 |
| 层级定义 | 填写层级的个数，支持最多9个层级。 |
| 生成叶子维度 | 选择是否生成叶子维度。如果选择是，系统会自动生成叶子维度，如果选择否，您需定义叶子维度的生成逻辑。 |
| 层级根节点定义 | 基于来源主表字段，设定根层级（即无父层级）的判断条件，例如 is_parent='ture'，其中 is_parent来源主表字段。 |

| 参数 | 说明 |
|--------|--|
| 时间分区逻辑 | 基于来源主表字段，设定数据更新及获取时间周期的判断条件，例如 <code>ds=\${bizdate}</code> 。 |

- 如果您选择枚举维度，根据页面提示完成参数配置。



在设置枚举信息的左侧框中，填写枚举维度成员与对应成员的信息，并以英文逗号“,”隔开，例如。



右侧的codevalue，根据您在左侧填写的信息自动解析填充。

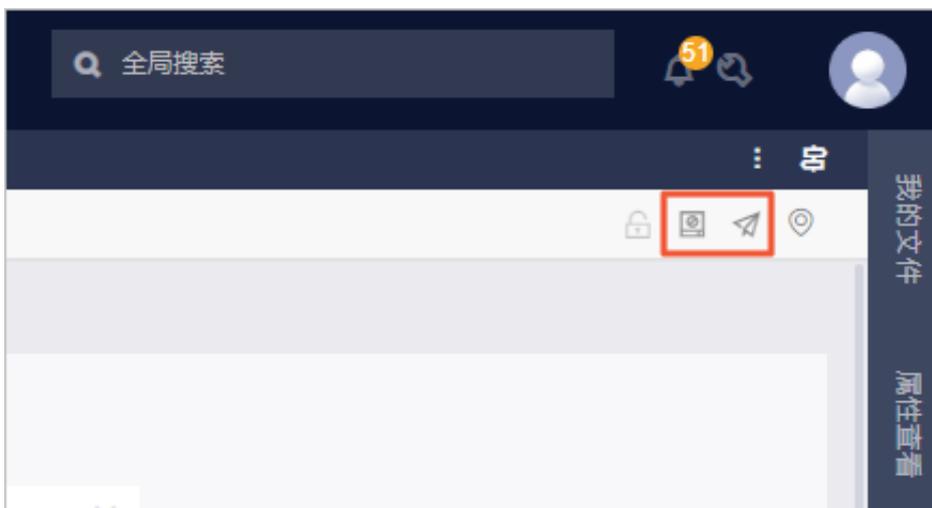
- 如果您选择虚拟维度，根据页面提示完成参数配置。



| 参数 | 说明 |
|-------|--|
| 主键英文名 | 填写维度主键英文名。英文名由字母、数字或下划线组合组成。 |
| 主键名称 | 填写主键名称。维度名称由汉字、字母、数字、下划线或-组合组成。 |
| 主键类型 | 选择主键类型。主键类型包括：STRING、BIGINT、DOUBLE、DATETIME。 |

 **说明:**
 如果虚拟维度为维度逻辑表或事实逻辑表的关联维度，则新建派生指标可以使用虚拟维度，同时作为统计粒度的虚拟维度的值从被关联的字段中选取。

7. 维度信息填写完成后，单击页面右上方的、保存和提交新建的维度。



8. 提交成功后，完成维度的新建。发布维度，请参见[发布管理](#)：

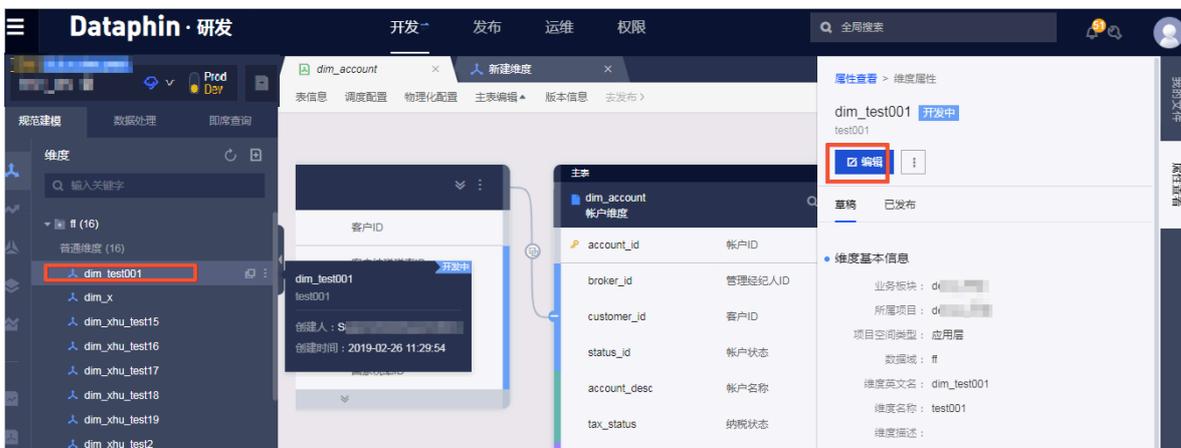
- 普通维度提交成功后，会自动生成对应的维度逻辑表。
- 普通维度（层级）提交成功后，会自动生成一组具有层级关联关系的维度逻辑表。
- 维度逻辑表详情请参见[编辑模型信息和表信息](#)。

3.2.2 修改维度

本文为您介绍如何修改维度信息。

修改维度

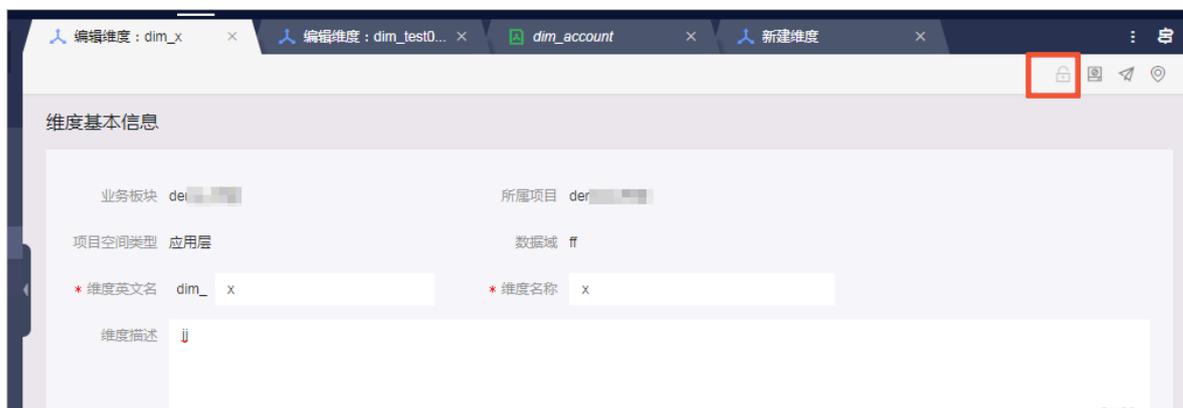
1. 进入研发 > 开发 > 规范建模 > 维度功能菜单。
2. 在左侧列表中，单击选择需要修改的维度，弹出右侧的维度信息框，单击编辑即可进入修改页面，如下图所示。



 说明：

Basic模式下的项目和Dev-Prod模式的开发环境项目可直接修改维度（Dev-Prod模式中的开发环境修改维度后，需要执行发布操作，发布成功后生产环境对应的维度才会更新）。

3. 偷锁。已发布的维度规范定义，如果创建人不是本账户，需要单击偷锁进行解锁，才可以进行维度定义修改，如下图所示。



4. 修改完成后，单击右上角的提交按钮即可。

维度主键信息修改后，如需要对应维度逻辑表也生效，需要对维度逻辑表处也执行一次提交。如果是Dev-Prod模式，且该维度和维度逻辑表已发布到生产环境，需要再执行一次发布操作，才会对生产环境的维度和维度逻辑表更改生效。

3.2.3 下线或删除维度

维度下线后，维度定义的当前状态变为草稿状态。维度删除后，维度定义被删除。本文为您介绍维度的下线和删除操作。

维度下线时，对应维度逻辑表会被删除，以保证生产环境无维度表可使用。

- Basic模式和Dev-Prod模式的开发环境可以直接下线或删除。
- Dev-Prod模式的生产环境，需要在开发环境中下线或删除操作完成后，再执行发布操作，才会下线或删除成功（发布时，维度与维度逻辑表都需要发布）。

下线或者删除前提是无依赖引用，包括对应维度逻辑表无使用、维度没有被关联、无子维度等。

1. 进入研发 > 开发 > 规范建模 > 维度功能菜单。

2. 下线或删除维度共有3个操作入口，具体如下：

- 鼠标悬停于某个维度上，单击右侧浮现的“”图标，即可显示下线、删除功能。
- 单击左侧列表中的某个维度，页面右侧会弹出该维度的属性信息框，单击“”图标，即可下线或删除维度。
- 在“维度对象列表页”，单击维度后面的“”图标，即可显示下线、删除功能。

3. 单击下线或删除功能，即可实现维度的下线或删除。如果下线、删除需要在生产环境生效，则相关操作需要提交并[发布管理](#)。



说明：

- 维度下线时，会使对应的维度逻辑表被删除，以保证生产环境无维度表可使用。
- 下线或者删除维度的前提是无依赖引用，包括对应维度逻辑表无使用、维度没有被关联、无子维度等。

3.3 规范定义-业务过程

3.3.1 新建业务过程

业务过程即业务活动中所有的事件，为不可拆分事件。创建业务过程，即从顶层视角，规范业务活动中事件的内容类型及唯一性。

1. 登录Dataphin。

2. 单击研发 > 开发 > 规范建模 > 业务过程，进入业务过程页面。



3. Basic模式下的Baisc项目和Dev-Prod模式下的Dev项目可以直接新建业务过程。在Dev项目中完成业务过程的新建并提交后，再执行发布操作，Prod项目中才会新增该业务过程。您可以通过以下任何一种方式进入新建业务过程页面：

- 单击业务过程后的。



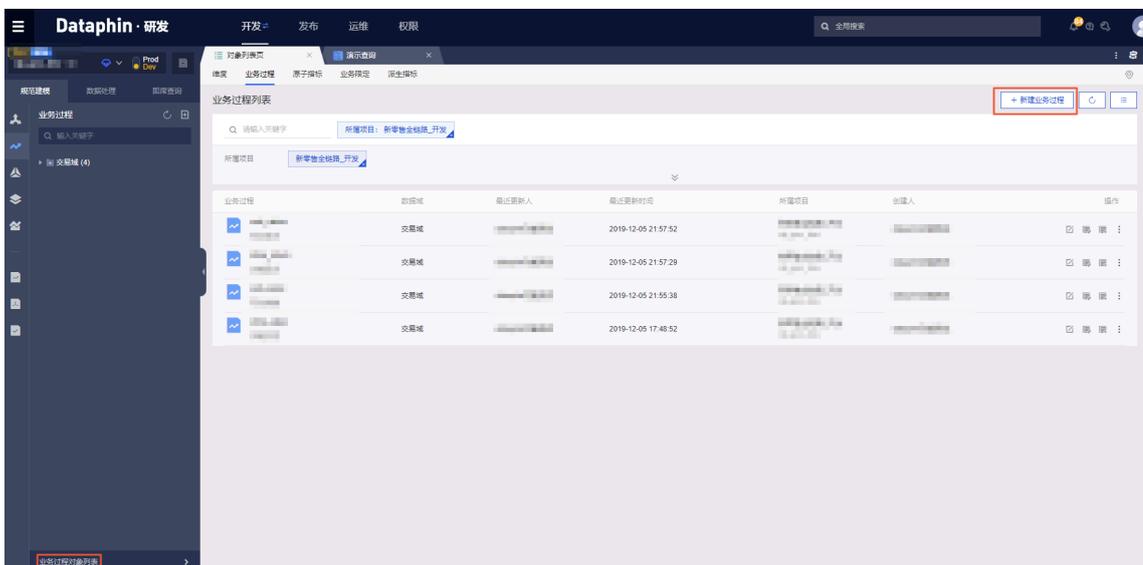
- 在右侧的操作指引图上，单击业务过程下方的。



- 单击项目名称右侧的按钮，选择规范定义 > 业务过程。



- 单击左侧功能区下方的业务过程对象列表，然后单击页面右上方的新建业务过程。



4. 在新建业务过程页面，根据页面提示配置参数。

| 参数 | 说明 |
|-----|---------------------------------|
| 数据域 | 选择数据域。 |
| 英文名 | 填写业务过程英文名。英文名由字母、数字、下划线组合组成。 |
| 名称 | 填写业务过程名称。名称由汉字、数字、字母、下划线或-组合组成。 |
| 描述 | 填写简单的描述。 |

5. 完成参数配置并单击提交。在弹出的提交备注页面，填写备注以说明此次提交的变更内容，并单击确定并提交，即可完成业务过程的新建。

3.3.2 管理业务过程

本文为您介绍业务过程的修改、删除、克隆，帮助您更好的管理业务过程。

修改业务过程

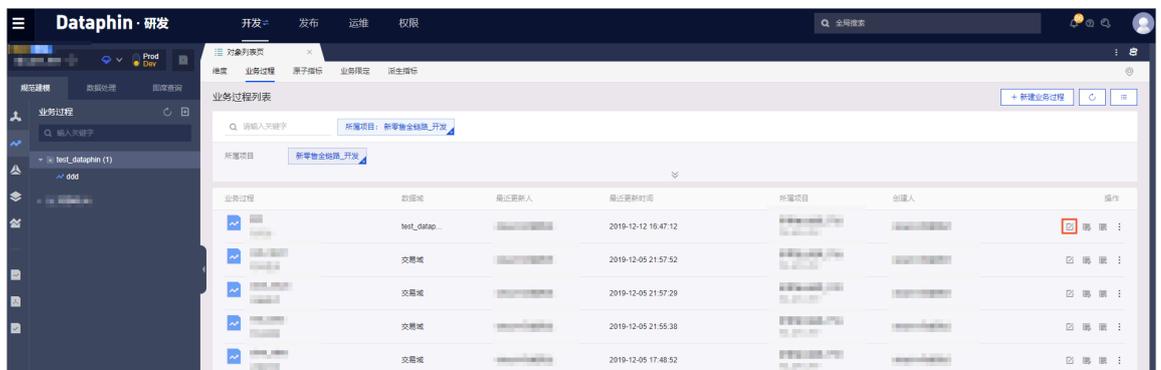
1. 登录Dataphin。
2. 单击研发 > 开发 > 规范建模 > 业务过程，进入业务过程页面。

3. Basic模式下的Baisc项目和Dev-Prod模式下的Dev项目可以直接修改业务过程。在Dev项目中完成业务过程的修改并提交后，再执行发布操作，Prod项目中才会修改该业务过程。您可以通过以下任何一种方式进入编辑业务过程页面：

- 将鼠标悬浮在 ，单击编辑。



- 单击左侧导航栏下方的业务过程对象列表，然后单击操作列下的 .



- 单击需编辑的业务过程，右侧弹出该业务过程的属性框，单击编辑。



4. 在编辑业务过程页面，可修改数据域、英文名、名称、描述，参数的详细内容请参见[新建业务过程](#)。



- 5. 完成参数配置并单击提交。在弹出的提交备注页面，填写备注以说明此次提交的变更内容，并单击确定并提交，即可完成业务过程的修改。

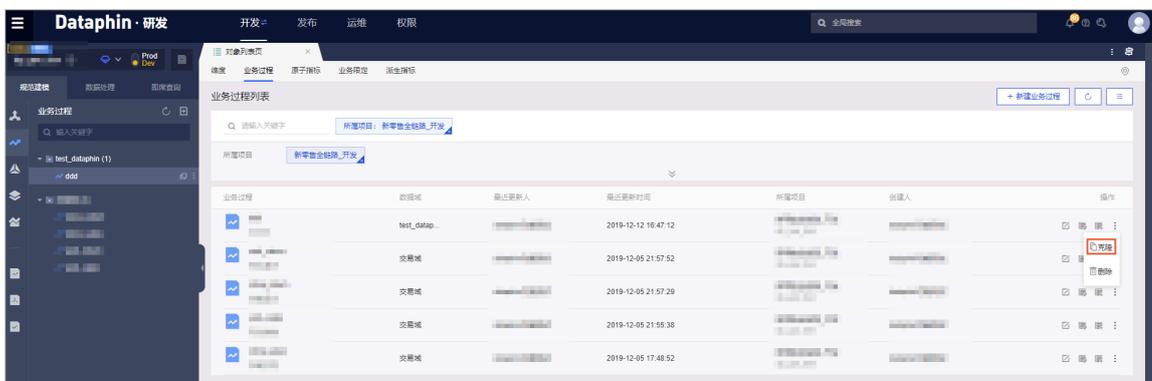
克隆业务过程

- 1. Basic模式下的Baisc项目和Dev-Prod模式下的Dev项目可以直接克隆业务过程。在Dev项目中完成业务过程的克隆并提交后，再执行发布操作，Prod项目中才会有该业务过程。您可以通过以下任何一种方式进入新建业务过程页面：

- 单击需克隆的业务过程，右侧弹出该业务过程的属性框，将鼠标悬浮在 ，单击克隆。



- 单击左侧导航栏下方的业务过程对象列表，将鼠标悬浮在  后，单击克隆。



2. 在新建业务过程页面，可修改数据域、英文名、名称、描述，参数的详细内容请参见[新建业务过程](#)。

新建业务过程

业务板块 [模糊] 所属项目 [模糊] 项目空间类型 公共层

* 数据域 test_data... ▾

* 英文名 ddd

* 名称 对对对

描述

0/128

取消 提交

3. 完成参数配置并单击提交。在弹出的提交备注页面，填写备注以说明此次提交的变更内容，并单击确定并提交，即可完成业务过程的克隆。

删除业务过程

1. Basic模式下的Baisc项目和Dev-Prod模式下的Dev项目可以直接删除业务过程。在Dev项目中完成业务过程的删除后，再执行发布操作，Prod项目中才会删除该业务过程。您可以通过以下任何一种方式进入删除业务过程的提示操作页面：

- 将鼠标悬浮在 ，单击删除。



- 单击需删除的业务过程，右侧弹出该业务过程的属性框，将鼠标悬浮在 ，单击删除。



- 单击左侧导航栏下方的业务过程对象列表，将鼠标悬浮在 ，单击删除。



2. 在提示框中，填写备注信息并单击确定并提交。

提示 ×

 确认删除吗?

* 备注 请输入此次提交的相关备注说明，如变更信息

0/128

特殊说明：当前项目，提交成功后，请前往待发布列表，以发布对象至生产环境。

取消 确定并提交



说明:

仅限无依赖引用（即无相关事实逻辑表）的业务过程支持删除操作。

3.4 逻辑表-维度逻辑表

3.4.1 编辑模型信息和表信息

维度逻辑表与维度一一对应，是通过丰富维度中的属性信息构建而成的。维度发布成功后，系统会自动创建对应的维度逻辑表，并支持对维度逻辑表进行添加属性、关联维度、添加子维度、物理化配置等操作。

注意事项

编辑维度逻辑表的核心目标是丰富维度之上的相关属性信息。需要注意的是：

- 抽象提炼维度自身具备且相对持久的属性，无需冗余非维度关联的对象信息。
- 子维度中不需要重复定义已发布的父维度中相同属性字段，系统会自行引用。

操作步骤

1. 登录Dataphin。
2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。

3. 选择开发 > 规范建模 > 维度逻辑表。



4. 单击左侧列表中的某个维度逻辑表，进入逻辑表编辑页面。



说明:

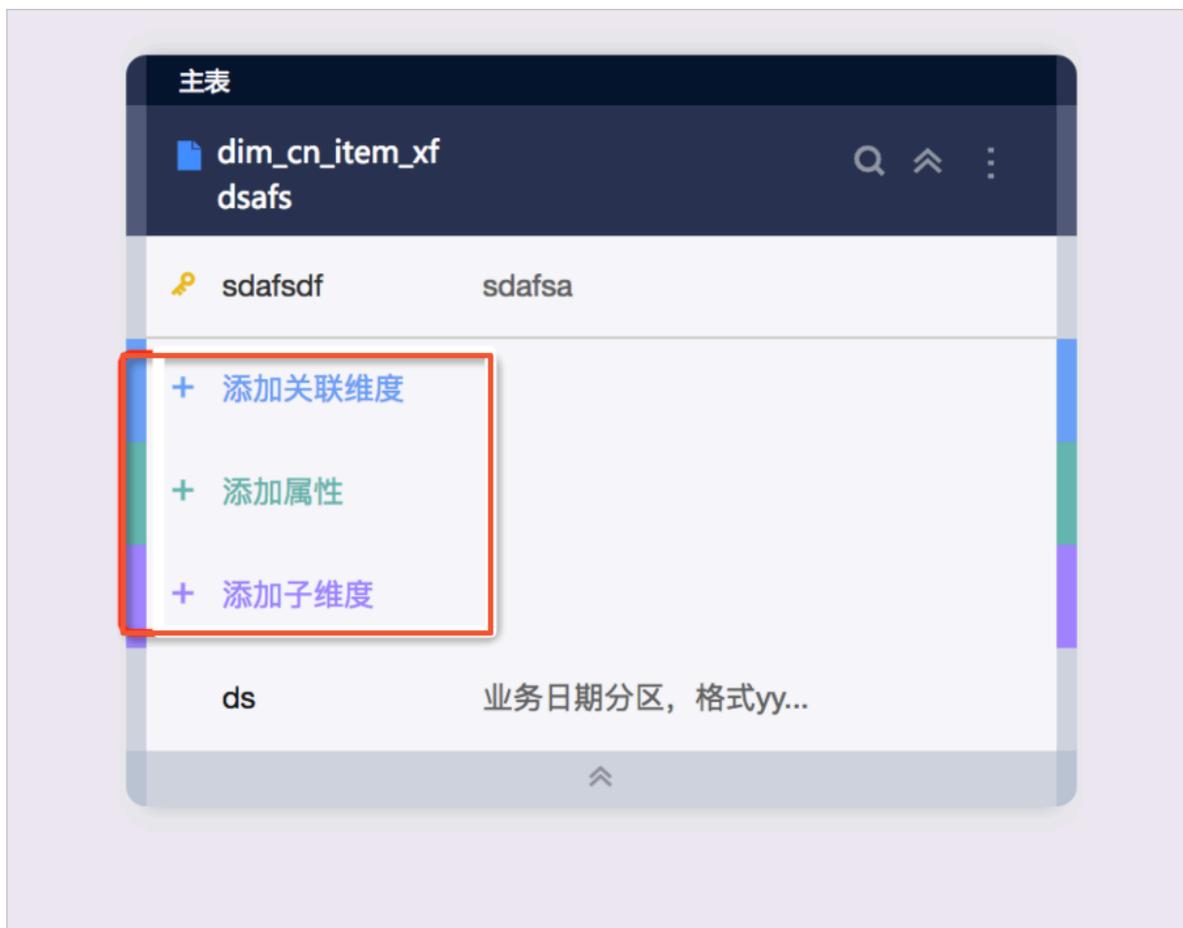
Basic模式项目和Dev-Prod模式的开发环境项目可以直接编辑维度逻辑表。如果需要修改Dev-Prod模式的生产环境中的维度逻辑表，需要开发环境修改维度逻辑表后，提交发布才会修改成功。

5. 逻辑表编辑页面分为如下几个区域:

- 在左上方菜单栏，您可以查看维度逻辑表的表信息、配置调度参数等。
- 在右上方区域，您可以对维度逻辑表进行保存、提交等操作。
- 在中间区域，您可以查看逻辑表信息、添加关联维度、添加子维度等。

6. 编辑模型信息。

模型信息为您展示维度逻辑表的表名称、字段清单，以及与其他逻辑表的关系。如下图所示，您可以快速实现添加属性、添加关联维度、添加子维度。同时，您也可以直接在逻辑表编辑页面，单击上方菜单栏的主表编辑，编辑逻辑表相关内容。



a. 添加属性。

单击模型信息中的添加属性，进入新建属性页面。系统支持通过引入字段和SQL自定义字段两种方式创建属性。

- 引入字段方式会自动生成计算逻辑，生成后用户可以编辑计算逻辑。
- 如果字段具有复杂的计算逻辑，您可以选择SQL自定义字段方式，自行编写自定义SQL。以as k标识关联键，其他字段默认为引入的属性。单击参考示例可以查看SQL编写样例，如果您不确定，可以单击规范性校验验证您编写的SQL语句是否正确。

b. 添加关联维度（构建星型模型或雪花模型）。

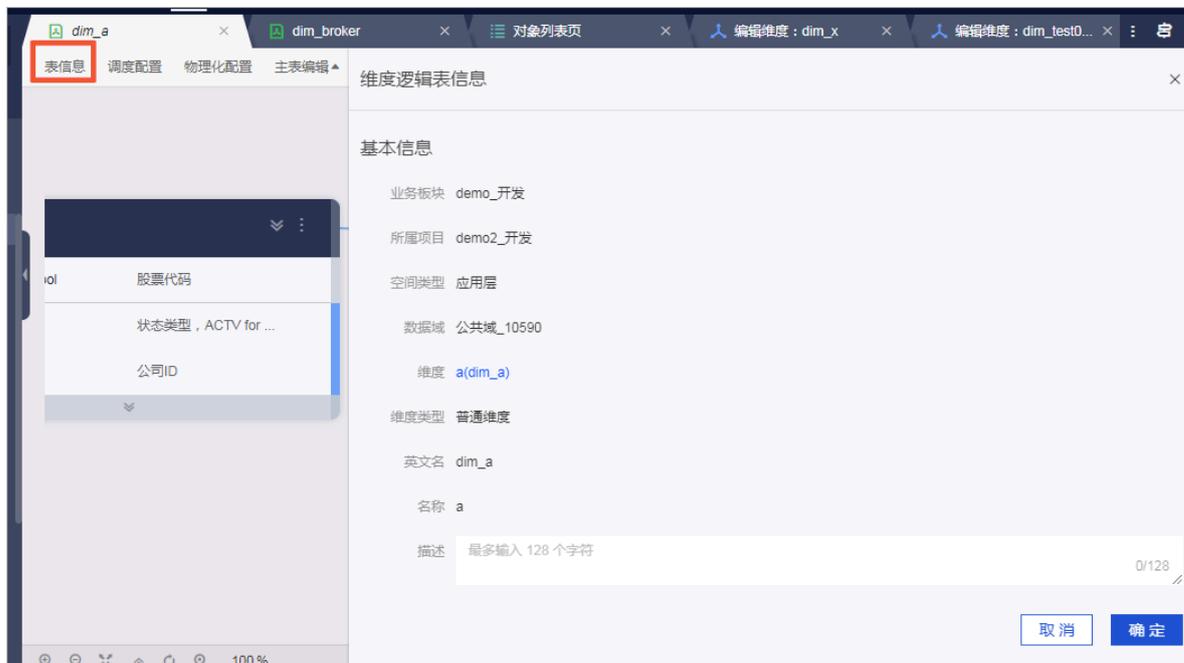
单击模型信息中的添加关联维度，选择需要关联的维度和属性、编辑维度角色信息之后，即可完成与其它逻辑表的关联。

c. 添加子维度。

单击模型信息中的添加子维度，系统自动跳转至新建维度页面，并完成父维度绑定。提交、发布新建的维度之后，即完成子维度添加。

7. 编辑表信息。

在逻辑表编辑页面，单击上方菜单栏的表信息，页面右侧会弹出窗口，为您展示维度逻辑表信息。您可以修改描述，完成后单击确定即可。



3.4.2 主表编辑

本文为您介绍如何进行主表编辑，并查看表计算逻辑。

1. [进入编辑逻辑表页面](#)，单击上方菜单栏中主表编辑，打开主表编辑页面。
2. 在主表编辑页面，可以编辑属性、编辑与其他逻辑表的关联关系、添加属性、添加关联维度、定义公用计算逻辑。

定义公用计算逻辑：支持逻辑表内可公共使用的SQL片段沉淀并定义为公用计算逻辑，以保证相同的数据处理逻辑得到快捷、方便的复用，达到临时表或者中间表效果。

3. 单击维度逻辑表名称旁边的“”图标，可以查看表计算逻辑，同时支持进行本地与已发布版本的代码对比。



说明：

- 系统会自动将生产业务板块、生产项目生成对应的变量。例如您拥有业务板块LD_Trade，则系统自动生成业务板块变量\${LD_Trade}。该变量在开发环境执行时默认被替换为

LD_Trade_dev，在生产环境执行时默认被替换为LD_Trade。您也可以在执行时设置固定的值，提高代码在不同环境执行时的灵活性。

- 逻辑表编辑时，来源表建议选择生产环境的表（如果有生产环境）。代码里表名的命名空间前缀建议取项目或业务板块的变量（或者直接配置生产环境表），以保证逻辑表发布至生产环境，生产数据时不引用开发环境的数据，避免出现数据错误。

3.4.3 物理化配置

本文为您介绍维度逻辑表的物理化配置。旨在对维度逻辑表进行自动化代码生成并物理表化前，获得用户基于业务的考虑输入。

1. **进入编辑逻辑表页面**，单击上方菜单栏中物理化配置，右侧弹出物理化配置窗口，可以根据界面提示配置相关信息，配置完成后单击确定。

- 存储生命周期为必选参数，生命周期默认值为36000天，也可通过快速选择进行修改。



说明:

Dataphin系统升级后，生命周期默认值由原来的30天更新36000天。如果您在升级之前进行的物理化配置，升级后生命周期由原来的30天更新为36000天。

- 选择分区字段为必选参数，默认值为ds分区字段。逻辑表自动创建的ds分区字段。
 - 无需物理化字段为可选参数，分区字段不可做为无需物理化字段。
 - 自定义参数设置为可选参数，自定义参数配置请参见[#unique_31](#)。
2. 维度逻辑表发布后，后台代码会自动化生成并进行周期调度。逻辑表发布需要生产账号，而非个人账号，并且对所有表有权限。

3.4.4 调度配置

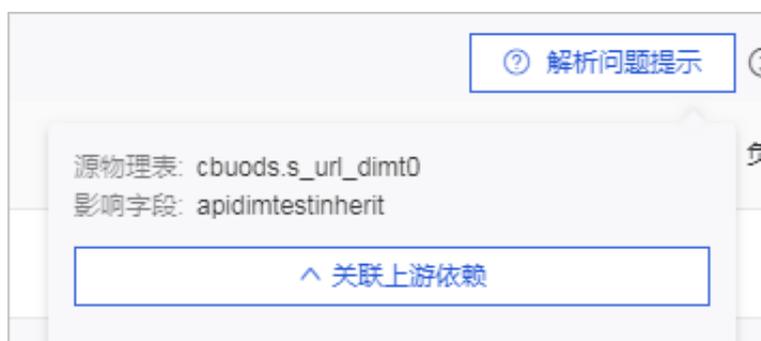
逻辑表支持查看和自定义调度配置信息。

进入编辑逻辑表页面，单击上方菜单栏中调度配置，可以查看及配置逻辑表的基本信息、调度周期、调度依赖三大项内容。配置信息会直接作用于逻辑表的调度生产、代码的自动化生成以及执行。

各项配置具体功能说明如下：

- **基本信息**：根据提示填写内容即可。其中，参数配置支持自定义参数输入，以英文分号（;）作为区隔。
- **调度周期**：可以选择是否依赖上周期，如果选择依赖，则表示此逻辑表本周期将于上周期完成后执行调度。
- **调度依赖**：分为自动解析、上游依赖和逻辑表输出三部分。
 - **自动解析**：通过代码内容自动解析上游依赖和逻辑表输出两部分。上游依赖解析的内容根据节点是否为黑盒生成的物化节点而分为：系统解析-黑盒依赖和系统解析-常规依赖。重复单击解析输入输出按钮，将刷新上游依赖和逻辑表输出两部分结果。
 - **上游依赖**：用于展现该逻辑表内的物理节点所依赖的父节点。

单击解析问题提示按钮将出现窗口，提示源代码中出现的源物理表，以及未被解析出对应的关联节点ID的节点，多组纵向展示。可以通过下方的关联上游依赖按钮手动关联。



- **逻辑表输出**：用于展现此逻辑表全部的输出物化节点信息。支持展开下游，单击可以下拉展现该节点全部的一级下游，展示信息包括下游节点名和下游节点

逻辑表的输出由系统自动生成。逻辑表的上游：基于解析出的逻辑表代码里的引用表，找到输出名与表名一致的节点，然后添加为上游依赖的节点。生产环境和开发环境的常规任务节点ID一致，逻辑表任务的节点ID会有差异，您可以通过运维中心的逻辑表调度功能了解逻辑表节点及实例，详情参见[运维调度概述](#)。

3.4.5 下线或删除维度逻辑表

本文为您介绍维度逻辑表的下线和删除操作。

1. 进入研发 > 开发 > 规范建模 > 维度逻辑表功能菜单。
2. 在左侧列表中，鼠标悬停于某个维度逻辑表上，单击右侧浮现的“”图标，即可显示下

线、删除功能。



说明：

- 草稿状态的维度逻辑表不支持下线操作。
- 下线或删除维度逻辑表的前提是无依赖引用，包括维度逻辑表无使用、对应的维度没有被关联、对应的维度无子维度等。删除维度逻辑表之后，对应的维度会变为草稿状态。
- Basic模式项目和Dev-Prod模式的开发环境项目可以直接操作后下线或删除。Dev-Prod模式的生产环境项目，需在开发环境对应项目中下线或删除操作完成后，再执行发布操作，发布成功后生产环境对应的维度逻辑表才会下线或删除（发布时，维度与维度逻辑表需同时发布）。

3.5 逻辑表-事实逻辑表

3.5.1 新建与修改事实逻辑表

归属于某个业务过程的事实逻辑表，可以描述具体业务过程所对应事务的详细信息。创建事实逻辑表即完成公共的事务明细数据沉淀，以便提取业务中事务相关的明细数据。



说明：

事实逻辑表是抽象提炼事务的描述性信息，无需冗余事务相关联维度的属性信息。

新建事实逻辑表

1. 进入研发 > 开发 > 规范建模 > 事实逻辑表功能菜单。

2. 新建事实逻辑表有4个操作入口：

- 项目名称右侧的  逻辑表 > 事实逻辑表。



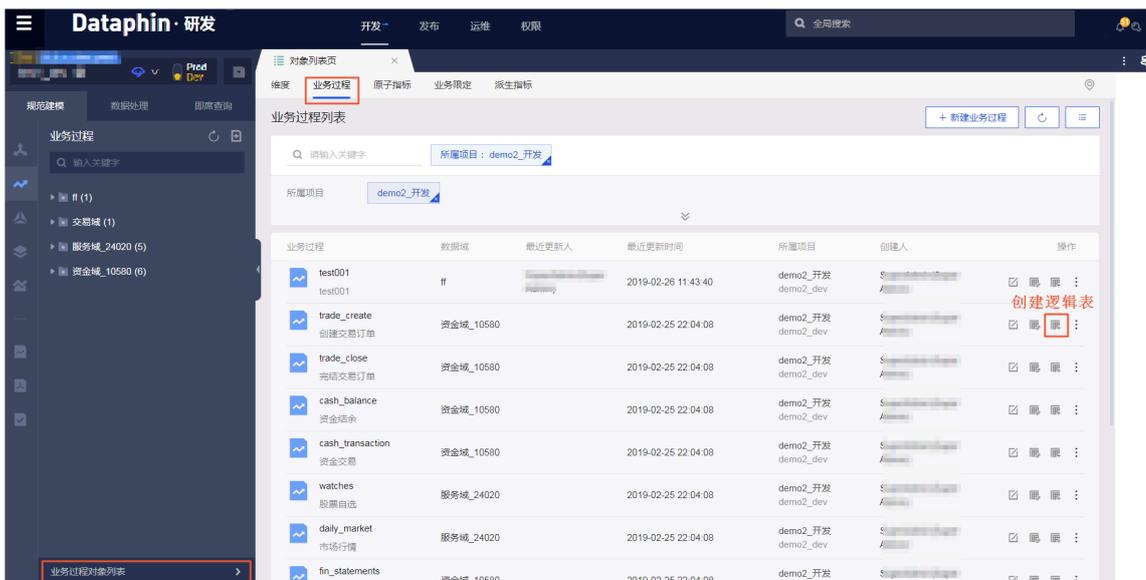
- 左侧功能区右上方的新建文件图标。



- 操作指引图（详情参见[使用说明](#)）上事实逻辑表后的。



- 通过业务过程对象列表页面，选择某个具体的业务过程，单击操作列的创建逻辑表按钮。



3. 通过上述操作入口，进入新建事实逻辑表页面。

4. 首先需要填写事实逻辑表的基本信息，根据界面提示填写即可。

基本信息包括如下几部分：

- 所属业务板块及项目：不需要填写，系统默认。
- 所属业务过程及对应的数据域：必填项。
- 明细类型：必填项，可以选择事务型和周期快照型。
 - 事务型：针对业务发生过程中的事件（即业务过程）进行跟踪和度量。例如，针对订单业务发生过程中的支付订单事件进行设计。
 - 周期快照型：主要针对实体的状态进行周期性的度量，例如账户余额、商品库存等。
- 中、英文名称及描述：
 - 英文名命名格式为fct_业务过程_[自定义部分]_di/df，其中自定义部分为非必填。
 - 后缀di/df：针对事务型为di，表示天增量数据。针对周期快照型为df，表示天全量数据。
- 来源主表：确定事实逻辑表的来源表，可以是物理表，也可以是逻辑表。

5. 事实逻辑表的主键定义。

- 不设定主键，表示事实逻辑表没有主键，此时可以填写主表的筛选条件。不需要填写where关键词，直接填写筛选条件，此项为非必填项。
- 设定主键，表示事实逻辑表有主键，此时需要填写主键信息和主键计算逻辑。

 **说明：**

- 事实逻辑表如果有唯一的主键（例如订单ID），后续添加的字段需要关联至事实逻辑表的主键，即字段计算逻辑里需要将被关联的字段写为系统标识关联键的关键字k，例如order_id as k。
- 事实逻辑表如果没有唯一的主键（例如浏览明细表），所添加字段的计算逻辑中如果有union、join，则需要保证SQL中引用了该来源表，且SQL计算逻辑加工出的记录需要在事实逻辑表的来源表记录范围内。

6. 通过上述步骤的基本信息和主键定义之后，进入事实逻辑表初始状态。

操作区域主要包含三个：

- 左上方区域的表信息、物理化配置、主表编辑等。
 - 物理化配置：可以参照维度逻辑表的[物理化配置](#)。
 - 主表编辑：可以参照维度逻辑表的[主表编辑](#)。
- 主表区域中间区域的添加关联维度、添加度量、添加事实属性。
- 主表区域右上区域的搜索字段、收起/折叠字段、主表详情和业务过程查看。

7. 调度配置。

操作同维度逻辑表的[调度配置](#)。

8. 提交发布。

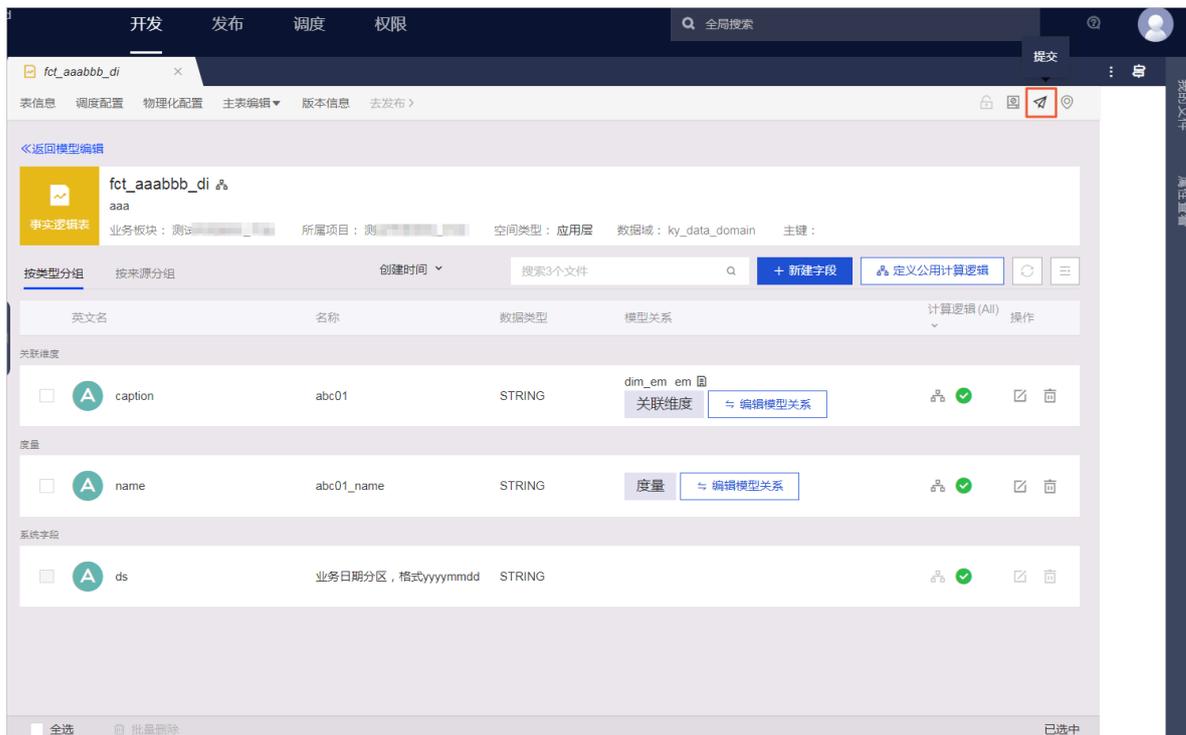
完成模型设计之后，在模型的右上操作区，单击提交。



说明：

- 该操作区的偷锁功能同其他模块，不再赘述。

· 与维度逻辑表相同，事实逻辑表编辑界面也支持列表筛选和过滤、定义公用计算逻辑、查看表计算逻辑、物理化配置、逻辑表调度配置等。



9. 事实逻辑表发布之后，后台代码会自动化生成并进行周期调度。



说明:

- 逻辑表发布需要生产账号，而非个人开发账号，并且要对所用表有权限。
- Basic模式项目和Dev-Prod模式的开发环境项目可以直接新建事实逻辑表。在开发环境新建完成并提交后，再执行发布操作，生产环境中对应的项目才会新增事实逻辑表。

修改事实逻辑表

事实逻辑表的修改操作与新建操作相同。

3.5.2 下线或删除事实逻辑表

本文为您介绍事实逻辑表的下线和删除操作。

1. 进入研发 > 开发 > 规范建模 > 事实逻辑表功能菜单。
2. 在左侧列表中，选择需要下线或删除的事实逻辑表，单击右侧浮现的“”图标，选择下线、删除即可。



说明:

- 下线及删除事实逻辑表的前提是无依赖引用，如使用事实逻辑表定义原子指标等。

- Basic模式项目和Dev-Prod模式的开发环境项目可以直接操作后下线或删除，Dev-Prod模式的生产环境，需要在开发环境中下线或删除操作完成后，再执行发布操作，才会下线或删除成功。

3.6 规范定义-原子指标

3.6.1 新建原子指标

原子指标是对指标统计口径、具体算法的抽象。Dataphin创新性地提出了设计即开发的理念，指标定义同时也明确了设计统计口径（即计算逻辑），提升了研发效率，并保证了统计结果的一致性。

根据计算逻辑的差异性，Dataphin将原子指标分为2种：

- 原生的原子指标，例如支付金额。
- 衍生原子指标，基于原子指标组合构建，例如客单价为支付金额除以买家数。

新建原子指标

1. 登录Dataphin。

2. 在Dataphin首页，选择研发 > 开发 > 规范建模 > 原子指标，进入原子指标页面。



3. 您可以通过以下任何一种方式进入新建原子指标。

- 单击项目名称后的 ，选择规范定义 > 原子指标，进入新建原子指标页面。



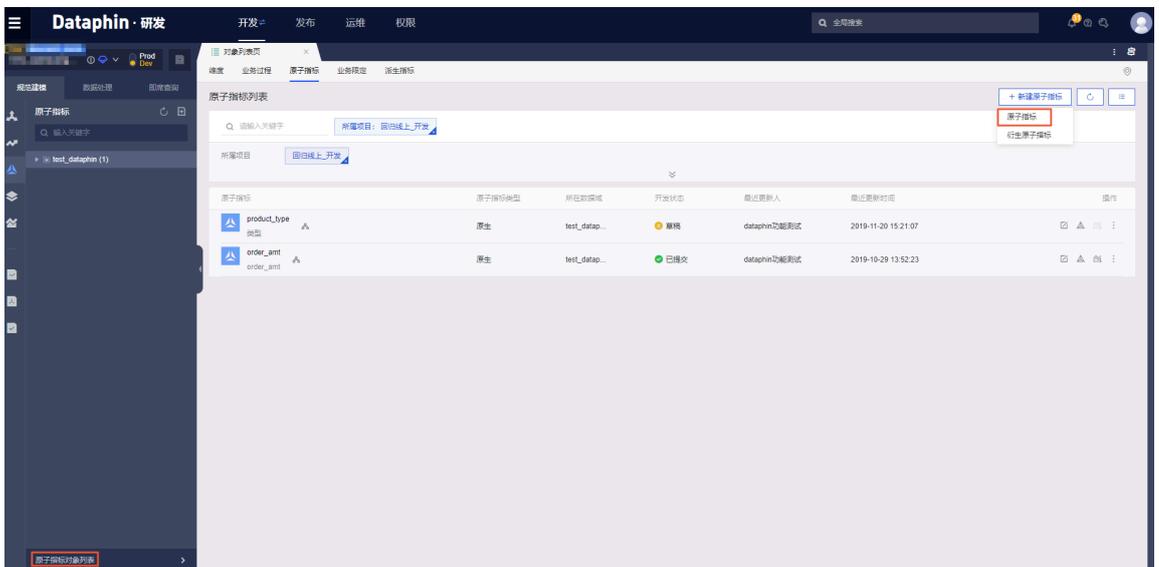
- 单击原子指标 ，然后单击新建原子指标，进入新建原子指标页面。



- 单击原子指标下的 ，然后单击新建原子指标，进入新建原子指标页面。



- 单击页面左下角的原子指标对象列表，进入对象列表页的原子指标页面，单击页面右上的新建原子指标，然后单击原子指标，进入新建原子指标页面。



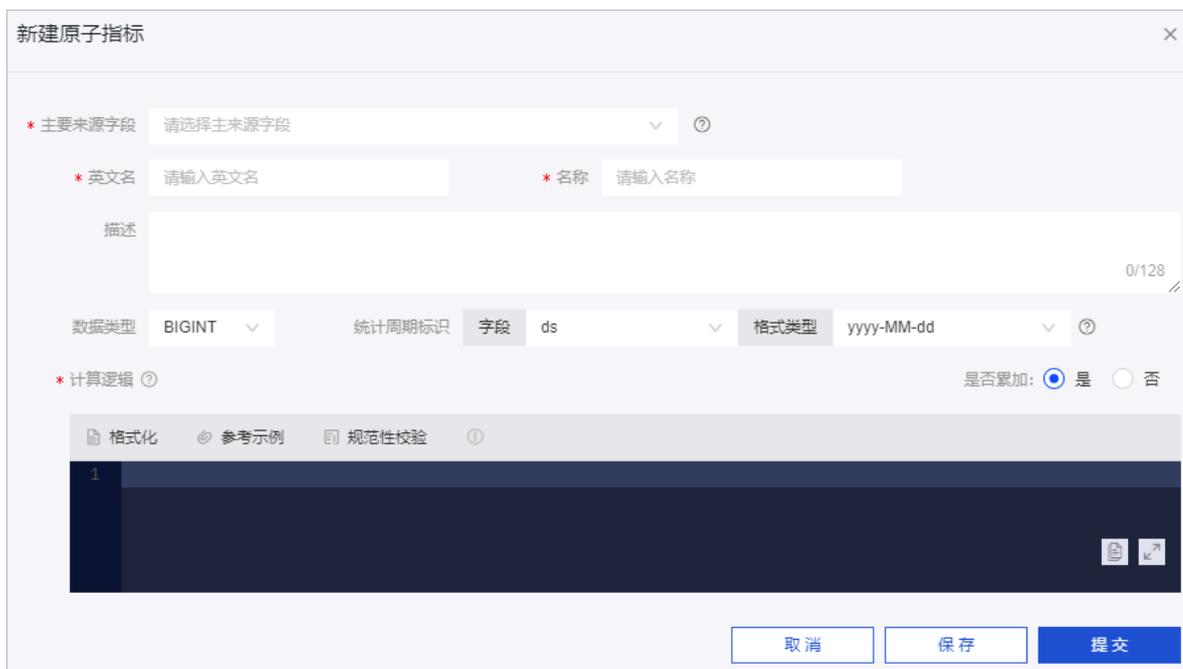
4. 在新建原子指标页面，根据页面提示选择数据域和来源表。



| 参数 | 说明 |
|-----|---|
| 数据域 | <p>根据原子指标来源表所属的业务过程，选择汇总该业务过程所在的数据域。</p> <p>说明： 数据域的划分请参见#unique_21。</p> |

| 参数 | 说明 |
|-----|---|
| 来源表 | <p>选择原子指标计算逻辑中，所用字段的来源表。为保证规范建模过程中，建设过程标准规范，来源表仅支持维度逻辑表、事实逻辑表。</p> <p> 说明： 选择来源表之后，页面会展示该来源表下已有的原子指标列表。</p> |

5. 在新建原子指标页面，单击新建原子指标，进入新建原子指标，完成参数配置后单击提交。



新建原子指标

* 主要来源字段 请选择主要来源字段

* 英文名 请输入英文名 * 名称 请输入名称

描述 0/128

数据类型 BIGINT 统计周期标识 字段 ds 格式类型 yyyy-MM-dd

* 计算逻辑 是否累加: 是 否

格式化 参考示例 规范性校验

1

取消 保存 提交

| 参数 | 说明 |
|--------|--|
| 主要来源字段 | 基于所选来源表，选择事实表或维度表中的字段。 |
| 英文名 | 基于所选主要来源字段，英文名自动填充为字段名。您也可以根据需要修改英文名。 |
| 名称 | 填写名称。名称由汉字、字母、数字、下划线或-组合组成。 |
| 描述 | 对原子指标的简单描述。 |
| 数据类型 | 选择数据类型。类型包括：STRING、BIGINT、DOUBLE、DATETIME。 |
| 字段 | 选择统计周期所限制的字段，默认选择所属业务板块设置的时间分区字段。 |
| 格式类型 | 统计周期绑定的字段，在计算时使用的格式类型。 |

| 参数 | 说明 |
|------|--|
| 计算逻辑 | <p>定义基于来源逻辑表模型定义所需指标的SQL计算表达式，例如 <code>count(distinct order_id)</code>，其中<code>order_id</code>为来源逻辑表模型中的字段。</p> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p> 说明:</p> <p>a. 单击主键计算逻辑后面的参考示例，可以查看SQL编写示例。</p> <p>b. SQL编写完成后，单击主键计算逻辑后面的规范性校验，可以校验您编写的SQL是否符合规范。</p> </div> |

6. 提交成功后，完成原子指标的新建。发布新建的原子指标，请参见[发布管理](#)。



说明:

Basic模式下的Basic项目和Dev-Prod模式下的Dev项目均可以直接新建原子指标。在开发环境中的Dev项目新建、提交、发布原子指标成功后，生产环境中的Prod项目才会新增该原子指标。

新建衍生原子指标

基于已提交的原子指标进行引用，设定组装逻辑，生成衍生原子指标。例如，其中一个已提交的原子指标为下单金额，另一个已提交的原子指标为下单人数，您可以通过新建衍生原子指标的方式统计人均下单金额：

1. 请参见[新建原子指标](#)进入新建衍生原子指标页面。

2. 根据页面提示完成参数的配置，并单击提交。

| 参数 | 说明 |
|------|---|
| 英文名 | 填写英文名。英文名由字母、数字、下划线组合组成。 |
| 名称 | 填写名称。名称由汉字、字母、数字、下划线或-组合组成。 |
| 描述 | 对衍生指标的简单描述。 |
| 数据类型 | 选择数据类型。类型包括：STRING、BIGINT、DOUBLE、DATETIME。 |
| 计算逻辑 | <p>基于已提交原子指标，组合设定所需指标的统计计算方式：</p> <ul style="list-style-type: none"> 在SQL语句输入框中输入衍生指标的计算公式，其中计算公式中引用的原子指标均为已提交的原子指标。 a. 单击引用原子指标的，选择数据域及来源表。 b. 单击筛选出来的原子指标后的添加，将原子指标添加在右侧的SQL语句输入框中。 <p>或在后输入原子指标关键字进行快速搜索，单击搜索出来的原子指标后的添加，将原子指标添加在右侧的SQL语句输入框中。</p> <p>例如，计算人均下单金额<code>crt_amt/usr_cnt</code>，其中<code>crt_amt</code>为下单金额对应的已提交原子指标，<code>usr_cnt</code>为用户数对应的已提交原子指标。</p> |

3. 提交成功后，完成衍生原子指标的新建。发布新建的衍生原子指标，请参见[发布管理](#)。



说明:

Basic模式下的Basic项目和Dev-Prod模式下的Dev项目均可以直接新建衍生原子指标。在开发环境中的Dev项目新建、提交、发布衍生原子指标成功后，生产环境中Prod项目才会新增该衍生原子指标。

3.6.2 管理原子指标

本文为您介绍如何进行修改、克隆、下线、下线并删除原子指标及查看相关派生指标，帮助您更好的管理原子指标。

修改原子指标

1. 登录Dataphin。
2. 在Dataphin首页，选择研发 > 开发 > 规范建模 > 原子指标，进入原子指标页面。

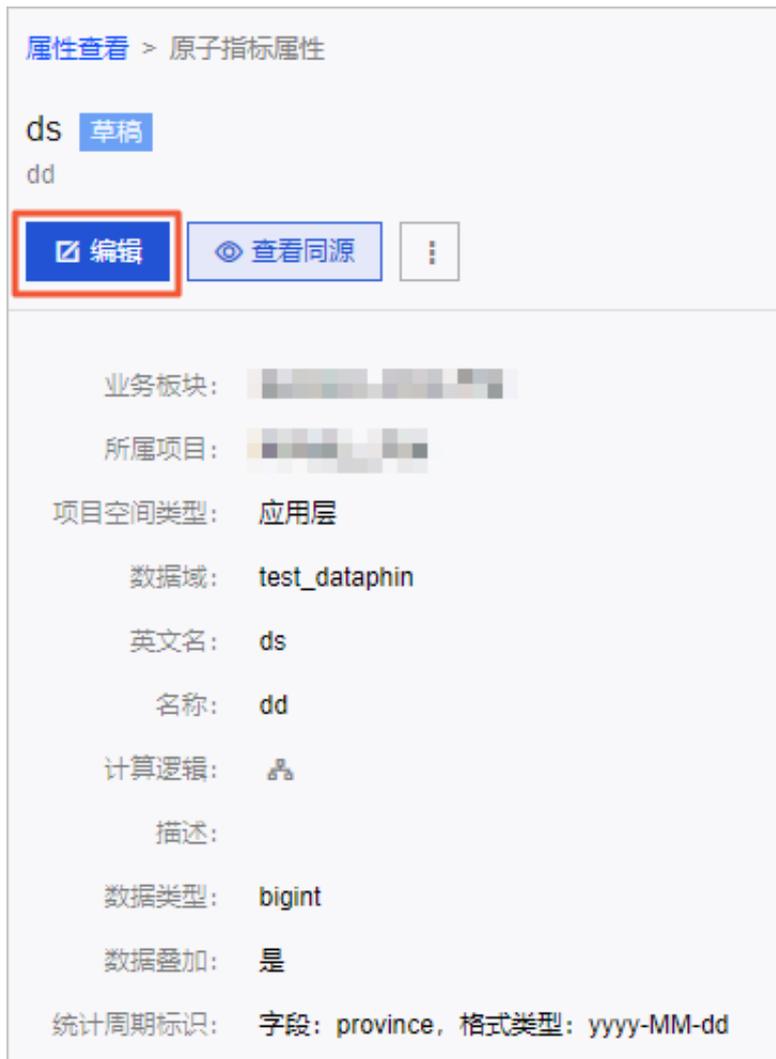


3. 您可以通过以下3种方式中的任何1种方式，进入编辑原子指标页面：

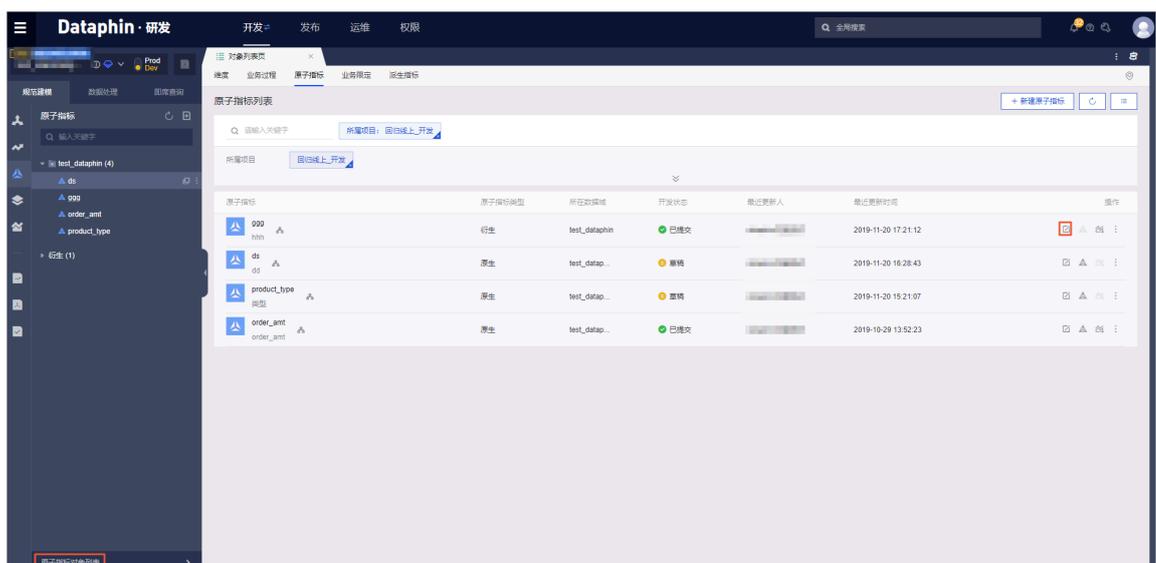
- 选择需修改的原子指标，将鼠标悬浮在 ，单击编辑。进入编辑原子指标页面。



- 单击需修改的原子指标，页面右侧弹出的该原子指标的属性框。单击属性框中的编辑，进入编辑原子指标页面。



- 单击页面左下角的原子指标对象列表，进入对象列表页的原子指标页面，单击需修改原子指标所在行的操作列下的 ，进入编辑原子指标页面。



- 在编辑原子指标页面，根据页面提示配置参数，参数配置的详细内容请参见[新建原子指标](#)。

5. 完成参数配置后，单击提交。提交成功后，完成该原子指标的修改。发布该原子指标，请参见发布管理。



说明:

- Basic模式下的Basic项目和Dev-Prod模式下的Dev项目可以直接修改原子指标。在开发环境中的Dev项目修改、提交、发布原子指标成功后，生产环境中的Prod项目该原子指标才会修改。
- 原子指标的统计口径变更发布后，会影响下游所有涉及的派生指标，因此需要充分考虑影响。

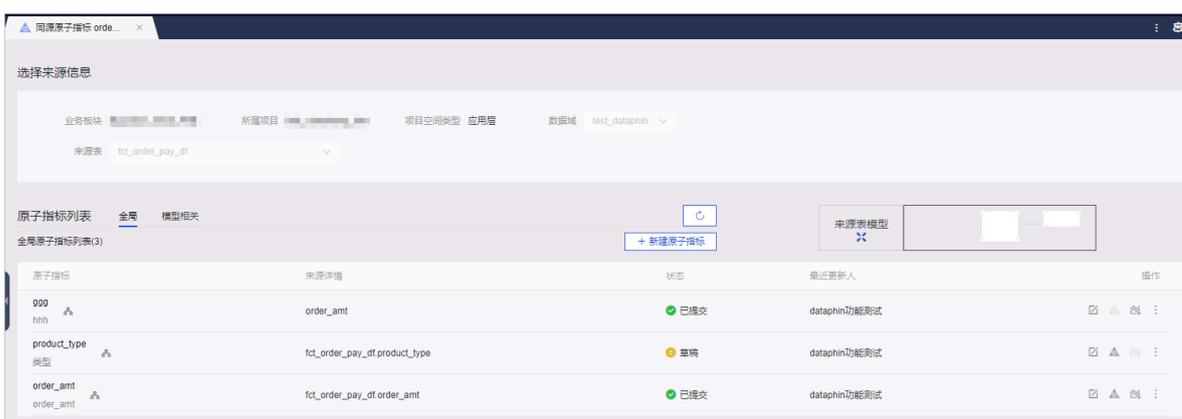
查看同源

您可以查看与该原子指标归属为同一个来源逻辑表的原子指标:

1. 在原子指标页面，单击需查看的原子指标，右侧弹出该原子指标的属性框。



2. 单击查看同源，进入同源原子指标页面。



如果您需基于该原子指标的来源逻辑表再次创建原子指标，您可以在新建原子指标时选择该来源逻辑表。或在同源原子指标页面，单击新建原子指标，进入新建原子指标页面，参数配置的内容请参见新建原子指标。



说明:

衍生原子指标不存在同源原子指标。

克隆原子指标

1. 您可以通过以下2种方式中的任何1种方式，进入克隆原子指标页面：

- 单击需克隆的原子指标，页面右侧弹出的该原子指标的属性框。单击属性框中的，然后单击克隆。



- 单击页面左下角的原子指标对象列表，进入对象列表页的原子指标页面，单击需克隆原子指标所在行的操作列下的，然后单击克隆，进入克隆原子指标页面。



| 原子指标 | 原子指标类型 | 所在数据域 | 开发状态 | 最近更新人 | 最近更新时间 | 操作 |
|--------------|--------|---------------|------|-------|---------------------|----|
| 999_hnn | 衍生 | test_dataphin | 已提交 | | 2019-11-20 17:21:12 | 克隆 |
| ds_dd | 原生 | test_datap... | 草稿 | | 2019-11-20 16:28:43 | 克隆 |
| product_type | 原生 | test_datap... | 草稿 | | 2019-11-20 15:21:07 | 克隆 |
| order_amt | 原生 | test_datap... | 已提交 | | 2019-10-29 13:52:23 | 克隆 |

2. 在克隆原子指标页面，根据页面提示完成参数配置，并单击提交。参数配置的详细内容请参见[新建原子指标](#)。

3. 提交成功后，完成原子指标的克隆。发布克隆的原子指标，详细内容请参见[发布管理](#)。



说明：

Basic模式下的Basic项目和Dev-Prod模式下的Dev项目可以直接克隆原子指标。在开发环境中的Dev项目克隆、提交、发布原子指标成功后，生产环境中的Prod项目才会新增该原子指标。

下线原子指标

1. 您可以通过以下3种方式中任何1种，进入下线原子指标提示页面。

- 选择需下线的原子指标，将鼠标悬浮在，单击下线，进入下线原子指标提示页面。



- 单击需下线的原子指标，页面右侧弹出的该原子指标的属性框。单击属性框中的，然后单击下线，进入下线原子指标提示页面。



· 单击页面左下角的原子指标对象列表，进入对象列表页的原子指标页面，单击需下线原子指标所在行的操作列下的，然后单击下线，进入下线原子指标提示页面。



2. 在下线原子指标提示页面，填写下线备注信息后，单击确定并提交，即可下线原子指标。原子指标的开发状态变更为草稿状态。。

提示 ×

 是否下线原子指标

* 备注 请输入此次提交的相关备注说明，如变更信息

0/128

特殊说明：当前项目，提交成功后，请前往待发布列表，以发布对象至生产环境。



说明:

- 原子指标的开发状态为草稿状态，不支持下线操作。
- 当且仅当原子指标无依赖引用时（派生指标引用及其他原子指标引用），才支持下线操作。
- Basic模式下的Basic项目和Dev-Prod模式下的Dev项目可以直接执行下线操作。Dev项目下的原子指标下线、发布成功后，Prod项目中与该原子指标相对应的任务将被删除。

下线并删除原子指标

针对开发状态为已提交和开发中的标签，通过以下步骤删除原子指标：

1. 您可以通过以下3种方式中任何1种，进入下线并删除原子指标提示页面。

- 选择需下线并删除的原子指标，将鼠标悬浮在，单击下线并删除，进入下线并删除原子指标提示页面。



- 单击需下线并删除的原子指标，页面右侧弹出的该原子指标的属性框。单击属性框中的，然后单击下线并删除，进入下线并删除原子指标提示页面。



- 单击页面左下角的原子指标对象列表，进入对象列表页的原子指标页面，单击需下线并删除原子指标所在行的操作列下的，然后单击下线并删除，进入下线并删除原子指标提示页面。



2. 在下线原子指标提示页面，填写下线并删除备注信息后，单击确定并提交，即可删除原子指标。

提示 ×

 是否完全删除原子指标

* 备注 请输入此次提交的相关备注说明，如变更信息

0/128

特殊说明：当前项目，提交成功后，请前往待发布列表，以发布对象至生产环境。

针对开发状态为草稿状态的原子指标，通过以下操作完成原子指标的删除：

1. 您可以通过以下3种方式中任何1种，进入下线并删除原子指标提示页面。

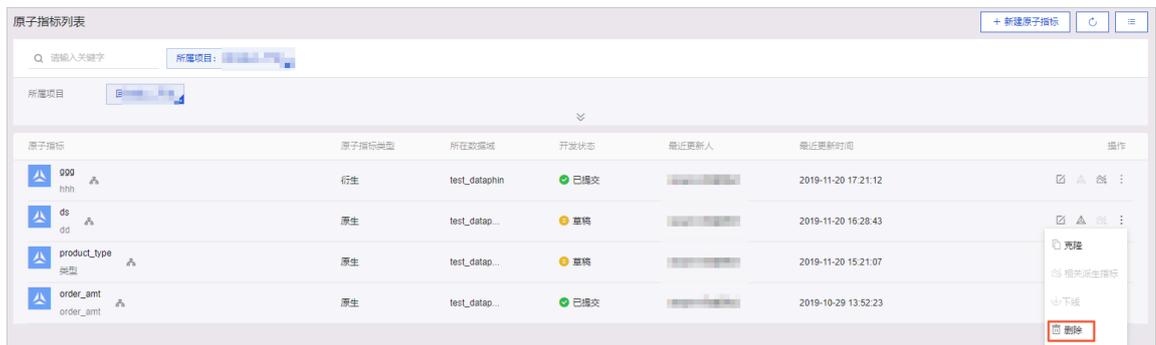
- 选择需删除的原子指标，将鼠标悬浮在 ，单击删除，进入删除原子指标提示页面。



- 单击需删除的原子指标，页面右侧弹出的该原子指标的属性框。单击属性框中的 ，然后单击删除，进入删除原子指标提示页面。



- 单击页面左下角的原子指标对象列表，进入对象列表页的原子指标页面，单击需删除原子指标所在行的操作列下的 ，然后单击删除，进入删除原子指标提示页面。



2. 在下线原子指标提示页面，填写下线并删除备注信息后，单击确定并提交，即可删除原子指标。

提示 ×

 是否完全删除原子指标

* 备注

0/128

特殊说明：当前项目，提交成功后，请前往待发布列表，以发布对象至生产环境。



说明:

- 当且仅当原子指标无依赖引用时（派生指标引用及其他原子指标引用），才支持下线并删除操作。
- Basic模式下的Basic项目和Dev-Prod模式下的Dev项目可以直接执行删除操作。Dev项目下的原子指标删除、发布成功后，Prod项目中与该原子指标相对应的任务将被删除。

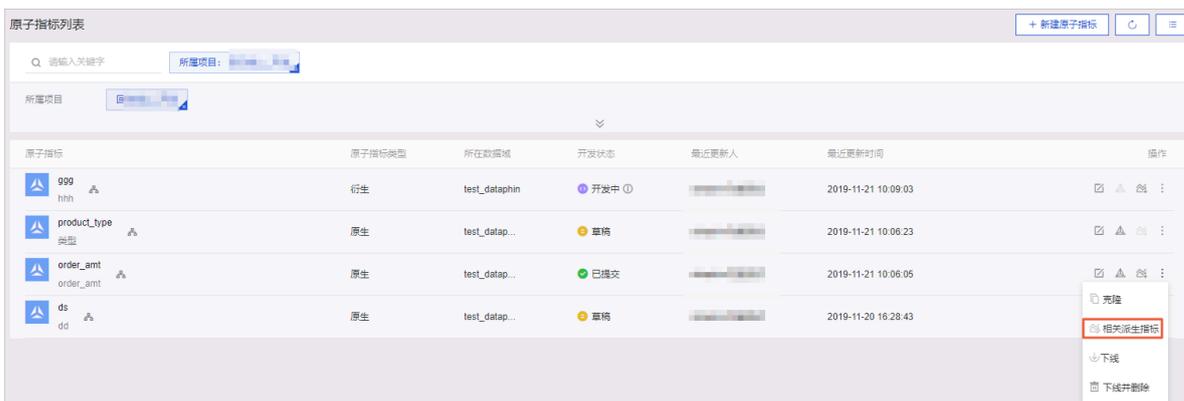
查看相关派生指标及创建派生指标

您可以通过以下2种方式中任何1种方式，查看依赖某个原子指标的派生指标：

- 单击需查看的原子指标，页面右侧弹出的该原子指标的属性框。单击属性框中的 ，然后单击相关派生指标，进入派生指标页面。



- 单击页面左下角的原子指标对象列表，进入对象列表页的原子指标页面，单击需查看该原子指标所在行的操作列下的 ，然后单击相关派生指标，进入派生指标页面。



您也可以基于某个原子指标创建派生指标。

在原子指标的属性框中单击 ，然后单击创建派生指标，进入新建派生指标页面，根据页面提示

完成参数配置，详细内容请参见[新建派生指标](#)。



原子指标与派生指标的关系：

- 原子指标的计算逻辑修改生效后，会直接更新应用于相关的派生指标。
- 原子指标被删除或者变更英文名，需要校验下游是否有派生指标使用，如果有，则无法删除或变更。
- 原子指标的数据类型更改会影响下游派生指标。



说明：

仅开发状态为已提交和开发中，支持创建派生指标及查看相关派生指标。

3.7 规范定义-业务限定

3.7.1 新建业务限定

原子指标是计算逻辑的标准化定义，业务限定则是条件限制的标准化定义。为保障所有统计指标统一、标准、规范地构建，业务限定在业务板块内唯一，并唯一归属于一个来源逻辑表，计算逻辑也以该来源逻辑表模型的字段为基础进行定义。由于一个业务限定的定义可能来自于归属不同数据域的多个逻辑表，因此一个业务限定可能归属于多个数据域。

1. 进入研发 > 开发 > 规范建模 > 业务限定功能菜单。
2. 新建业务限定有三个操作入口。



说明:

Basic模式项目和Dev-Prod模式的开发环境项目可以直接新建业务限定。在开发环境新建完成并提交后，再执行发布操作，生产环境中对应的项目才会新增业务限定。

3. 单击上述操作入口之后，在选择来源信息区域，根据页面提示，选择数据域、来源表。



说明:

选择来源表之后，页面会展示该来源表已生成的业务限定列表。

4. 在业务限定所在来源区域，单击新建业务限定，按需填写信息，然后单击提交即可。
5. 页面返回提交成功后，业务限定即为已上线状态。

3.7.2 修改业务限定

本文为您介绍如何修改业务限定。

- 业务限定的来源逻辑表说明:

业务限定是为构建应用统计分析所需的派生指标而定义的数据组件。因此，您只可以基于维度逻辑表、事实逻辑表这两类明细数据表来创建业务限定。

· 业务限定与派生指标的关系：

- 业务限定的计算逻辑修改生效后，会直接更新应用于相关的派生指标。
- 业务限定被删除或者变更英文名，需要校验下游是否有派生指标使用，如果有，则无法删除或变更。

1. 进入研发 > 开发 > 规范建模 > 业务限定功能菜单。

2. 在左侧列表中，选择需要修改的业务限定，单击编辑即可进入修改页面。



说明：

Basic模式项目和Dev-Prod模式的开发环境项目可以直接修改业务限定（Dev-Prod模式中开发环境修改业务限定后，需要执行发布操作，才能修改生产环境中的业务限定）。

3. 按需修改业务限定，然后提交即可。和原子指标相同，业务限定的统计口径变更发布后，会影响下游所有涉及的派生指标，因此需要充分考虑影响。

4. 查看同源。单击查看同源，您可以查看与该业务限定归属为同一个来源逻辑表的业务限定。

如果您需要基于该业务限定的来源逻辑表再次创建业务限定，您可以新建业务限定选择该来源逻辑表，或者单击查看同源打开对应页面，新建业务限定。

3.7.3 下线或删除业务限定

本文为您介绍如何下线、删除业务限定。

1. 登录Dataphin。

2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。

3. 选择开发 > 规范建模 > 业务限定。



4. 在左侧列表中，将鼠标悬停于要下线或删除的业务限定上，单击右侧浮现的按钮，选择下

线、删除，如下图所示。

 **说明：**

- 下线、删除业务限定的前提是无依赖引用，即业务限定没有被派生指标引用。

- Basic模式项目和Dev-Prod模式的开发环境项目可以直接操作后下线或删除，Dev-Prod模式的生产环境项目，需要在开发环境中下线或删除操作完成后，再执行发布操作，才会下线或删除成功。



3.8 规范定义-派生指标

3.8.1 新建派生指标

本文为您介绍派生指标的定义，以及如何新建派生指标。

派生指标简介

派生指标即常见的统计指标，派生指标=原子指标+业务限定+统计周期+统计粒度。为保证统计指标标准、规范、无二义性地生成，OneData方法论将派生指标抽象为如下四部分：

- 原子指标：明确统计口径，即计算逻辑。
- 业务限定：统计的业务范围，筛选出符合业务规则的记录（类似于SQL中Where后面的条件，不包括时间区间）。
- 统计周期：统计的时间范围，例如，最近一天、最近30天等（类似于SQL中Where后面的时间条件）。

- **统计粒度**：统计分析的对象或视角，定义数据需要汇总的程度，可理解为聚合运算时的分组条件（类似于SQL中的Group By的对象）。粒度是维度的一个组合，例如某个指标是某个卖家在某个省份的成交额，则粒度就是卖家、地区这两个维度的组合。

基于上述定义组合机制，即可批量、无重复地快速创建派生指标。同时，可以保证命名和计算逻辑明确且不重复，业务人员也可以自主定义生产。派生指标作为与字段同级的概念，在同一个统计粒度内唯一。

新建派生指标

1. 进入研发 > 开发 > 规范建模 > 派生指标功能菜单。
2. 新建派生指标，共有三个操作入口，如下图所示。



3. 通过上述操作入口，进入新建派生指标页面。根据界面提示填写内容，完成后单击预生成派生指标，再根据提示单击提交、确定并提交即可完成派生指标创建。



说明：

Basic模式项目和Dev-Prod模式的开发环境项目可以直接新建派生指标。在开发环境新建完成并提交后，再执行发布操作，生产环境中对应的项目才会新增派生指标。

4. 派生指标发布后，会挂载至统计粒度对应的汇总逻辑表（如果汇总逻辑表不存在，则系统自动生成），后台代码会自动化生成并进行周期调度。

3.8.2 修改派生指标

本文为您介绍如何修改已创建的派生指标。

1. 进入研发 > 开发 > 规范建模 > 派生指标功能菜单。
2. 在左侧派生指标列表中，找到需要修改的派生指标，单击选择编辑即可进入派生指标编辑页面。



说明：

- 派生指标仅支持修改基本信息，不能修改派生组合口径。

- 如果您需要修改业务限定等派生指标的构成元素，则需要新生成一个派生指标。请您新建对应组合的派生指标，并保证对象定义生成规范无二义性、前后一致。派生指标会随着构成元素的计算逻辑更新，而自动更新计算逻辑。
- Basic模式项目和Dev-Prod模式的开发环境项目可以直接修改派生指标（Dev-Prod模式中开发环境修改派生指标后，需要执行发布操作，才能修改生产环境中对应项目的派生指标）。

3. 修改派生指标基本信息。按需完成修改，保存、发布即可。



说明：

派生指标名称，基于所属的汇总逻辑表（即统计粒度，也即维度或维度组合）内唯一。

3.8.3 下线或删除派生指标

本文为您介绍如何下线、删除派生指标。

1. 进入研发 > 开发 > 规范建模 > 派生指标功能菜单。
2. 在左侧列表中，可搜索查找到需要下线或删除的派生指标，然后单击“”操作扩展图

标，选择下线或删除，即可完成操作。



说明：

Basic模式项目和Dev-Prod模式的开发环境项目可以直接操作后下线或删除，Dev-Prod模式的生产环境项目，需要在开发环境项目中下线或删除操作完成后，再执行发布操作，才会下线或删除成功。

3.8.4 查看相关

本文为您介绍如何查看派生指标的相关信息。

1. 进入研发 > 开发 > 规范建模 > 派生指标功能菜单。
2. 在左侧列表中，单击需要查看的派生指标，即可展开查看相关按钮。
3. 单击查看相关按钮，进入派生指标所属的汇总逻辑表（根据统计粒度唯一确定）页面。

在这里，可以对派生指标进行管理，包括新建/编辑/删除、查看详情、分类管理，还可以新建自定义指标等，具体请参见 [新增自定义指标与派生指标管理](#)。

3.9 逻辑表-汇总逻辑表

3.9.1 新建汇总逻辑表

汇总逻辑表是由一个特定的分析对象（如会员）及其相关的统计指标组成的。组成一个汇总逻辑表的统计指标都具有相同的统计粒度（如会员）。汇总逻辑表面向用户提供了以统计粒度（如会员）为主题的所有统计数据（如会员主题集市），是Dataphin支持的主题式数据查询工具。

组成汇总逻辑表的统计指标有两种来源：

- 系统按照相同统计粒度，自动汇聚。
- 通过非派生指标创建方式创建的普通物理表字段，根据同主键原则归一至相应的汇总逻辑表。

基于这种构建原理，汇总逻辑表只包含已提交状态。

1. 进入研发 > 开发 > 规范建模 > 汇总逻辑表功能菜单。

2. 新建汇总逻辑表有三个操作入口。



说明：

- 派生指标提交后，会自动生成新的汇总逻辑表。
- Basic模式项目和Dev-Prod模式的开发环境项目可以直接新建维度在开发环境新建完成并提交后，再执行发布操作，生产环境中对应的项目才会新增汇总逻辑表。

3. 通过上述操作入口，进入新建汇总逻辑表页面。根据界面提示填写信息，填写完成后单击提交、确定并提交即可完成汇总逻辑表创建。

3.9.2 物理化配置

汇总逻辑表的物理化配置为您提供自定义参数的配置功能，本文为您介绍如何对汇总逻辑表进行物理化配置。



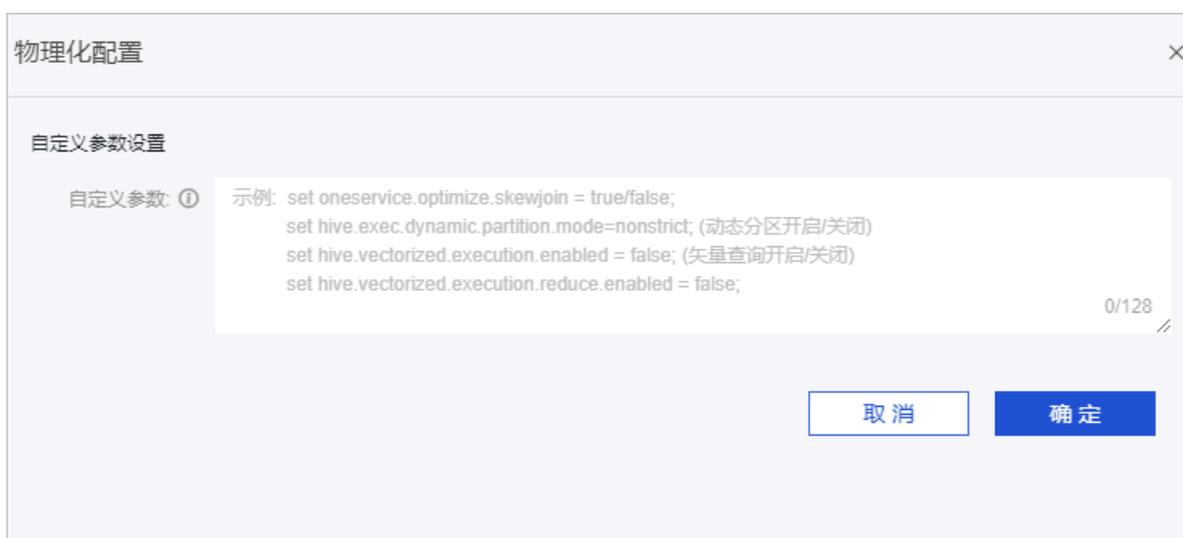
注意：

汇总逻辑表默认存储生命周期为36000天，暂不支持自定义修改。

1. 进入研发 > 开发 > 规范建模 > 汇总逻辑表功能菜单。
2. 在左侧导航栏中，单击需物理化配置的汇总逻辑表。在汇总逻辑表编辑页面，单击上方菜单栏中物理化配置。



右侧弹出汇总逻辑物理化配置窗口，可以根据界面提示配置相关信息，配置完信息后单击确定。



自定义参数设置功能会将配置的参数应用于当前逻辑表以及使用当前逻辑表的SQL代码，即会应用于汇总逻辑表对应的物理表的执行SQL，执行时自定义参数功能生效。例如，数据倾斜自优化等功能的开启关闭。您可以根据需求设置自定义参数，详情参见[#unique_31](#)。

3. 汇总逻辑表发布后，后台代码会自动化生成并进行周期调度。逻辑表发布需要生产账号，而非个人账号，并且对所有表有权限。

3.9.3 查看表详情

本文为您介绍如何查看和编辑汇总逻辑表的基本信息。包括主键信息和汇总逻辑表描述信息。

1. 进入研发 > 开发 > 规范建模 > 汇总逻辑表功能菜单。

2. 在左侧导航栏中，单击汇总逻辑表。在汇总逻辑表编辑页面，单击上方菜单栏中表详情。



右侧弹出列表，可以查看主键信息及汇总逻辑表描述信息。同时可编辑对汇总表的描述。单击组成统计粒度的某个具体维度，右侧弹出该维度详情窗口。



3.9.4 新增自定义指标与派生指标管理

本文为您介绍在汇总逻辑表功能模块可以对派生指标进行管理（如新建、删除、编辑派生指标等等），还可以新建自定义指标。

1. 进入研发 > 开发 > 规范建模 > 汇总逻辑表功能菜单。

2. 派生指标管理

在左侧列表中，单击汇总逻辑表，可以对派生指标进行管理。主要操作包括：新建/删除/编辑派生指标、查看派生指标详情等。

3. 新增自定义指标

除了派生指标，还可以添加自定义指标，即从用户已有的物理表中新增指标，仅限当前项目内、当前用户为owner的物理表。添加自定义指标的步骤如下：

- a. 切换到自定义指标页签，单击页面右侧的新建自定义指标。
- b. 在弹出的页面中，根据提示填写信息，填写完之后单击确定，即可完成自定义指标创建。

3.9.5 删除汇总逻辑表

本文为您介绍如何删除汇总逻辑表。

1. 进入研发 > 开发 > 规范建模 > 汇总逻辑表功能菜单。
2. 在左侧列表中，单击某个汇总逻辑表右侧的“”图标，选择删除按钮，即可删除该汇总逻辑表。



说明：

- 汇总逻辑表可以删除的前提是表内为空，无任何字段或指标。
- Basic模式项目和Dev-Prod模式的开发环境项目可以直接操作后删除，Dev-Prod模式的生产环境项目，需要在开发环境对应项目中删除操作完成后，再执行发布操作，才会删除成功。

3.10 规范建模-建模引擎

Dataphin逻辑表开发、提交和调度运行，会自行根据系统内置规则对逻辑表模型进行校验、物理化，提供查询使用。标准规范、可靠稳定及高效的逻辑表研发过程依赖于智能建模引擎服务。

逻辑表查询SQL

Dataphin支持逻辑表雪花模型的查询，即支持基于[逻辑模型.关联维度.关联维度.···.属性]（如订单.买家.会员类型.类型值）查询所需数据，提高查询类SQL的编码效率。

```

1 SELECT a.bob_cw_bob_ooo,
2       a.ds_1d__t_dts,
3       a.ds_cw_234444444556,
4       a.dim_cc.code
5 FROM
6       LD_retail.dws_cc a
7 where a.dim_cc.value is NOT NULL

```

- 事实逻辑表/维度逻辑表查询

- SQL格式说明如下。

```

select Column,
       LTable.Column, //逻辑表名.字段名
       LTable.Role(relation - dimension)....Dimension_Column,
       //逻辑表名.角色名(关联维度字段名).维度属性字段名
from   LD.LTable //业务板块名.逻辑表名
where  ...

```

```

select Column,
       LTable_Alias.Column,
       LTable_Alias.Role(relation - dimension)....Dimension_Column
from   LD.LTable Alias
where  ...

```

- SQL格式举例如下。

```

select a.dim_industry.in_name,
       a.dim_industry.IN_ID,
       name,
       ceo_name,
       company_id
from   LD_demo.dim_company.dim_industry.in_name a
where  ds = '${bizdate}'
limit  100;

```

- 汇总逻辑表查询

- SQL格式说明如下。

```

select Column,
       LTable.Column,
       LTable.Dimension(granularity)....Dimension_Column,
from   LD.LTable
where  ...

```

```

select Column,
       Alias.Column,
       Alias.Dimension (granularity)....Dimension_Column
from   LD.LTable Alias

```

```
where ...
```

- SQL格式举例如下。

```
select customer_id,
       c.dim_customer.address_line1,
       c.dim_customer.dim_tax_rate.tx_name,
       crt_trd_comm_amt_30d_trd_valid,
       acct_cnt_td_actv_account,
       watch_sec_cnt_td_actv_watches,
       avg_crt_trd_comm_7d,
       crt_trd_acct_cnt_7d,
       crt_trd_cnt_1d,
       avg_crt_trd_comm_1d,
       crt_trd_comm_amt_1d_trd_valid,
       crt_trd_acct_cnt_1d,
       watch_sec_cnt_td_inac_watches,
       watch_sec_cnt_td,
       acct_cnt_td,
       crt_trd_cnt_7d
from LD_DEMO.DWS_CUSTOMER c
where ds='${bizdate}'
limit 100;
```



说明:

- 如果维度逻辑表两个角色引用的字段有重名，需要给两个字段命别名，以防字段名冲突。
- 仅主维度的维度逻辑表可以被查询（子维度的维度逻辑表不可查询），包含被作为关联维度（取角色名）的情况。
- 子维度的维度逻辑表包含的字段，合并在主维度的维度逻辑表字段list中。
- 组成汇总逻辑表统计粒度的维度，可以通过类似关联维度方式引用，汇总逻辑表select * 可获取[维度_主键]数据。为防全表扫描过慢，建议使用select 汇总表.维度表.主键获取主键，select 汇总表.维度表.属性获取属性信息。

计算逻辑重复性校验

Dataphin核心价值之一是定义无二义性，核心要求为命名、计算逻辑无重复。提交规范定义、逻辑表对象时，Dataphin会校验对象的英文名、中文名、计算逻辑完全没有重复。如果计算逻辑存在重复，则进行系统提示，避免误创建同名不同义、或者同义不同名的对象。

新建或编辑之后，在您提交并发布时间周期、维度、维度逻辑表字段、事实逻辑表、事实逻辑表字段、原子指标、业务限定、汇总逻辑表自定义指标时，会基于AST抽象语法树结构进行解析，在命名空间内进行计算逻辑（或表达式）重复性校。如果校验有重复，则提示对象计算逻辑重复。

模型多版本动态刷新

动态刷新可以提高逻辑表提交及物理化的效率，增加逻辑表变更的灵活性，也可以降低物理表变更及历史数据变更迁移的计算引擎源资源消耗。在SQL任务调度时，系统会基于当前最新的逻辑表物理化情况自动识别和路由对应的物理表，从而获取查询所需的数据。

逻辑表的版本信息，可以在逻辑表编辑页面上方的菜单栏中，单击版本信息查看。

4 数据处理

4.1 数据处理概述

数据处理用于执行数据研发任务，管理数据研发过程中使用的函数和文件。

数据处理包含任务管理和文件管理两个模块：

- 任务管理用于执行数据研发任务，包括代码任务和同步任务。
- 文件管理用于管理数据研发过程中使用的文件，包括函数管理和资源管理。

1. 进入研发 > 开发 > 数据处理功能模块。

- 将鼠标悬停于左侧功能区，可以显示数据处理的相关功能分类，包括代码任务、同步任务、函数管理、资源管理。单击对应的分类名称，可以进入各个功能区。您可以像管理文件一样管理代码、函数和资源，按照文件夹分类存放。



- 单击左侧功能区右上角的加号图标，可以新建文件和文件夹。



- 在左侧功能区上方的搜索框内，可以输入关键字搜索文件。



- 将鼠标悬停在某个文件上的 ，会展示管理该文件的具体操作，例如编辑、下线、重命名、删除等。



页面框架详情请参见[研发页面概览](#)。

4.2 代码任务

4.2.1 新建代码任务

数据研发任务过程中，不同的计算引擎配置支持不同类型的代码任务。本文为您介绍如何基于MaxCompute计算引擎进行新建代码任务。

- 您需要首先在项目管理界面配置每个项目的IP白名单。通过这种方式，使项目在执行任务或调度任务时，可以访问对应的IP的资源等，保证任务运行的安全稳定。

- MaxCompute计算引擎支持MaxCompute SQL、MaxCompute MR、Spark on MaxCompute、Shell、Python等代码任务。如果配置的MaxCompute计算引擎源支持Spark（详情参见[#unique_61](#)），则Spark任务可正常执行。

新建代码任务

1. 进入研发 > 开发 > 数据处理 > 代码任务功能模块。
2. 在左侧功能区，单击右上角的+图标，选择需要的任务类型，新建文件夹和代码任务。该区域支持文件及文件夹目录移动。



- 支持2.7版本的Python。Python任务系统内置lib库可以通过pip list获取。Shell及Python任务支持设置内存大小，上限为8G。
- Shell、Python任务默认分配0.1核CPU、256MB内存。
- Shell、Python任务设置资源大小、引用资源（如DataX脚本），代码前可以添加如下内容：

```
@required_resource{required_memory=2Gb;required_cpus=1.0} @
resource_reference{"myDataXJob.json"}
python $DATAX_HOME/bin/datax.py --jvm '-Xms2g -Xmx2g' myDataXJob.
json
```

 说明:

使用`--jvm '-Xms2g -Xmx2g'`来指定DataX实际运行时JVM的内存时，建议将该内存大小与`required_resource`中的`required_memory`设置一致，充分使用申请的资源。

- Spark任务运行时，前缀需要加上如下内容，以保证执行通过。

```
--conf spark.hadoop.odps.task.major.version=cupid_v2
--conf spark.hadoop.odps.end.point=http://service.cn.maxcompute.aliyun.com/api
--conf spark.hadoop.odps.runtime.end.point=http://service.cn.maxcompute.aliyun-inc.com/api
--deploy-mode cluster
```

- 逻辑表查询时，必须带业务板块英文名为前缀，跨项目物理表使用需要带项目英文名为前缀。如果您需要查询开发环境数据，请在生产名称后加上`_dev`，系统会自动将生产业务板块、生产项目生成对应的变量。例如，您拥有业务板块`LD_Trade`，则系统自动生成业务板块变量`${LD_Trade}`。该变量在开发环境执行时默认被替换为`LD_Trade_dev`，在生产环境执行时默认被替换为`LD_Trade`。您也可以在执行时设置固定的值，提高代码在不同环境执行时的灵活性。

3. 在新键代码任务窗口中，填写任务名称、调度类型、描述等信息。信息填写完毕后，单击确定完成创建。



说明:

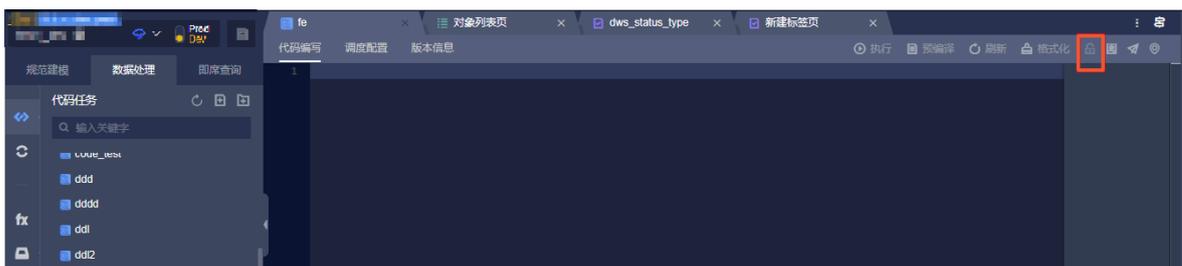
- 任务名称必须全局唯一。
- 调度类型分为周期性节点和手动节点。周期性节点定时调度，手动节点每次单击后运行。

- 生产开发隔离模式下，开发环境的数据处理对象发布至生产环境时，生产环境不会记录文件夹信息。请注意对象在项目内命名唯一，以免出现对象在对应生产环境的项目内重名，导致发布到生产环境时失败。

4.2.2 修改代码任务

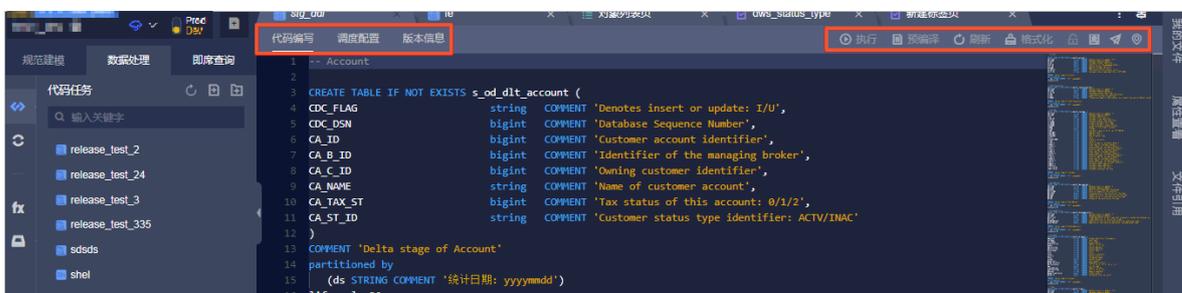
本文为您介绍如何修改代码任务。

1. 进入研发 > 开发 > 数据处理 > 代码任务功能模块。
2. 对于其他用户创建的代码任务，修改之前需先偷锁。



偷锁提高了代码开发过程中的稳定性。代码编写页面仅限解锁及被您自己锁定的状态方可编辑，如果他人锁定显示，单击进行偷锁，偷锁成功后显示.

3. 可以进行编辑代码，编辑配置调度参数，查看节点版本。编辑完成后，单击右上角功能区预编译、执行、、。同时您也可以单击页面右侧的文件引用，支持快速引用资源或函数。



- 执行即运行SQL。
- 预编译检验SQL语法是否符合规范。
- 单击格式化，系统自动调整您编写的SQL语句，使得其符合规范。
- 单击后，代码任务进入了待发布列表，详细请参见[发布管理](#)。

4. 单击调度配置，修改任务的调度信息。

- 如果您修改的调度类型是周期性节点，根据页面提示修改任务的调度信息，完成修改后单击确定。

• 基本信息

名称 dddf

节点id 暂无，请提交调度 [查看PROD节点](#)

节点类型 MAX_COMPUTE_SQL

负责人 dataphin功能测试

描述

优先级 中等优先级 ▼

参数配置

[节点参数配置说明](#)

• 调度配置

时间属性 正常调度 空跑调度 暂停调度

调度周期 日 ▼ 00:00 🕒

cron 表达式 0 0 0 ** ?

依赖上周期

• 依赖关系 开始解析

上游依赖 [新建上游依赖](#)

| 输出名称 | 节点名 | |
|------|-----|---|
| | | <div style="border: 1px solid #ccc; padding: 5px;"> <p>新建上游依赖</p> <p><input type="text" value="请输入关键字搜索节点"/></p> <p style="text-align: right;">取消 确定新增</p> </div> |

[新增](#)

当前节点

| 输出名称 | 节点名 | |
|------|-----|--|
| | | <div style="border: 1px solid #ccc; padding: 5px;"> <p>新增当前节点输出</p> <p><input type="text" value="请填写输出名称"/></p> <p style="text-align: right;">取消 确定新增</p> </div> |

| 区域 | 参数 | 说明 |
|------|----|--------------|
| 基本信息 | 描述 | 修改代码任务的基本描述。 |

| 区域 | 参数 | 说明 |
|------|-------|--|
| | 优先级 | 修改调度任务的优先级。 |
| | 参数配置 | 修改参数配置。 |
| 调度配置 | 时间属性 | 选择时间属性。时间属性包括正常调度、空跑调度、暂停调度。  说明: 空跑调度任务不执行脚本，直接把任务标记为执行成功。 |
| | 调度周期 | 配置调度周期。 |
| | 依赖上周期 | 选择是否依赖上周期。 |
| 依赖关系 | 上游依赖 | <p>a. 单击新增上游依赖，输入节点名称的关键字进行搜索。如果没有上游依赖输入virtual，选择租户默认的虚拟节点。</p> <p>b. 单击确定新增。</p>  说明: <ul style="list-style-type: none"> - 所有节点至少有一个上游依赖节点，如果没有，可以依赖一个虚拟任务节点。系统默认生成一个虚拟任务节点作为全局的根节点，即virtual_root_node。 - 调度的上游节点暂时仅支持按照节点输出名称模糊匹配、节点ID精确匹配，建议每个上游节点的节点名与输出名称保持一致，每个节点配置一个输出名称，以免调度管理出现误差，影响使用。生产环境和开发环境的常规任务节点ID一致。 |

| 区域 | 参数 | 说明 |
|----|------|--|
| | 当前节点 | <p>a. 单击 新增，填写当前节点输出名称。</p> <p>b. 单击 确定新增。</p> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p> 说明: 建议您将输出名称配置为 项目名.表名，因为输出名称全局唯一（生产、开发隔离，生产环境、开发环境各自独立校验输出名称的全局唯一性）。如果输出名称没有项目名前缀，可能会导致依赖节点识别不清晰，上游依赖节点解析失败，从而调度出错。</p> </div> |

- 如果您修改的调度类型是手动节点，根据页面提示修改任务的调度信息，修改完成后单击**确定**。

调度配置

- 基本信息
 - 名称
 - 节点id 暂无，请提交调度 [查看PROD节点](#)
 - 节点类型 MAX_COMPUTE_MR
 - 负责人 dataphin功能测试
 - 描述
 - 优先级
 - 参数配置
 - [节点参数配置说明](#)

| 参数 | 说明 |
|-----|--------------|
| 描述 | 修改代码任务的基本描述。 |
| 优先级 | 修改调度任务的优先级。 |

| 参数 | 说明 |
|------|---|
| 参数配置 | 修改配置的参数，支持自定义参数，例如 <code>keyA=123; keyB=\${yyyymmdd-7};</code> ，您也可以单击节点参数配置说明查看参数配置说明。 |

5. 单击页面右侧的文件引用，根据页面提示修改引用的函数和资源。

4.2.3 下线或删除代码任务

本文为您介绍代码任务的下线和删除操作。

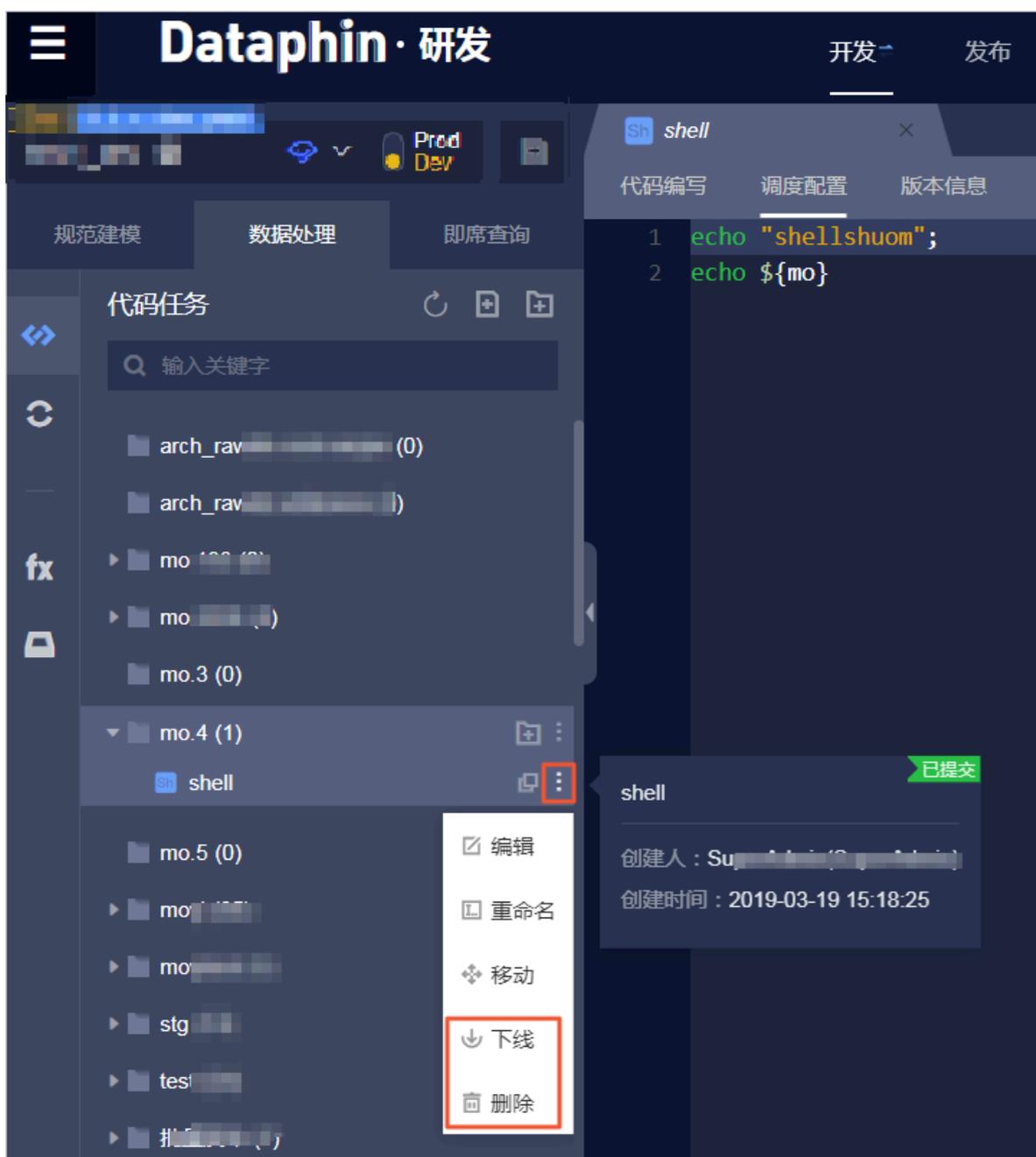
1. 进入研发 > 开发 > 数据处理 > 代码任务功能模块。
2. 在左侧列表中，鼠标悬停在需要下线或删除的代码任务上，单击右侧的“”图标，选择下

线或删除按钮，即可完成代码任务的下线、删除操作。



说明:

如果代码任务已被其他节点依赖，则不可下线或删除。

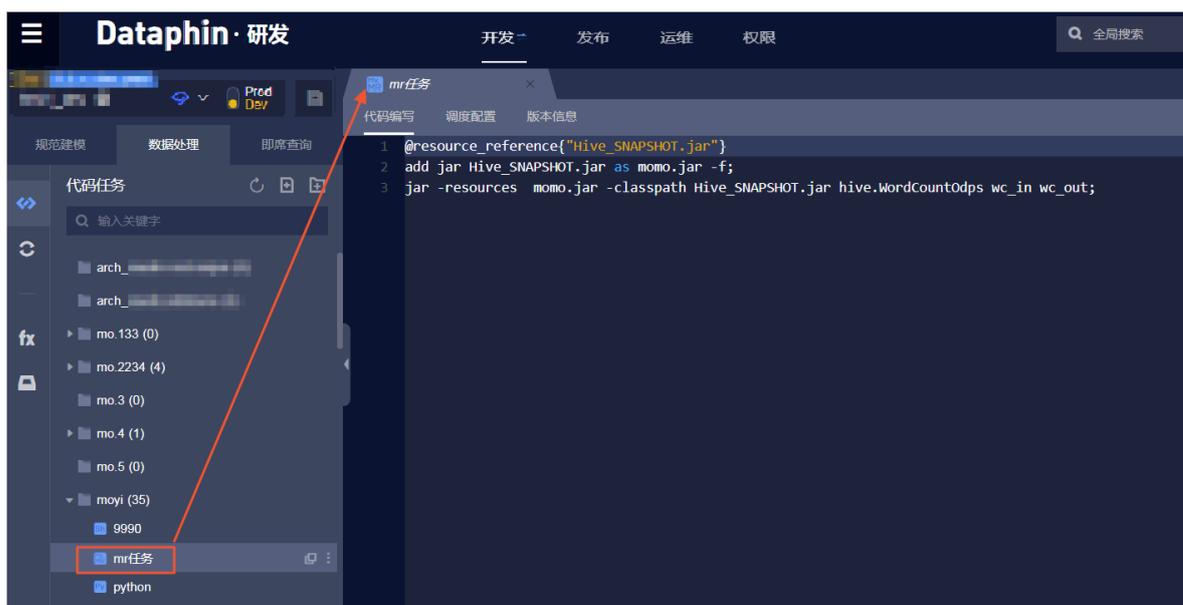


4.2.4 查看版本信息

本文为您介绍如何查看代码任务的版本信息。

1. 进入研发 > 开发 > 数据处理 > 代码任务功能模块。

2. 在左侧列表中，单击选择任意一个代码任务。



3. 单击版本信息，可以查看历史版本提交情况，并将当前编辑中的版本代码与所选历史版本提交代码进行对比。



4.3 同步任务

本文为您介绍同步任务及如何通过Shell代码任务实现同步任务。

同步任务

同步任务节点的配置，请参见数据引入模块的[配置同步任务](#)。

代码任务

Dataphin支持通过Shell配置同步任务。Shell任务中引用到的JSON文件需提前创建，因此需要您在资源管理中上传JSON资源包，然后在Shell任务中进行引用。

1. 本实例以MySQL数据库为例，[下载JSON文件实例](#)。根据JSON文件提示将您的数据库信息填写至JSON文件中。

2. 登录Dataphin。
3. 在Dataphin首页，单击研发 > 数据处理 > 资源管理，进入资源管理页面。
4. 单击资源管理后的，进入新建资源页面。



5. 完成参数配置后，单击提交。在弹出的提交备注页面，填写备注信息并单击确定并提交，即可完成DataX资源的创建。

新建资源
✕

* 类型 ▼

others

* 名称

DataX.json

* 描述

DataX

6/128

上传文件

📁 请点击选择文件上传

📎 datax.json

文件大小 1.60KB

选择目录 ▼

资源管理

取消

提交

| 参数 | 说明 |
|------|---|
| 类型 | 类型包括： <ul style="list-style-type: none"> · file · jar · python · others 此处选择others。 |
| 名称 | 填写新建资源的名称，例如DataX.json。 |
| 描述 | 填写资源的描述。 |
| 上传文件 | 上传DataX.json文件。 |
| 选择目录 | 选择资源文件的所放置的目录。 |

6. 在数据处理页面，单击代码任务，进入代码任务页面。

7. 单击代码任务后的 \oplus ，单击SHELL。



8. 根据页面提示完成参数配置，单击确定，进入DataX任务代码编写页面。

新建文件 ×

* 名称

* 调度类型 周期性节点 手动节点

描述

选择目录

| 参数 | 说明 |
|----|--------------------|
| 名称 | 填写代码任务的名称，例如DataX。 |

| 参数 | 说明 |
|------|-------------|
| 调度类型 | 选择周期性节点。 |
| 描述 | 填写对任务的简单描述。 |
| 选择目录 | 选择代码任务的目录。 |

9. 在DataX任务代码编写页面，进行编写代码任务。



编写的代码如下。

```
@required_resource{required_memory=2Gb;required_cpus=1.0}
@resource_reference{"Datax.json"}
python $DATA_HOME/bin/datax.py --jvm '-Xms2g -Xmx2g' Datax.json
```

其中，

- `required_resource{}`，对所需资源大小进行自定义配置。
- 系统已经内置DataX_Home为DataX的安装目录，DataX的入口类在DataX安装的bin文件下。
- `--jvm '-Xms2g -Xmx2g'`用来指定DataX实际运行时的JVM内存，建议您将内存的大小设置成与`required_resource`中的`required_memory`一致。

对于所需资源较小的DataX任务，您可以使用如下代码示例。

```
@resource_reference{"Datax.json"}
```

```
python $DATA_HOME/bin/datax.py Datax.json
```

10完成代码编写后，单击页面右上角的进行保存，然后单击进行提交代码。发布该代码任务请参见[发布管理](#)，发布后的Shell代码任务参与生产环境的调度。

4.4 函数管理

4.4.1 新建用户自定义函数

函数管理主要用于管理代码开发过程中用到的SQL函数，包括计算引擎源常用且系统默认支持的函数和用户自定义函数，前者不支持编辑操作。本文为您介绍如何新建用户自定义函数。

背景信息

如果系统提供的函数不能满足您的诉求，您可以将函数的代码文件通过新建资源的方式上传到Dataphin控制台中（请参见[新建资源](#)），再通过引用该资源的方式，注册函数。

操作步骤

1. 进入研发 > 开发 > 数据处理 > 函数管理功能模块。
2. 您可以通过下图中3种方式中的任何1种方式，进入新建函数页面。



3. 在新建函数页面，根据页面提示配置参数。

新建函数
✕

* 名称

* 选择资源

* 类名

* 类型

* 命令格式

* 使用文档

选择目录

| 参数 | 说明 |
|------|---|
| 名称 | 填写名称。名称由字母，数字，下划线组合组成，且首字符仅限字母。 |
| 选择资源 | 选择资源文件。下拉框中提供的是匹配当前项目的资源名称。 <div style="background-color: #f0f0f0; padding: 5px; margin-top: 5px;"> 说明: <ul style="list-style-type: none"> 选择多个资源时，仅限但选择同一种资源类型。 如果您还没有资源，请参见新建资源创建。 </div> |
| 类名 | 填写类名。针对Maxcompute的计算类型中的资源，提取资源中类的内容，例如test_udf.UDFGETSrcId。 |
| 类型 | 选择类型，下拉框中可供选择的类型包含窗口、统计、数值、字符串、时间、ip地址相关函数、URL、编解码、业务、其他。 |

| 参数 | 说明 |
|------|--|
| 命令格式 | 填写命令格式。命令格式即函数引用格式，例如， <code>bigintweekday (datetime date)</code> 。 |
| 使用文档 | 填写函数使用说明，例如。 <pre>select get_week_date("20170810",0,2),--获取8月10日本周周二日期 from cndata.dual</pre> |
| 选择目录 | 选择目录。目录默认选择您新建函数时选择的目录，您也可以单击下拉箭头选择其他目录。 |

4. 参数配置完成后，单击提交完成函数创建。您可以在即席查询中通过执行SQL语句（SQL语句中引用该函数）验证函数是否符合预期效果。关于即席查询，请参见[即席查询概述](#)。

4.4.2 管理用户自定义函数

本文为您介绍如何进行修改、删除、移动用户自定义函数，帮助您管理用户自定义函数。

前提条件

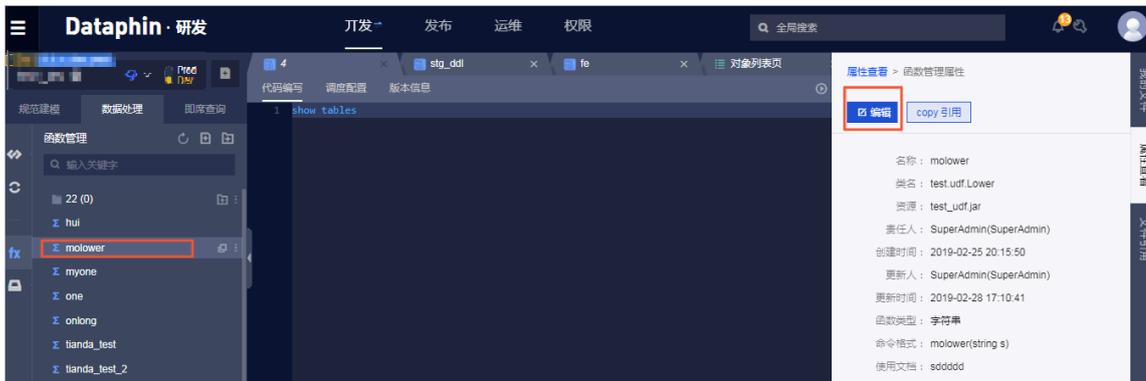
已完成用户自定义函数的创建。

操作步骤

1. 进入研发 > 开发 > 数据处理 > 函数管理功能模块。

2. 您可以通过以下2种方式中的任何1种方式，进入编辑函数页面。

- 单击某个自定义函数，右侧弹出该函数的属性框，单击编辑。



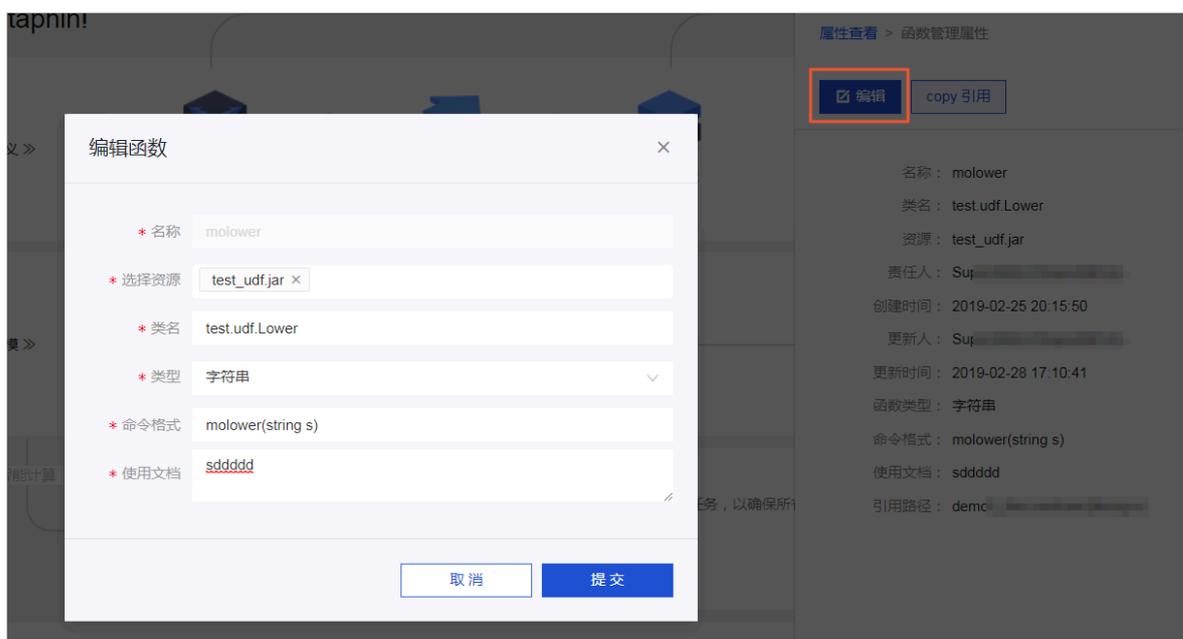
- 将鼠标悬浮在需编辑的自定义函数后的 ，单击编辑。



您也可以单击删除，删除自定义函数。

同时，您也可以单击移动，将自定义函数移动到其他的自定义函数目录中。

3. 在编辑函数页面，根据页面提示修改参数，详细内容请参见[新建用户自定义函数](#)。



说明：

如果该函数引用的资源文件有更新，资源文件更新后，需重新提交函数，函数引用的资源文件才会更新。

4.5 资源管理

4.5.1 新建资源

本文为您介绍如何创建资源。

背景信息

- 资源管理用于将代码研发所需的各种文件集中存储并管理。例如，创建UDF需要的Jar包以及Shell、Python等脚本文件。
- MaxCompute计算引擎类型下，使用OSS存储文件。

操作步骤

1. 登录Dataphin。
2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。
3. 选择开发 > 数据处理 > 资源管理。

4. 您可以通过以下方式新建资源：

- 如下图中编号1所示，单击项目名称右侧的  图标，选择数据开发 > 资源。
- 如下图中编号2所示，单击左侧功能区的新建文件按钮。
- 如下图中编号3所示，在操作指引图上，单击资源右侧的  图标。



5. 在新建资源对话框中，填写资源类型、名称、描述，并上传文件、选择目录，单击提交。

新建资源 ✕

* 类型

* 名称

* 描述 0/128

上传文件

选择目录



说明:

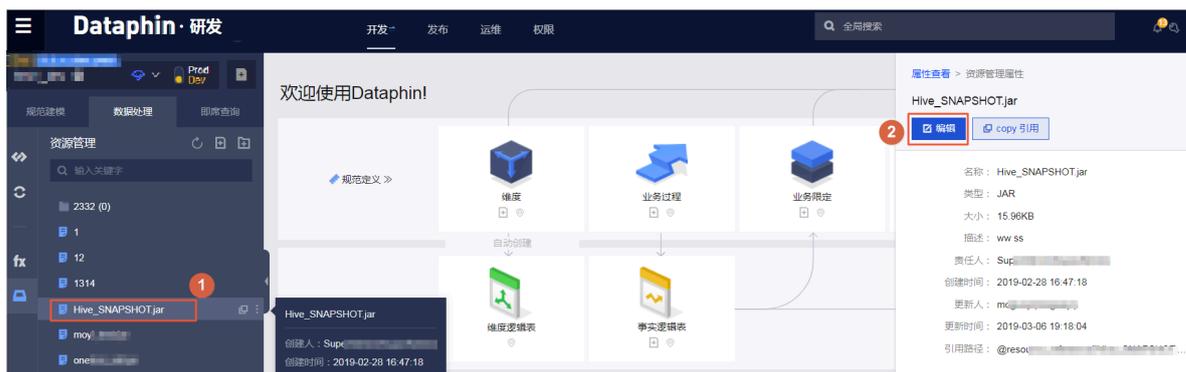
- 资源名称在Dataphin系统内全局唯一。
- 默认支持上传file、Jar、Python等类型的文件，每个文件最大50MB，最多上传1000个文件。

4.5.2 修改资源

本文为您介绍如何修改资源。

1. 进入研发 > 开发 > 数据处理 > 资源管理功能模块。

2. 在左侧列表中，单击选择需要修改的资源文件，右侧弹出资源属性信息框。



3. 单击编辑按钮，即可进入资源修改页面。



4. 修改描述信息，重新上传文件，单击提交完成资源修改。



说明:

默认提供file、Jar、Python及其他文件类型上传。每个文件最大50MB。如果该文件被函数引用，文件更新后，需要重新提交函数，函数引用的文件才会更新。

4.5.3 删除资源

本文为您介绍如何删除资源。

1. 进入研发 > 开发 > 数据处理 > 资源管理功能模块。
2. 在左侧列表中，选择需要删除的资源，单击右侧浮现的“”图标，选择删除即可。



5 即席查询

5.1 即席查询概述

即席查询提供数据查询功能。MaxCompute计算引擎类型下支持MaxCompute SQL。系统会根据您所设置的计算类型自动识别切换SQL类型。

因为Dataphin独有的逻辑表模型以业务板块为命名空间，所以针对逻辑表，Dataphin语法如下：

- 引用逻辑表的方式为：业务板块.逻辑表。
- 支持[业务板块.逻辑表.关联维度角色.关联维度角色.维度逻辑表字段]的查询逻辑。如果作为SELECT或者WHERE等设置的内容，则仅需填写[逻辑表.关联维度角色.关联维度角色.维度逻辑表字段]，即可实现关联维度对象的字段查找使用。



说明：

虚拟维度的主键字段不能用于即席查询，虚拟维度仅用于构建统计粒度。

在Dataphin首页，单击研发 > 即席查询，进入即席查询页面。



5.2 新建即席查询

本文为您介绍如何新建即席查询任务。

1. 进入研发 > 开发 > 即席查询功能模块。
2. 创建即席查询任务，共有3个操作入口。



3. 单击创建即席查询任务之后，会弹出信息窗口，填写任务名称、描述，选择目录，然后单击确定即可。

新建文件

* 名称

描述

选择目录

5.3 管理即席查询

本文为您介绍如何进行编辑、执行、删除即席查询任务及编辑器，帮助您更好的管理即席查询任务。

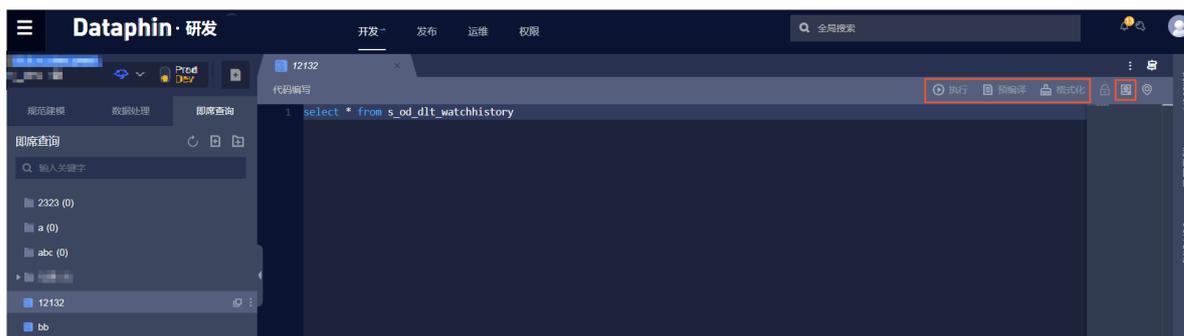
编辑及执行即席查询任务

1. 进入研发 > 开发 > 即席查询功能模块。
2. 在左侧列表中，单击已创建的即席查询任务，会出现代码编写窗口。单击偷锁后开始编辑SQL，编辑SQL请参见[编辑器](#)。



- 偷锁提高了代码开发过程中的稳定性。代码编写页面仅限解锁及被您自己锁定的状态方可编辑，如果他人锁定显示，单击进行偷锁，偷锁成功后显示。
- 使用跨项目物理表时，名称需以项目英文名为前缀。例如`select * from cloudtest_dev.table4`，其中`table4`为物理表。
- 逻辑表查询时，名称需以业务板块的英文名为前缀。例如`select province from ld_practice.dim_province`，其中`dim_province`为逻辑表。
- 查询开发环境数据时，需生产环境名称后加上`_dev`，系统会自动根据生产环境中的业务板块及项目生成对应的变量。例如，业务板块`LD_Trade`，则系统自动生成业务板块变量`${LD_Trade}`。该变量在开发环境执行时默认被替换为`LD_Trade_dev`，在生产环境执行时默认被替换为`LD_Trade`。您也可以执行时设置固定的值，提高代码在不同环境执行的灵活性。

3. SQL代码编辑完成后，可执行单击右上角功能区预编译、格式化、。



- 执行即运行SQL。
- 预编译检验SQL语句是否符合规范。
- 单击格式化，系统自动调整您编写的SQL语句，使得其符合规范。

删除即席查询任务

在即席查询页面，鼠标悬浮在需删除的即席查询任务后的，单击删除即可删除即席查询。

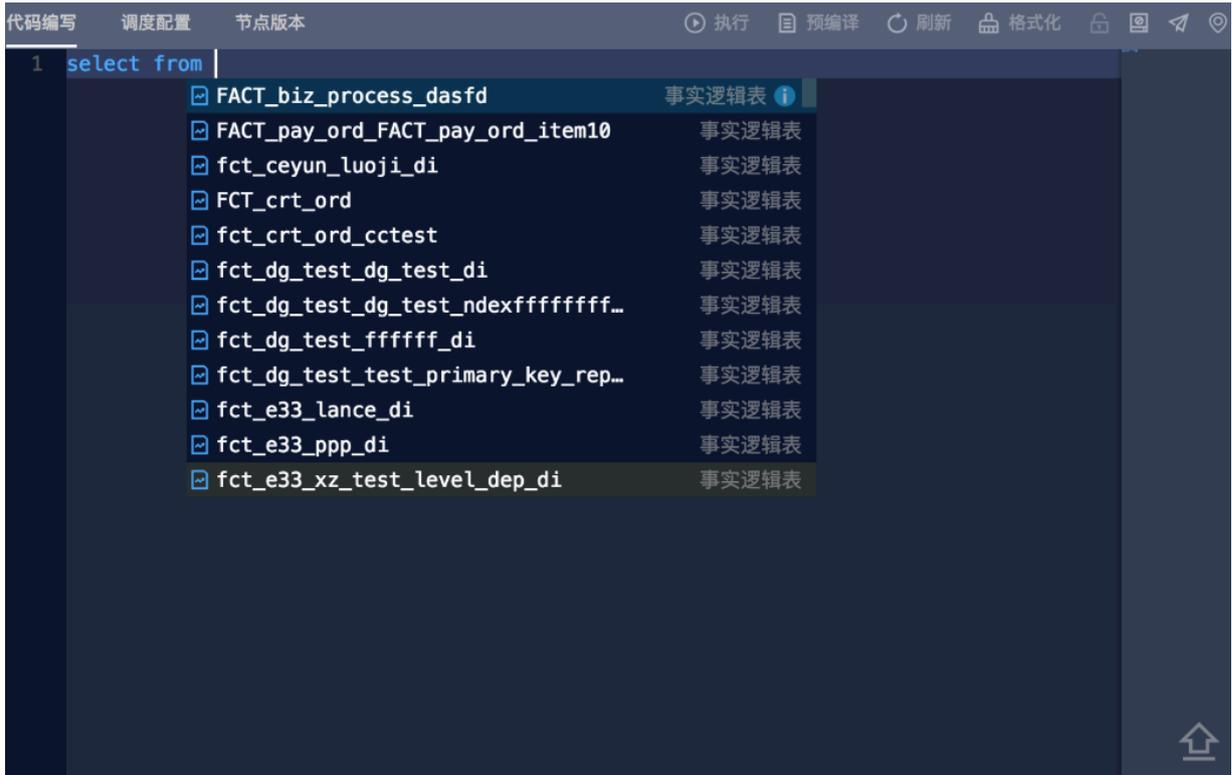


编辑器

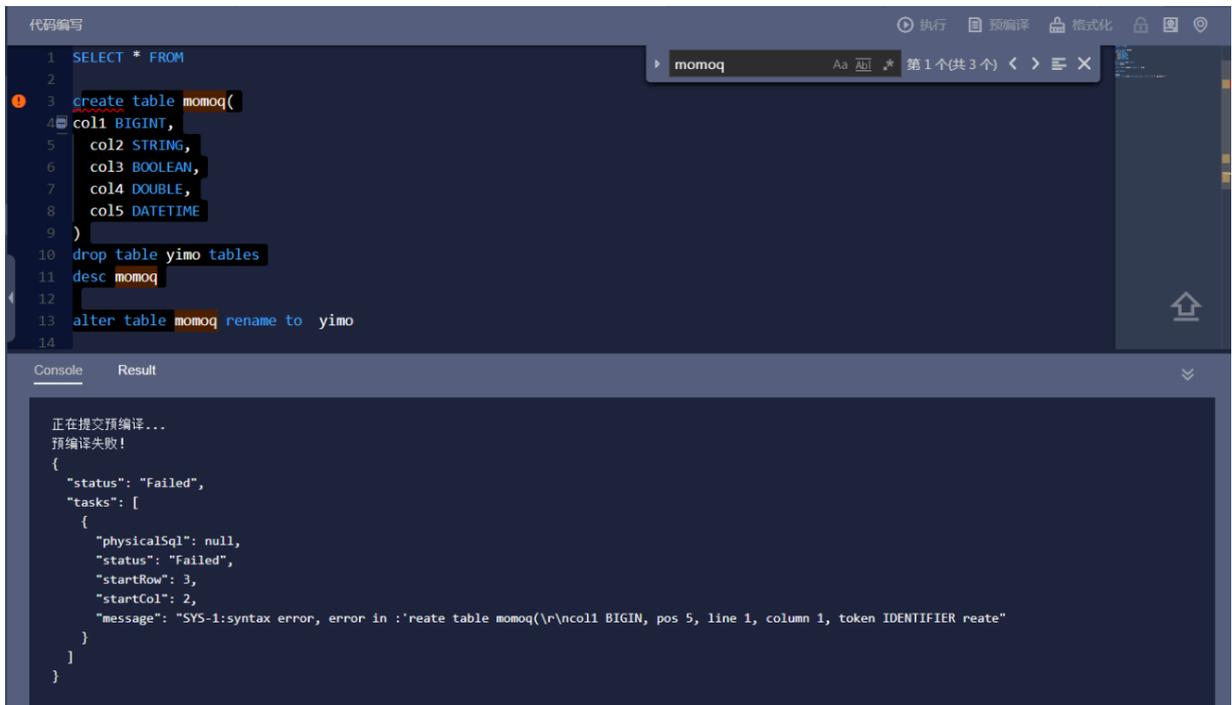
Dataphin的编辑器通过分析用户输入，构建AST抽象语法树，从而提供精确的语法提示、语法补全和语法纠错功能。同时对接元数据，使UDF详情、表详情一键可查，提供智能化、沉浸式的开发体验。支持标准SQL及其他语法，如Shell和Python语法。

```
1 insert overwrite table s_od_watchlist partition(ds='${bizdate}')
2 select
3     W_C_ID as CUSTOMER_ID
4     ,W_S_SYMB as SECURITY_SYMBOL
5     ,SUBSTR(w_actv_dts, 1, 10) as PLACED_DATE_ID
6     ,SUBSTR(w_cncl_dts, 1, 10) as REMOVED_DATE_ID
7     ,CASE
8     WHEN w_actv_dts IS NOT NULL
9     AND w_cncl_dts IS NULL |
10    THEN
11    WHEN w_actv_dts IS NOT NU &&
12    AND w_cncl_dts IS NOT &&
13    AND w_actv_dts >= w_c &&
14    THEN 'ACTV' IS
15    WHEN w_actv_dts IS NOT NU OR
16    AND w_cncl_dts IS NOT XOR
17    AND w_actv_dts < w_cn && ||
18    THEN 'CNCL'
19 END as WATCH_STATUS
20 FROM (
21     SELECT W_C_ID, W_S_SYMB
22     , CASE
23     WHEN W_ACTION = 'ACTV' THEN w_dts
24     ELSE NULL
25     END AS w_actv_dts
```

支持预编译下的语法检查、代码自动补全、智能提示（关键字、项目名、表名、字段名的自动补全，以及逻辑表SQL的雪花模型关联提示）及支持子查询的字段透传等。



支持代码格式化、代码折叠、代码缩略图、代码搜索功能等。



6 发布管理

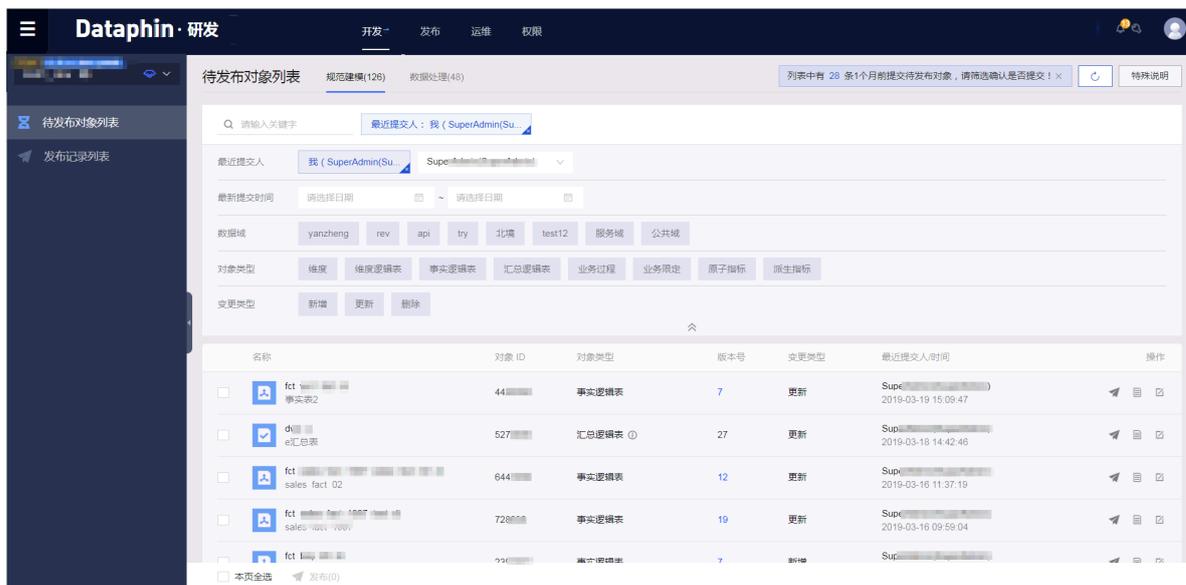
开发环境生成的规范建模、数据处理类任务提交后，只有发布成功，任务才会同步到生成环境进行调度。本文为您介绍如何进行发布任务及查询发布记录。

开发环境生成的规范建模、数据处理类任务提交后，都会进入待发布列表。需要在发布页面进行任务发布，这些任务才会在生产环境上线。开发环境与生产环境相对独立，任务在开发环境编辑，进行调试、执行、验证后，再复制到生产环境，保证生产环境和开发环境数据隔离，且生产环境数据、任务稳定不受影响。

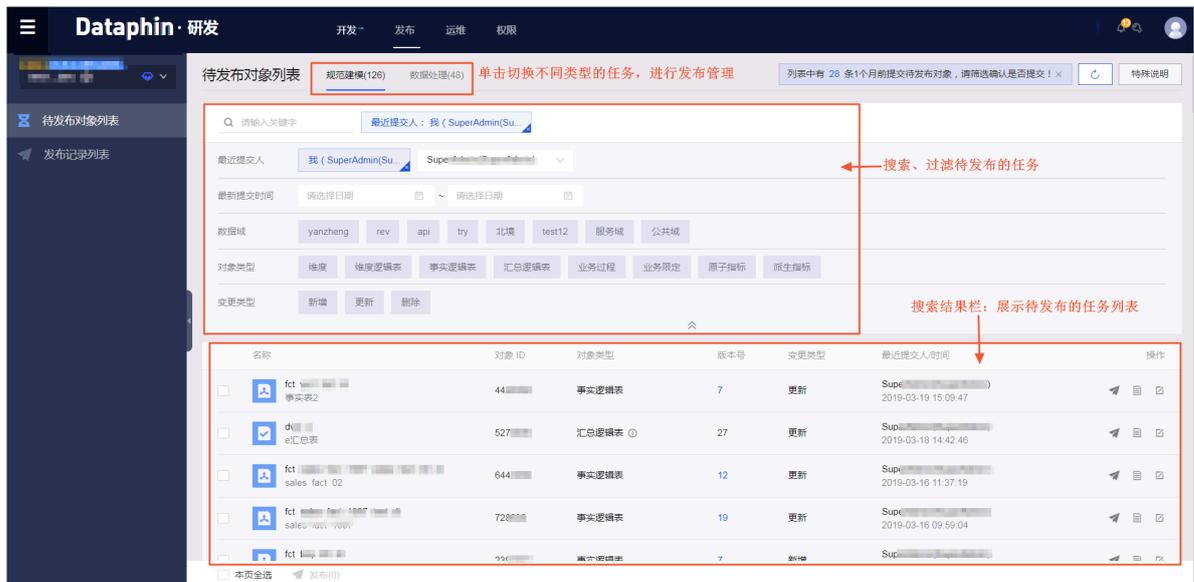
 **说明:**
发布管理功能仅开发环境可见。

待发布对象列表

- 1. 进入研发页面。
- 2. 单击发布页签，切换到发布管理页面，如下图所示。



3. 在待发布对象列表页，可以搜索、查看和发布已提交的任务。



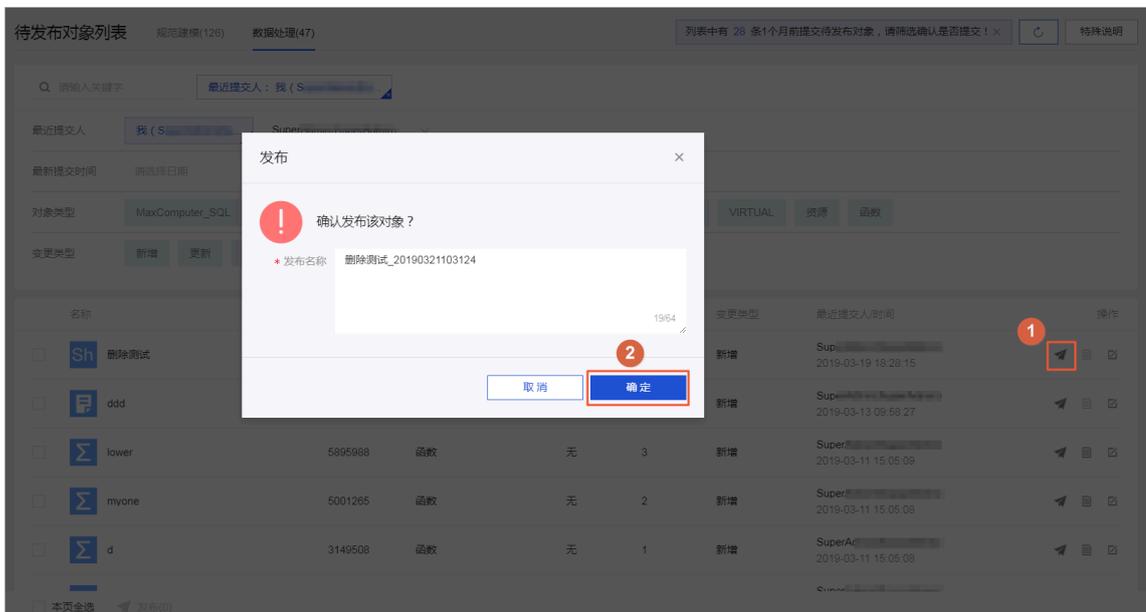
待发布的任务，按类型不同（规模建模、数据处理），分页展示。

- 用户可以在右侧面板上方的搜索框中，输入关键字搜索待发布的任务。单击下拉箭头可展开筛选项，支持通过选择最近提交人、提交时间、对象类型、变更类型等来过滤搜索结果。
- 下方显示搜索结果，会展示待发布任务的名称、对象ID、对象类型、版本号、变更类型、最近提交人/时间等信息。

4. 发布任务。

- 单独发布

在下方的搜索结果展示区，单击任务最右侧操作栏的发布按钮，可以将任务发布到生产环境。

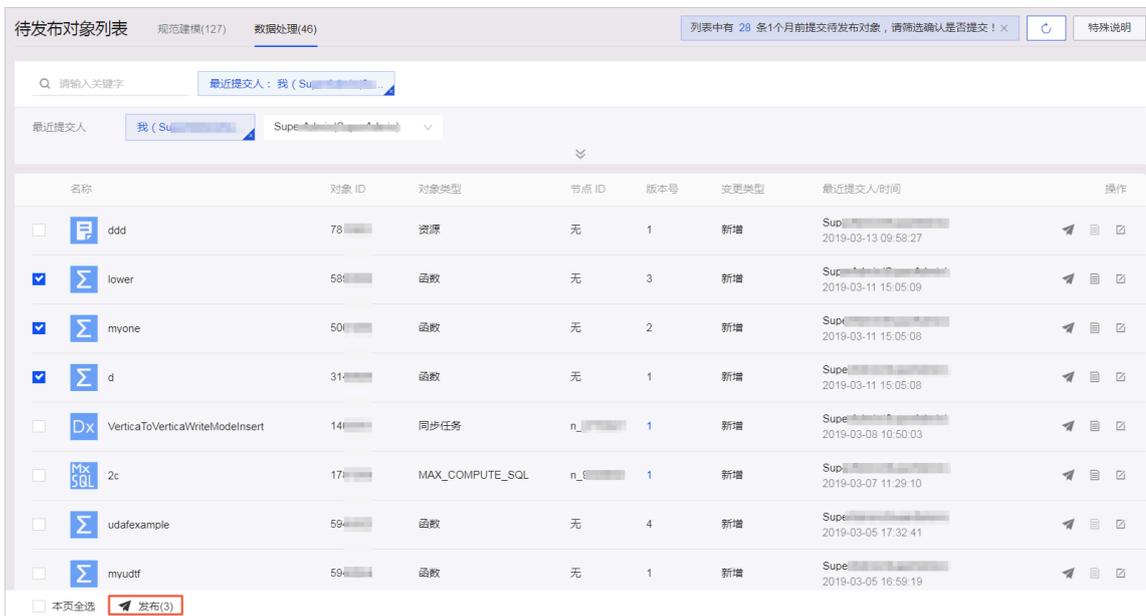


- 批量发布

在下方的搜索结果展示区，勾选需要发布的任务，单击左下角的发布按钮，即可批量发布选中的任务。

 说明:

批量发布有数量限制，批量发布超过10个，有可能会出错，请您根据发布出错原因，查看是否发布重试，或者修改提交后再发布。详情请参见界面右上角的特殊说明。



5. 单击任务后面操作栏的查看发布记录按钮，可以跳转到发布记录列表页面，获取该对象历史的发布记录信息。单击编辑跳转按钮，可以跳转到任务的开发编辑页面。



说明:

- 发布时，部分对象有些特殊的约束操作，以保证能完整地发布至生产环境及正常运行，详情请见待发布对象列表页面右上角的特殊说明。
- 开发环境的数据处理对象发布至生产环境时，生产环境不会记录文件夹信息，请保证对象在对应生产环境的项目内没有重名对象，以免出现对象重名，发布失败。

发布记录列表

单击左侧菜单栏的发布记录列表，切换到发布记录列表页面，搜索、查看任务的发布状态、发布详情，同时支持编辑跳转至该对象的开发编辑页面。

发布记录列表 规范建模(1038) 数据处理(618)

Q 请输入关键字 最新发布人: 我 (SuperAdmin)

最新发布人: 我 (SuperAdmin) SuperAdmin

最新发布时间: 请选择日期 请选择日期

数据域: yanzheng rev api try 北境 test12 服务域 公共域

对象类型: 维度 维度逻辑表 事实逻辑表 汇总逻辑表 业务过程 业务限定 原子指标 衍生指标

变更类型: 新增 更新 删除

发布状态: 失败 成功 发布中

| 发布名 | 发布 ID | 名称 | 对象 ID | 对象类型 | 版本号 | 变更类型 | 发布人时间 | 发布状态/完成时间 | 操作 |
|---------------------|------------|----------------------|-------|-------|-----|------|--|--------------------------|-----------|
| 叶子 ce1_201903201... | 20002_... | dim ce1 leaf 叶子ce1 | 63 | 维度 | 3 | 更新 | SuperAdmin(SuperAdmin) 2019-03-20 16:20:29 | 发布成功 2019-03-20 16:20:39 | 编辑跳转 发布详情 |
| x_20190320161922 | 20002_x... | dim x | 61 | 维度 | 3 | 删除 | SuperAdmin(SuperAdmin) 2019-03-20 16:19:24 | 发布成功 2019-03-20 16:19:25 | 发布详情 |
| 20002_2... | | dim cewd leaf 叶子cewd | 81 | 维度逻辑表 | 3 | 新增 | SuperAdmin(SuperAdmin) 2019-03-19 16:57:27 | 发布成功 2019-03-19 16:57:42 | 发布详情 |
| | | dim cewd_M3 | | | | | SuperAdmin(SuperAdmin) | 发布成功 | 详情 |

7 数据萃取

7.1 数据萃取概述

基于Dataphin数据建模研发沉淀的数据，数据萃取提供以目标对象为中心的数据打通和深度挖掘功能，并生成代码与调度任务，完成实体对象识别、连接及标签生产，可快速应用于各类业务。

初始化配置

1. [登录Dataphin](#)。
2. 将鼠标悬停在开发上后，单击萃取，进入初始化配置页面。



3. 根据页面提示选择计算引擎源后，单击测试连接性。

如果您没有计算引擎源，请参见[计算引擎源](#)创建计算引擎源。

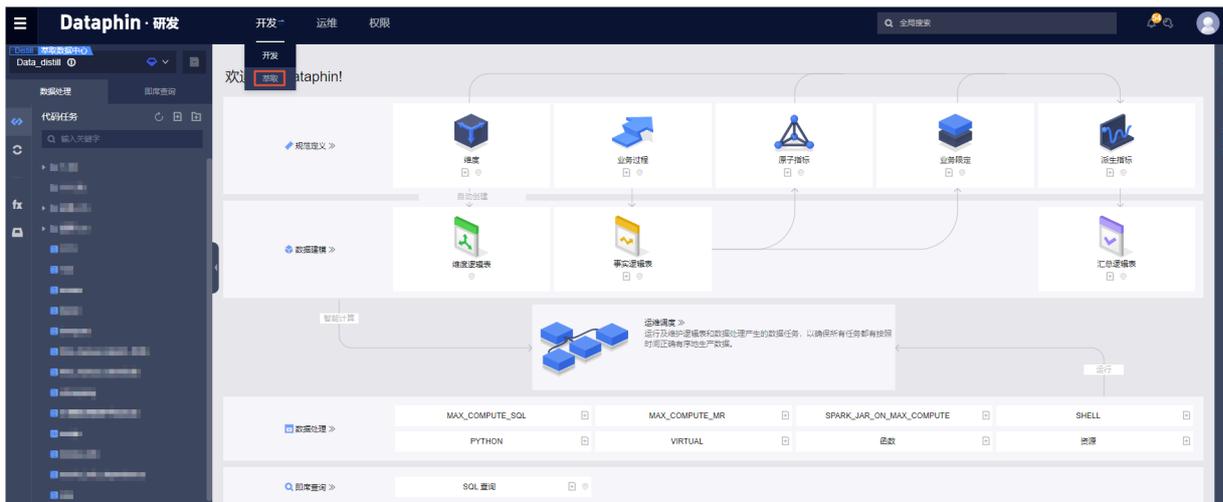
4. 测试连接通过后，单击确定并开始初始化，初始化成功后即可在萃取模块进行数据的开发。

萃取简介

萃取由行为中心和标签中心组成：

- 行为中心提供明细行为的配置及行为数据统计信息。
- 标签中心支持标签可视化配置并自动生成标签调度任务，对不同开发状态的标签任务进行管理。

在Dataphin控制台首页，单击研发后，鼠标悬浮在开发上选择萃取，即可进入萃取首页。



注意事项

请勿删除萃取模块中的系统表，如下所示。

| 表名 | 名称 | 描述 |
|-----------------------------|------------|--------------------------------------|
| oi_behavior_detail_d | 行为中心源数据接入 | 抽取源数据的行为大宽表。 |
| oi_behavior_detail_d_sample | 行为中心抽样表 | oi_behavior_detail_d的抽样表，用于行为看板查看数据。 |
| oi_behavior_label_value_d | 行为中心打标结果 | 行为打标的中间表。 |
| oi_behavior_label_value_all | 行为中心打标结果 | 行为打标的累积表。 |
| oi_person_label_value_all | 行为中心打标结果 | 行为中心打标结果的表。 |
| oi_label_value_rule_d | 标签打标规则表 | 标签打标的中间表。 |
| oi_dim_behavior_id | 行为定义维表 | 行为定义的表，从源数据同步而来。 |
| oi_behavior_stat_d | 行为数据统计天表 | 行为数据统计表。 |
| oi_behavior_stat_sync | 行为数据统计90天表 | 行为统计90天的数据。 |
| oi_behavior_object_all | 行为数据对象累积年表 | 行为对象，对象属性的累积表。 |
| oi_behavior_property_value | 行为数据属性表 | 对象值表，存储行为、对象以及对象值的表，供您在标签打标时参考。 |

7.2 行为中心

7.2.1 行为中心概述

行为中心由行为元素、行为规则、行为看板组成。

行为元素

行为元素管理功能提供行为数据的标准化定义与归类。行为元素分为行为域、业务线、动作、对象、对象属性，您可以对行为元素进行归类管理。

- **行为域**：聚合业务含义一致的行为数据，如电商域、文娱域。
- **业务线**：基于行为域将行为数据进一步细分，各业务线之间相对独立，如淘宝业务线、天猫业务线。
- **动作**：行为主体发出的操作，如购买、浏览。
- **对象**：行为主体操作的具体事物，如商品、电影。
- **对象属性**：对象的描述性信息，如名称、品牌、年份。

例如，在电商领域的行为数据中，您可以将电商抽象为一个行为域。在行为域中按照区域可分为国内业务线和国际业务线，动作为支付，对象是商品，对商品的描述（类目、品牌、价格、尺寸、颜色等）属于对象属性。

在萃取首页，单击行为中心后，鼠标滑动到左侧导航栏可显示行为元素。通过单击具体的行为元素即可进入其管理页面。

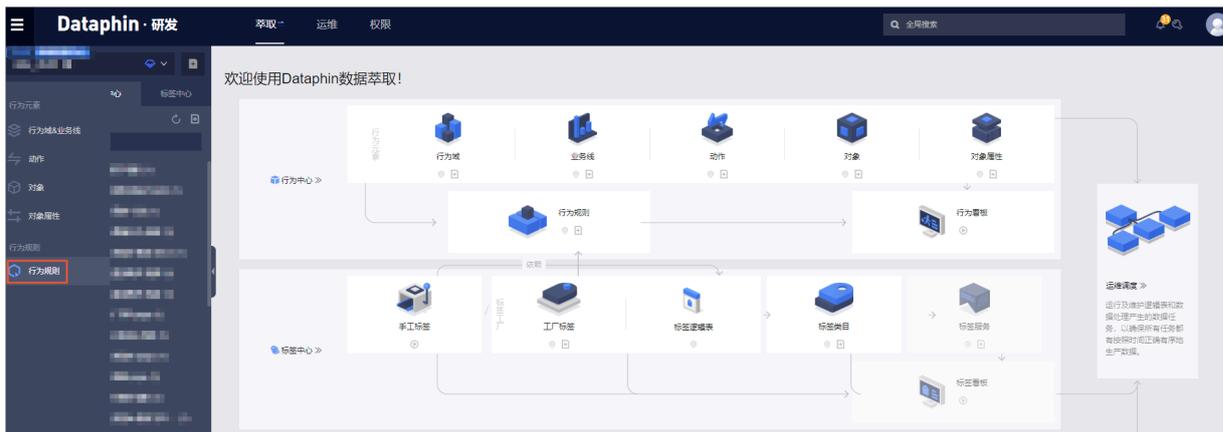


行为规则

行为规则是基于定义的行为元素，用于进一步配置对应的数据来源及其加工清洗规则。

行为规则根据行为和来源表唯一确定，将规范结构化的行为明细（行为域、业务线、动作、对象）与实际数据进行匹配。

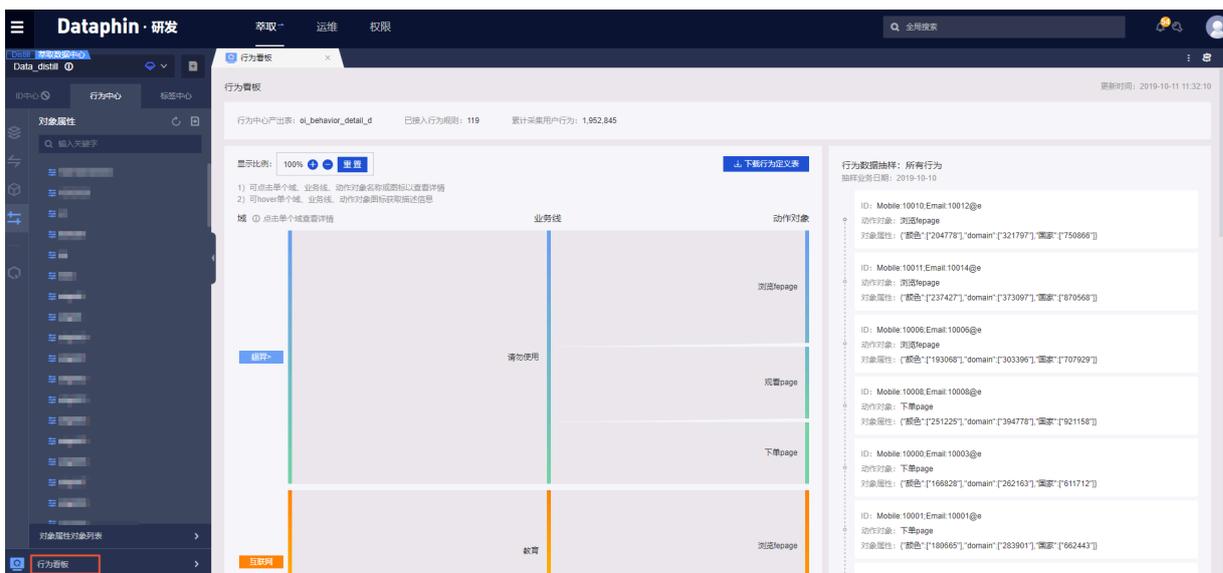
在萃取首页，单击行为中心后，鼠标滑动到左侧导航栏可显示行为规则。单击行为规则可进入其管理页面。



行为看板

行为看板为您提供行为数据分布及抽样数据展示。帮助您动态获取以目标对象为中心的所有行为的规范化结构化聚集。

在萃取首页，单击行为中心 > 行为看板，即可进入行为看板页面。



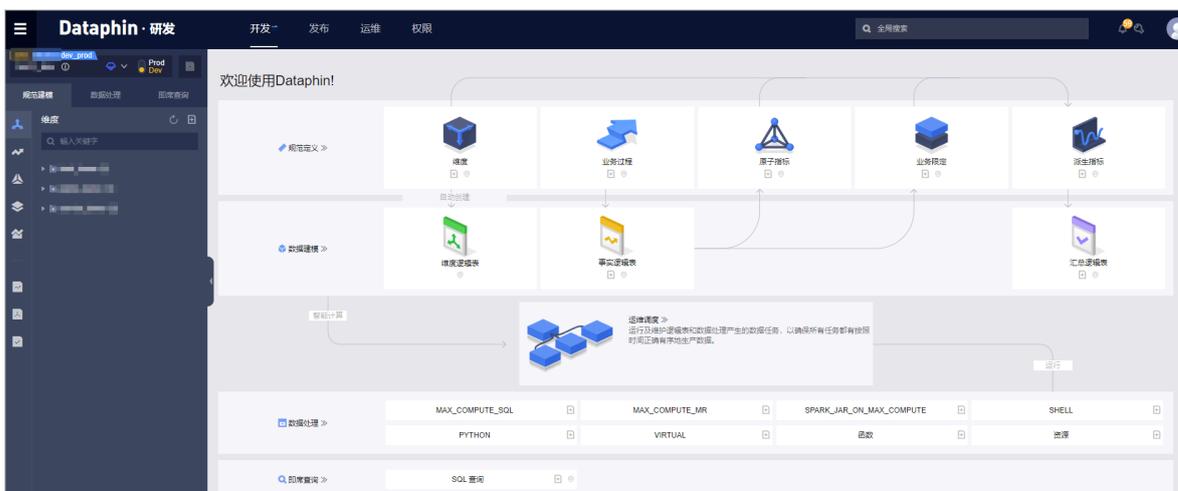
7.2.2 管理行为域

行为域是聚合业务含义一致的行为数据，如电商域、文娱域。本文为您介绍如何新建、编辑、删除行为域。

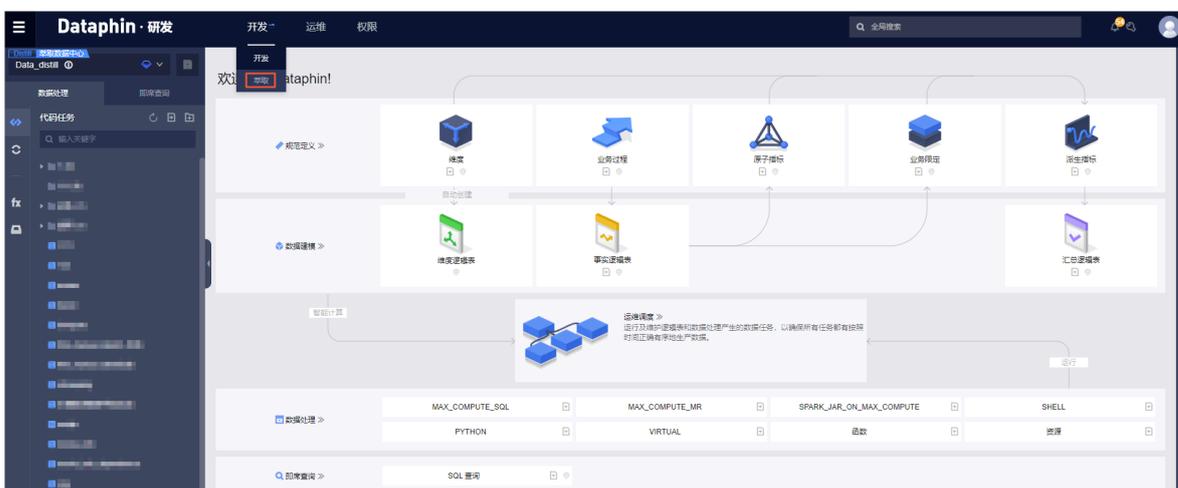
新建行为域

1. 登录Dataphin。

2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。



3. 将鼠标悬停在开发上，单击萃取，进入数据萃取研发页面。



4. 您可以通过以下任何一种方式进入新建行为域页面：

- 在数据萃取研发页面，单击左侧导航栏中的  后，单击行为域&业务线后的 ，然后单击行为域，进入新建行为域弹框。



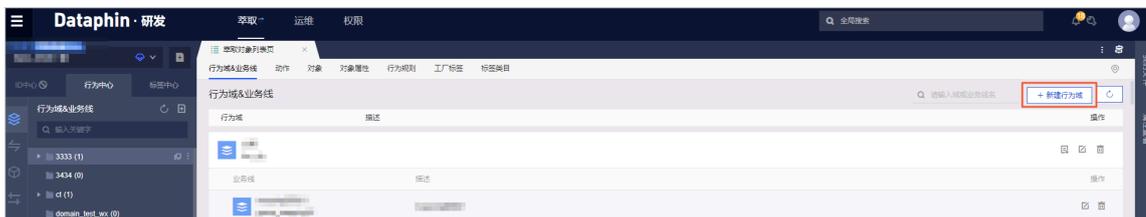
- 在数据萃取研发页面，单击项目名称后 ，选择行为中心 > 行为域，进入新建行为域弹框。



- 在萃取首页，单击行为域下的，进入新建行为域弹框。



- 在数据萃取研发页面，单击左侧导航栏下的行为域&业务线对象列表，进入萃取对象列表页面。在行为域&业务线页面，单击右上角的新建行为域，进入新建行为域弹框。



5. 根据页面提示配置行为域英文名、行为域名称、描述参数。配置完参数后单击提交。

新建行为域 ×

* 行为域英文名

* 行为域名称

描述



说明:

行为域英文名、行为域名称不支持与已有的行为域英文名、行为域名称重复。

6. 在弹出的提交备注页面，填写备注以说明此次提交的变更内容。单击确定并提交，即可完成行为域的创建。

* 提交备注

0/128

填写备注以说明此次提交的变更内容
当前项目，提交成功后，即更新至生产环境，请谨慎操作!



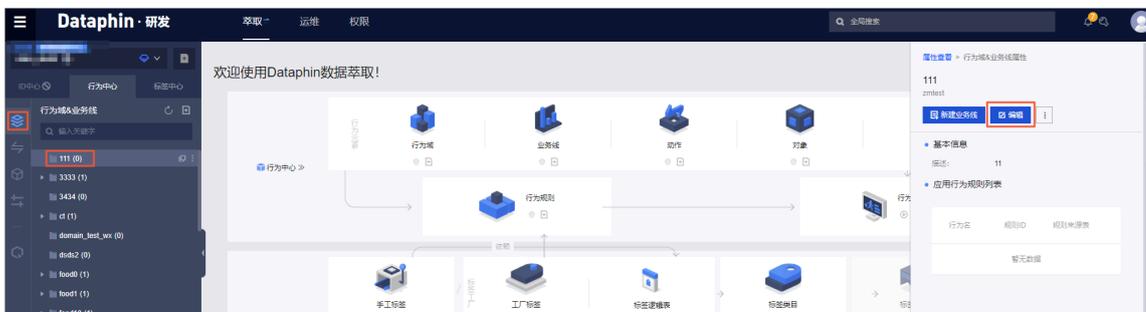
说明:

行为元素新建成功即发布到生产环境，因此需填写提交备注。

编辑行为域

1. 您可以通过以下任何一种方式，进入编辑行为域页面：

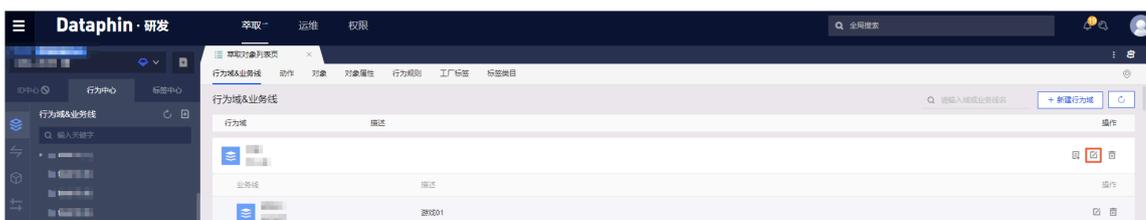
- 在数据萃取研发页面，单击左侧导航栏中的，然后单击需编辑的行为域。右侧弹出该行为域的属性框，单击编辑，进入编辑行为域弹框。



- 在数据萃取研发页面，单击左侧导航栏中的，将鼠标悬浮在需编辑行为域后的，单击编辑，进入编辑行为域弹框。



- 在数据萃取研发页面，单击左侧导航栏下方的行为域&业务线对象列表，进入萃取对象列表页面。在行为域&业务线页面，单击操作列下的，进入编辑行为域弹框。



2. 根据页面提示可修改行为域英文名、行为域名称、描述参数。完成修改后单击提交。

编辑行为域

① 不建议您对元素本意进行大幅变动，大幅变动将对已生成行为数据含义产生影响。如：您可以把“看”改为“观赏”，但不建议把“看”改为“听”，如需“听”建议您另新建一个元素。

* 行为域英文名

* 行为域名称

描述



说明:

行为域英文名、行为域名称不支持与已有的行为域英文名、行为域名称重复。

3. 在弹出的提交备注页面，填写备注以说明此次提交的变更内容。单击确定并提交，即可完成行为域的编辑。

* 提交备注

请输入此次提交的相关备注说明，如变更信息

0/128

填写备注以说明此次提交的变更内容
当前项目，提交成功后，即更新至生产环境，请谨慎操作!



说明:

行为元素编辑成功即发布到生产环境。因此需填写提交备注。

删除行为域

1. 您可以通过以下任何一种方式，进入删除行为域页面：

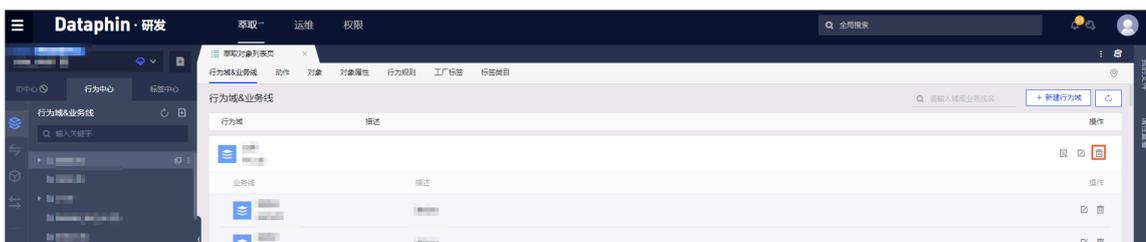
- 在数据萃取研发页面，单击左侧导航栏中的，然后单击需删除的行为域。右侧弹出该行为域的属性框，单击后，单击删除，进入删除行为域弹框。



- 在数据萃取研发页面，单击左侧导航栏中的，将鼠标悬浮在需删除行为域后的，单击删除，进入删除行为域弹框。



- 在数据萃取研发页面，单击左侧导航栏下方的行为域&业务线对象列表，进入萃取对象列表页面。在行为域&业务线页面，单击操作列下的，进入删除行为域弹框。



2. 在弹出的删除行为域页面，填写备注以说明此次删除的变更内容。单击确定并提交，即可完成行为域的删除。

删除行为域 ×

 确定删除吗？

* 备注

0/128

特殊说明：当前项目，提交成功后，即更新至生产环境，请谨慎操作！



说明：

- 行为元素删除成功即从生产环境清理相关数据，因此删除行为元素时需填写提交备注。
- 被行为规则依赖的行为元素不支持删除。

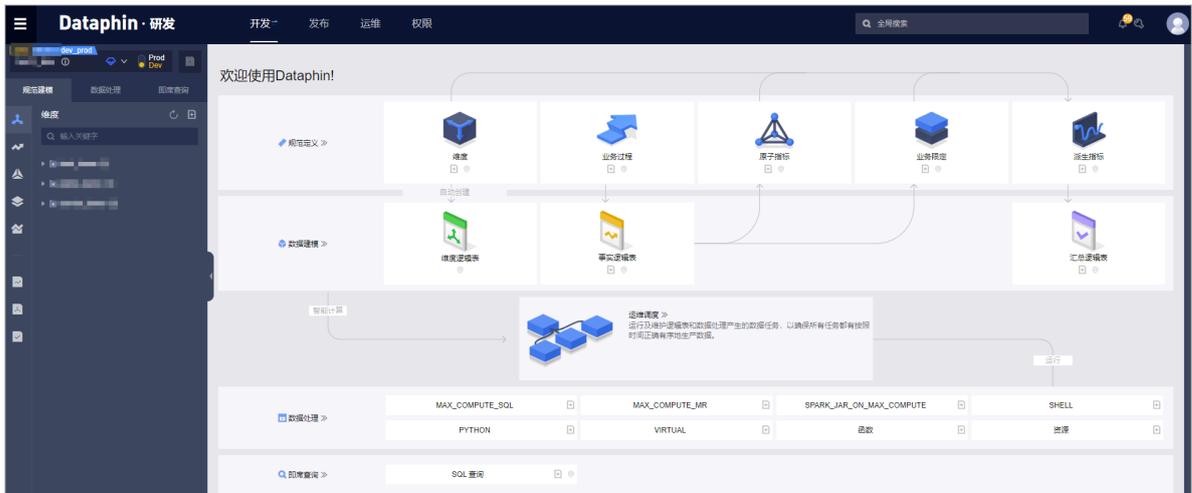
7.2.3 管理业务线

本文为您介绍如何进行新建、编辑、删除业务线，帮助您更好地管理业务线。

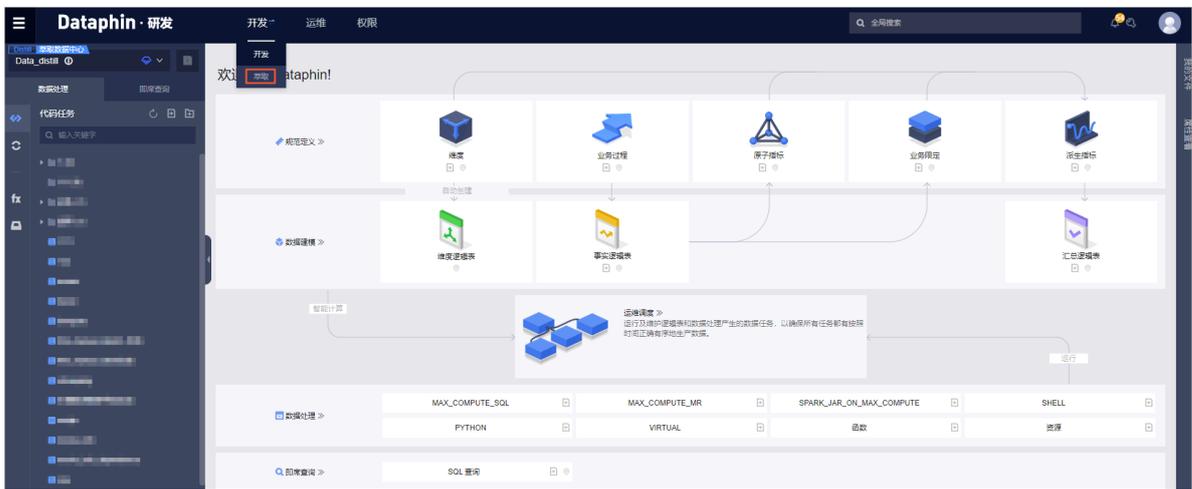
新建业务线

1. 登录Dataphin。

2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。



3. 将鼠标悬停在开发上，单击萃取，进入数据萃取研发页面。



4. 您可以通过以下任何一种方式，进入新建业务线页面：

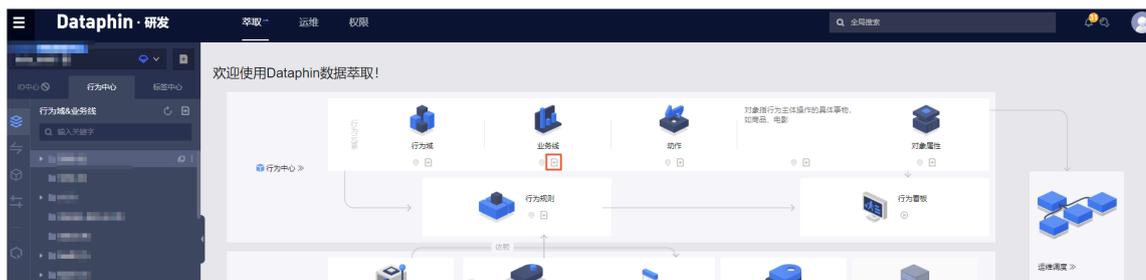
- 在数据萃取研发页面，单击左侧导航栏中的后，单击，然后单击业务线，进入新建业务线页面。



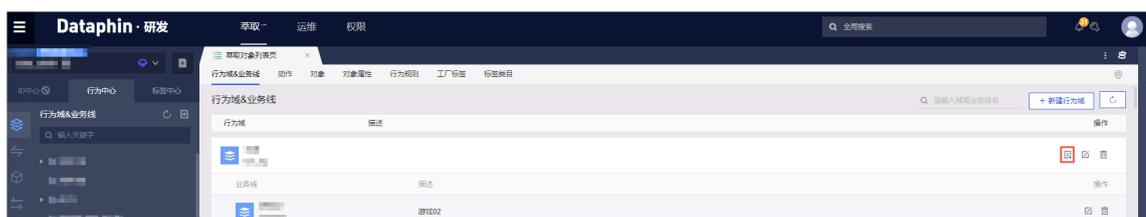
- 在数据萃取研发页面，单击项目名称后，选择行为中心 > 业务线，进入新建业务线页面。



- 在萃取首页，单击业务线下的，进入新建业务线页面。



- 在数据萃取研发页面，单击左侧导航栏下的行为域&业务线对象列表，进入萃取对象列表页面。在行为域&业务线页面，单击行为域所在行的操作列的 ，进入新建业务线页面。



- 根据页面提示选择行为域及配置业务线英文名、业务线名称、描述参数。配置完参数后单击提交。

新建业务线 ✕

*** 行为域**

*** 业务线英文名**

*** 业务线名称**

描述



说明:

- 业务线英文名、业务线名称不允许重复。
- 业务线依赖于行为域而存在。新建业务线时，在下拉列表中选择新建的业务线所在的行为域。

6. 在弹出的提交备注页面，填写备注以说明此次提交的变更内容。单击确定并提交，即可完成业务线的创建。



说明：

行为元素新建成功即发布到生产环境，因此需填写提交备注。

编辑业务线

1. 您可以通过以下任何一种方式，进入编辑业务线页面。

- 在数据萃取研发页面，单击左侧导航栏中的，然后单击需编辑的业务线。右侧弹出该业务线的属性框，单击编辑，进入编辑页面。



- 在数据萃取研发页面，单击左侧导航栏中的，将鼠标悬浮在需编辑业务线后的，单击编辑，进入编辑页面。



- 在数据萃取研发页面，单击左侧导航栏下方的行为域&业务线对象列表，进入萃取对象列表页面。在行为域&业务线页面，单击操作列下的，进入编辑页面。



2. 根据页面提示可修改业务线英文名、业务线名称、描述参数。完成修改后单击提交。

编辑业务线 ×

① 不建议您对元素本意进行大幅变动，大幅变动将对已生成行为数据含义产生影响。如：您可以把“看”改为“观赏”，但不建议把“看”改为“听”，如需“听”建议您另新建一个元素。

行为域

* 业务线英文名

* 业务线名称

描述



说明：

业务线英文名、业务线名称不允许重复。

3. 在弹出的提交备注页面，填写备注以说明此次提交的变更内容。单击确定并提交，即可完成业务线的编辑。



说明：

行为元素编辑成功即发布到生产环境。因此需填写提交备注。

删除业务线

1. 您可以通过以下任何一种方式，进入删除业务线页面。

- 在数据萃取研发页面，单击左侧导航栏中的 ，然后单击需编辑的业务线。右侧弹出该业务线的属性框，单击  后，单击删除，进入删除业务线页面。



- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需删除业务线后的 ，单击删除，进入删除业务线页面。



- 在数据萃取研发页面，单击左侧导航栏下方的行为域&业务线对象列表，进入萃取对象列表页面。在行为域&业务线页面，单击操作列下的 ，进入删除业务线页面。



- 2. 在弹出的删除业务线页面，填写备注以说明此次删除的变更内容。单击确定并提交，即可完成业务线的删除。

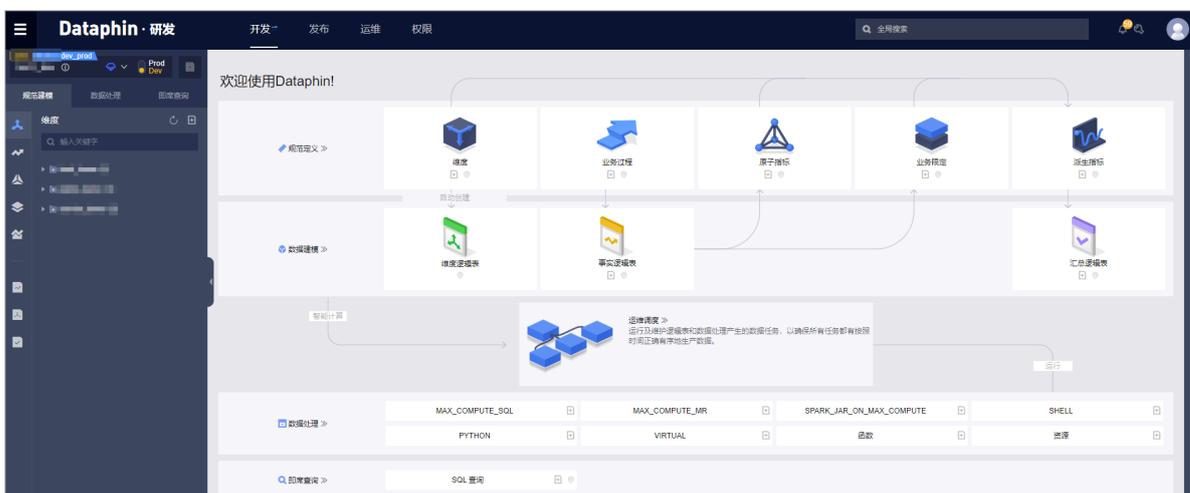
 **说明：**
被行为规则依赖的行为元素不支持删除。

7.2.4 管理动作

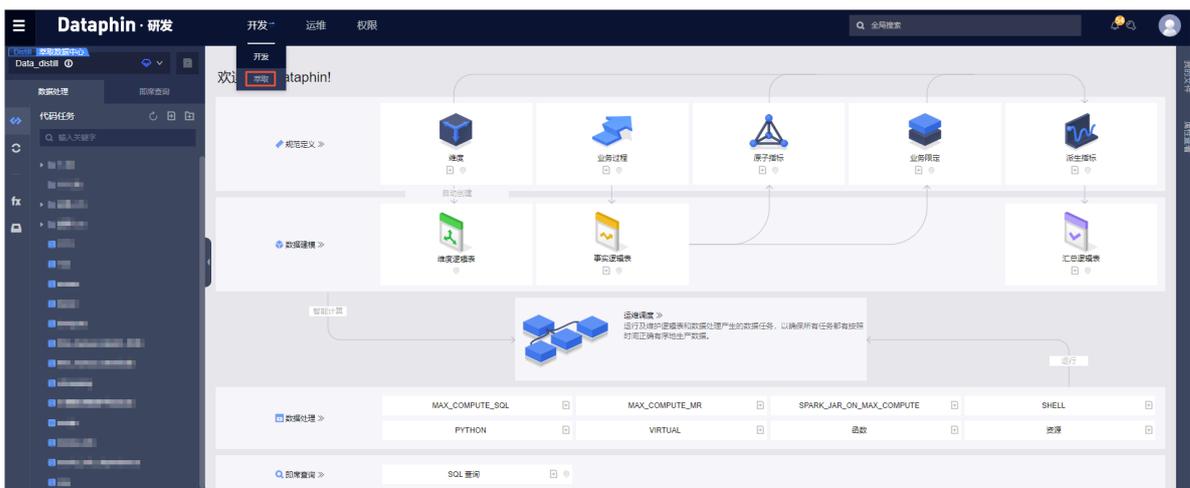
动作即行为主体发出的操作，如购买、浏览。本文为您介绍如何新建、编辑、删除动作。

新建动作

- 1. [登录Dataphin](#)。
- 2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。



- 3. 将鼠标悬停在开发上后，单击萃取，进入数据萃取研发页面。



4. 您可以通过以下任何一种方式，进入新建动作页面：

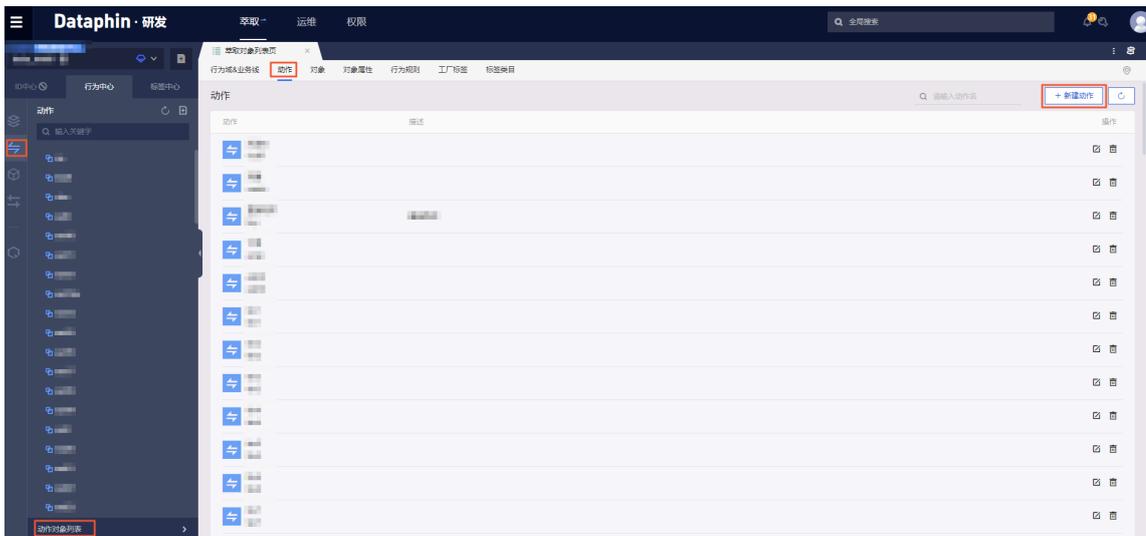
- 在数据萃取研发页面，单击左侧导航栏中的，然后单击，进入新建动作弹框。



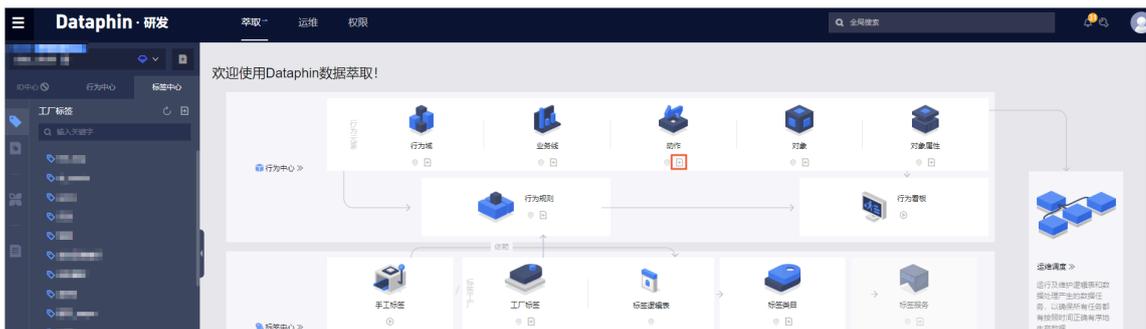
- 在数据萃取研发页面，单击项目名称后，选择行为中心 > 动作，进入新建动作弹框。



- 在数据萃取研发页面，单击左侧导航栏下的动作对象列表，进入萃取对象列表页页面。在动作页面，单击页面右上角的新建动作，进入新建动作弹框。



- 在萃取首页，单击动作下的 ，进入新建动作弹框。



5. 根据页面提示配置动作英文名、动作名称、描述参数。配置完参数后单击提交。

新建动作

* 动作英文名

* 动作名称

描述

 说明:

动作英文名、动作名称不支持与已有的动作英文名、动作名称重复。

6. 在弹出的提交备注页面，填写备注以说明此次提交的变更内容。单击确认并提交，即可完成动作的新建。

*** 提交备注**

请输入此次提交的相关备注说明，如变更信息

0/128

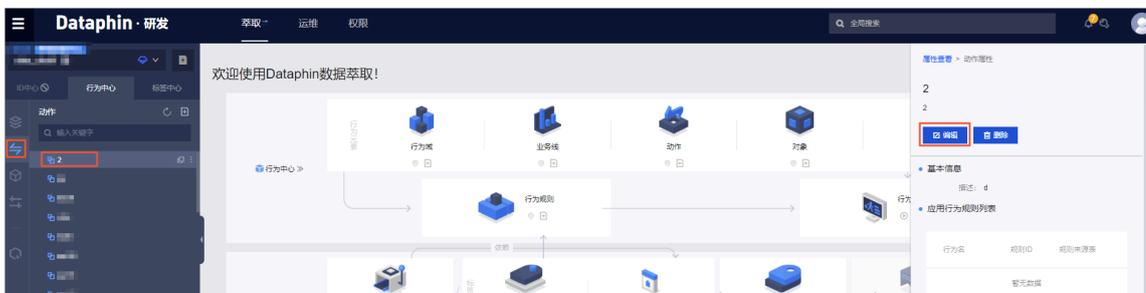
填写备注以说明此次提交的变更内容
当前项目，提交成功后，即更新至生产环境，请谨慎操作!

 **说明:**
行为元素新建成功即发布到生产环境，因此需填写提交备注。

编辑动作

1. 您可以通过以下任何一种方式，进入编辑动作页面：

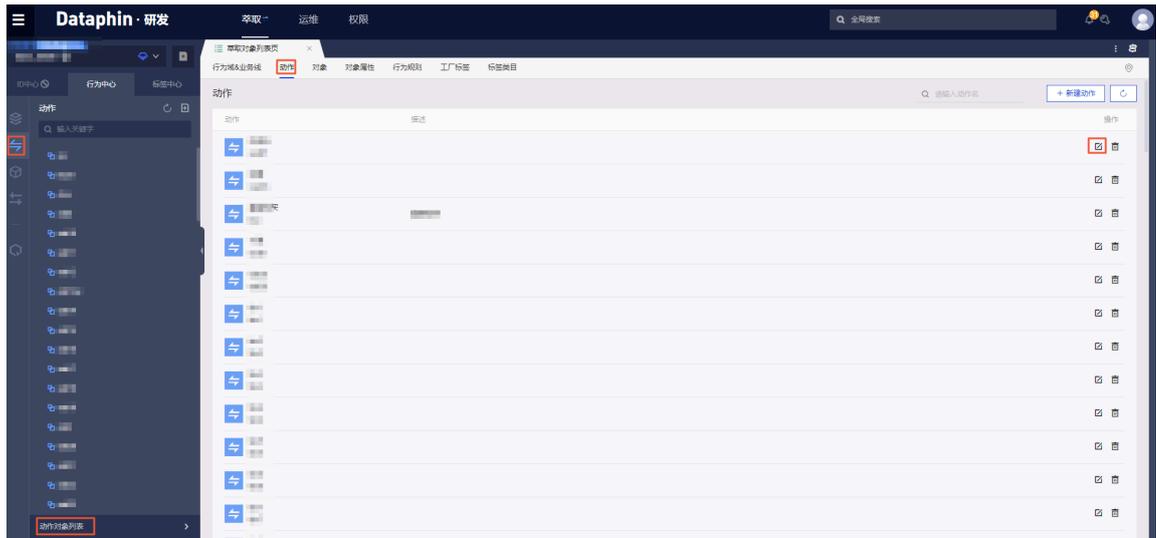
- 在数据萃取研发页面，单击左侧导航栏中的 ，然后单击需编辑的动作，右侧弹出该动作的属性框。在属性框中单击编辑，进入编辑动作弹框。



- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需编辑动作后的 ，单击编辑，进入编辑动作弹框。



- 在数据萃取研发页面，单击左侧导航栏下的动作对象列表，进入萃取对象列表页页面。在动作页面，单击操作列下的 ，进入编辑动作弹框。



2. 根据页面提示可修改动作英文名、动作名称、描述参数。完成修改后单击提交。

编辑动作

① 不建议您对元素本意进行大幅变动，大幅变动将对已生成行为数据含义产生影响。如：您可以把“看”改为“观赏”，但不建议把“看”改为“听”，如需“听”建议您另新建一个元素。

* 动作英文名

* 动作名称

描述



说明:

动作英文名、动作名称不支持与已有的动作英文名、动作名称重复。

3. 在弹出的提交备注页面，填写备注以说明此次提交的变更内容。单击确认并提交，即可完成动作的编辑。

*** 提交备注**

请输入此次提交的相关备注说明，如变更信息

0/128

填写备注以说明此次提交的变更内容
当前项目，提交成功后，即更新至生产环境，请谨慎操作!

[取消](#) [确定并提交](#)



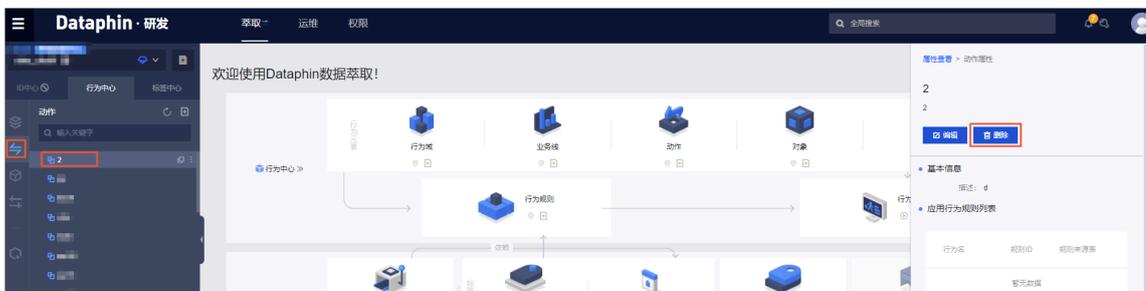
说明:

行为元素编辑成功即发布到生产环境。因此需填写提交备注。

删除动作

1. 您可以通过以下任何一种方式，进入删除动作页面：

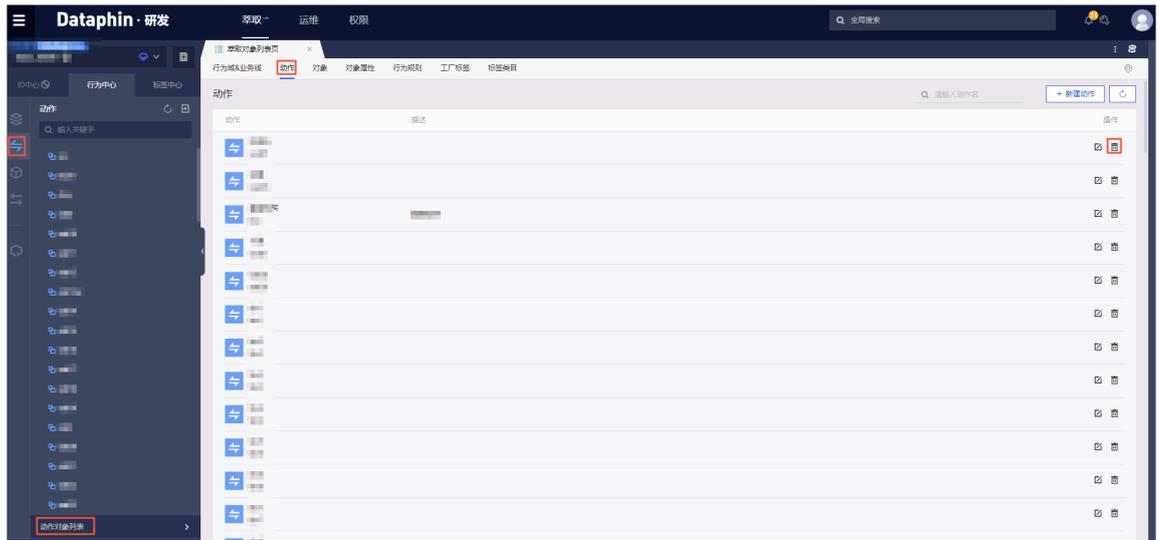
- 在数据萃取研发页面，单击左侧导航栏中的 ，然后单击需删除的动作，右侧弹出该动作的属性框。在属性框中，单击删除，进入删除动作弹框。



- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需删除动作后的 ，单击删除，进入删除动作弹框。



- 在数据萃取研发页面，单击左侧导航栏下的动作对象列表，进入萃取对象列表页页面。在动作页面，单击操作列下的 ，进入删除动作弹框。



2. 在弹出的删除动作页面，填写备注以说明此次删除的变更内容。单击确认并提交，即可完成动作的删除。

删除动作

确定删除2动作吗?

* 备注

0/128

特殊说明：当前项目，提交成功后，即更新至生产环境，请谨慎操作!



说明:

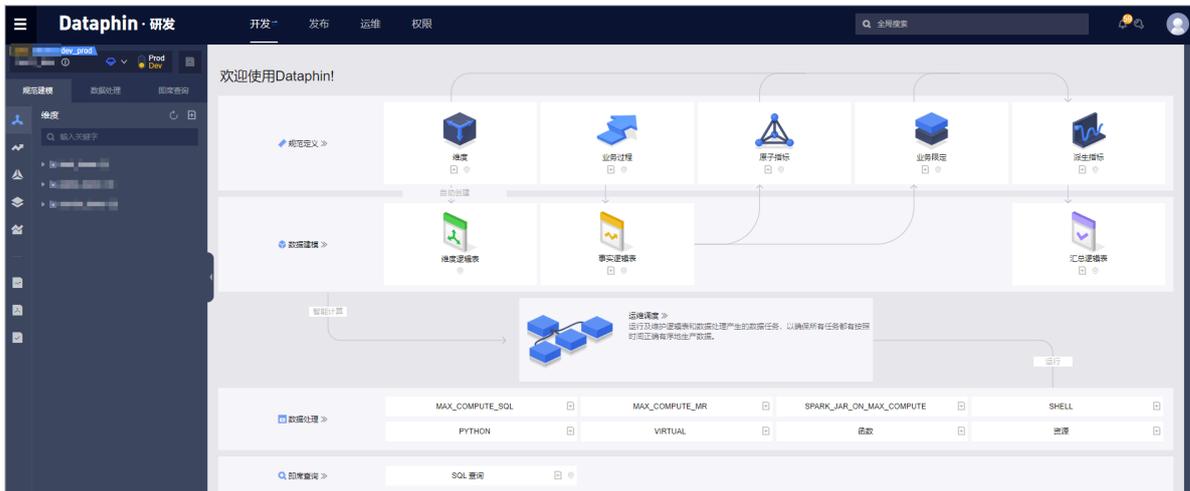
- 行为元素删除成功即从生产环境清理相关数据，因此删除行为元素时需填写提交备注。
- 被行为规则依赖的行为元素不可删除。

7.2.5 管理对象

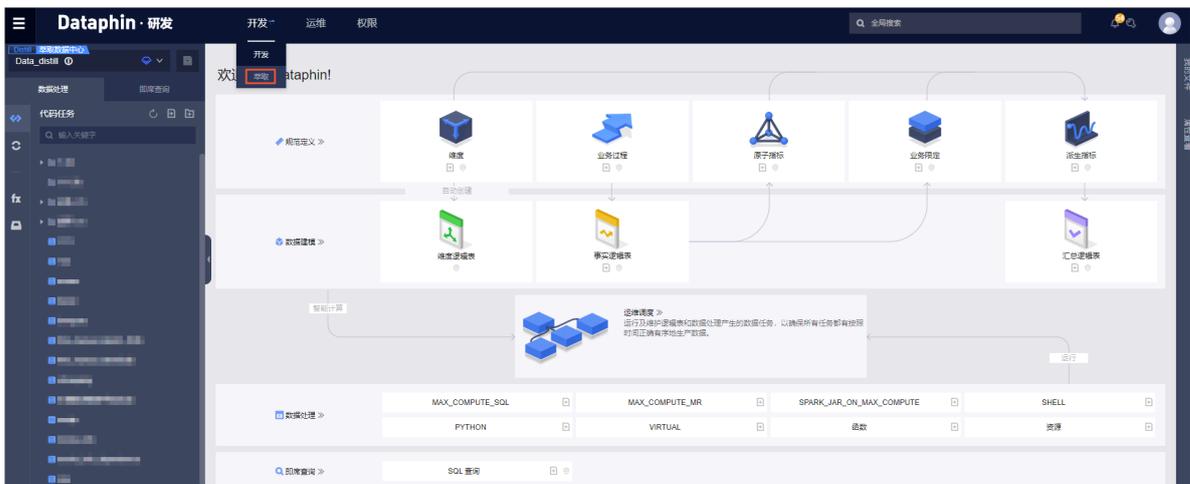
对象即行为主体操作的对象，如商品、电影。本文为您介绍如何新建、编辑、删除对象。

新建对象

1. [登录Dataphin](#)。
2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。



3. 将鼠标悬停在开发上后，单击萃取，进入数据萃取研发页面。



4. 您可以通过以下任何一种方式，进入新建对象页面：

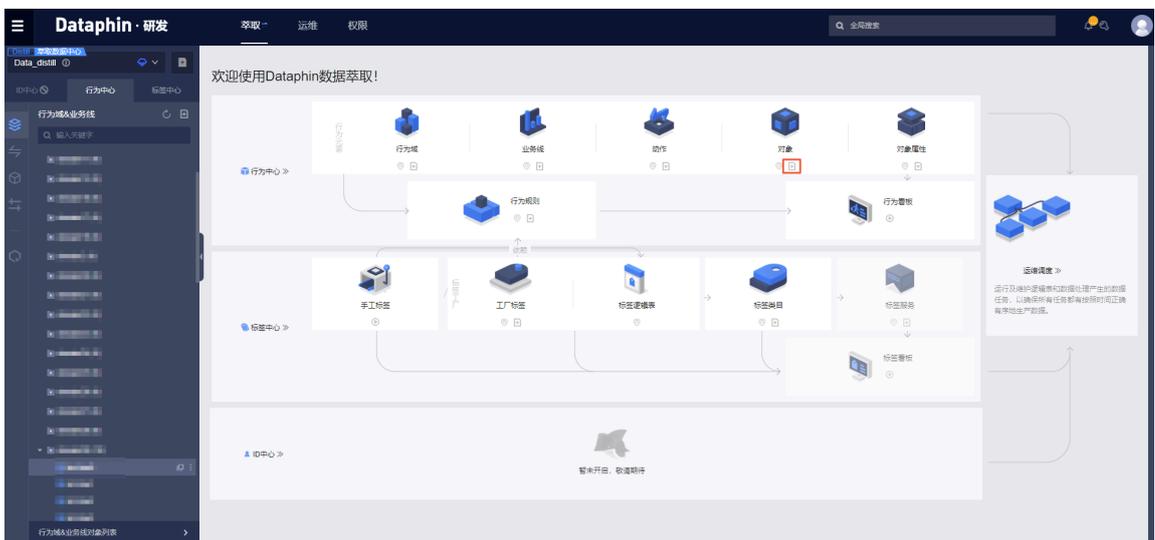
- 在数据萃取研发页面，单击左侧导航栏中的  后，然后单击对象后的 ，进入新建对象弹框。



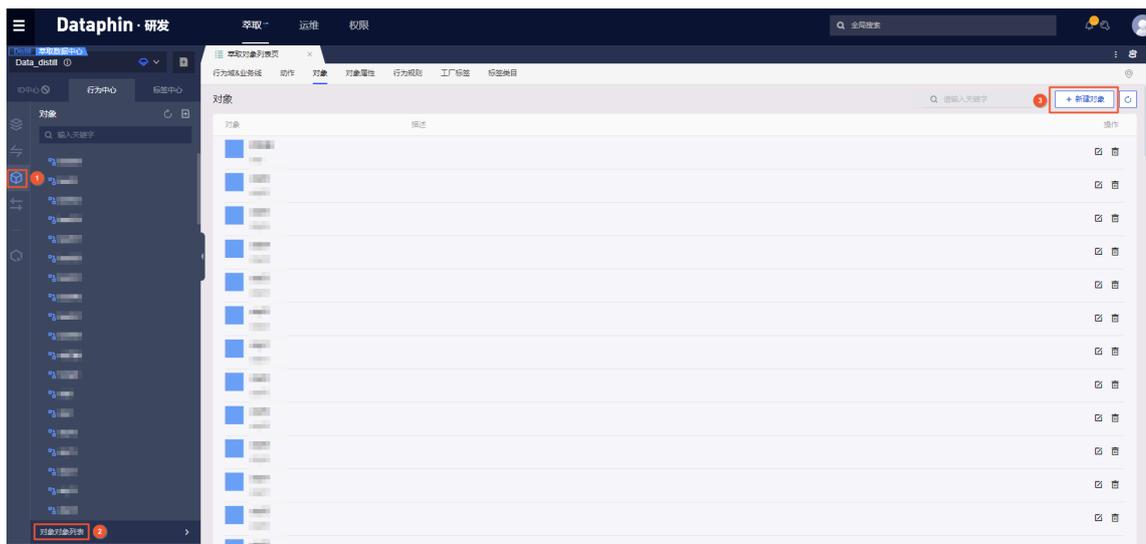
- 在数据萃取研发页面，单击项目名称后 ，选择行为中心 > 对象，进入新建对象弹框。



- 在萃取首页，单击对象下的，进入新建对象弹框。



- 在数据萃取研发页面，单击左侧导航栏下的对象列表，进入萃取对象列表页面。在对象页面，单击右上角的新建对象，进入新建对象弹框。



5. 根据页面提示配置对象英文名、对象名称、描述参数。配置完参数后单击提交。

新建对象

* 对象英文名 英文名

* 对象名称 名称

描述 描述

取消 提交



说明:

对象英文名、对象名称不支持与已有的对象英文名、对象名称重复。

6. 在弹出的提交备注页面，填写提交备注。单击确认并提交，即可完成对象的创建。

*** 提交备注**

请输入此次提交的相关备注说明，如变更信息

0/128

填写备注以说明此次提交的变更内容
当前项目，提交成功后，即更新至生产环境，请谨慎操作!



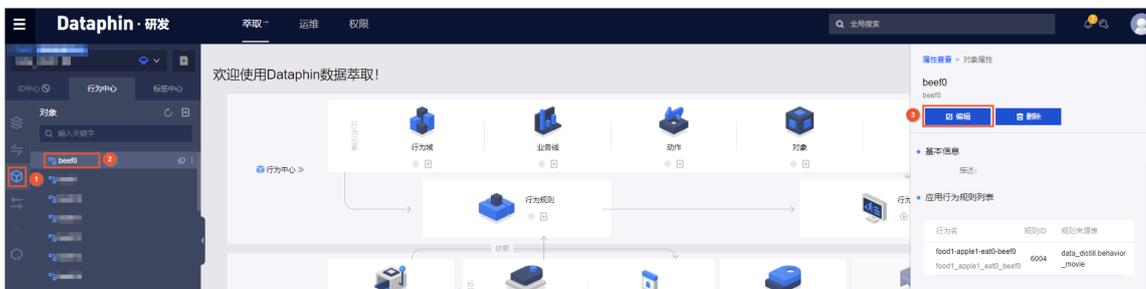
说明:

行为元素新建成功即发布到生产环境，因此需填写提交备注。

编辑对象

1. 您通过以下任何一种方，进入编辑对象页面：

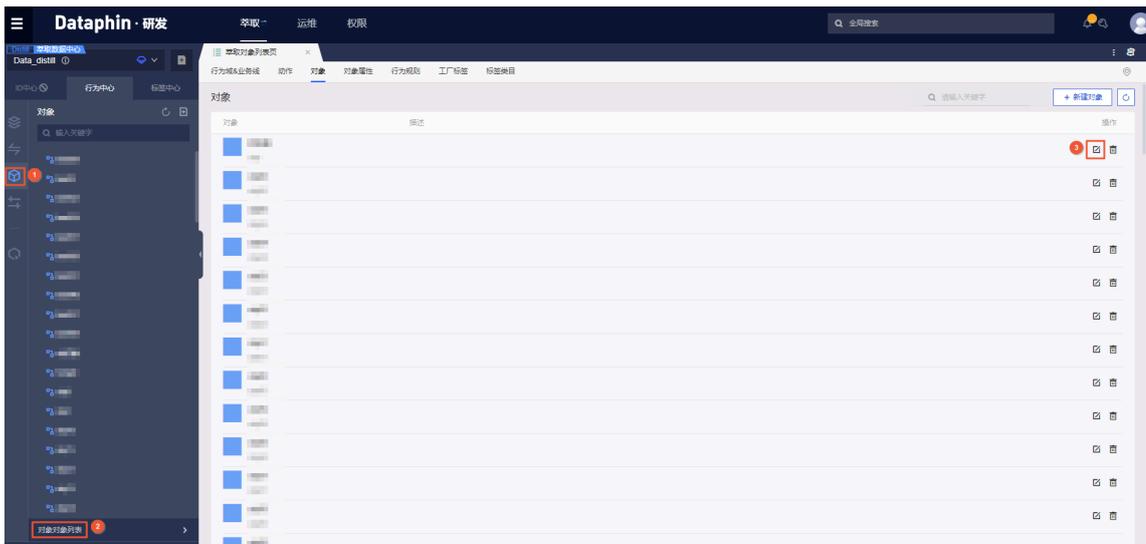
- 在数据萃取研发页面，单击左侧导航栏中的 ，然后单击需编辑的对象。右侧弹出该对象的属性框，单击编辑，进入编辑对象弹框。



- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需编辑对象后的 ，单击编辑，进入编辑对象弹框。



- 在数据萃取研发页面，单击左侧导航栏下方的对象列表，进入萃取对象列表页页面。在对象页面，单击操作列下的 ，进入编辑对象弹框。



2. 根据页面提示可修改对象英文名、对象名称、描述参数。完成修改后单击提交。

编辑对象

ⓘ 不建议您对元素本意进行大幅变动，大幅变动将对已生成行为数据含义产生影响。如：您可以把“看”改为“观赏”，但不建议把“看”改为“听”，如需“听”建议您另新建一个元素。

* 对象英文名

* 对象名称

描述

 **说明：**
对象英文名、对象名称不支持与已有的对象英文名、对象名称重复。

3. 在弹出的提交备注页面，填写备注以说明此次提交的变更内容。单击确认并提交，即可完成对象的编辑。

*** 提交备注**

请输入此次提交的相关备注说明，如变更信息

0/128

填写备注以说明此次提交的变更内容
当前项目，提交成功后，即更新至生产环境，请谨慎操作!

[取消](#) [确定并提交](#)



说明:

行为元素编辑成功即发布到生产环境。因此需填写提交备注。

删除对象

1. 您可以通过以下任何一种方式进入删除对象页面：

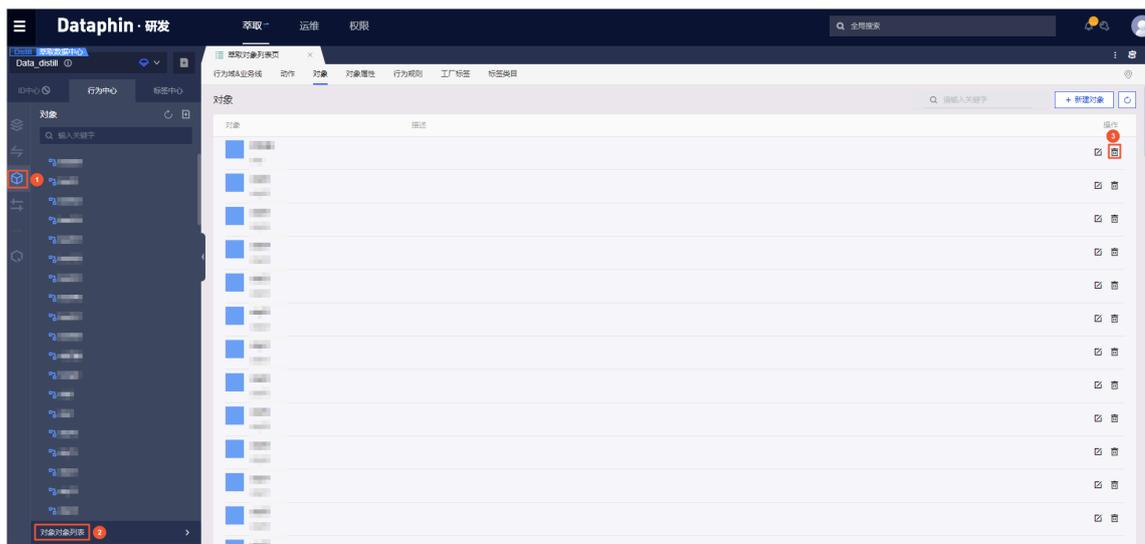
- 在数据萃取研发页面，单击左侧导航栏中的 ，然后单击需删除的对象。右侧弹出该的对象属性框，单击删除，进入删除对象弹框。



- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需删除对象后的 ，单击删除，进入删除对象弹框。



- 在数据萃取研发页面，单击左侧导航栏下方的对象列表，进入萃取对象列表页页面。在对象页面，单击操作列下的 ，进入删除对象弹框。



2. 在弹出的删除对象页面，填写备注以说明此次删除的变更内容。单击确认并提交，即可完成对象的删除。

删除对象

确定删除吗?

* 备注 请输入此次提交的相关备注说明, 如变更信息

0/128

特殊说明: 当前项目, 提交成功后, 即更新至生产环境, 请谨慎操作!

取消 确定并提交



说明:

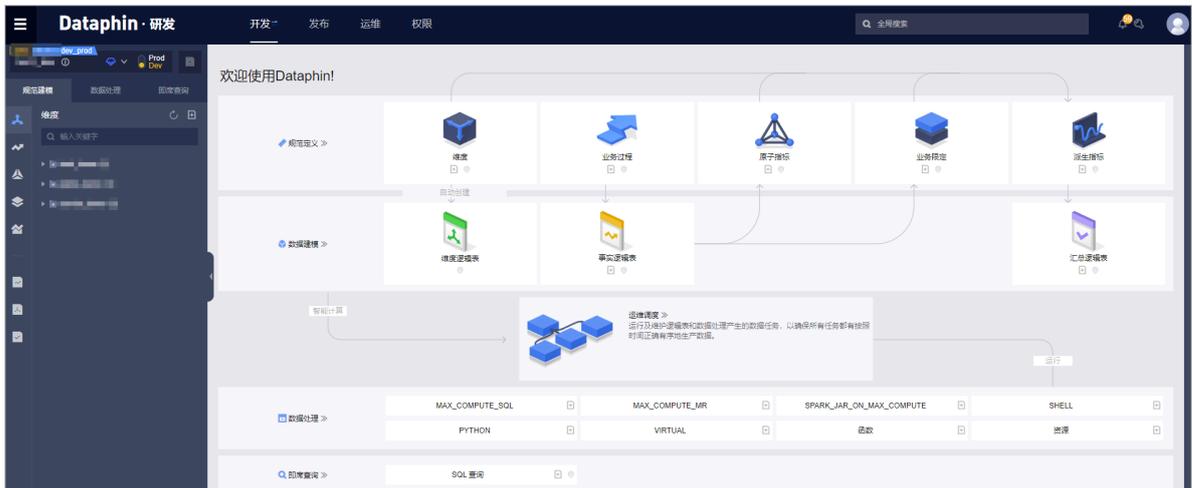
- 行为元素删除成功即从生产环境清理相关数据，因此删除行为元素时需填写提交备注。
- 被行为规则依赖的行为元素不可删除。

7.2.6 管理对象属性

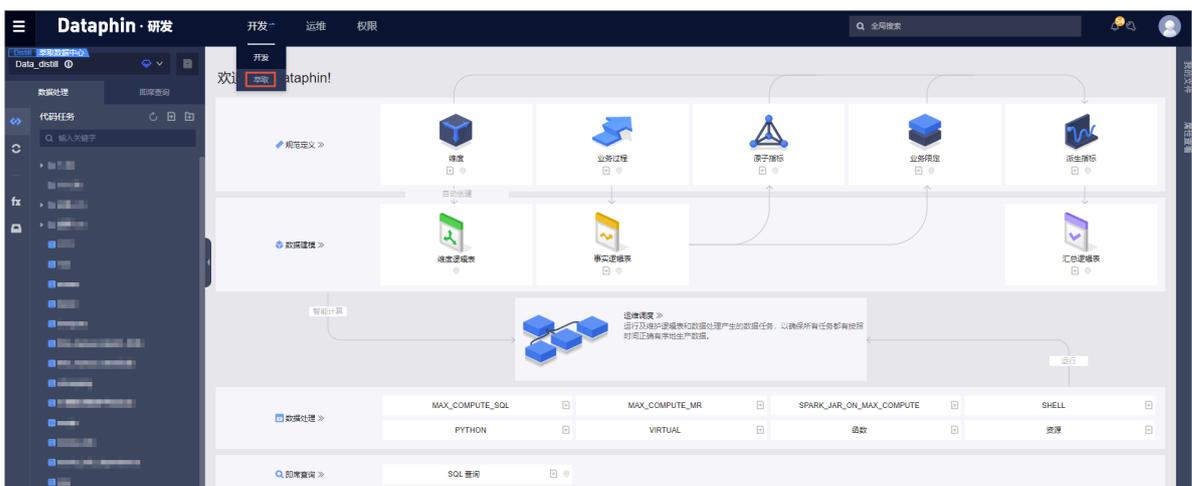
本文为您介绍如何进行新建、编辑、删除对象属性。

新建对象属性

1. 登录Dataphin。
2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。



3. 单击萃取，进入数据萃取研发页面。



4. 您可以通过以下任何一种方式，进入新建对象属性页面：

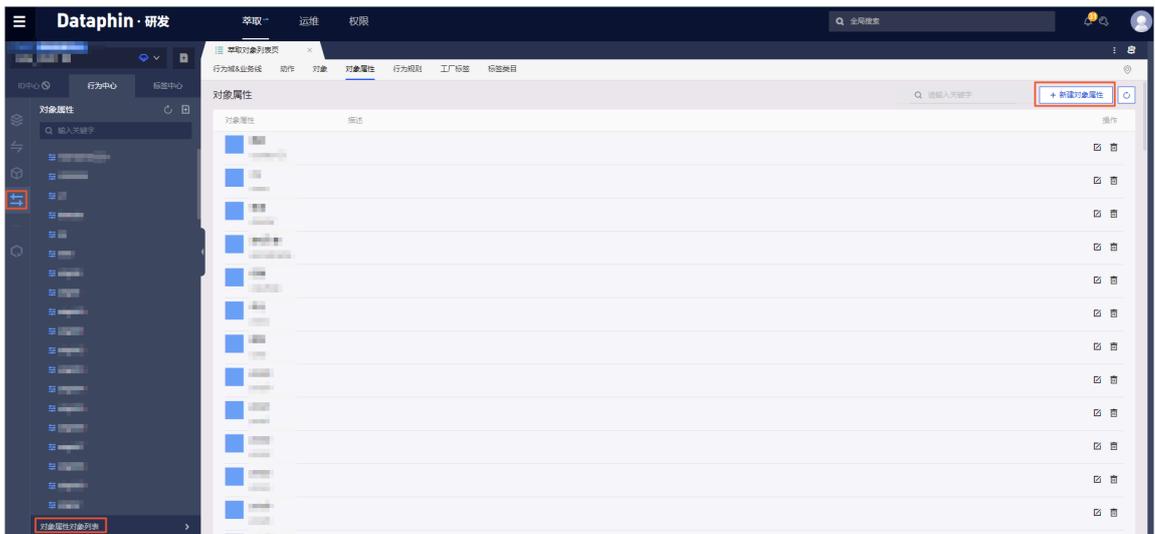
- 在数据萃取研发页面，单击左侧导航栏中的，然后单击，进入新建对象属性页面。



- 在数据萃取研发页面，单击项目名称后，选择行为中心 > 对象属性，进入新建对象属性页面。



- 在数据萃取研发页面，单击左侧导航栏下的对象属性对象列表，进入萃取对象列表页页面。在对象属性页面，单击页面右上角的新建对象属性，进入新建对象属性页面。



- 在萃取首页，单击对象属性下的，进入新建对象属性页面。



5. 根据页面提示配置对象属性英文名、对象属性名称、描述参数。配置完参数后单击提交。

新建对象属性 ×

| | |
|-----------|-----|
| * 对象属性英文名 | 英文名 |
| * 对象属性名称 | 名称 |
| 描述 | 描述 |

取消 提交



说明:

对象属性英文名、对象属性名称不允许重复。

6. 在弹出的提交备注页面，填写备注以说明此次删除的变更内容。单击确认并提交，即可完成对象属性的创建。

* 提交备注

请输入此次提交的相关备注说明，如变更信息

0/128

填写备注以说明此次提交的变更内容
当前项目，提交成功后，即更新至生产环境，请谨慎操作!

取消 确定并提交



说明:

行为元素新建成功即发布到生产环境，因此需填写提交备注。

编辑对象属性

1. 您可以通过以下任何一种方式，进入编辑对象属性页面：

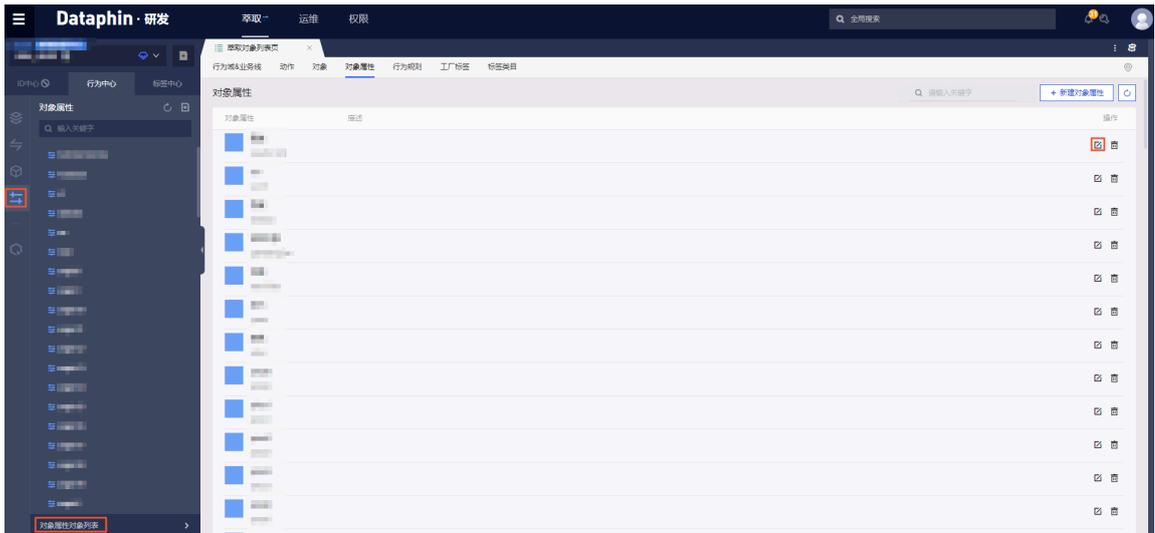
- 在数据萃取研发页面，单击左侧导航栏中的，然后单击需编辑的对象属性，右侧弹出该对象属性的属性框。在属性框页面，单击编辑，进入编辑页面。



- 在数据萃取研发页面，单击左侧导航栏中的，将鼠标悬浮在需编辑对象属性后的，单击编辑，进入编辑页面。



- 在数据萃取研发页面，单击左侧导航栏下的对象属性对象列表，进入萃取对象列表页页面。在对象属性页面，单击操作列下的，进入编辑页面。



2. 根据页面提示可修改对象属性英文名、对象属性名称、描述参数。完成修改后单击提交。

编辑对象属性

① 不建议您对元素本意进行大幅变动，大幅变动将对已生成行为数据含义产生影响。如：您可以把“看”改为“观赏”，但不建议把“看”改为“听”，如需“听”建议您另新建一个元素。

* 对象属性英文名

* 对象属性名称

描述



说明:

对象属性英文名、对象属性名称不允许重复。

3. 在弹出的提交备注页面，填写备注以说明此次提交的变更内容。单击确认并提交，即可完成对象属性的编辑。

*** 提交备注**

请输入此次提交的相关备注说明，如变更信息

0/128

填写备注以说明此次提交的变更内容
当前项目，提交成功后，即更新至生产环境，请谨慎操作!

[取消](#) [确定并提交](#)



说明:

行为元素编辑成功即发布到生产环境。因此需填写提交备注。

删除对象属性

1. 您可以通过以下任何一种方式，进入删除对象属性页面：

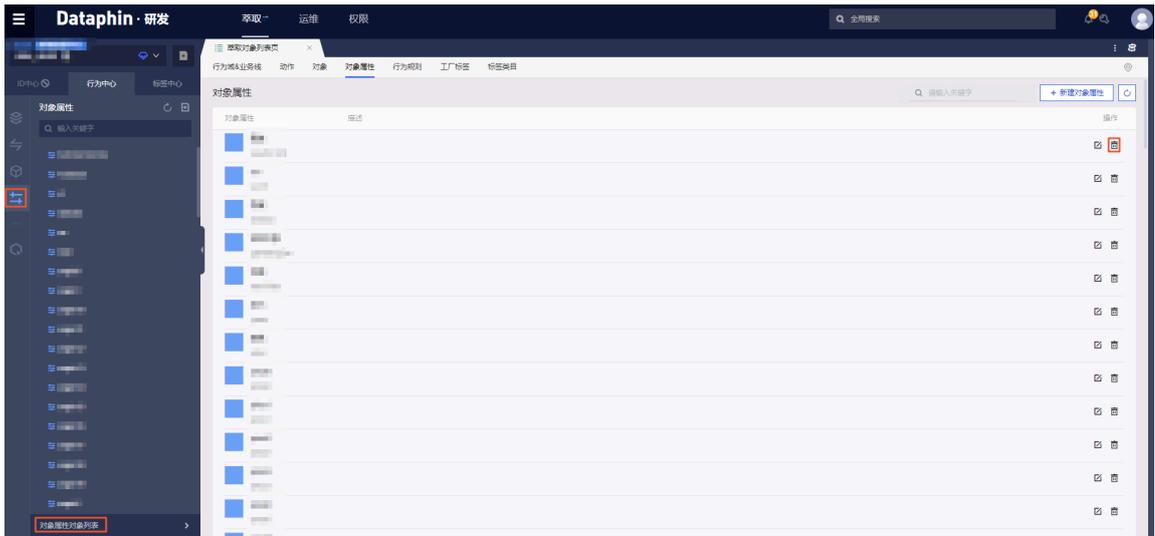
- 在数据萃取研发页面，单击左侧导航栏中的，然后单击需删除的对象属性。右侧弹出该对象属性的属性框，单击删除，进入删除对象属性页面。



- 在数据萃取研发页面，单击左侧导航栏中的，将鼠标悬浮在需删除对象属性后的，单击删除，进入删除对象属性页面。



- 在数据萃取研发页面，单击左侧导航栏下的对象属性对象列表，进入萃取对象列表页页面。在对象属性页面，单击操作列下的，进入删除对象属性页面。



2. 在弹出的删除对象页面，填写备注以说明此次提交的变更内容。单击确认并提交，即可完成对象属性的删除。

删除对象属性

确定删除123123123123对象属性么？

* 备注 请输入此次提交的相关备注说明，如变更信息

0/128

特殊说明：当前项目，提交成功后，即更新至生产环境，请谨慎操作！

取消 确定



说明：

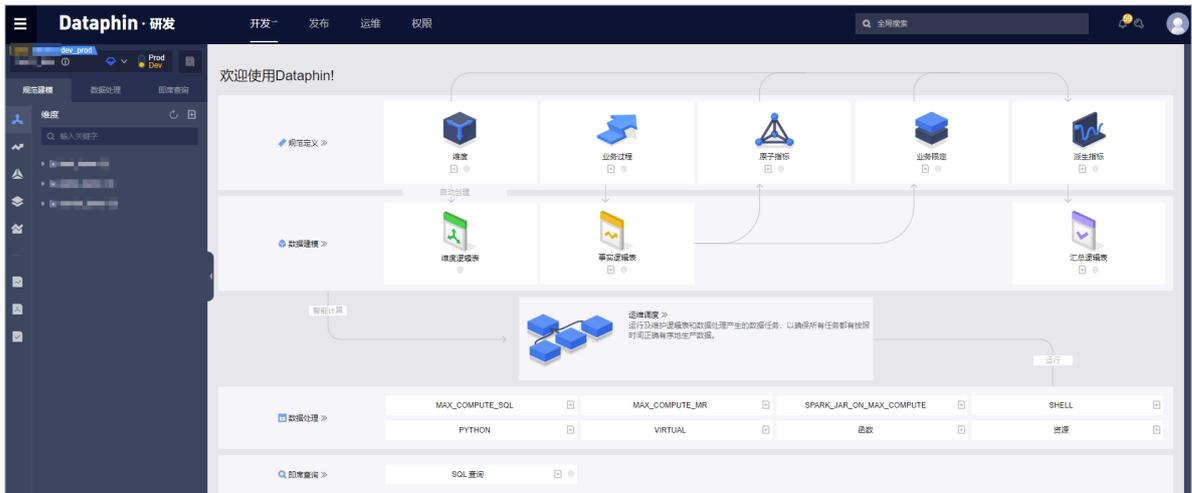
- 行为元素删除成功即从生产环境清理相关数据，因此删除行为元素时需填写提交备注。
- 被行为规则依赖的行为元素不可删除。

7.2.7 新建行为规则

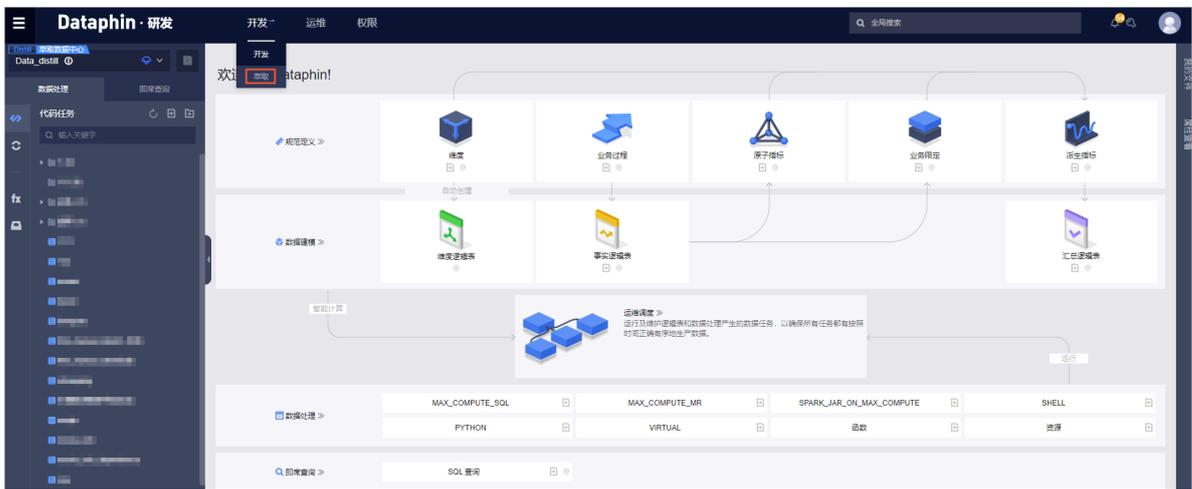
本文为您介绍如何进行行为规则的基本配置、调度配置和物理化配置。

1. 登录Dataphin。

2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。



3. 将鼠标悬停在开发上，单击萃取，进入数据萃取研发页面。



4. 您可以通过以下任何一种方式，进入新建行为规则弹框。

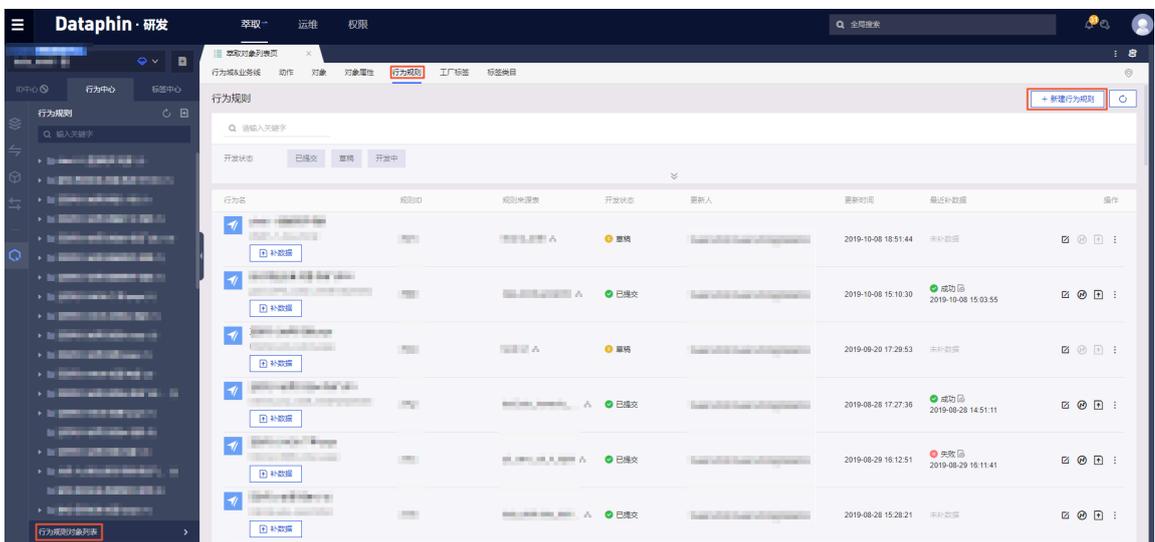
- 在数据萃取研发页面，单击左侧导航栏中的 ，然后单击 ，进入新建行为规则页面。



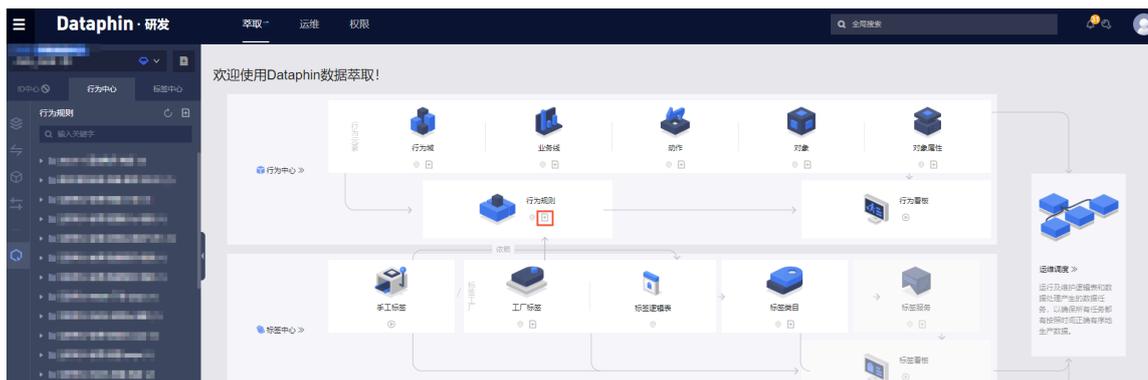
- 在数据萃取研发页面，单击项目名称后 ，选择行为中心 > 行为规则，进入新建行为规则页面。



- 在数据萃取研发页面，单击左侧导航栏下的行为规则对象列表，进入萃取对象列表页页面。在行为规则页面，单击页面右上角的新建行为规则，进入新建行为规则页面。



- 在萃取首页，单击行为规则下的 ，进入新建行为规则页面。



5. 在新建行为规则页面，根据页面提示配置参数。

基本信息

业务板块 萃取数据中心 空间类型 应用层 所属项目 数据萃取

* 行为域&业务线 请选择行为域 请选择业务线

* 动作 请选择动作 * 对象 对象

来源表设置

来源表类型 物理表 逻辑表

* 来源主表 请选择来源主表 ① 缺少数据源? 前往 [同步任务](#) 或 [代码任务](#) 加工

筛选条件 ①

格式化 规范性校验 参考示例

```
1 ds='${bizdate}'
```

取消
确定

| 分区 | 参数 | 说明 |
|-------|---------|--|
| 基本信息 | 行为域&业务线 | 根据页面提示选择行为域和业务线。 |
| | 动作 | 选择动作。 |
| | 对象 | 选择对象。 |
| 来源表设置 | 来源表类型 | 选择物理表或逻辑表。 |
| | 来源主表 | <ul style="list-style-type: none"> 如果您来源表类型选择物理表，则选择来源主表。 如果您来源表类型选择逻辑表，根据页面提示选择业务板块及来源主表。 |

| 分区 | 参数 | 说明 |
|----|------|--|
| | 筛选条件 | 填写筛选条件，即对应来源表的具体清洗规则，例如ds=\${bizdate}。 |

 **说明:**

- 若您选择了开发环境的来源主表，则无法在调度配置部分进行调度依赖关系的自动解析。
- 您需关注行为数据所属来源表及筛选条件，保证来源表符合行为数据要求，否则无法新建行为规则。

6. 配置完参数后单击确定，进入规则配置页面。



在规则配置页面，可查看已配置行为规则的基本信息。若需修改，您可以单击页面上方的行为规则信息，在页面右侧的行为规则信息弹框内进行修改。

7. 根据页面提示配置参数。

| 分区 | 参数 | 说明 |
|------|-------|---|
| 行为主体 | ID原始值 | <p>选择ID_Type和ID_value:</p> <ul style="list-style-type: none"> 您可以选择ID类型管理 > 新建ID类型新增自定义ID类型，系统内置的ID类型包括OneID、Mobile、UserID、Email和IPAddress。 您可以单击新增行为主体，新增多个行为主体ID。您也可以删除已添加的行为主体ID，但当仅剩一个行为主体时不支持删除。 <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p> 说明: 自定义ID_type支持删除，系统内置ID_type不支持删除和修改。</p> </div> |
| | 对象标识 | 选择 标识值。 |
| 对象 | 对象名称 | 选择对象名称。 |
| | | <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p> 说明: 此处标识值 (object_id) 和对象名称 (object_name) 仅作为对象标识，不作为对象属性处理。若您需要使用标识值或对象名称作为对象属性，请在对象属性中添加。</p> </div> |

| 分区 | 参数 | 说明 |
|------|------|--|
| 对象属性 | 对象属性 | <p>对象属性是对象的事实性描述信息，如视频对象的名 称、年份、导演等。选择属性及属性值：</p> <ul style="list-style-type: none"> 您可以通过单击对象属性管理，进入对象属性页面。 系统支持对已有的对象属性进行管理，同时支持新建 对象属性，详细内容请参见管理对象属性。 您也可以单击新增对象属性，新增多个对象属性，同 时也可以删除已添加的对象属性。但当仅存在最后 一个对象属性时，不支持删除。 <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p> 说明： 请保持属性值来源字段取值稳定性，防止属性值波动 对后续标签计算结果的影响。一般情况下同一对象的 同一属性取值稳定且唯一，且可从下拉列表选择已在 行为元素部分定义的对象属性。</p> </div> |
| 行为属性 | 发生次数 | <p>非必填信息。若您选择的来源表中有表示行为发生次数 的字段时，您可以单击  后，选择属性值。</p> |

| 分区 | 参数 | 说明 |
|----|----|---------------------------------------|
| 系统 | ds | 系统自动为行为中心产出表创建系统业务日期分区字段，格式为yyyymmdd。 |

- 若您已有标准化的ID倒排表，可以通过单击ID归一化配置选择ID倒排表。ID归一化开启后对所有规则全局生效。

您可以通过单击页面上ID倒排表，查看ID倒排表展示示意图。如果您已选择了ID类型，通过单击ID类型，查看倒排表详情。

| 字段 | 描述 | 参考取值 |
|----------|--|--|
| oneid | 唯一识别对象的id值。id类型为Oneid，可根据oneid找到对应任意key_type值的key_id | 生成规则#关键ID类型#随机生成不重复24位编码， [blurred] |
| key_type | 系统中存在的ID类型，根据业务系统中存在的可识别对象的数据，设定的枚举值 | OneID（系统默认的归一后ID的ID类型），Mobile，IDCard， 等大小写不敏感 |
| key_id | 该key_type对应的具体id值 | 具体的ID取值，如OneID类型取值为#关键ID类型#随机生成 不重复24位编码，Mobile类型的ID取值为+86具体手机号 |



说明:

- 仅超级管理员可以开启和关闭ID归一配置。
- ID倒排表基于OneID方法论，将同一用户的不同ID通过算法识别，用OneID连接起来。

- ID归一化配置开启后，若已选行为主体ID_type映射成功，结果表会按照OneID和原始ID_type分别生成记录。若映射失败，仅按照原始ID_type生成记录。
- 若您没有ID倒排表，您也可以在Data_distill项目下，通过单击开发 > 数据处理，引入ID倒排表。但需保证ID倒排表任务节点输出名称为Data_distill.倒排表名称（即项目名.表名）。

8. 单击页面上方的调度配置，右侧弹出调度配置的页面，根据界面提示配置参数。配置完成后单击确定。

调度配置
✕

基本信息

节点名称 oi_behavior_detail_d_8

节点ID

节点类型 Maxcompute_SQL

负责人

描述 行为规则8的调度配置

优先级 中等优先级 ▾

参数配置 对筛选条件部分填写的系统无法自动解析的参数进行赋值，以保证调度任务正常运行 [节点参数配置说明](#)

调度配置

时间属性 正常调度 暂停调度

调度周期 日 ▾ 00:00 🕒

依赖关系 📎 依赖解析

上游依赖 + 新增上游依赖

| 父节点输出名称 | 节点名 | 节点ID(实例ID) | 负责人 | 操作 |
|---------|-----|------------|-----|----|
| 暂无数据 | | | | |

当前节点

| 输出名称 | 节点名 | 节点ID (实例ID) | 负责人 | 操作 |
|-----------------------|-----------------------|---|---|----|
| oi_behavior_detail... | oi_behavior_detail... | | | 🔄 |

取消
确定

| 分区 | 参数 | 说明 |
|------|-----|---|
| 基本信息 | 优先级 | 系统将自动生成节点名称、节点ID、节点类型、负责人和描述，不支持修改。您可以选择优先级，优先级包括： <ul style="list-style-type: none"> • 最低优先级。 • 低优先级。 • 中等优先级。 • 高优先级。 • 最高优先级。 |

| 分区 | 参数 | 说明 |
|------|------|---|
| | 参数配置 | 单击节点参数配置说明，参见节点参数配置说明对来源表筛选条件部分的变量参数进行配置。 |
| 调度配置 | 时间属性 | 选择时间属性，时间属性包括正常调度和暂停调度。 |
| | 调度周期 | <p>调度周期可选择日、月、周：</p> <ul style="list-style-type: none"> · 如果您选择日，则单击  选择时间。 · 如果您选择月，则需先选择具体得日期再单击  选择时间。 · 如果您选择周，则需先选择具体的日期再单击  选择时间。 |
| 依赖关系 | 上游依赖 | <ul style="list-style-type: none"> · 您可以单击依赖解析，系统根据选择的来源表自动进行依赖解析。 · 您可通过单击新建上游依赖，新增生产环境的节点（来自Basic项目、Prod项目）为上游依赖。 <div style="background-color: #f0f0f0; padding: 5px; margin-top: 10px;"> <p> 说明：</p> <ul style="list-style-type: none"> · 若来源表为逻辑表，则系统解析的上游依赖节点不支持删除。 · 若来源表为物理表，则系统解析的上游依赖节点支持删除。 · 若您切换来源表，则系统已解析的上游依赖节点将被清空，您可能需要重新手动解析或修改上游依赖节点。 · 若系统自动解析失败，可单击依赖解析按钮重新进行解析，重新解析不会删除您新增的上游依赖节点。 </div> |
| | 当前节点 | 您可以单击  ，查看下游节点。 |

9. 单击页面上方的物理化配置，右侧弹框弹出物理化配置页面，根据界面提示配置参数。

物理化配置 ×

- 物理化配置
 - * 存储生命周期
快速选择 180 | 365 | 3600 | 36000 天
 - * 选择分区字段

备注：存储生命周期对全局行为规则生效，默认36000天，无最大限制，仅超级管理员可修改；
建议您设定的存储生命周期不小于下游标签选择的的来数据源数据时间跨度，否则标签数据可能出错

10 参数配置完成后，通过单击页面上方的 ，进行预览。通过单击 ，进行参数的保存。通过单击 ，进行提交新建的行为规则。

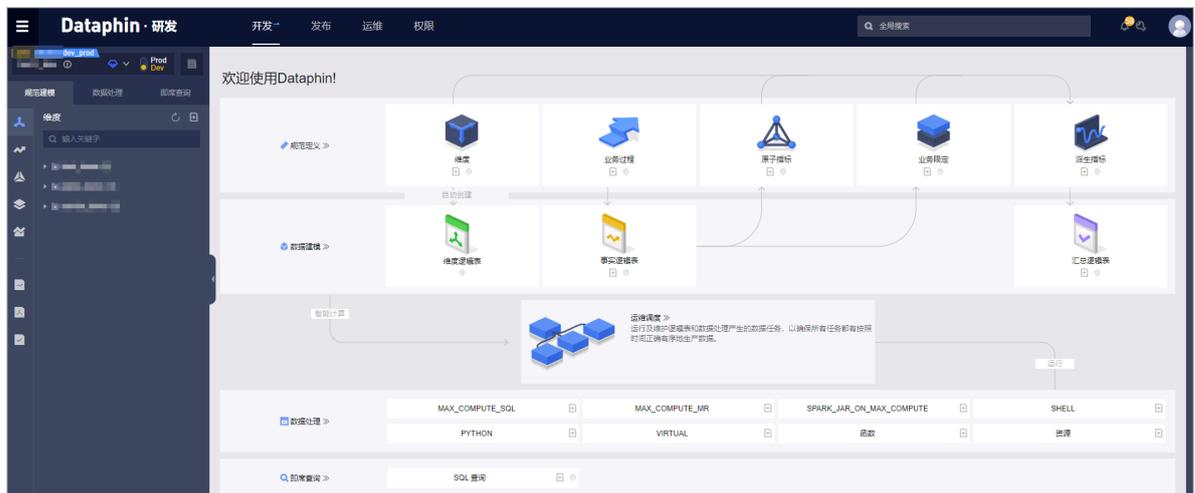


7.2.8 管理行为规则

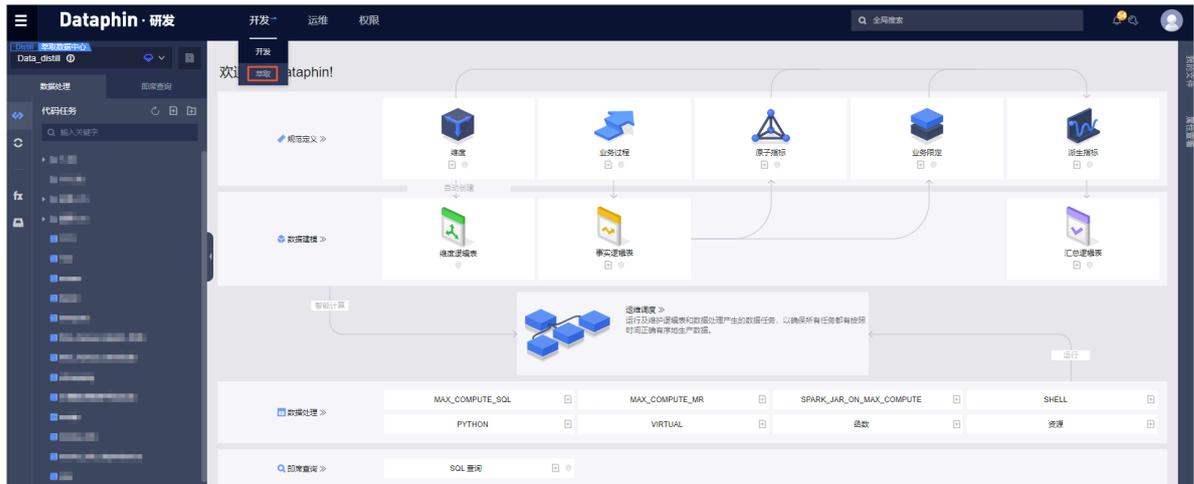
本文为您介绍如何进行行为规则的查看、编辑、补数据、下线等操作。

编辑行为规则

1. 登录Dataphin。
2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。



3. 单击萃取，进入数据萃取研发页面。



4. 您可以通过以下3种方式中的任何一种方式，进入规则配置页面，进行编辑行为规则。

- 在数据萃取研发页面，单击左侧导航栏中的，然后单击需编辑的行为规则，进入行为规则属性框。单击编辑，进入规则配置页面。

属性查看 > 行为规则属性

test2.a1 开发中

行为规则ID: 7940

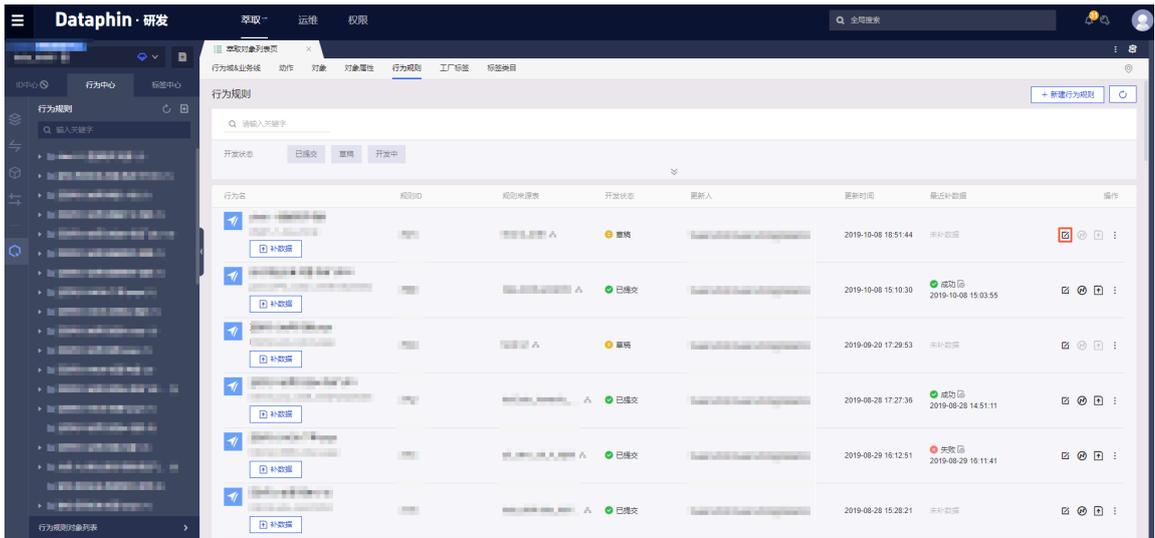
行为: 文娱-大文娱之游戏-观看-极品飞车02

编辑 补数据 ⋮

- 基本信息
 - 行为英文名: fun_dg_game_watch_needForSpeed02
 - 来源表筛选条件: test2.a1 🔍
 - 更新人: SuperAdmin(SuperAdmin@dataphin)
 - 更新时间: 2019-09-19 16:32:27
 - 最近补数据: 无补数据
 - 补数据记录:
- 消费信息
 - 标签依赖:
 - 资产详情: oi_behavior_detail_d
- 行为主体信息

| 行为主体 | ID_Type | ID_Value |
|-----------|-----------|----------|
| ID归一化配置:Y | IPAddress | ds |
| | OneID | ds |
| | UserID | ds |
| | Mobile | ds |
- 对象信息

- 在数据萃取研发页面，单击左侧导航栏下的行为规则对象列表，进入萃取对象列表页页面。在行为规则页面，单击行为规则所在行的操作列的 ，进入规则配置页面。



- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需编辑行为规则后的 ，单击编辑，进入规则配置页面。



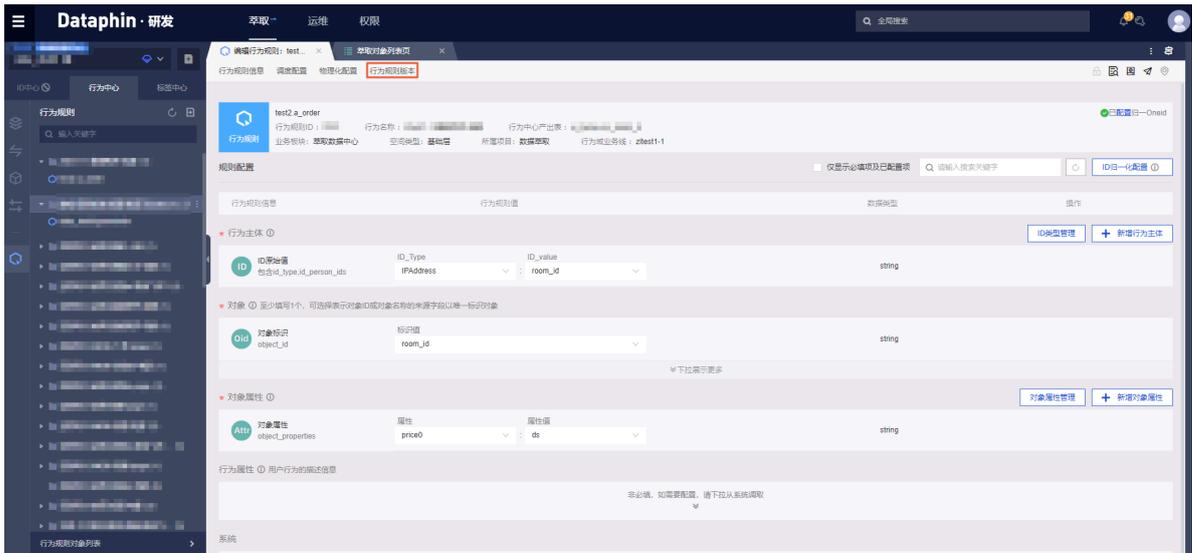
5. 在规则配置页面，您可通过单击页面上方的行为规则信息，编辑行为规则基本信息和来源表设置。您也可以在该页面编辑规则配置信息。详细内容请参见[新建行为规则](#)。



6. 参数配置完成后，您可通过单击页面上方的进行预览。通过单击进行参数的保存。通过单击进行提交编辑后的行为规则。



7. 在规则配置页面，单击页面上方的行为规则版本，可查看该行为规则上一次提交的版本信息。



查看行为规则

1. 您可以通过以下2种方式查看行为规则不同维度的信息：

- 在数据萃取研发页面，您可以单击左侧导航栏中的，然后单击需查看的行为规则，进入行为规则属性框。您可查看该行为规则的基本信息、消费信息、行为主体信息、对象信息、对象属性信息、行为属性信息。

[属性查看](#) > 行为规则属性

test2.a1 开发中

行为规则ID: 7940
行为: 文娱-大文娱之游戏-观看-极品飞车02

编辑 补数据 

- 基本信息**

行为英文名: fun_dg_game_watch_needForSpeed02

来源表筛选条件: test2.a1 

更新人: SuperAdmin(SuperAdmin@dataphin)

更新时间: 2019-09-19 16:32:27

最近补数据: 无补数据

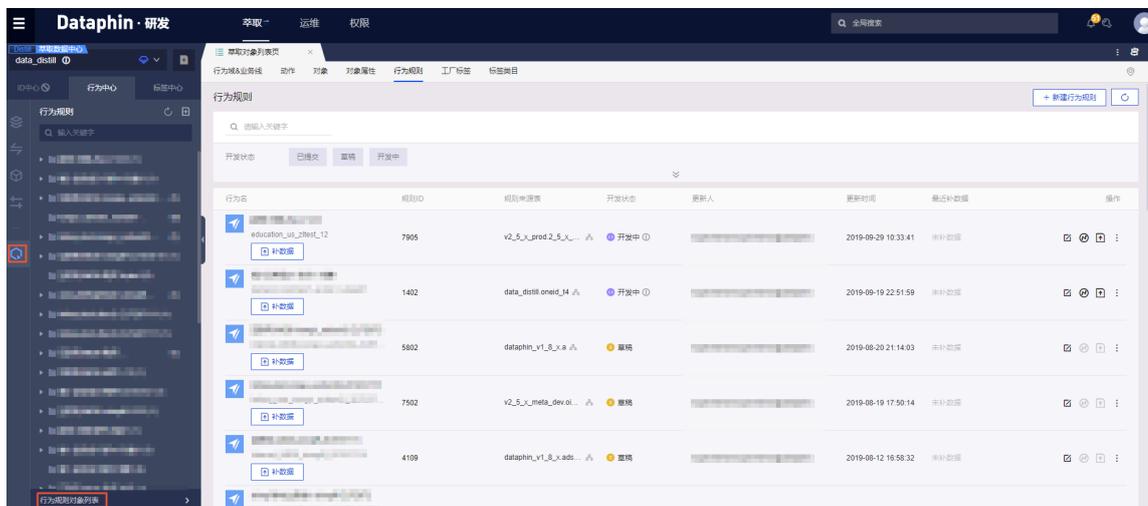
补数据记录:
- 消费信息**

标签依赖:

资产详情: [oi_behavior_detail_d](#)
- 行为主体信息**

| 行为主体 | ID_Type | ID_Value |
|-----------|-----------|----------|
| ID归一化配置:Y | IPAddress | ds |
| | OneID | ds |
| | UserID | ds |
| | Mobile | ds |
- 对象信息**

- 在行为规则列表页面，单击页面左下方的 行为规则对象列表，进入萃取对象列表页面。在萃取对象列表页面，您可以通过筛选批量查看规则信息，也可以通过单击 规则ID下的具体ID，在右侧属性弹框中查看规则信息。



- 行为名：行为域名称。
- 规则ID：与行为名对应的行为规则ID。
- 规则来源表：行为规则来源表。
- 开发状态：行为规则的开发状态，包括草稿、已提交、开发中3种状态。开发中表示已提交行为规则重新编辑后仅保存未提交，仍根据修改前的配置在生产环境参与任务调度。已提交表示行为规则已在生产环境中参与任务调度。草稿表示新建行为规则未提交的状态。
- 最近补数据：包含未补数据、补数据中、成功和失败4种。鼠标悬浮在最近补数据状态后面可查看该行为规则的所有补数据记录。在补数据记录页面，单击结果列下的，进入萃取运维页面，查看该次补数据生成的实例及DAG图。

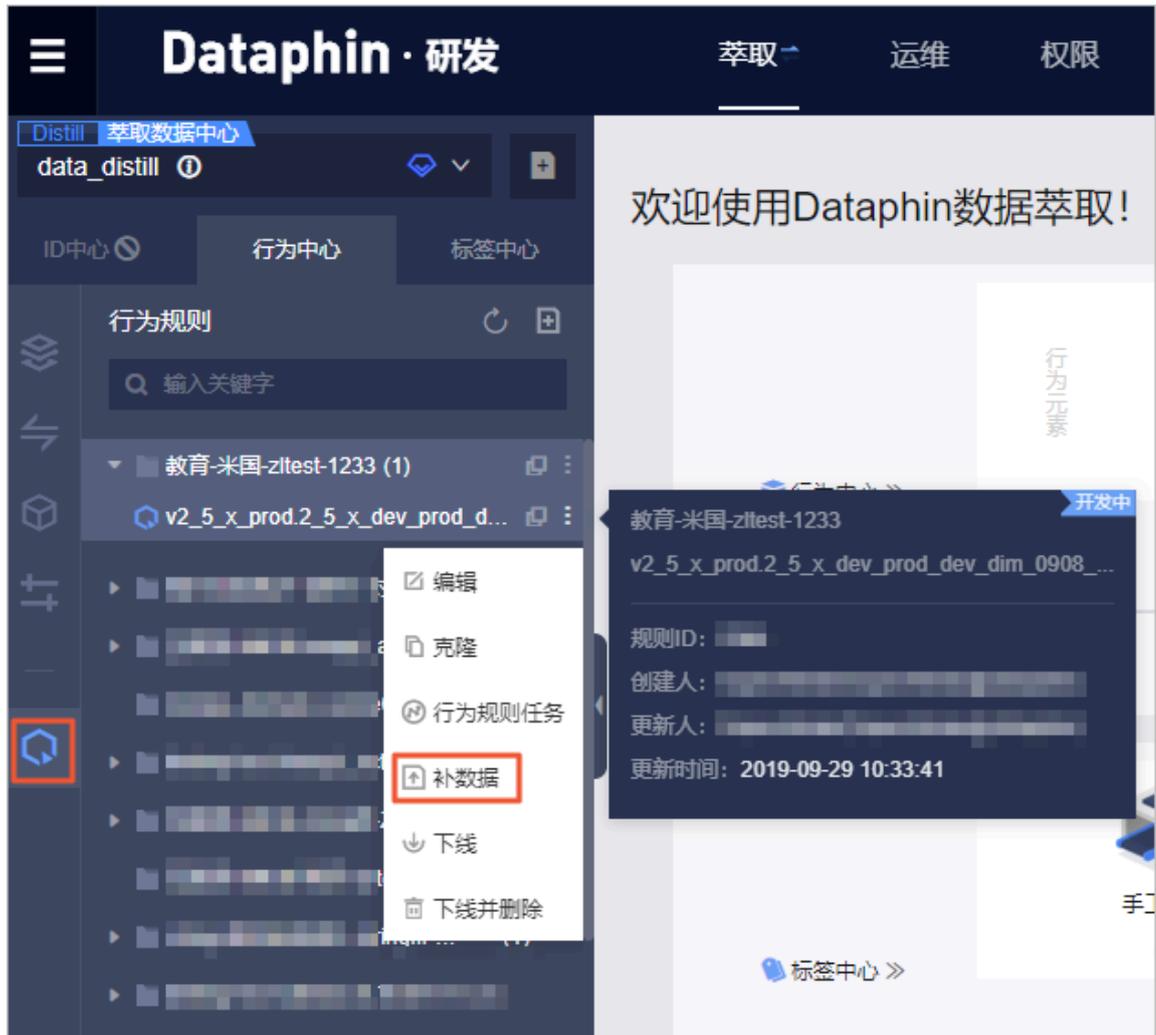
补数据

1. 您可以通过以下4种方式中的任何一种方式，进入行为规则补数据页面：

- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需补数据的行为规则所在的行为名后 ，单击 补数据，进入行为规则补数据页面。



- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需补数据的行为规则后 ，单击 补数据，进入行为规则补数据页面。



- 在数据萃取研发页面，单击左侧导航栏中您需补数据的行为规则，右侧弹出该行为规则的属性框。在属性框中单击 补数据，即可进入行为规则补数据页面。

属性查看 > 行为规则属性

v2_5_x_prod.2_5_x_dev_prod_dev_dim_0... 开发中

行为规则ID: 7905
行为: 教育-米国-zltest-1233

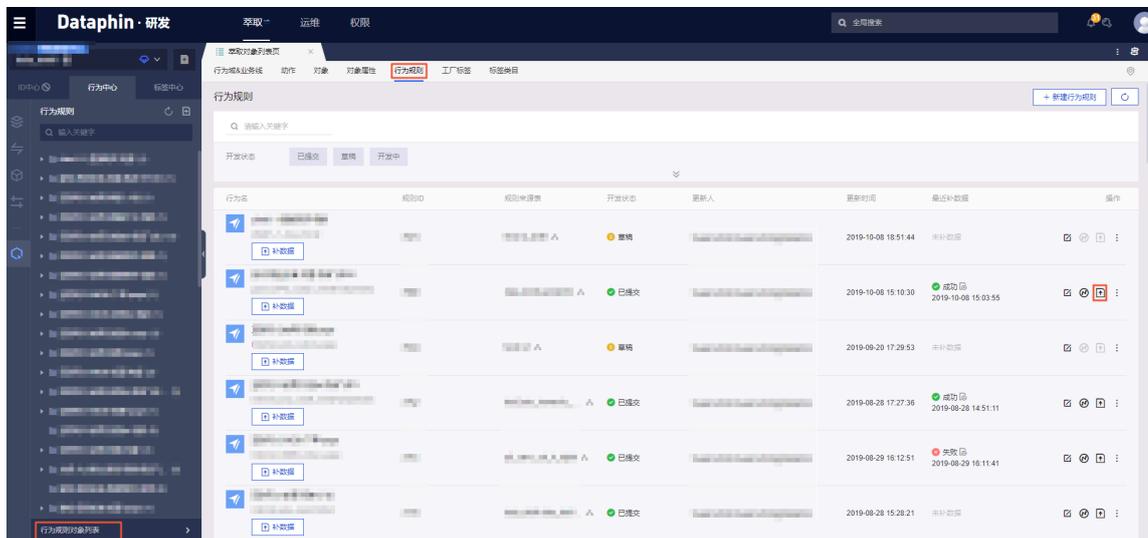
编辑 补数据 ⋮

- 基本信息
 - 行为英文名: education_us_zltest_12
 - 来源表筛选条件: v2_5_x_prod.2_5_x_dev_prod_dev... 🔗
 - 更新人: SuperAdmin(SuperAdmin@dataphin)
 - 更新时间: 2019-09-29 10:33:41
 - 最近补数据: 无补数据
 - 补数据记录:
- 消费信息
 - 标签依赖:
 - 资产详情: [oi_behavior_detail_d](#)
- 行为主体信息

| 行为主体 | ID_Type | ID_Value |
|-----------|---------|------------|
| ID归一化配置:N | UserID | 0908_03_id |
- 对象信息

| 对象 | 标识值 |
|-------------------|------------|
| 对象标识 object_id | 0908_03_id |

- 在数据萃取研发页面，单击左侧导航栏下的行为规则对象列表，进入萃取对象列表页面。在行为规则页面，单击需补数据行为规则所在行的操作列下的 ，进入行为规则补数据页面。



- 在行为规则补数据页面，根据界面提示配置补数据生效时段和实例名称，参数配置完成后单击确定。

行为规则补数据

选择行为规则 已选中1条(1/10)

| 行为名 | 规则ID | 规则来源表 | 更新时间 | 状态 | 最近补数据 |
|--|-------------------------------------|-----------------------|---------------------|-----|-------|
| 域2-业务线21-动作1-对象1 domain2_bizline21_action1_... | <input checked="" type="checkbox"/> | data_distill.oneid_t4 | 2019-09-19 22:51:59 | 开发中 | 未补数据 |

补数据生效时段

* 开始日期: 2019-10-01 自 * 结束日期: 2019-10-08 自

实例名称

* 实例名称: 11 _行为1301_规则1402_20191001_20191008

取消 确定



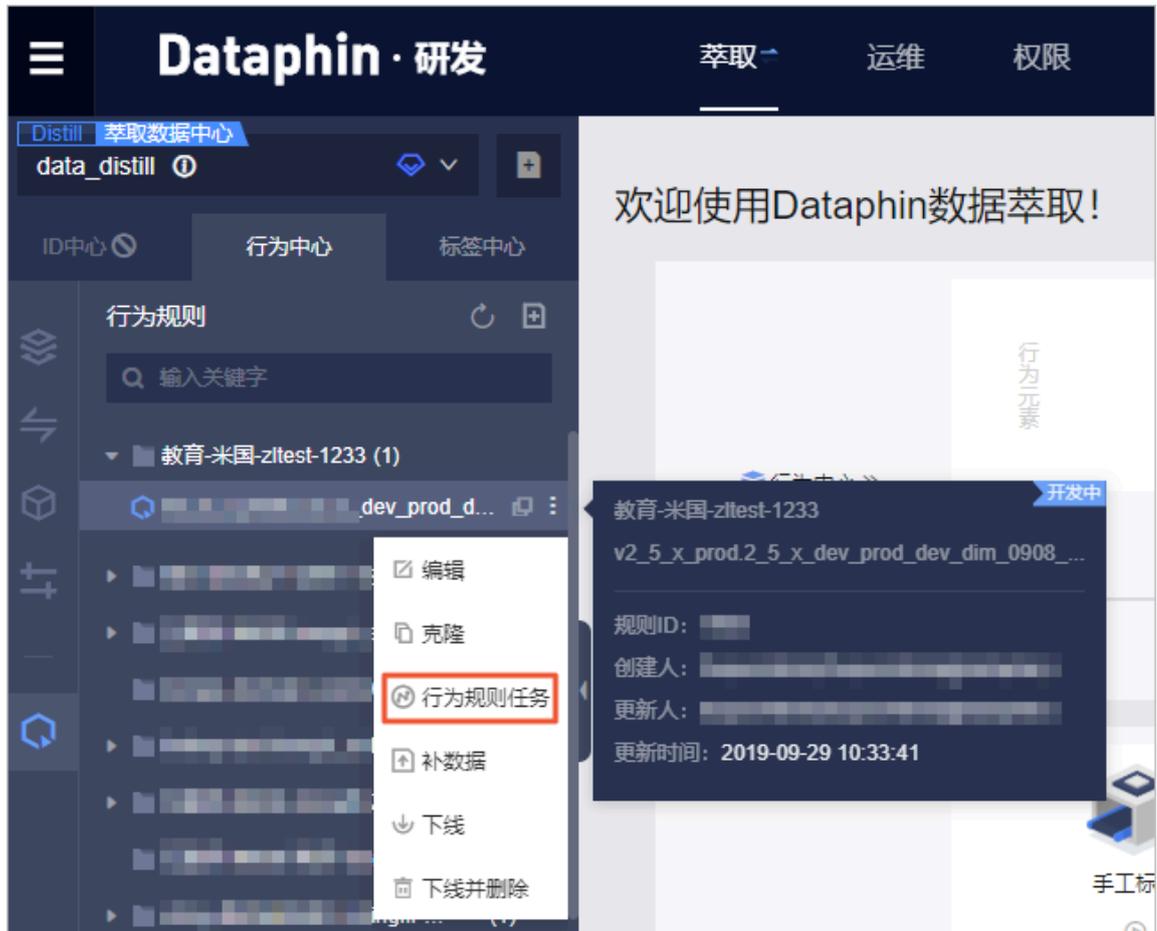
说明:

- 仅已提交和开发中状态的行为规则可进行补数据。
- 如果一个行为下有正在补数据的行为规则，则该行为下的其他行为规则无法进行补数据。即同一个行为下相同时间内有且只能有一个补数据任务。

查看行为规则任务

1. 您可以通过以下3种方式中的任何一种方式，进入萃取运维模块 行为规则任务页面：

- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需查看行为规则后的 ，单击 行为规则任务，进入萃取运维模块的行为规则任务页面。



- 在数据萃取研发页面，单击左侧导航栏中您需查看的行为规则，右侧弹出该行为规则的属性框。在属性框中单击  后，单击 行为规则任务，即可进入萃取运维模块的行为规则任务页面。

属性查看 > 行为规则属性

v2_5_x_prod.2_5_x_dev_prod_dev_dim_0... 开发

中

行为规则ID: 7905
行为: 教育-米国-zltest-1233

编辑 补数据 ⋮

- 基本信息
 - 行为英文: test_12
 - 来源表筛选条: 5_x_dev_prod_dev... 分享
 - 更新: perAdmin@dataphin)
 - 更新时间: 2019-09-29 10:33:41
 - 最近补数据: 无补数据
 - 补数据记录:
- 消费信息
 - 标签依赖:
 - 资产详情: [oi_behavior_detail_d](#)
- 行为主体信息

| 行为主体 | ID_Type | ID_Value |
|-----------|---------|------------|
| ID归一化配置:N | UserID | 0908_03_id |
- 对象信息

| 对象 | 标识值 |
|-------------------|------------|
| 对象标识 object_id | 0908_03_id |

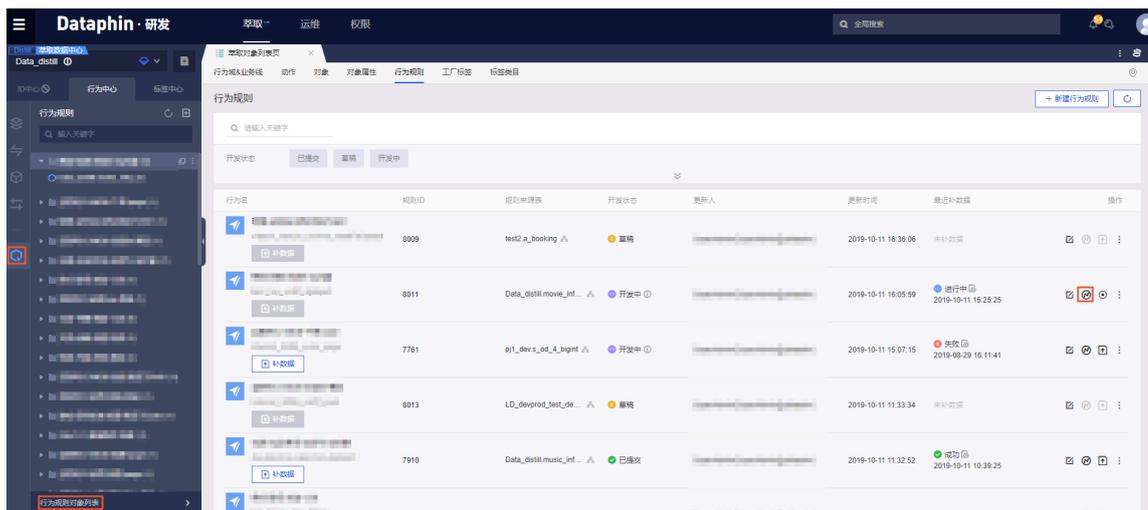
克隆

行为规则任务

下线

下线并删除

- 在数据萃取研发页面，单击左侧导航栏下方的行为规则对象列表，进入萃取对象列表页面。在行为规则页面，单击操作列下的，进入萃取运维模块的行为规则任务页面。



2. 在萃取运维模块中的行为规则任务页面，可查看生产数据运维。



说明:

仅已提交和开发中状态的行为规则可查看对应的行为规则任务。

克隆行为规则

1. 您可以通过以下3种方式的任何一种，进入 规则配置页面：

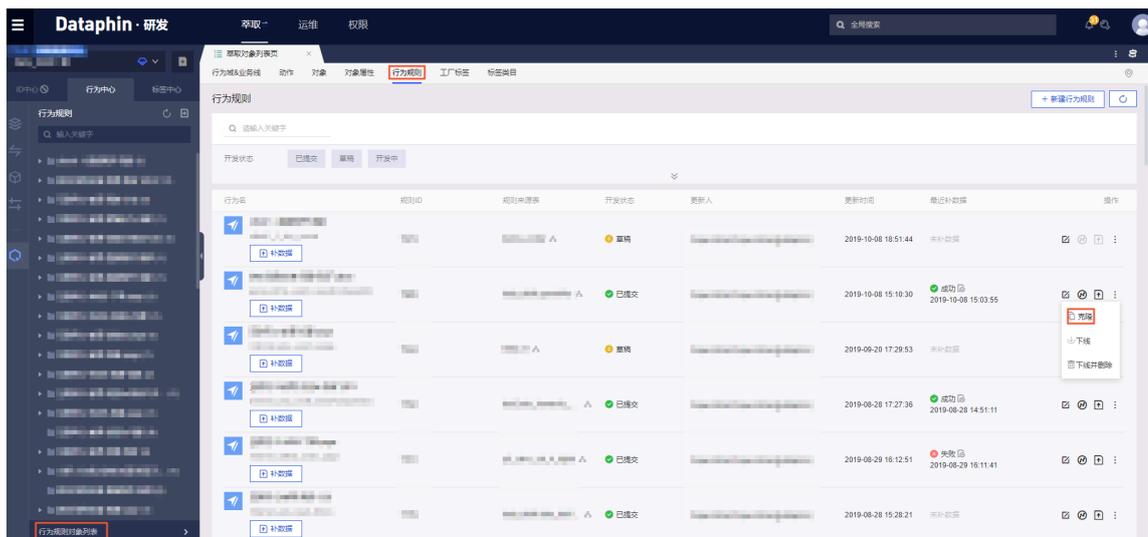
- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需克隆行为规则后的 ，单击 克隆，进入行为规则配置页面。



- 在数据萃取研发页面，单击左侧导航栏中您需克隆的行为规则，右侧弹出该行为规则的属性框。在属性框中单击  后，单击 克隆，即可进入行为规则配置页面。



- 在数据萃取研发页面，单击左侧导航栏下方的 行为规则对象列表，进入萃取对象列表页面。在 行为规则页面，单击 操作列下的 ⋮ 后，单击 克隆，即可进入行为规则配置页面。



2. 在规则配置页面，可查看克隆行为规则的详细配置信息，您可根据实际业务需求是否需要修改配置参数。若需要修改，请参见[编辑行为规则](#)进行修改。

下线行为规则

1. 您可以通过以下3种方式中的任何一种方式，进入下线提示页面：

- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需下线行为规则后的 ，单击 下线，进入下线提示页面。



- 在数据萃取研发页面，单击左侧导航栏中您需下线的行为规则，右侧弹出该行为规则的属性框。在属性框中单击  后，单击 下线，即可进入下线提示页面。

属性查看 > 行为规则属性

v2_5_x_prod.2_5_x_dev_prod_dev_dim_0... 开发中

行为规则ID: 7905
行为: 教育-米国-zltest-1233

编辑 补数据 ⋮

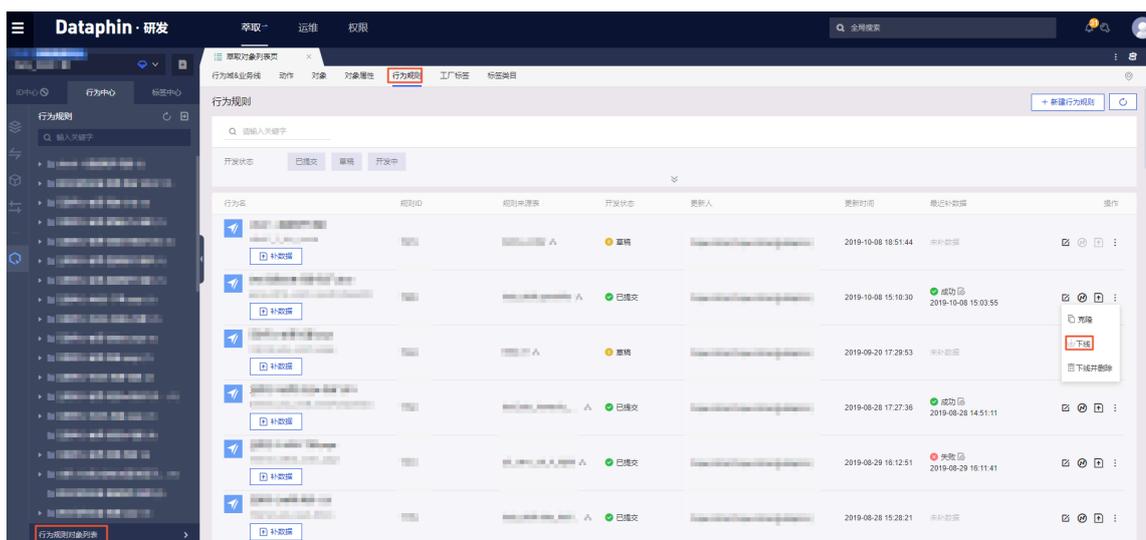
- 基本信息
 - 行为英文: test_12
 - 来源表筛选条: 5_x_dev_prod_dev... 🔗
 - 更新: perAdmin@dataphin)
 - 更新时间: 2019-09-29 10:33:41
 - 最近补数据: 无补数据
 - 补数据记录:
- 消费信息
 - 标签依赖:
 - 资产详情: [oi_behavior_detail_d](#)
- 行为主体信息

| 行为主体 | ID_Type | ID_Value |
|-----------|---------|------------|
| ID归一化配置:N | UserID | 0908_03_id |
- 对象信息

| 对象 | 标识值 |
|-------------------|------------|
| 对象标识 object_id | 0908_03_id |

克隆
行为规则任务
下线
下线并删除

- 在数据萃取研发页面，单击左侧导航栏下方的行为规则对象列表，进入萃取对象列表页面。在行为规则页面，您可通过单击操作列下的 ⋮ 后，单击下线，即可进入下线提示页面。



2. 在弹出的提示框中，填写备注信息，并单击确定并提交。下线成功后行为规则的开发状态更改为草稿。



说明:

- 若行为规则被已提交或开发中状态标签依赖时不支持下线操作。
- 下线行为规则后，系统不会自动为您删除行为中心产出表 `oi_behavior_detail_d` 中该规则的分区数据。如果您新建下游标签时无需使用该规则的数据，建议通过代码任务方式手动删除数据。删除语句详细内容请参见[相关说明](#)。

下线并删除行为规则

1. 您可以通过以下3种方式中的任何一种方式，进入下线并删除提示页面：

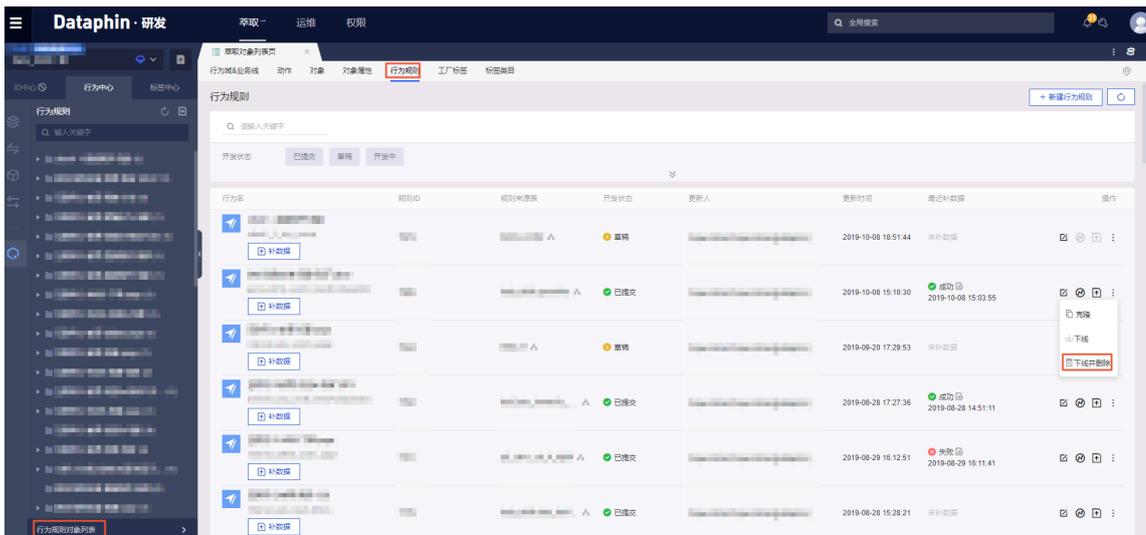
- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需下线并删除行为规则后的 ，单击 下线并删除，进入下线并删除提示页面。



- 在数据萃取研发页面，单击左侧导航栏中您需下线并删除的行为规则，右侧弹出该行为规则的属性框。在属性框中单击  后，单击 下线并删除，即可进入下线并删除提示页面。



- 在数据萃取研发页面，单击左侧导航栏下方的行为规则对象列表，进入萃取对象列表页面。在行为规则页面，您可通过单击操作列下的后，单击下线并删除，即可进入下线并删除提示页面。



2. 在弹出的提示框中，填写备注信息，单击确定并提交。



说明:

- 若行为规则被已提交、开发中状态标签依赖时不支持下线并删除操作。
- 下线行为规则后，系统不会自动为您删除行为中心产出表 oi_behavior_detail_d 中该规则的分区数据。如果您新建下游标签时无需使用该规则的数据，建议通过代码任务方式手动删除数据。删除语句详细内容请参见[相关说明](#)。

相关说明

删除行为规则分区数据。

在Data_distill项目中，物理表oi_behavior_detail_d是有3个分区，分别是ds, source_table, behavior_id。

1. 执行以下命令，查看物理表的分区。

```
show partitions oi_behavior_detail_d;
```

回显例如：

```
ds=20191020/source_table=data_distill.data_distill_movie_info/behavior_id=20
ds=20191021/source_table=data_distill.data_distill_movie_info/behavior_id=20
```

2. 执行以下命令，删除该规则相关的所有分区。

```
ALTER TABLE oi_behavior_detail_d
DROP IF EXISTS PARTITION(ds = 'ds', source_table =
```

```
'source_table', behavior_id = 'behavior_id');
```

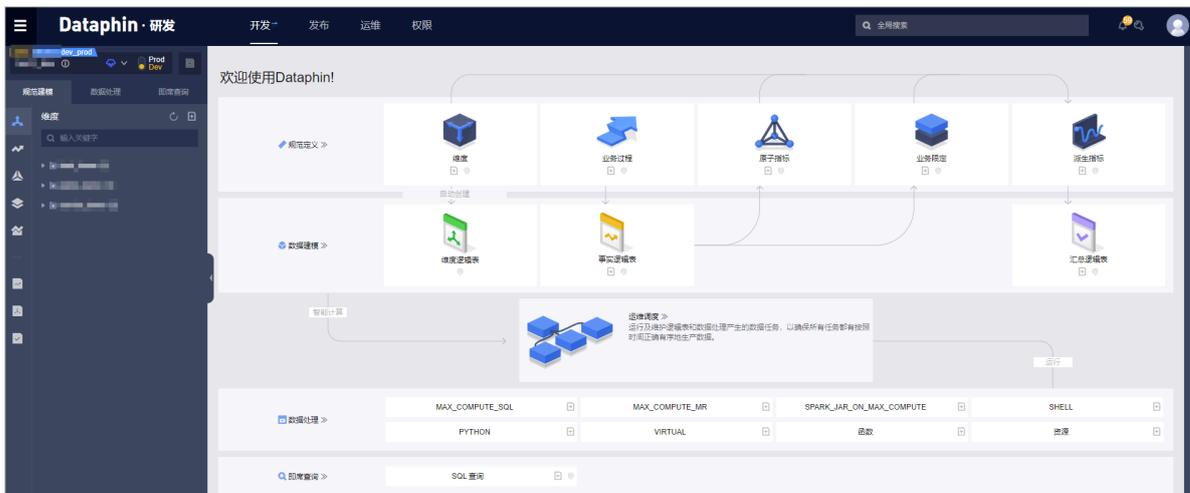
3. 执行以下命令，查看是否有遗留分区。如果有遗留分区，重复上述两步骤，直至删除完所有分区。

```
show partitions oi_behavior_detail_d;
```

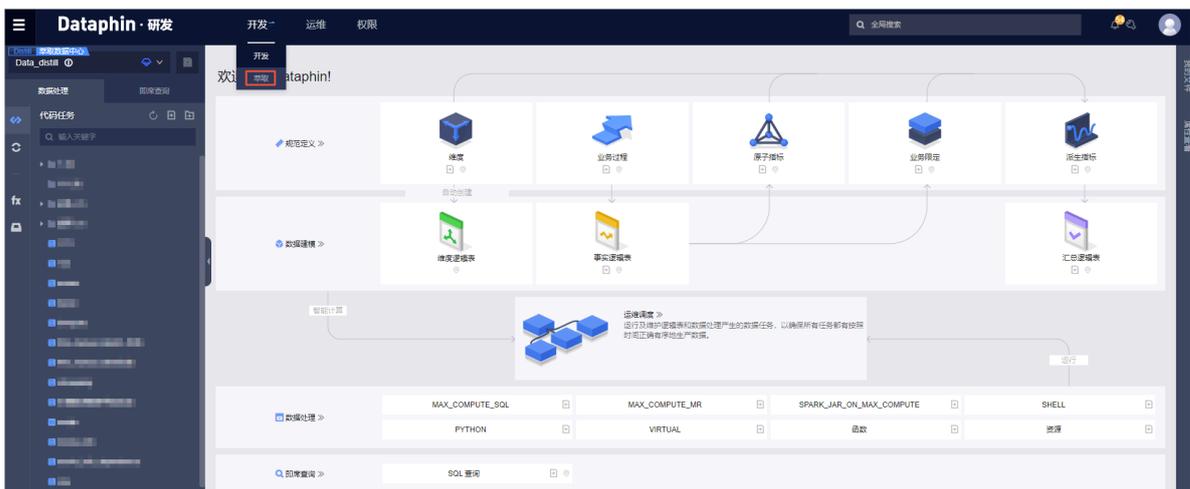
7.2.9 查看行为数据

本文为您介绍如何使用行为看板来查看行为数据分布及抽样数据。

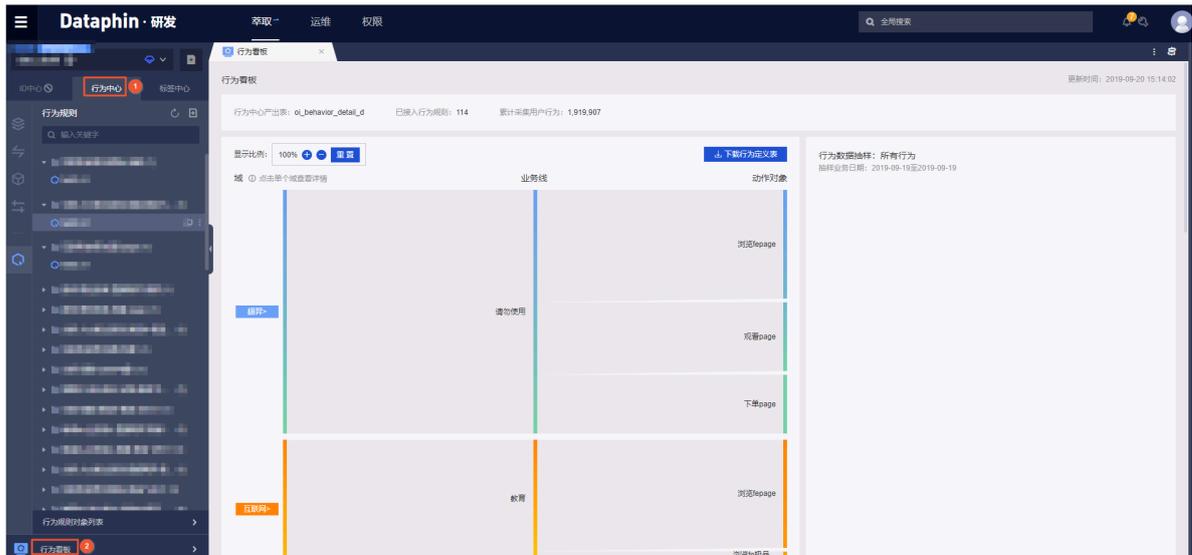
- 1. 登录Dataphin。
- 2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。



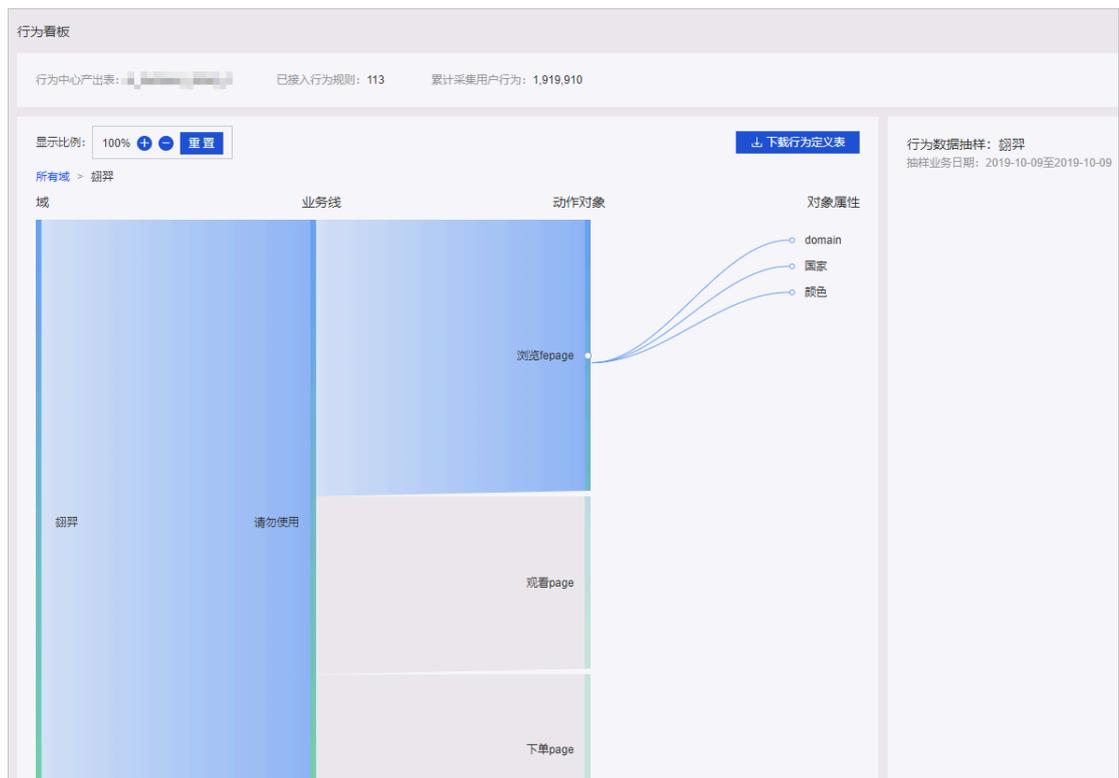
3. 将鼠标悬停在开发后，单击萃取，进入数据萃取研发页面。



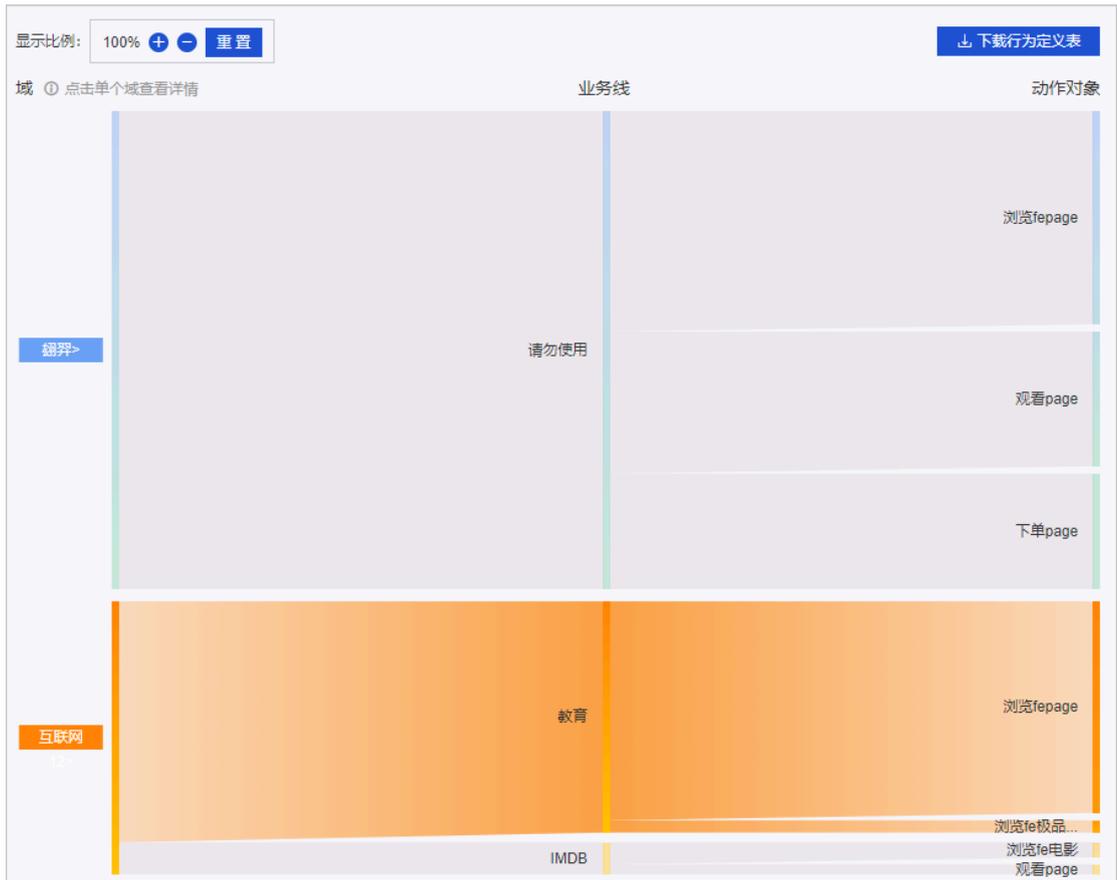
4. 单击行为中心 > 行为看板进入行为看板页面。



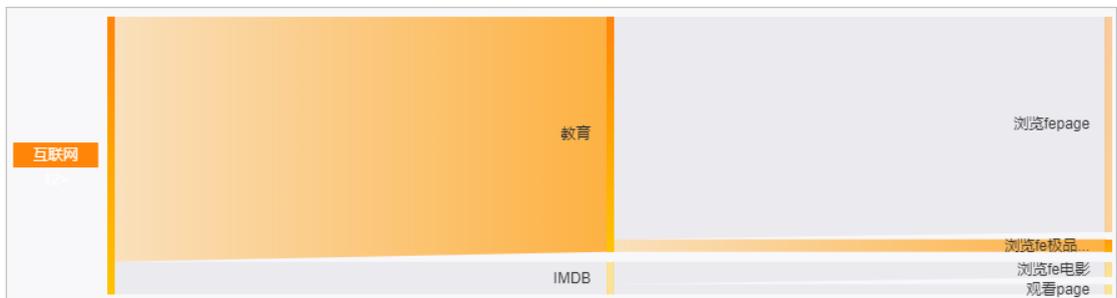
- 在行为看板左侧页面，从左到右依次是一级为行为域，二级为业务线，三级为动作对象。
 - 您可以将鼠标悬停在行为域名称或域下方的柱，查看使用某个行为域的行为数及该行为数占所有域行为总数百分比。



- 您可以将鼠标悬停在业务线下方的柱，查看使用某个业务线的行为数及该行为数占所有域行为总数百分比。



- 您也可以将鼠标悬停在动作对象下方的柱，查看使用某个动作对象的行为数及该行为数占所有域行为总数百分比。



- 在行为看板右侧页面，提供行为数据抽样展示，默认展示全局行为数据的抽样结果。若您已选中某个行为域，则展示该行为域下的抽样结果。抽样数据来源为行为规则周期任务产生的数据，抽样业务日期为昨天。



说明:

抽样业务日期格式为YYYY-MM-DD。

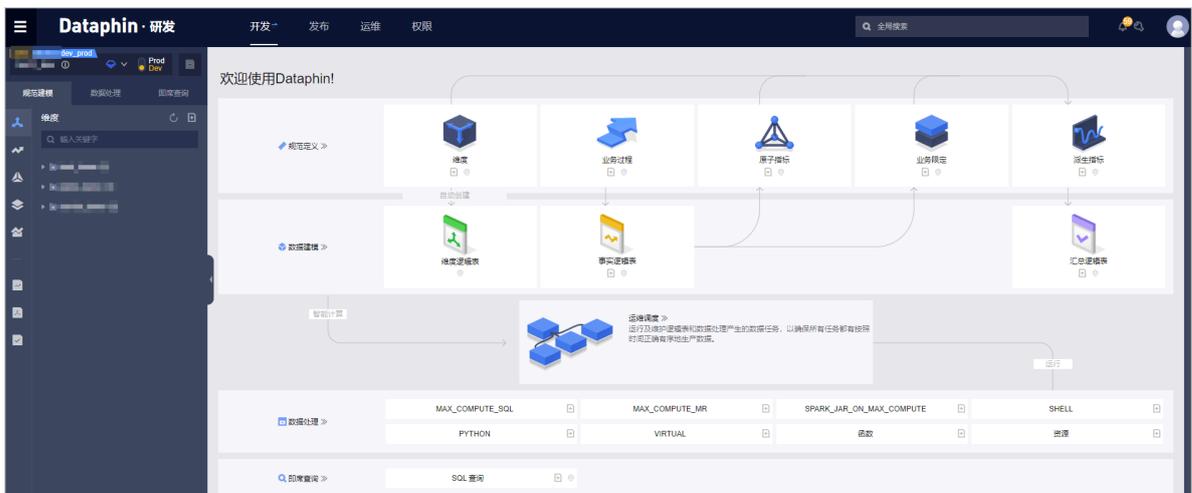
7.3 标签中心

7.3.1 新建工厂标签

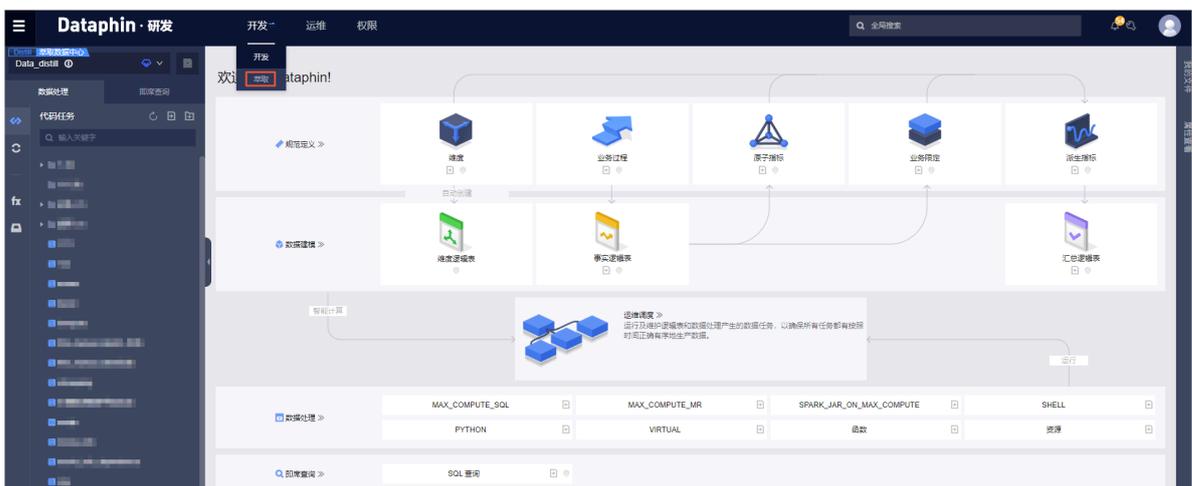
本文为您介绍如何新建工厂标签，包括新建基础指标、配置数据来源及标签值、物理化配置工程标签。帮助您完成行为偏好类的标签的创建。

新建基础指标

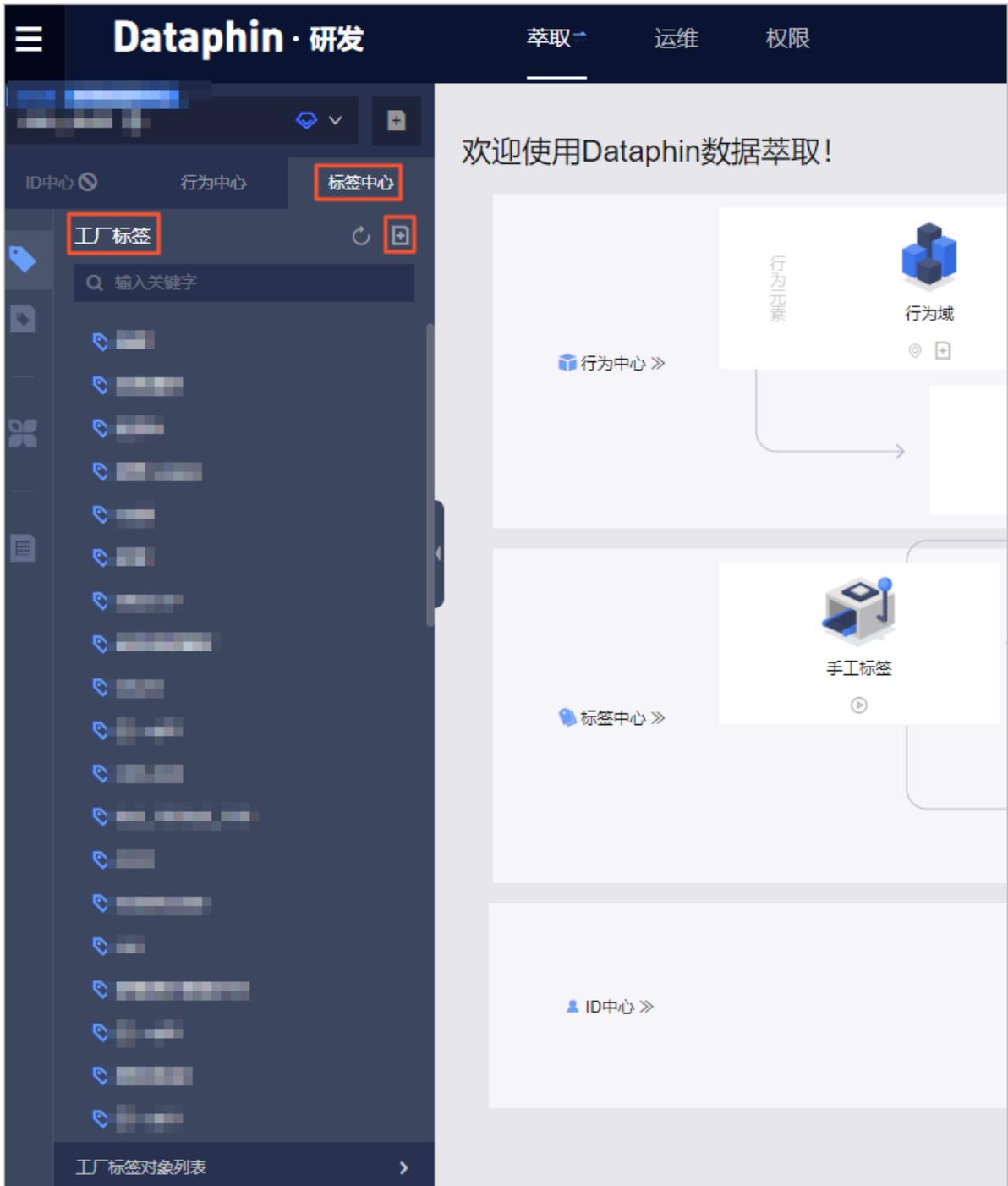
1. 登录Dataphin控制台。
2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。



3. 将鼠标悬停在开发上，单击萃取，进入数据萃取研发页面。



4. 单击标签中心 > 工厂标签进入工厂标签列表页面。



5. 您可以通过以下任何一种方式，进入新建工厂标签页面：

- 在数据萃取研发页面，单击页面右上角的 ，进入新建工厂标签页面。



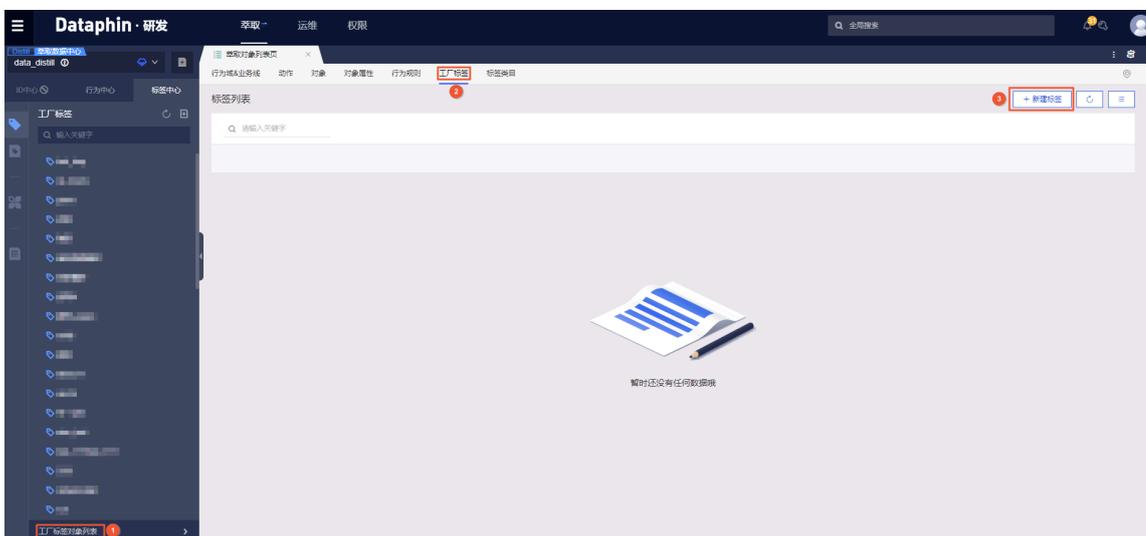
- 在数据萃取研发页面，单击项目名称后 ，选择标签中心 > 工厂标签，进入新建工厂标签页面。



- 在萃取首页，单击工厂标签下的 ，进入新建工厂标签页面。



- 在数据萃取研发页面，单击左侧导航栏下的工厂标签对象列表，进入萃取对象列表页页面。在工厂标签页面，单击页面右上角的新建标签，进入新建工厂标签页面。



6. 在新建工厂标签页面，根据页面提示完成参数配置，并单击确定。

新建工厂标签

* 标签英文名 请填写标签英文名

* 标签名称 请填写标签名称

描述 请填写描述 0/128

所属分类 其他标签

公开状态 公开

* 结果表名 Data_Distill_ 请填写结果表名

取消 确定

- 标签名称是由汉字、字母、数字、下划线、-组合组成，但不支持纯数字格式。
- 标签英文名是由数字、字母、下划线组合组成，但不支持纯数字格式。
- 结果表名是由汉字、字母、数字或下划线或-组合组成。

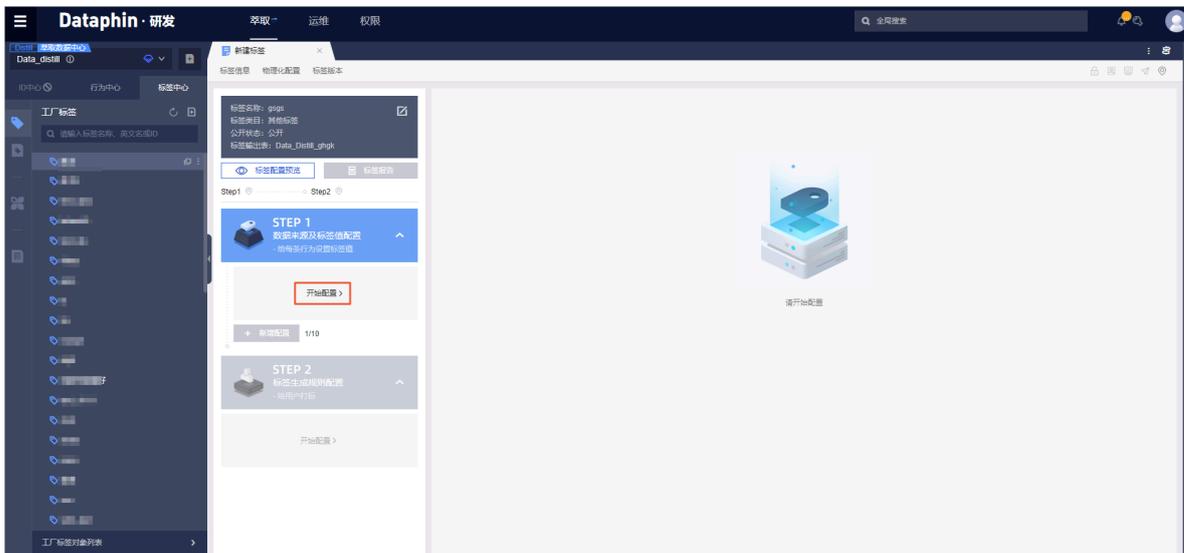


说明:

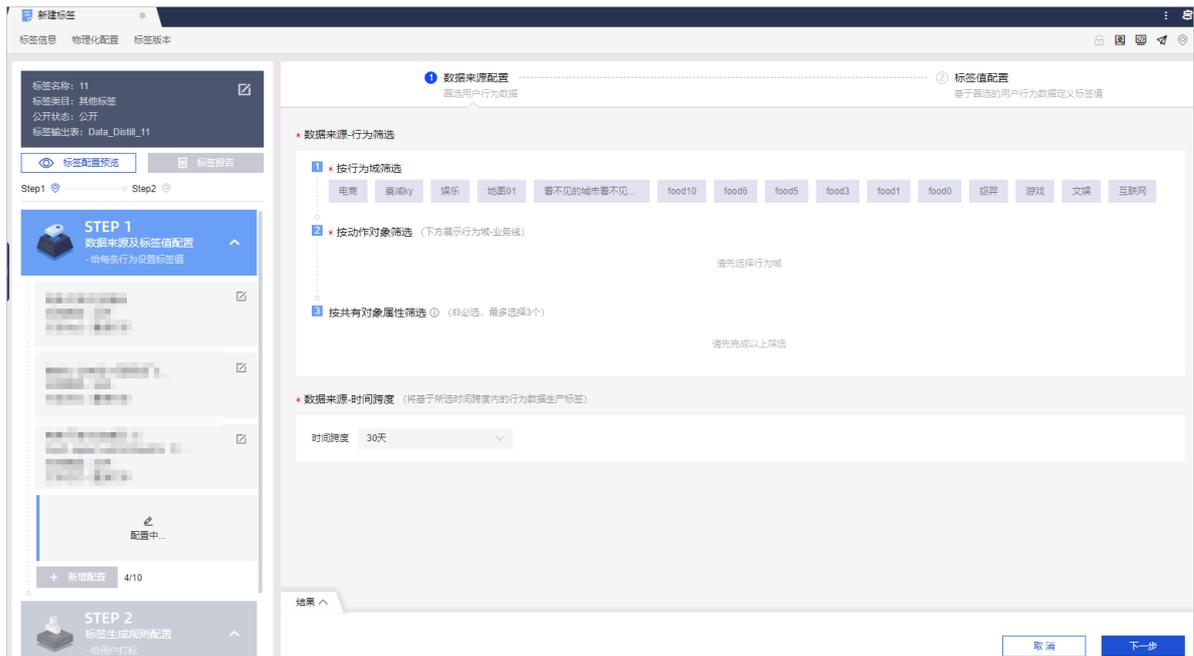
标签名称、标签英文名、结果表名均不支持与已有的标签名称、标签英文名、结果表明重复。

配置数据来源及标签值

1. 在新建标签页面，单击数据来源及标签配置下方的开始配置，进入数据来源配置页面。

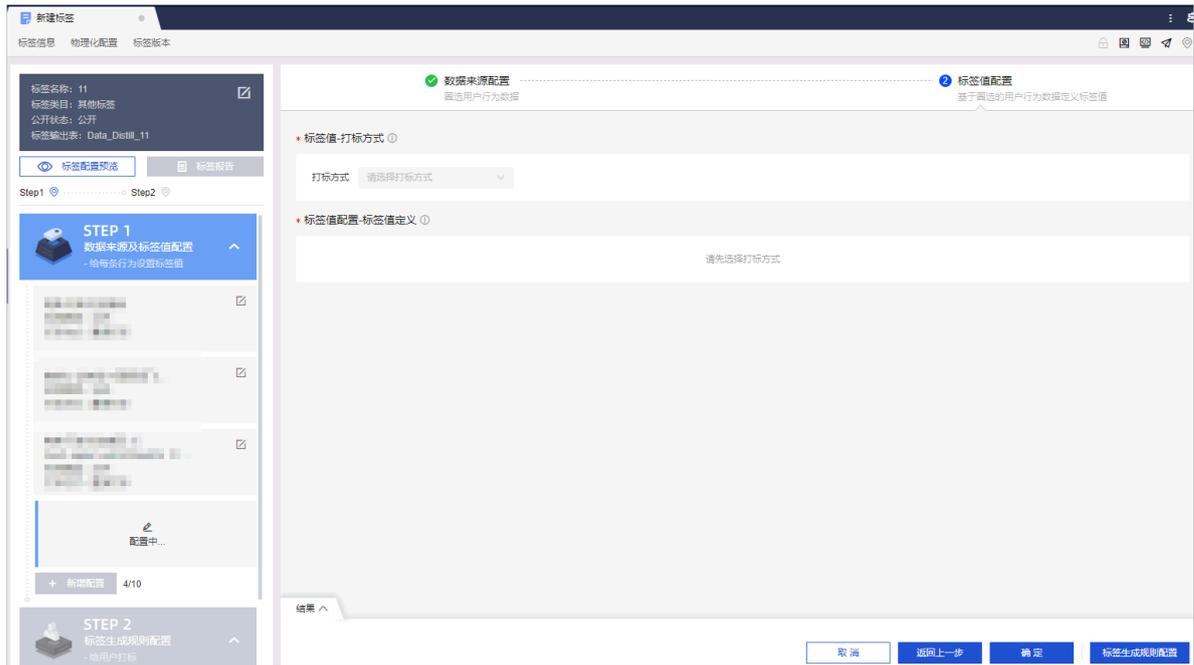


2. 根据页面提示完成参数的配置，并单击下一步。



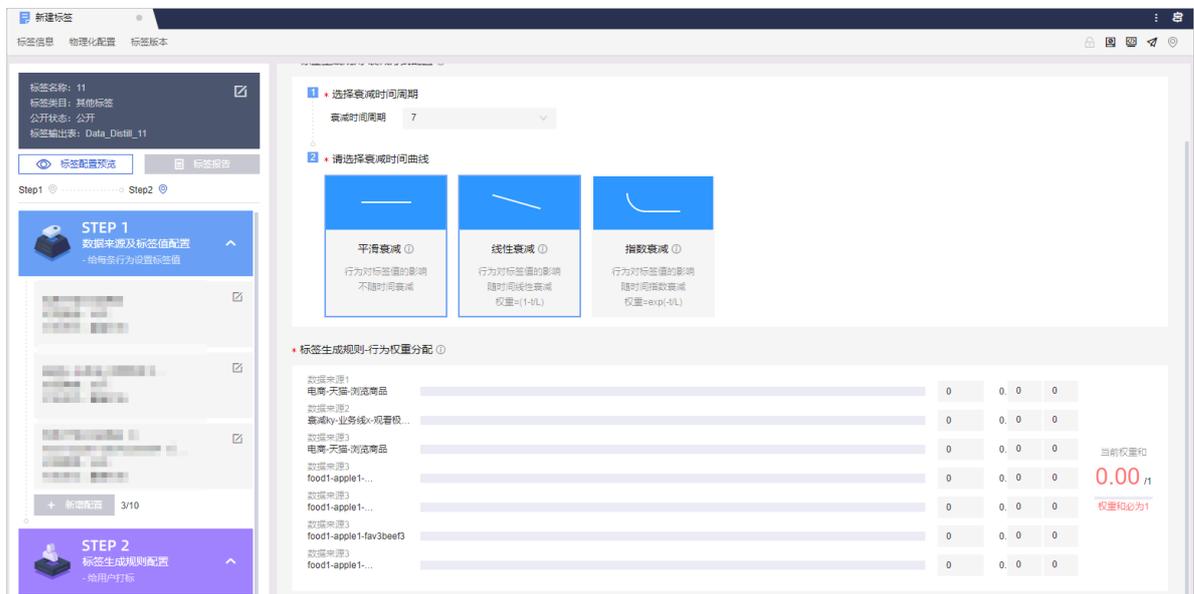
| 参数 | 说明 |
|-----------|---|
| 按行为域筛选 | 选择来源数据所属的行为域，支持多选。 |
| 按动作对象筛选 | 按照行为域下具体的业务线选择动作对象，支持多选。所有数据集选择的对象数量总数不能超过10个。 |
| 对象属性 | 如果已选行为数据有共有对象属性，还可以对共有对象属性设置筛选规则以进一步筛选需要的来源数据，最多选择3个共有对象属性。 |
| 数据来源-时间跨度 | Dataphin将基于所选时间跨度内的行为数据进行后续计算以生成最终标签。系统支持选择30天、60天、90天、180天、365天。 |

3. 进入标签值配置页面，根据页面提示完成参数的配置，单击确定。



- 标签值-打标方式：包括直接打标和对象属性映射打标。
- 标签值配置-标签值定义：根据标签值-打标方式配置标签值定义。

4. 单击标签生成规则配置下方的开始配置，进入标签生成规则配置页面。根据页面提示完成参数配置，并单击确定。



| 参数 | 说明 |
|---------|--|
| 返回标签值个数 | 定义每个ID最多能返回的标签值个数。每个ID可能对应多条行为数据，因此可能对应多个标签值，按照设置个数优先返回偏好度较大的标签值（标签值偏好度可在标签结果表查看）。 |

| 参数 | 说明 |
|---------------|---|
| 选择衰减时间周期 | 不同时间的行为对标签值的影响程度不同。同一周期内行为的影响力相同，不同周期行为的影响力按照设置的衰减周期和曲线随时间递减。对于标签值会随着行为快速变化的，建议选择较短的时间周期（例如纸尿裤型号偏好）。对于标签值相对稳定，不随行为快速变化的，建议选择较长的时间周期（例如预测个人肤质）。 |
| 请选择衰减时间曲线 | 支持对所选数据来源-时间跨度（例如90天）行为数据，根据不同时间周期（如7天）切片（ $90/7=13$ 个片段）设置衰减模式，不同时间周期行为数据标签值的影响力与您选择的衰减模式相关。例如，个人肤质是长期稳定的，因此一个消费者7天前收藏美妆商品的行为和最近7天收藏美妆商品的行为，对刻画这个消费者的护肤品功效偏好无影响，则选择衰减模式为平滑型。 |
| 标签生成规则-行为权重分配 | 基于业务上的判断对多个行为进行权重分配，要求最多两位小数，且各项权重和为1。权重越大，表示该行为对最后标签值的影响越大。 |
| 标签生成规则-覆盖用户范围 | 选择覆盖用户范围。 |

完成新建基础指标及配置数据来源及标签值操作步骤后，即可生成草稿状态的标签。

物理化配置工厂标签

1. 在编辑标签页面，单击页面物理化配置，进入物理化配置页面。

2. 根据页面提示完成参数配置，并单击确认。

物理化配置 ×

• 存储配置

* 存储生命周期

快速选择 天

* 选择分区字段

3. 完成以上配置后单击页面右上方的 ，可完成保存当前配置。

4. 单击页面右上方的  进行测试运行。必填信息配置完整后，提交标签前建议您先进行测试运行，以预览标签值分布及标签结果表数据抽样。

测试运行的状态包括未测试运行、测试运行中、测试运行成功及测试运行失败。测试运行中的标签不支持修改标签配置，如需修改配置，请您先终止标签的测试运行再进行修改配置。终止测试运行后测试运行状态将更改为测试运行失败。

如果您的标签无需测试运行，可以直接提交标签。

5. 单击页面右上方的 ，进行提交标签。

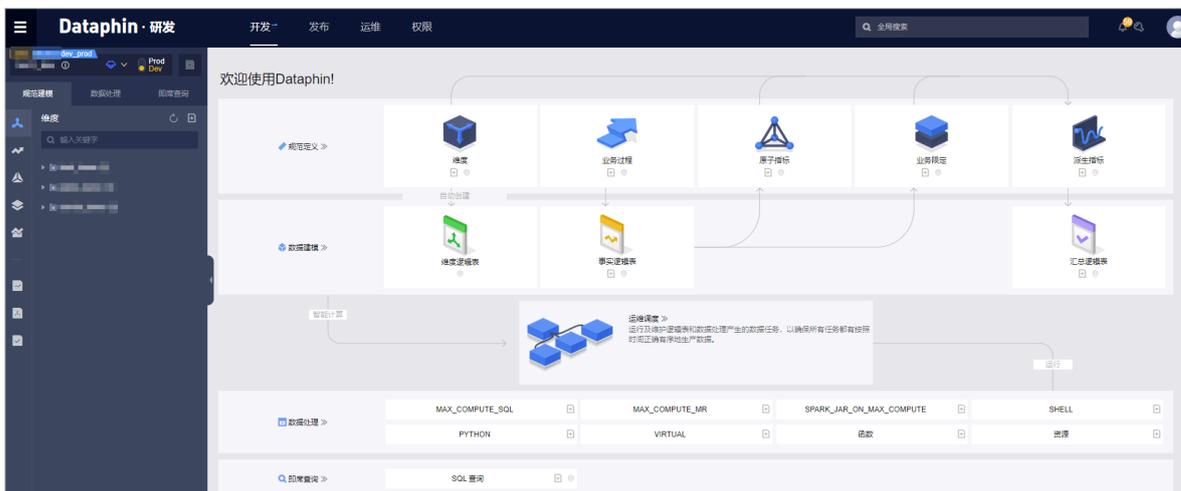
 **说明:**
提交后需对该标签进行补数据操作，以获取所选时间跨度内的历史行为数据作为来源数据。如果您未进行补数据操作，则标签结果表数据可能为空。

7.3.2 管理工厂标签

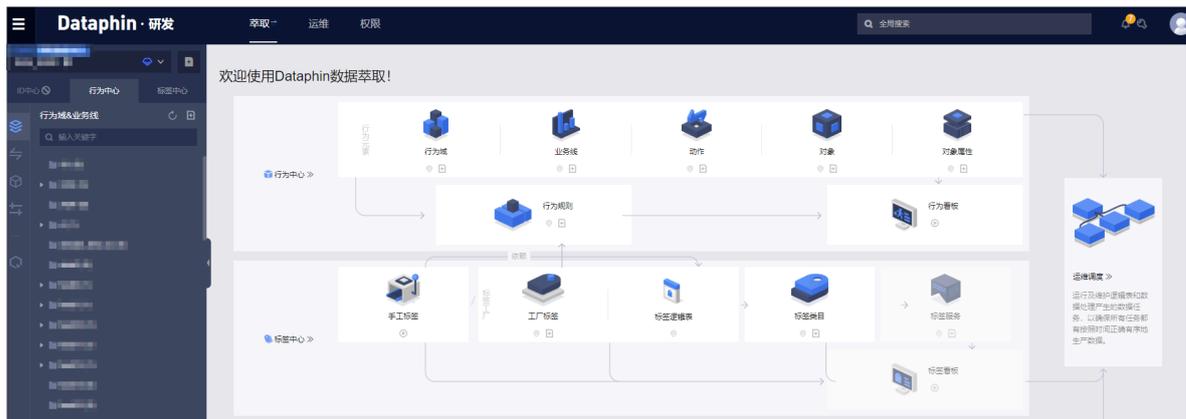
本文为您介绍如何进行工厂标签的查看、编辑、补数据、下线、下线并删除等操作，帮助您管理工厂标签。

查看工厂标签

1. 登录Dataphin控制台。
2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。



3. 单击萃取，进入数据萃取研发页面。



4. 单击标签中心 > 工厂标签，进入工厂标签列表页面。

5. 您可以通过以下2种方式查看工厂标签不同维度的信息：

- 在左侧导航栏单击您需查看的标签，右侧弹出该标签的属性框。在属性框中可查看该标签的基本信息：

属性查看 > 工厂标签属性

_____tag 开发中 其他标签 公开

needForSpeed04

编辑 查看标签报告 补数据 ⋮

• 数据来源及标签值

行为: _____

筛选条件: 游戏引擎包含engine 且 视角包含D

打标方式: 属性值映射打标

标签值: 共有对象属性: 视角、游戏引擎; 映射方式: 属性值应用为标签值、属性值...

时间跨度: 30

• 标签生成规则

对用户返回标签值个数 (最多) : 2

衰减时间周期: 7

时间衰减曲线: 平滑: 行为对标签值的影响不随时间衰减

行为权重:

| 行为 | 权重 |
|------------------------|----|
| 地图01-流水线03-检查01-极品飞车04 | 1 |

覆盖用户范围: 0% ~ 100%

• 标签结果表

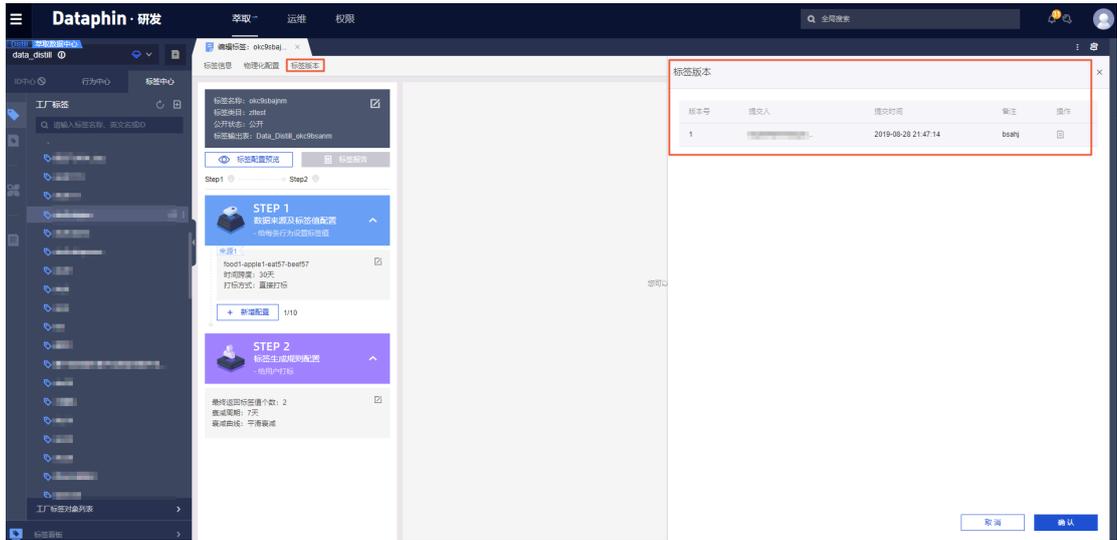
标签结果表: Data_Distill_needForSpeed04Tag01

• 依赖关系

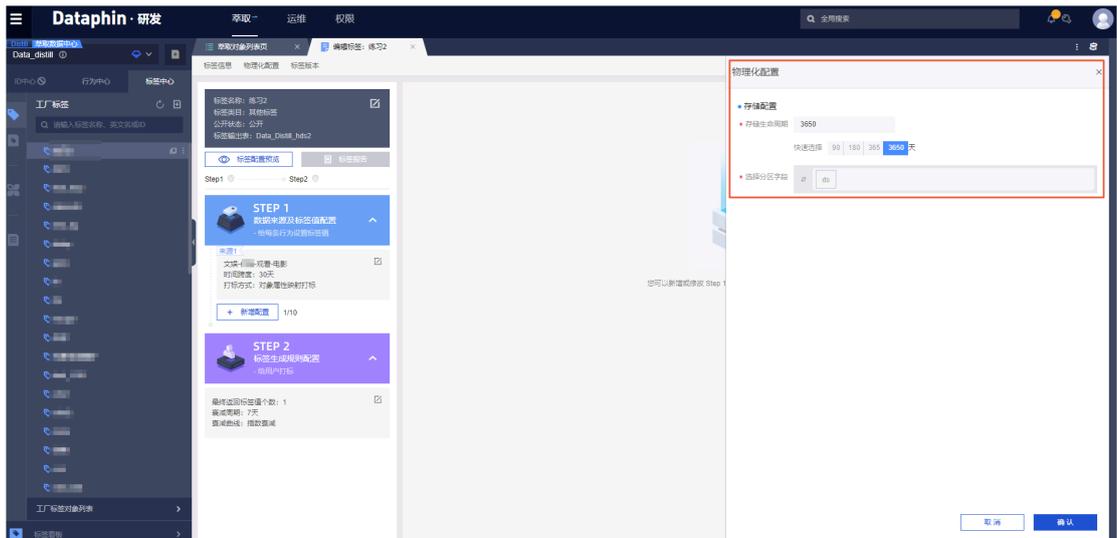
行为: 7803

行为规则: 7804

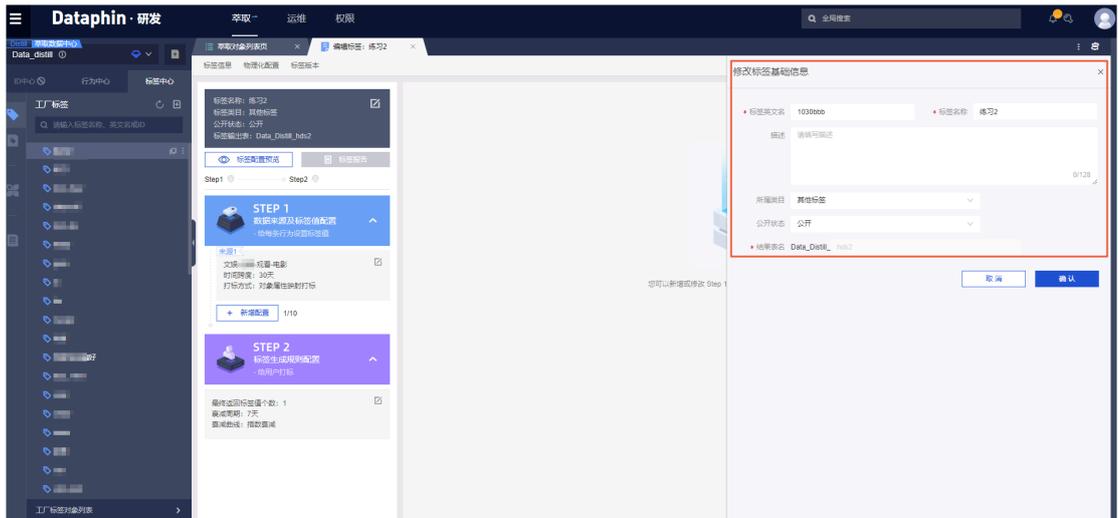
- 在标签的属性框页面，同时您可以单击编辑，进入编辑标签页面。在标签页面页面上方，单击标签版本，查看标签版本。



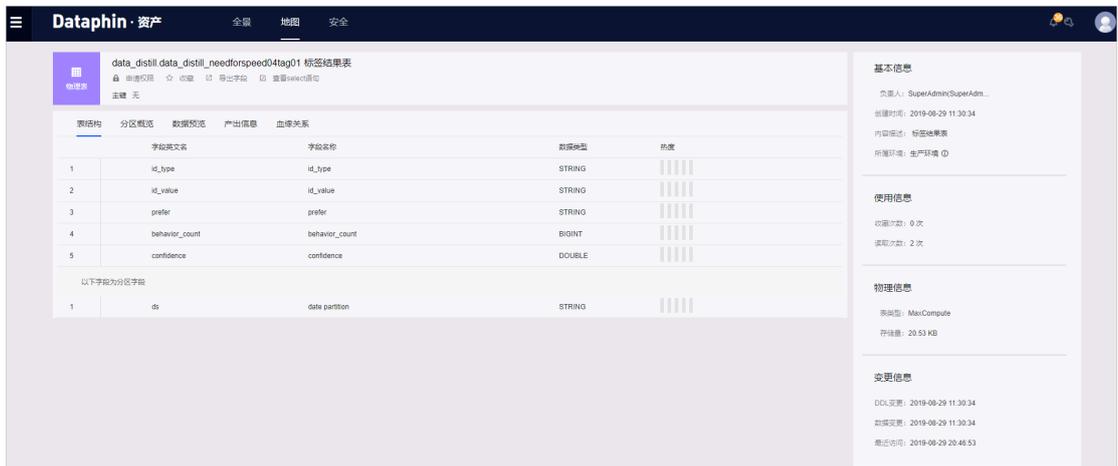
- 在标签的属性框页面，同时您可以单击编辑，进入编辑标签页面。在标签页面页面上方，单击物理化配置，查看物理化配置信息。



- 在标签的属性框页面，同时您可以单击编辑，进入编辑标签页面。在标签页面页面上方，单击标签信息，查看标签的基础信息。



- 在标签的属性框页面，您可以通过单击标签结果表下的结果表表名，可查看结果表的详细信息。



- 在标签的属性框页面，您可以通过单击依赖关系下的行为规则ID，可查看该标签依赖的行为规则的详细信息。

data_distill.behavior_movie
已提交

行为规则ID: 6118
行为: food1-apple1-eat57-beef57

编辑
补数据
⋮

● 基本信息

行为英文名: food1_apple1_eat57_beef57

来源表筛选条件: data_distill.behavior_movie

更新人: SuperAdmin(SuperAdmin@dataphin)

更新时间: 2019-08-25 17:59:10

最近补数据: ✔ 成功

补数据记录:

● 消费信息

标签依赖: [7325:okc24](#) [7793:okc911](#) [7798:okc91111](#)
[7794:okc9111](#) [7797:sajbjanm](#)

资产详情: [oi_behavior_detail_d](#)

● 行为主体信息

| 行为主体 | ID_Type | ID_Value |
|-----------|---------|----------|
| ID归一化配置:Y | UserID | user_id |

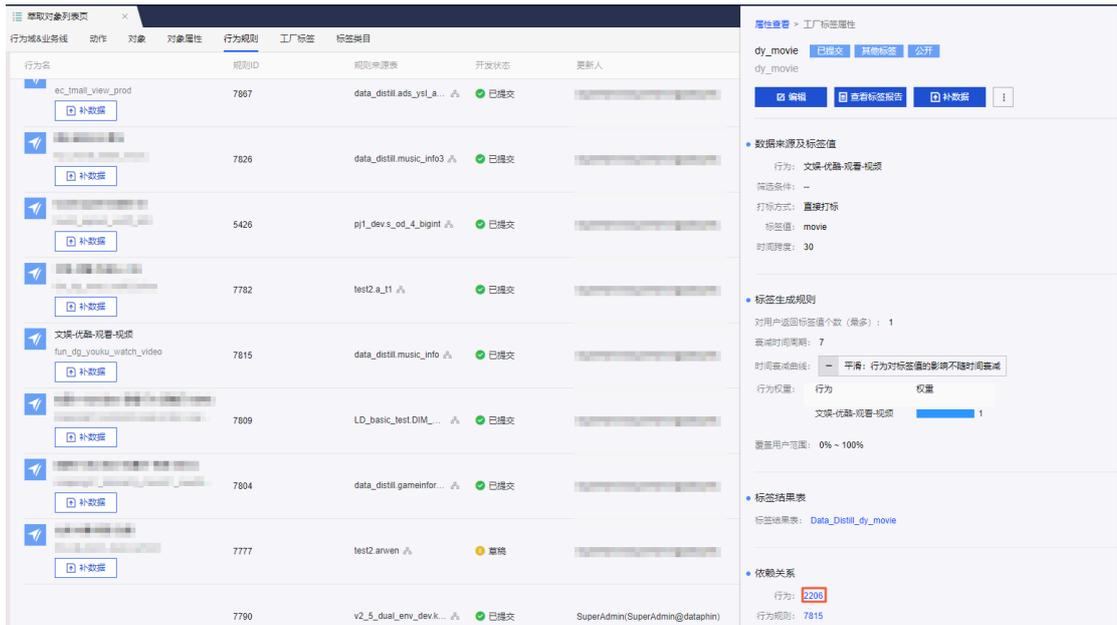
● 对象信息

| 对象 | 标识值 |
|-------------------|-----|
| 对象标识 object_id | ip |

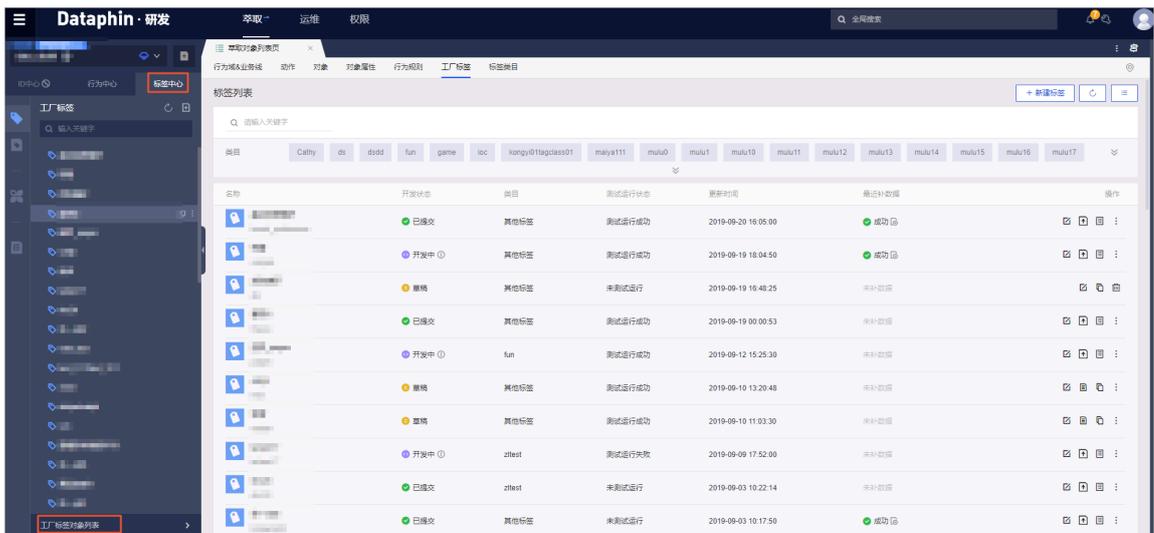
● 对象属性信息

| 对象属性 | 属性 | 属性值 |
|------|----|-----|
|------|----|-----|

- 在标签的属性框页面，您也可以通过单击依赖关系下的行为ID，进入萃取对象列表页面。在行为规则页面可查看行为基本信息。



- 在工厂标签列表页面，单击页面左下方的工厂标签对象列表，进入抽取对象列表页面。在抽取对象列表页面，您可以通过筛选批量查看标签信息，也可以通过单击标签名称查看单击标签信息。



- 开发状态：包括草稿、已提交、开发中3种状态。开发中表示已提交标签重新编辑后仅保存未提交，仍根据修改前的配置在生产环境参与任务调度。已提交表示标签已在生产环境中参与任务调度。草稿表示新建标签未提交的状态。
- 类目：标签所属类目。
- 测试运行状态：未测试运行、测试运行中、测试运行成功、测试运行失败4种状态。
- 最近补数据：包含未执行、运行中、成功和失败4种。鼠标悬浮在最近补数据状态后面可查看该标签的所有补数据记录。在补数据记录页面，单击结果列下的，进入萃取运维页面，查看该次补数据生成的实例及DAG图。

编辑工厂标签

1. 您可以通过一下3种方式中的任何一种方式，进入编辑标签页面：

- 在数据萃取研发页面，单击左侧导航栏中的  后，将鼠标悬浮在需编辑标签后的 ，单击编辑，进入编辑标签页面。



- 在数据萃取研发页面，单击左侧导航栏中您需编辑的标签，右侧弹出该标签的属性框。在属性框中单击编辑，即可进入编辑标签页面。

属性查看 > 工厂标签属性

极品飞车04_tag 开发中 其他标签 公开

needForSpeed04

编辑 查看标签报告 补数据 ⋮

- 数据来源及标签值
 - 行为： 地图01-流水线03-检查01-极品飞车04
 - 筛选条件： 游戏引擎包含engine 且 视角包含D
 - 打标方式： 属性值映射打标
 - 标签值： 共有对象属性：视角、游戏引擎；映射方式：属性值应用为标签值、属性值...
 - 时间跨度： 30

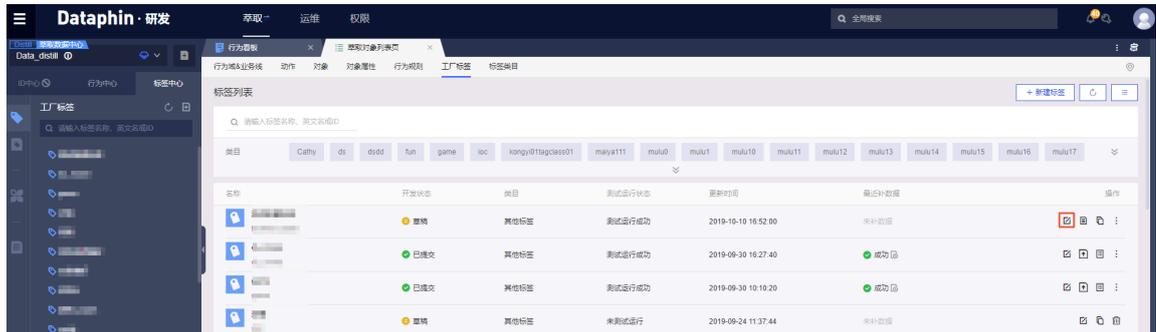
- 标签生成规则
 - 对用户返回标签值个数（最多）： 2
 - 衰减时间周期： 7
 - 时间衰减曲线： 平滑：行为对标签值的影响不随时间衰减
 - 行为权重：

| 行为 | 权重 |
|------------------------|----|
| 地图01-流水线03-检查01-极品飞车04 | 1 |
 - 覆盖用户范围： 0% ~ 100%

- 标签结果表
 - 标签结果表： [Data_Distill_needForSpeed04Tag01](#)

- 依赖关系
 - 行为： [7803](#)
 - 行为规则： [7804](#)

- 在数据萃取研发页面，单击左侧导航栏下的工厂标签对象列表，进入萃取对象列表页面。在工厂标签页面，单击标签所在行的操作列的 ，进入编辑标签页面。



2. 在编辑标签页面，完成参数配置，详细内容请参见[新建工厂标签](#)。



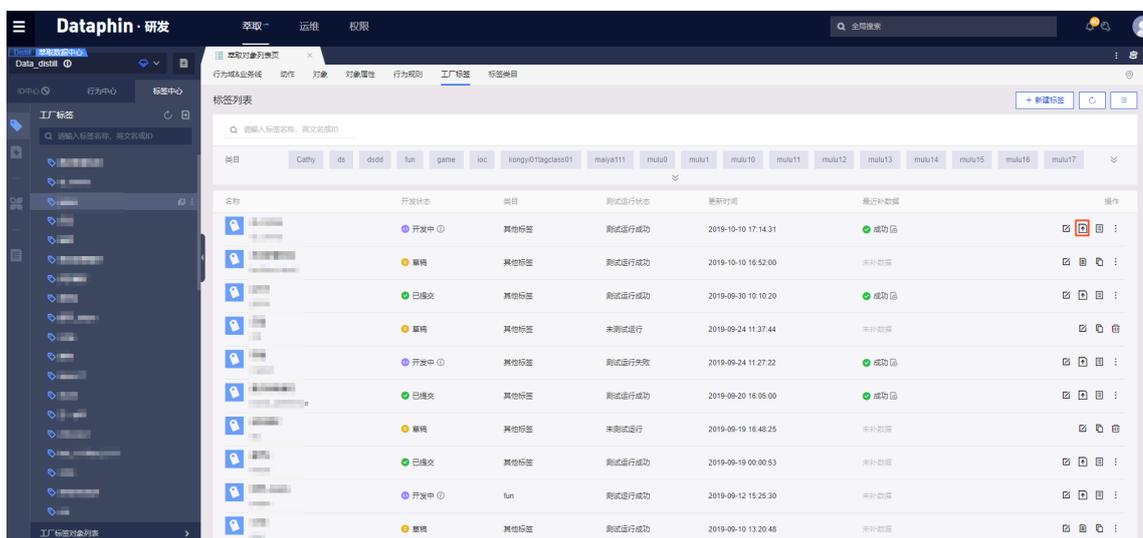
说明：

已提交和开发中状态的标签，结果表名称不可修改。

补数据

1. 您可以通过以下2种方式中的任何一种方式，进入标签补数据页面：

- 在数据萃取研发页面，单击左侧导航栏下的工厂标签对象列表，进入萃取对象列表页页面。在工厂标签页面，单击需补数据标签所在行的操作列下，进入标签补数据页面。



- 在数据萃取研发页面，单击左侧导航栏中您需补数据的标签，右侧弹出该标签的属性框。在属性框中单击补数据，即可进入标签补数据页面。

属性查看 > 工厂标签属性

极品飞车04_tag 开发中 其他标签 公开

needForSpeed04

编辑 查看标签报告 补数据 ⋮

- 数据来源及标签值
 - 行为: 地图01-流水线03-检查01-极品飞车04
 - 筛选条件: 游戏引擎包含engine 且 视角包含D
 - 打标方式: 属性值映射打标
 - 标签值: 共有对象属性: 视角、游戏引擎; 映射方式: 属性值应用为标签值、属性值...
 - 时间跨度: 30

- 标签生成规则
 - 对用户返回标签值个数 (最多): 2
 - 衰减时间周期: 7
 - 时间衰减曲线: 平滑: 行为对标签值的影响不随时间衰减
 - 行为权重:

| 行为 | 权重 |
|------------------------|----|
| 地图01-流水线03-检查01-极品飞车04 | 1 |
 - 覆盖用户范围: 0% ~ 100%

- 标签结果表
 - 标签结果表: [Data_Distill_needForSpeed04Tag01](#)

- 依赖关系
 - 行为: 7803
 - 行为规则: 7804

2. 根据界面提示配置补数据生效时段和实例名称，参数配置完成后单击确定。

标签补数据 ×

标签信息

| 标签名 | 标签 ID | 最近更新时间 | 开发状态 | 最近补数据 |
|----------------------|-------|---------------------|--------------------|-------------------|
| dy_movie dy_movie | 7971 | 2019-10-10 17:14:31 | 开发中 ① | 成功 ② |

补数据生效时段

* 结束日期 自
基于该标签配置的数据来源时间跨度自动计算开始时间，无需选择

实例名称

* 实例名称 _7971_20191009



说明:

仅已提交和开发中状态的标签可进行补数据。

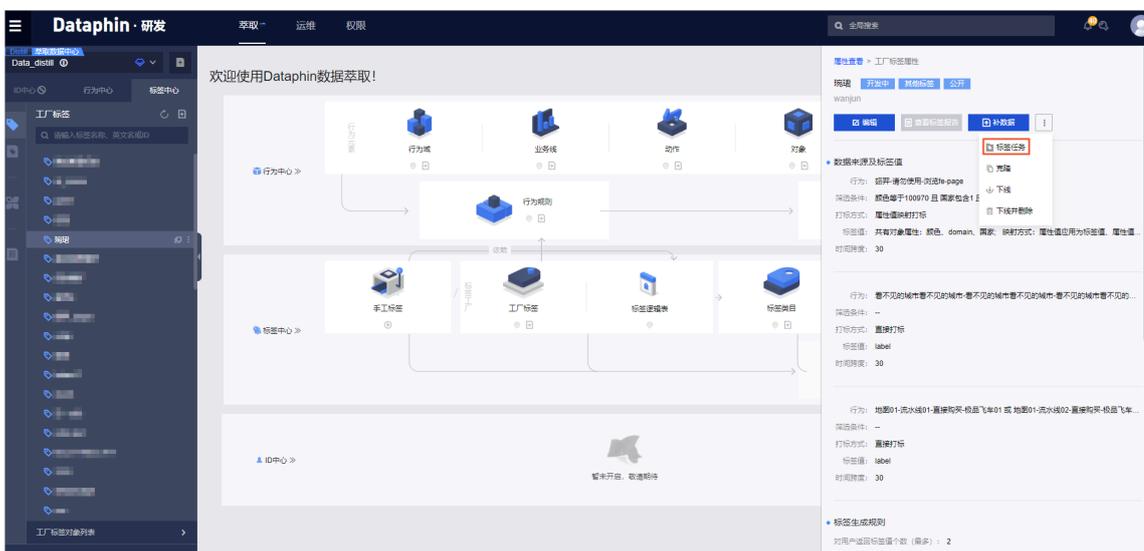
查看标签任务

1. 您可以通过以下3种方式中的任何一种方式，进入萃取运维模块标签任务页面：

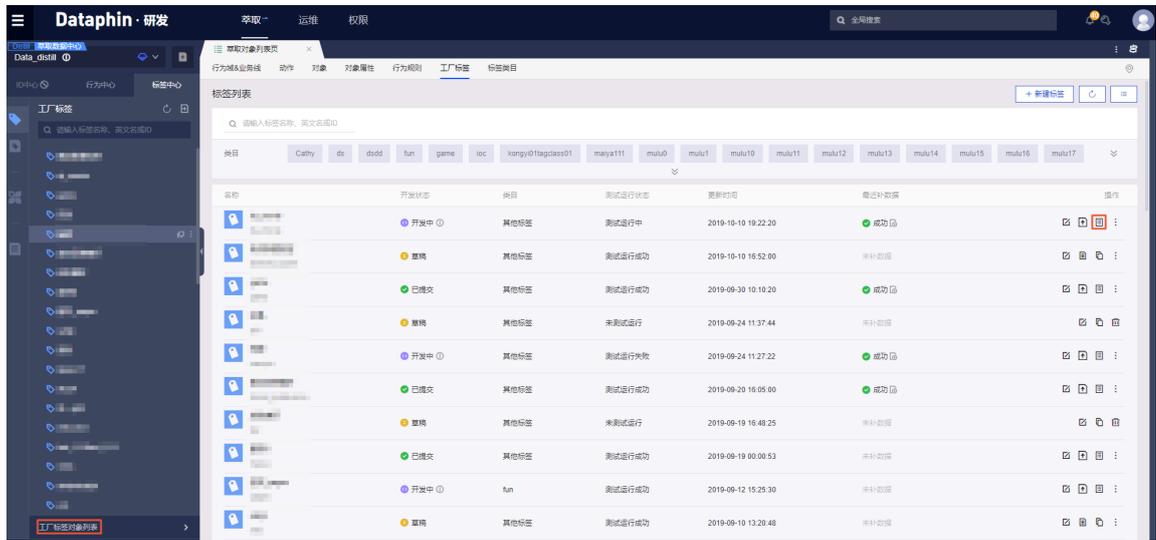
- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需查看标签后的 ，单击标签任务，进入萃取运维模块的标签任务页面。



- 在数据萃取研发页面，单击左侧导航栏中您需查看的标签，右侧弹出该标签的属性框。在属性框中单击  后，单击标签任务，即可进入萃取运维模块的标签任务页面。



- 在数据萃取研发页面，单击左侧导航栏下方的工厂标签对象列表，进入萃取对象列表页面。在工厂标签页面，单击操作列下的 ，进入萃取运维模块的标签任务页面。



说明：

仅已提交和开发中状态的标签可查看对应的标签任务。

- 在萃取运维模块中的标签任务页面，可查看生产数据运维。

查看标签报告

1. 您可以通过以下3种方式中的任何一种方式，进入标签报告详情页面：

- 在数据萃取研发页面，单击左侧导航栏中的，将鼠标悬浮在需查看标签后的，单击标签报告，进入标签报告详情页面。



- 在数据萃取研发页面，单击左侧导航栏中您需查看的标签，右侧弹出该标签的属性框。在属性框中单击后，单击查看标签报告，即可进入标签报告详情页面。

属性查看 > 工厂标签属性

game 已提交 其他标签 公开

game

编辑 查看标签报告 补数据 ⋮

- 数据来源及标签值
 - 行为: 游戏-角色扮演-观看-极品飞车02
 - 筛选条件: --
 - 打标方式: 直接打标
 - 标签值: gamedy
 - 时间跨度: 30

- 标签生成规则
 - 对用户返回标签值个数 (最多): 1
 - 衰减时间周期: 7
 - 时间衰减曲线: 平滑: 行为对标签值的影响不随时间衰减

| 行为 | 权重 |
|-------------------|----|
| 游戏-角色扮演-观看-极品飞车02 | 1 |

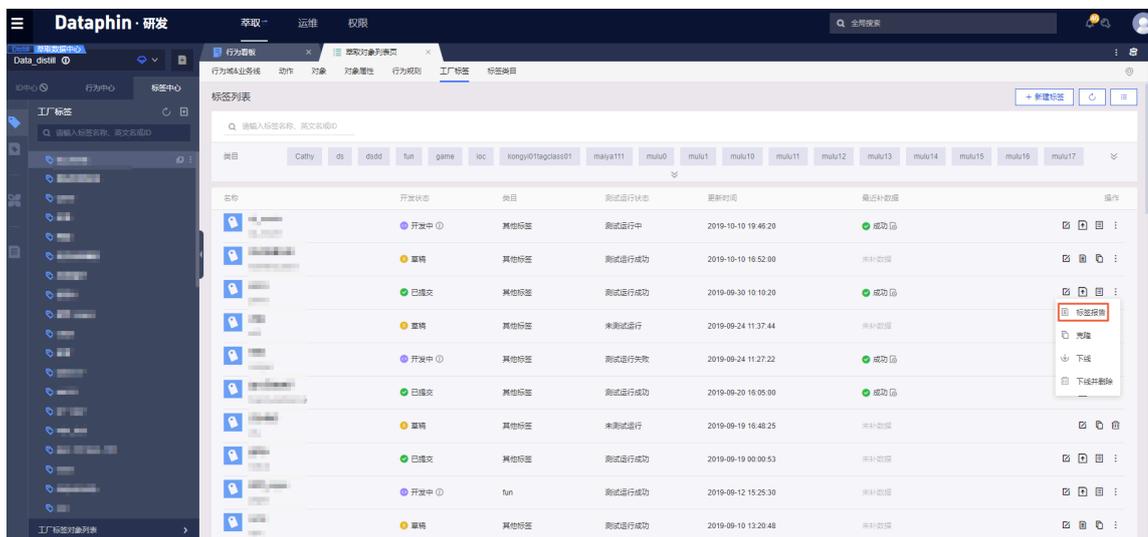
 - 覆盖用户范围: 0% ~ 100%

- 标签结果表
 - 标签结果表: [Data_Distill_game2](#)

- 依赖关系
 - 行为: 7968
 - 行为规则: 7969

- 在数据萃取研发页面，单击左侧导航栏下方的工厂标签对象列表，进入萃取对象列表页面。

在工厂标签页面，若您查看的标签的开发状态为已提交或开发中，可通过单击操作列下的  后，单击标签报告，即可进入标签报告详情页面。



若您查看标签的开发状态为草稿，可通过单击操作列下的  后，即可进入标签报告详情页面。

面。



2. 在标签报告详情页面，查看标签报告的信息。

报告生成时间: -

基本信息

标签英文名 brand_preference 标签名称 [模糊]

命中ID数 - 结果表表名 Data_Distill_preference

标签数据来源

行为 [模糊]

对象属性 [模糊]

时间跨度 30天

行为 [模糊]

对象属性

时间跨度 30天

标签分布

仅展示覆盖用户百分比最多的前 30 个标签值及分布

标签分布获取失败，请检查来源表是否有数据

结果表数据抽样

抽样展示 5 条标签结果表数据

| | 指标 | 记录1 | 记录2 | 记录3 | 记录4 | 记录5 |
|---|--------------|-----|-----|-----|-----|-----|
| 1 | ID_type ⓘ | - | - | - | - | - |
| 2 | ID_value ⓘ | - | - | - | - | - |
| 3 | confidence ⓘ | - | - | - | - | - |
| 4 | prefer ⓘ | - | - | - | - | - |
| 5 | count ⓘ | - | - | - | - | - |

克隆标签

1. 您可以通过以下3种方式的任何一种，进入克隆标签页面：

- 在数据萃取研发页面，单击左侧导航栏中的，将鼠标悬浮在需克隆标签后的，单击克隆，进入克隆标签页面。



- 在数据萃取研发页面，单击左侧导航栏中您需克隆的标签，右侧弹出该标签的属性框。在属性框中单击后，单击克隆，即可进入克隆标签页面。

属性查看 > 工厂标签属性

vbdhnsn 已提交 其他标签 公开

bxhdjsnm

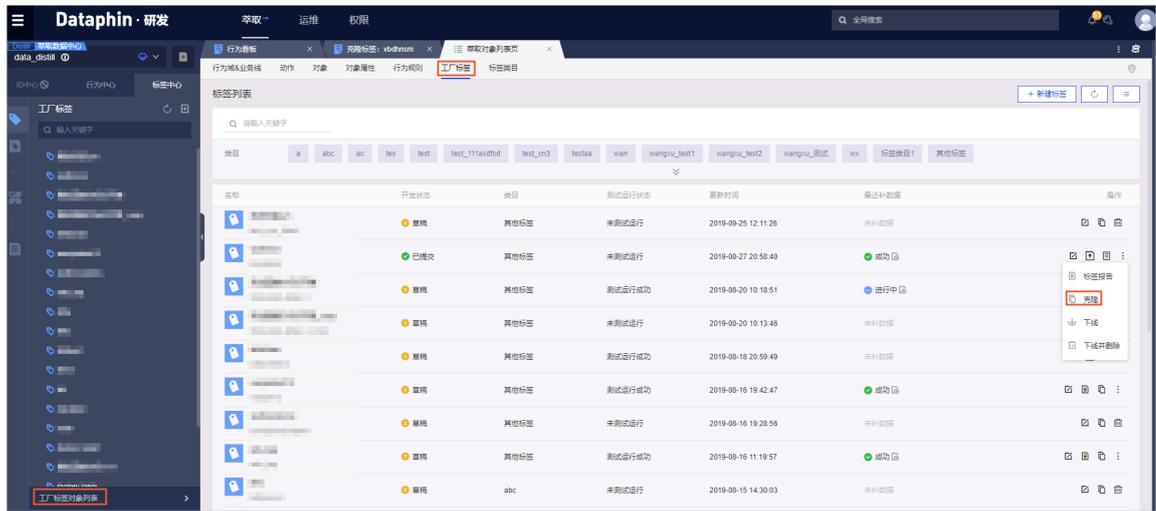
编辑 查看标签报告 补数据 ⋮

- 数据来源及标签值
 - 行为: 域2-业务线21-动作1-对象1
 - 筛选条件: --
 - 打标方式: 直接打标
 - 标签值: 1
 - 时间跨度: 30
- 标签生成规则
 - 对用户返回标签值个数 (最多): 2
 - 衰减时间周期: 7
 - 时间衰减曲线: 平滑: 行为对标签值的影响不随时间衰减

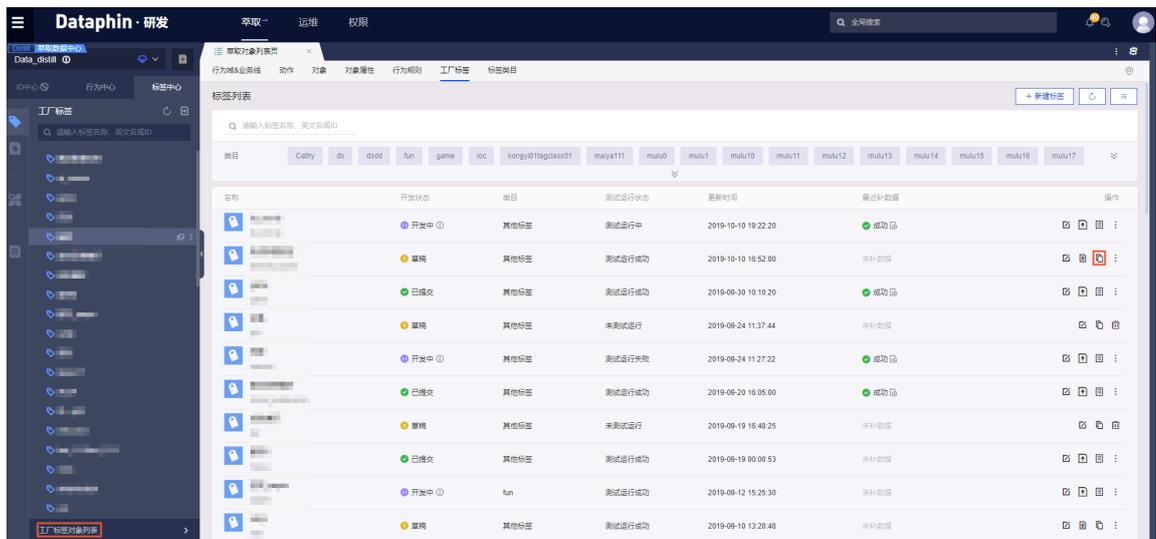
| 行为 | 权重 |
|------------------|----|
| 域2-业务线21-动作1-对象1 | 1 |

 - 覆盖用户范围: 0% ~ 100%
- 标签结果表
 - 标签结果表: [Data_Distill_cbbdhjsk](#)
- 依赖关系
 - 行为: 1201

- 在数据萃取研发页面，单击左侧导航栏下方的工厂标签对象列表，进入萃取对象列表页面。在工厂标签页面，若您需克隆的标签的开发状态为已提交或开发中，可通过单击操作列下的  后，单击克隆，即可进入克隆标签页面。



若您需克隆标签开发状态为草稿，可通过单击操作列下的 ，进入克隆标签页面。



- 在克隆标签页面，可查看克隆标签的详细配置信息，您可根据实际业务需求是否需要修改配置参数。若需要修改，请参见[编辑工厂标签](#)进行修改。

下线标签

1. 您可以通过以下3种方式中的任何一种方式，进入下线提示页面：

- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需下线标签后的 ，单击下线，进入下线提示页面。



- 在数据萃取研发页面，单击左侧导航栏中您需下线的标签，右侧弹出该标签的属性框。在属性框中单击  后，单击下线，即可进入下线提示页面。

属性查看 > 工厂标签属性

wanjunTest1 已提交 其他标签 公开

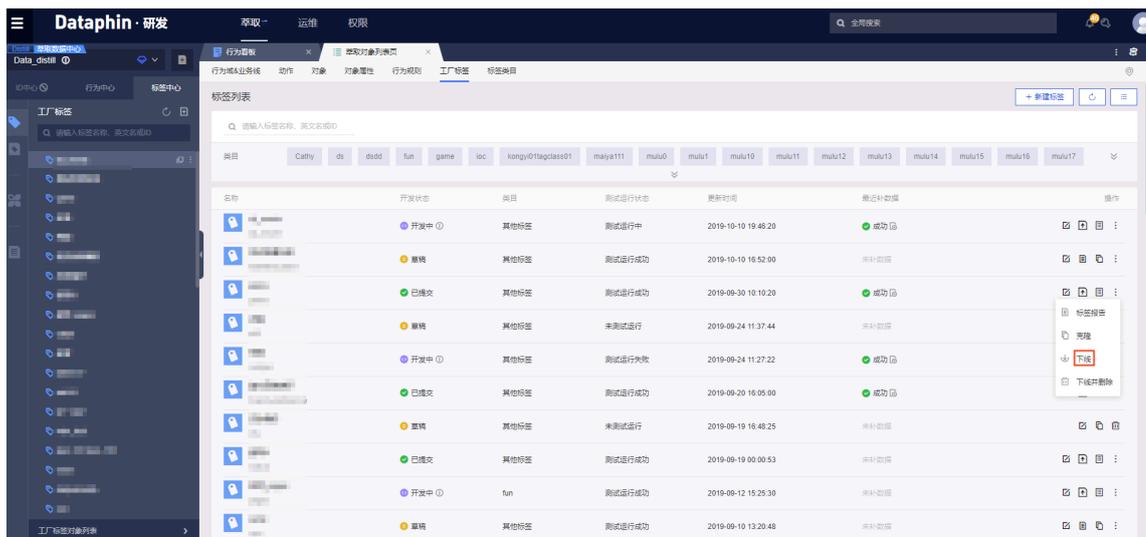
编辑 查看标签报告 补数据 ⋮

- 数据来源及标签值
 - 行为：互联网12-教育-浏览-商品
 - 筛选条件：--
 - 打标方式：属性值映射打标
 - 标签值：共有对象属性：xczvzx；映射方式：设计映射规则
 - 时间跨度：30
- 标签生成规则
 - 对用户返回标签值个数（最多）：1
 - 衰减时间周期：7
 - 时间衰减曲线：平滑：行为对标签值的影响不随时间衰减

| 行为 | 权重 |
|----------------|----|
| 互联网12-教育-浏览-商品 | 1 |

 - 覆盖用户范围：0% ~ 100%
- 标签结果表
 - 标签结果表：[Data_Distill_wanjun1](#)

- 在数据萃取研发页面，单击左侧导航栏下方的工厂标签对象列表，进入萃取对象列表页面。在工厂标签页面，您可通过单击操作列下的 ⋮ 后，单击下线，即可进入下线提示页面。



2. 在弹出的提示框中，填写备注信息，并单击确定并提交。下线成功后标签的开发状态更改为草稿。



说明:

仅限开发状态为已提交和开发中的标签支持下线操作。

下线并删除标签

1. 您可以通过以下3种方式中的任何一种方式，进入下线并删除提示页面：

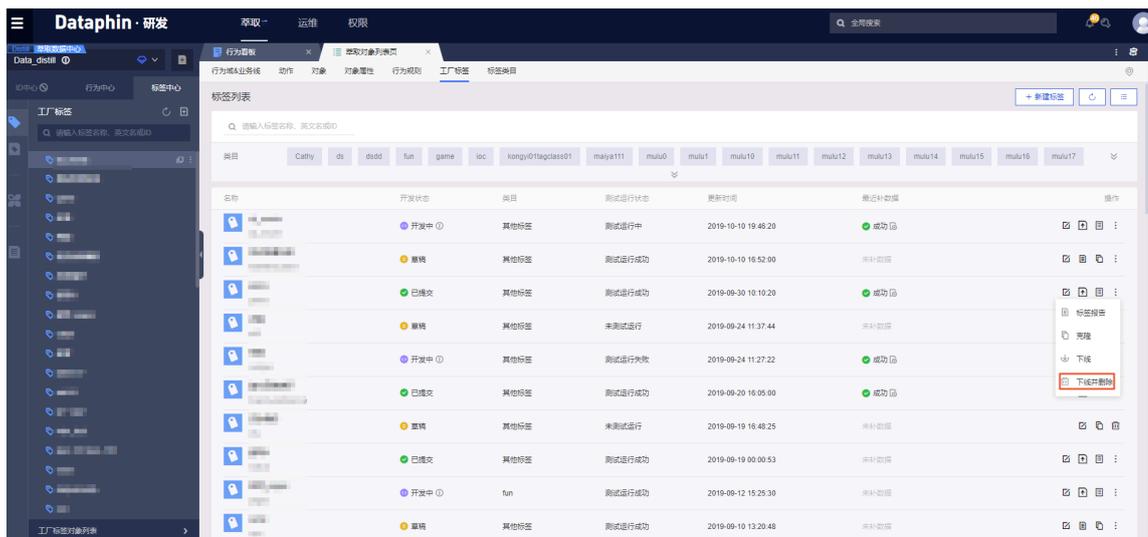
- 在数据萃取研发页面，单击左侧导航栏中的 ，将鼠标悬浮在需下线并删除标签后的 ，单击下线并删除，进入下线并删除提示页面。



- 在数据萃取研发页面，单击左侧导航栏中您需下线并删除的标签，右侧弹出该标签的属性框。在属性框中单击  后，单击下线并删除，即可进入下线并删除提示页面。



- 在数据萃取研发页面，单击左侧导航栏下方的工厂标签对象列表，进入萃取对象列表页面。在工厂标签页面，您可通过单击操作列下的  后，单击下线并删除，即可进入下线并删除提示页面。



2. 在弹出的提示框中，填写备注信息，单击确定并提交。

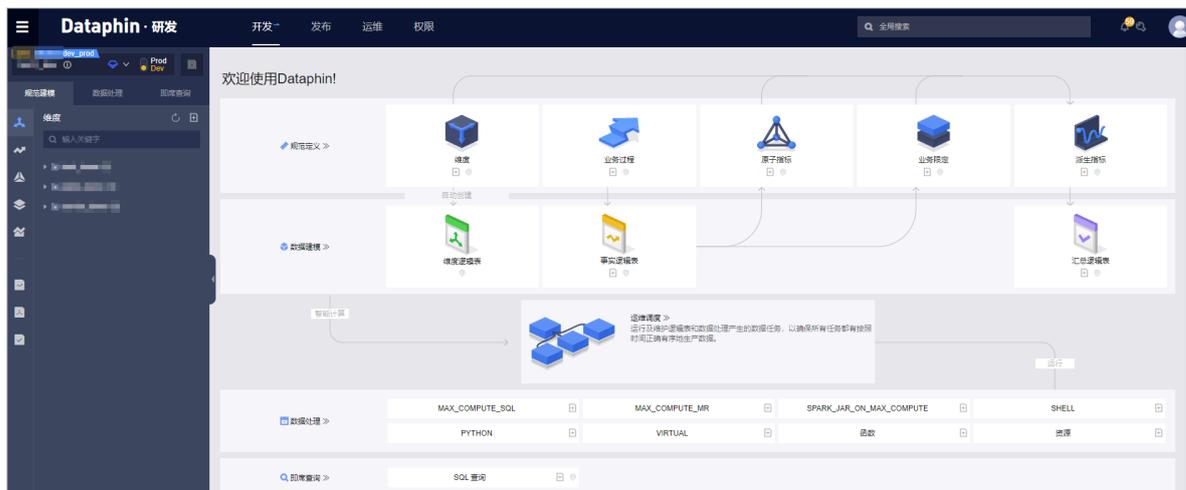
7.3.3 标签逻辑表

本文为您介绍标签逻辑表列表页面内容。

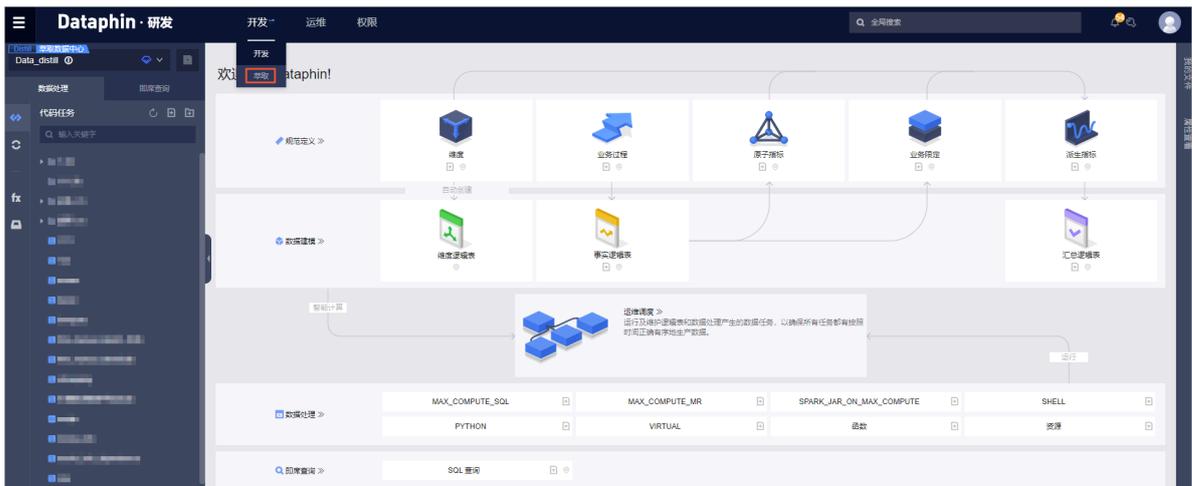
标签逻辑表由系统在萃取项目初始化时自动创建。以id_type和id_value为统计粒度，汇总所有已提交的工厂标签和物理表的手工标签（非标签工厂开发）。英文名固定为Ads_distill_标签逻辑表，不可修改。

查看标签逻辑表

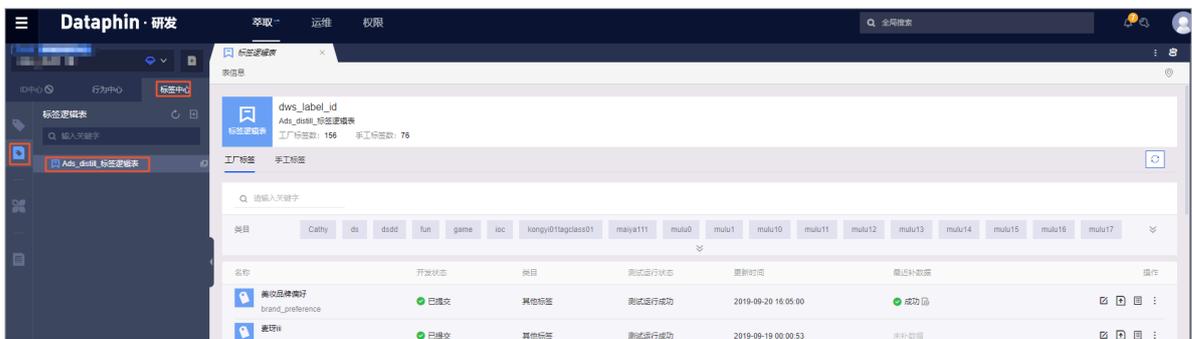
1. 登录Dataphin控制台。
2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。



3. 单击萃取，进入数据萃取研发页面。



4. 单击标签中心左侧导航栏中的 , 进入标签逻辑列表页面。

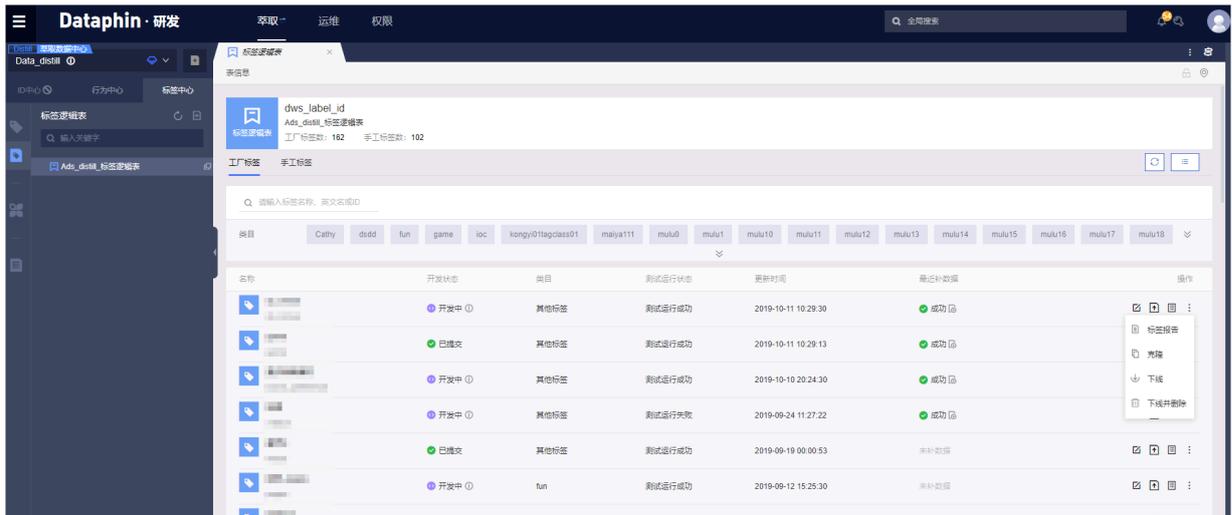


说明:

- 标签逻辑表列表页面分为工厂标签和手工标签。
- 所有已提交和开发中的工厂标签自动上挂到标签逻辑表，同时支持从物理表引入通过编写代码等方式已开发好的手工标签。

管理工厂标签

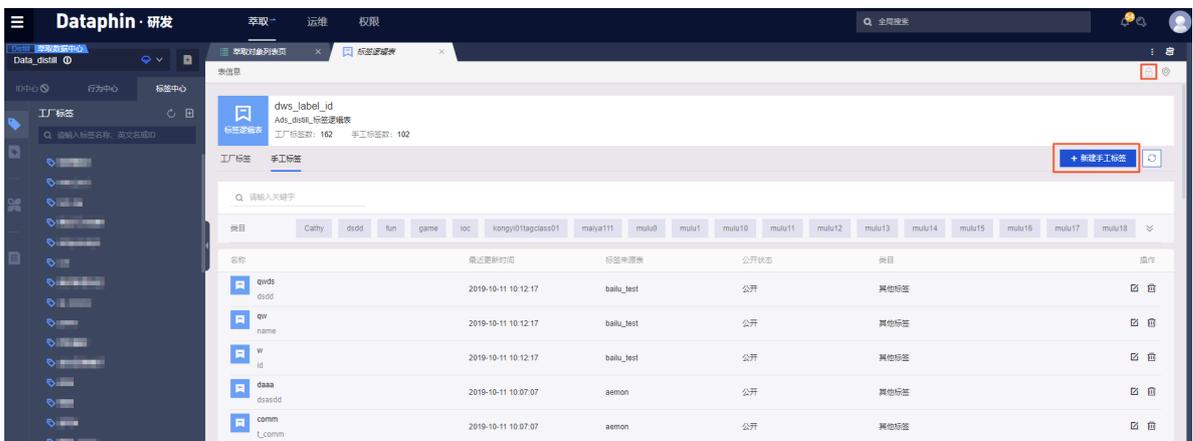
参见[标签逻辑表](#)，进入标签逻辑表页面。在工厂标签页面可对标签执行编辑、补数据、查看任务标签等操作。详细请参见[管理工厂标签](#)。



管理手工标签

新建手工标签

1. 参见**标签逻辑表**，进入**标签逻辑表**页面，单击**手工标签**，进入**手工标签**页面。如果当前页被其他用户锁定，请单击页面右上方进行偷锁，偷锁后单击**新建手工标签**，进入**新建手工标签**页面。



说明:

在手工标签页面，默认初始化系统锁定，若您未偷锁不能新建手工标签。

2. 在新建手工标签页面，根据页面提示完成参数的配置，并单击 关联并下一步。

3. 在 添加字段页面，根据页面提示完成参数的配置。并单击 确定。

选择左侧来源表的 字段及描述，单击 ，将左侧的 字段及描述同步到右侧的 新增

列表中。标签英文名为字段的名称，标签名称为字段的描述。

同时您也可以单击 新增列表中 操作列下的 ，删除不需的字段。



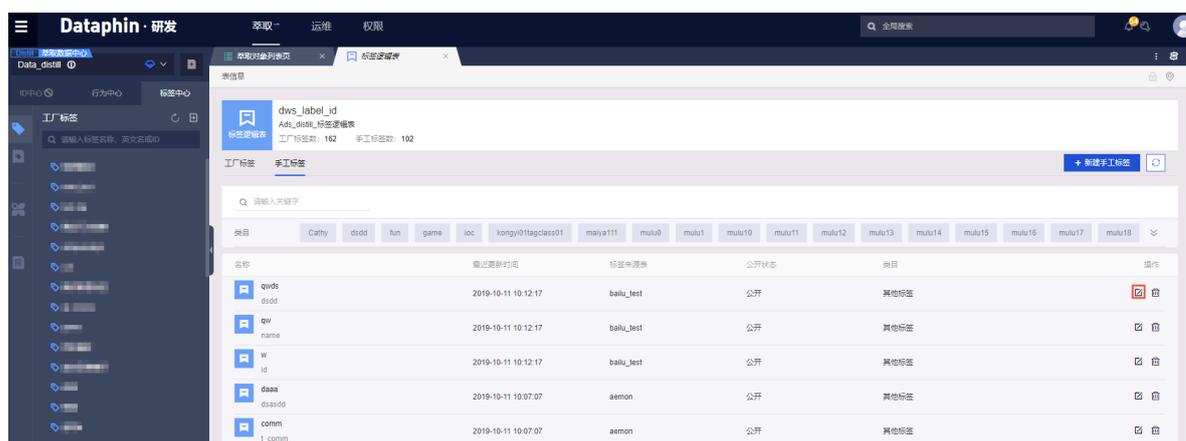
说明:

- 标签英文名是由数字、字母、下划线组合。但不支持纯数字格式。
- 标签名称是由汉字、字母、数字、下划线、-组合。但不支持纯数字格式。

4. 在弹出的提交备注页面填写备注信息。并单击确定并提交，即可完成手工标签的创建。

编辑手工标签

1. 在手工标签列表页面，单击需编辑的标签所在行的操作列下，单击，进入编辑页面。



2. 在编辑手工标签页面，根据页面提示可修改标签英文名、标签名称、描述、所属类目、公开状态。

编辑手工标签 ✕

*** 标签英文名**

*** 标签名称**

描述

0/128

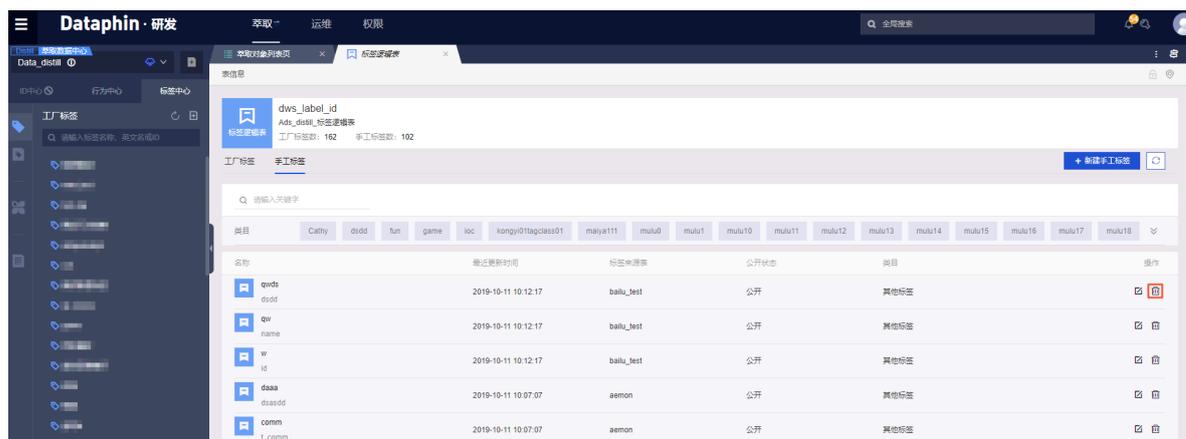
所属类目

公开状态

3. 在弹出的提交备注页面填写备注信息。并单击确定并提交，即可完成手工标签的编辑。

删除手工标签

1. 在手工标签列表页面，单击需删除的标签所在行的操作列下，单击，进入删除提示页面。



2. 在删除提示页面，填写备注信息后，单击确认并提交。



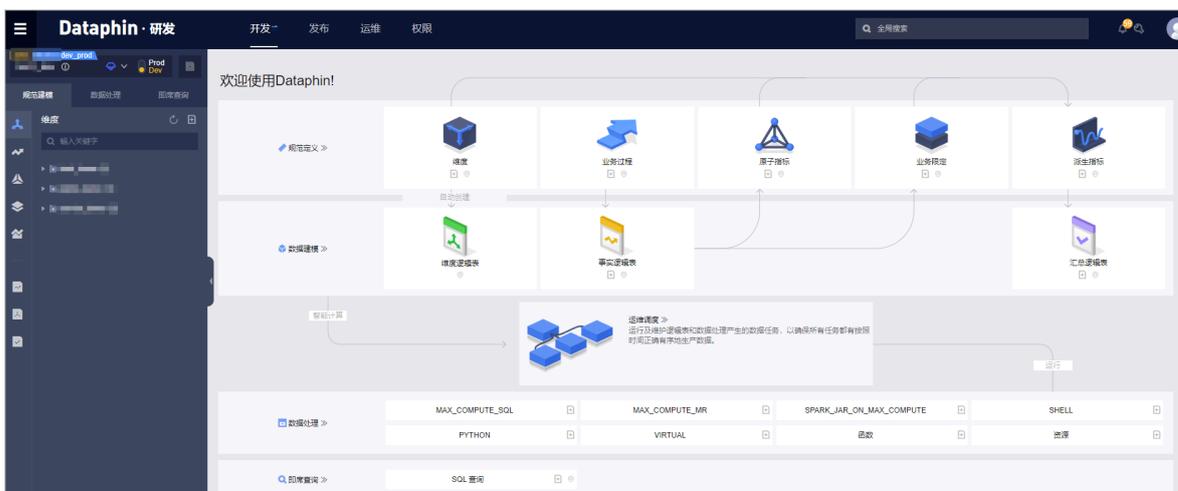
7.3.4 管理标签类目

本文为您介绍如何进行对标签类目的新建、编辑和删除。

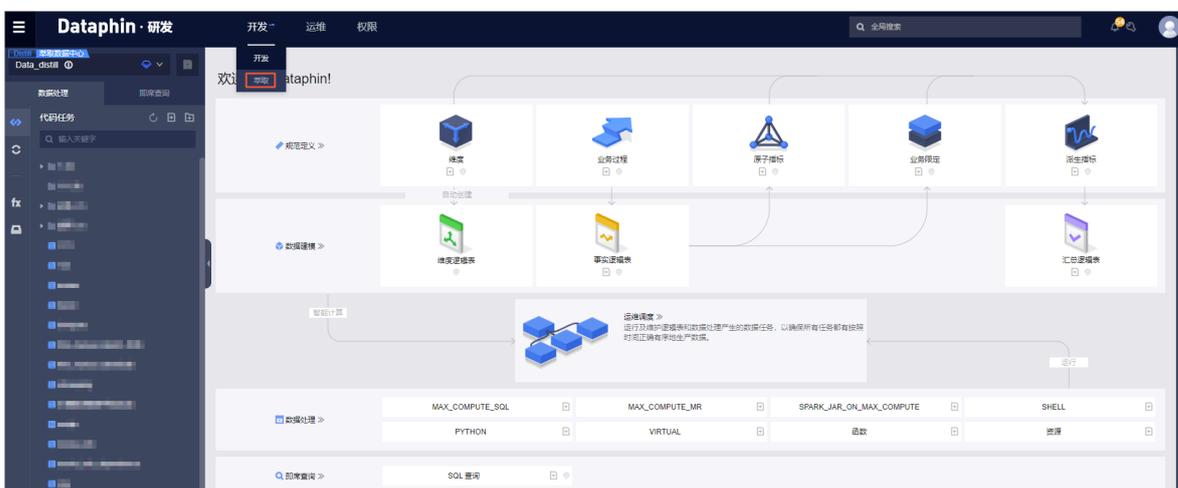
新建标签类目

1. 登录Dataphin控制台。

2. 在Dataphin首页，单击顶部菜单栏的研发，进入研发页面。



3. 单击萃取，进入数据萃取研发页面。



4. 您可以通过以下4种方式中的任何一种方式，进入新建类目页面。

- 在数据萃取研发页面，单击左侧导航栏中的后，单击，进入新建类目页面。



- 在数据萃取研发页面，单击项目名称后，选择标签中心 > 标签类目，进入新建类目页面。

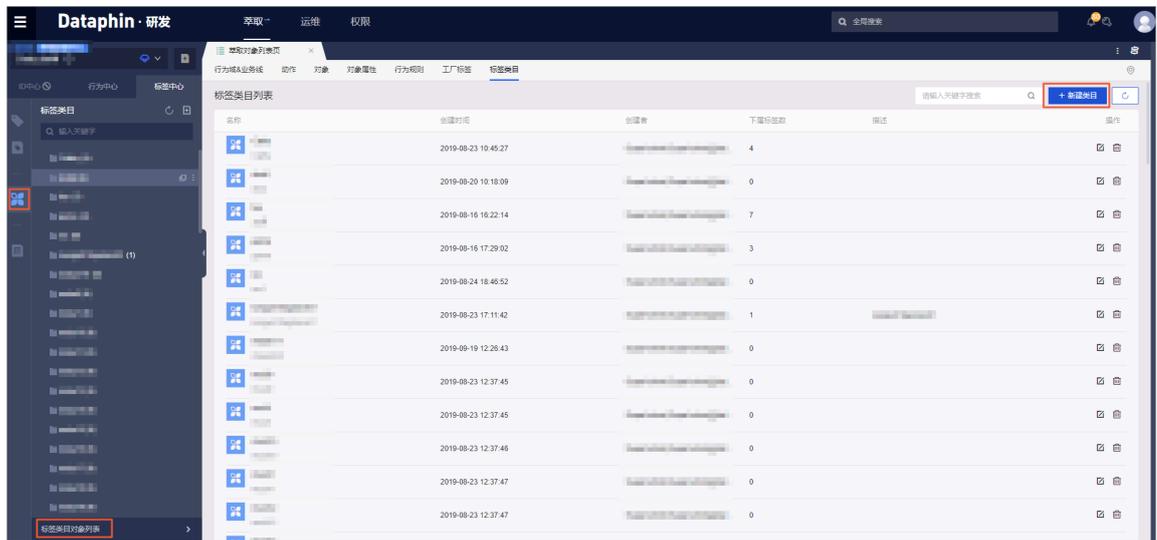
面。



- 在萃取首页，单击标签类目下的，进入新建类目页面。



- 在数据萃取研发页面，单击左侧导航栏下的标签类目对象列表，进入萃取对象列表页面。单击页面右上方的新建类目，进入新建类目页面。



5. 在新建类目页面，根据页面提示完成配置参数，并单击确定。

新建类目 ✕

* 类目名称

* 类目英文名

描述

0/128 ↗



说明:

- 类目名称由汉字、字母、数字、下划线或-组成，但不能完全由数字组合。
- 类目英文名由字母、数字或下划线组成，但不能完全由数字组合。

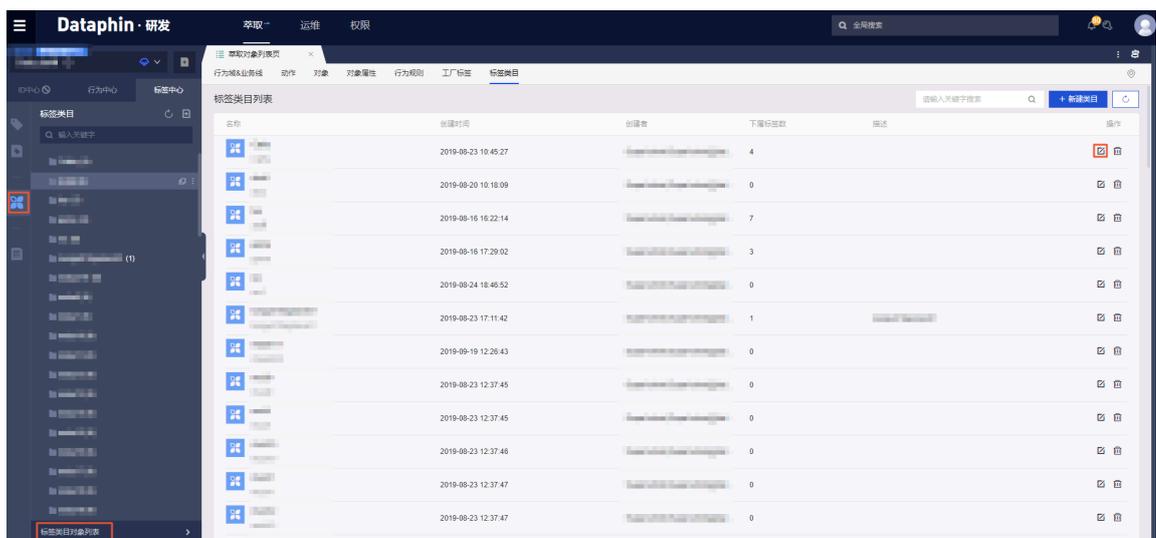
编辑标签类目

1. 您可以通过以下任何一种方式，进入编辑类目页面。

- 在标签分类列表页，选择需编辑的类目，将鼠标悬停在  上，单击编辑，进入编辑标签类目页面。



- 在数据萃取研发页面，单击左侧导航栏下的标签类目对象列表，进入萃取对象列表页页面。单击需编辑类目所在行的操作列下的 ，进入编辑类目页面。



2. 在编辑类目页面，根据页面提示完成参数修改，并单击确定。

编辑类目 ×

* 类目名称

* 类目英文名

描述

0/128

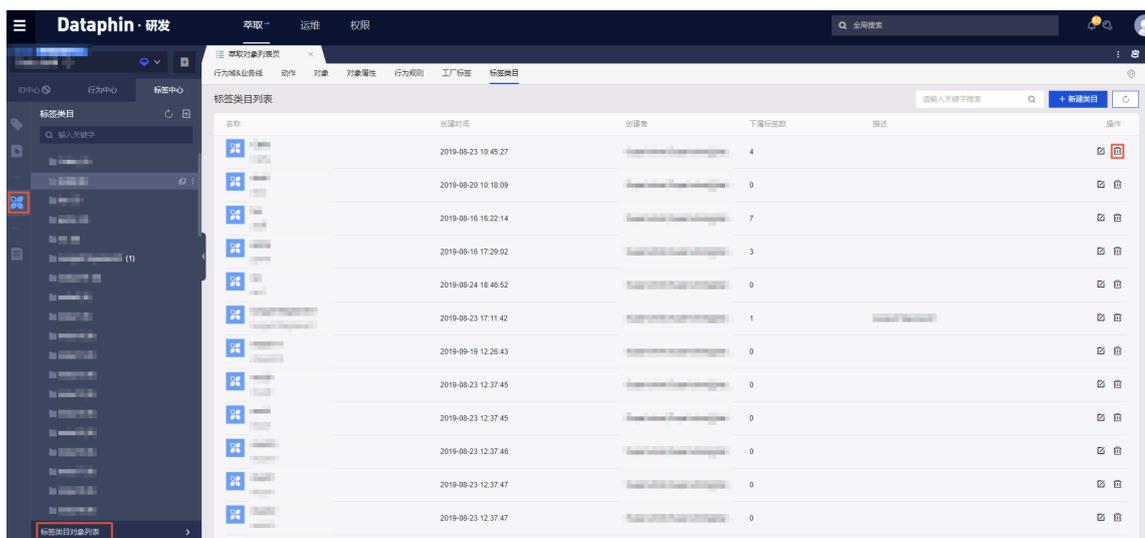
删除标签类目

1. 您可以通过以下2任何一种方式，进入删除提示页面。

- 在标签分类列表页，选择需编辑的类目，将鼠标悬停在  上，单击删除，进入删除标签类目提示页面。



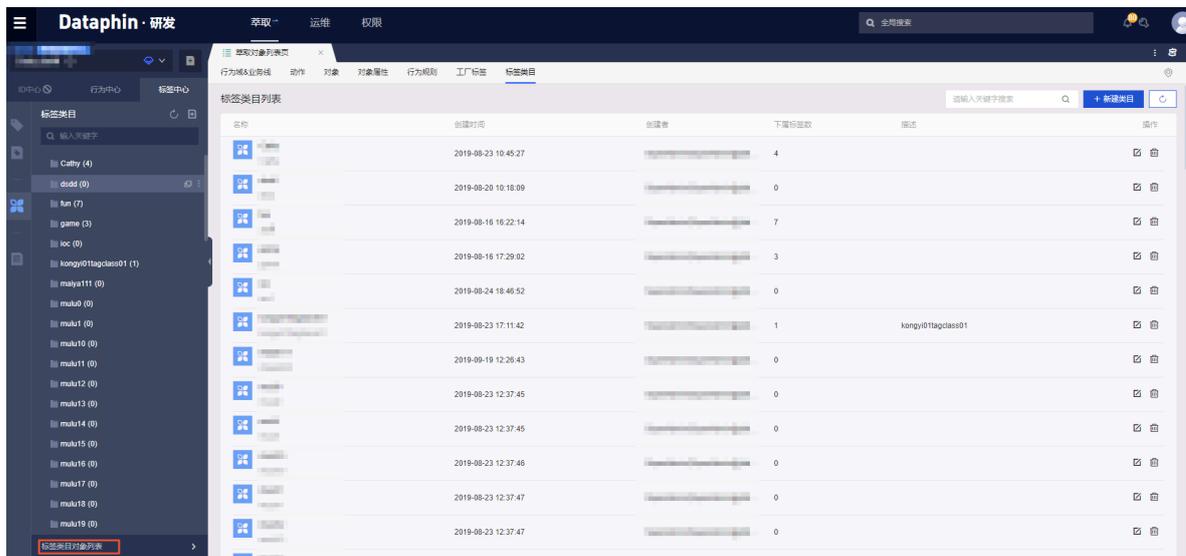
- 在数据萃取研发页面，单击左侧导航栏下的标签类目对象列表，进入萃取对象列表页页面。单击需编辑类目所在行的操作列下的 ，进入删除标签类目提示页面。



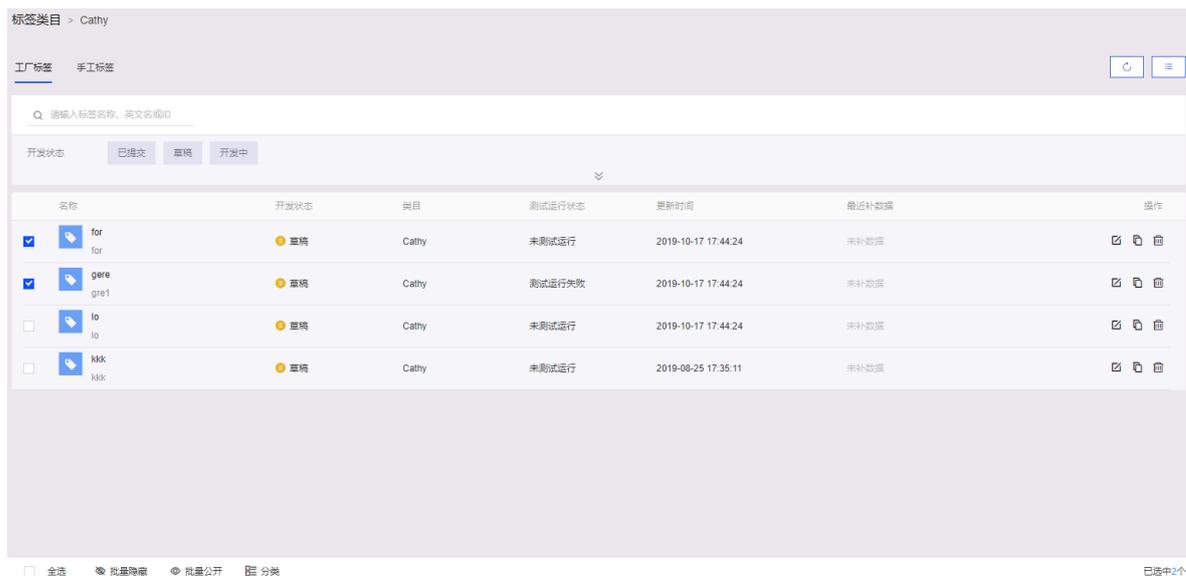
2. 在弹出的提示框中填写备注信息，并单击确认并提交，即可完成类目的删除。

类目详情页

1. 在数据萃取研发页面，单击左侧导航栏下的标签类目对象列表，进入萃取对象列表页面。

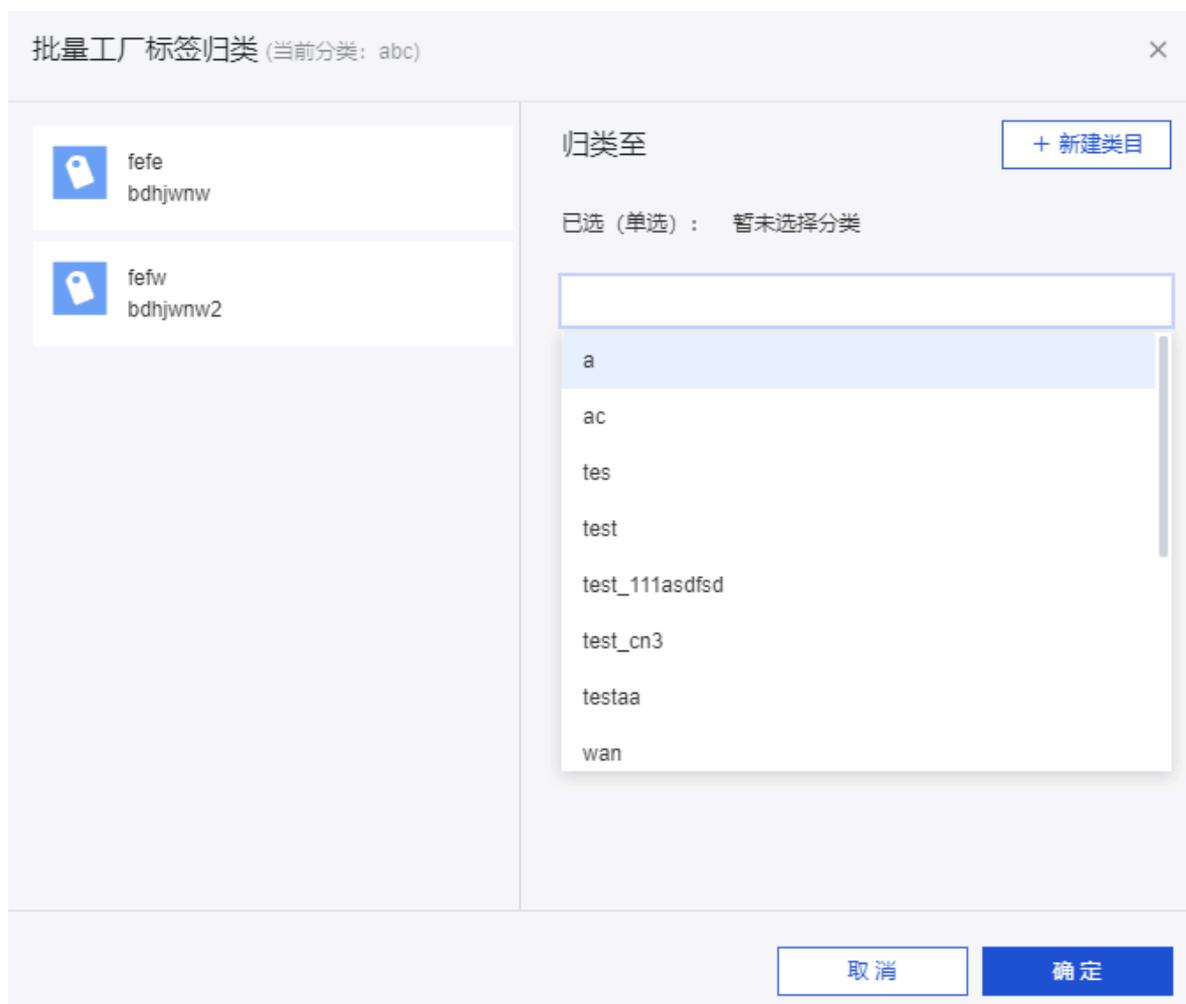


2. 单击名称列下的具体名称（此处以cathy为例），进入cathy类目页面。



3. 在cathy类目页面，选择多个类目后，单击页面下方的批量隐藏或批量公开进行批量隐藏和公开。

4. 在cathy类目页面，您可以通过单击页面下方的分类，进入批量工厂标签归类页面。



- 您选择已有的类目，将多个工厂标签移进去。
- 同时您也可以单击页面右上角的新建类目，进行新建类目。将工厂标签移动到新建的类目下。

在新建类目页面中，根据页面提示配置相关参数，配置完后单击确定并提交。

新建类目

* 类目名称

* 类目英文名

描述

0/128

当前项目，提交成功后，即更新至生产环境，请谨慎操作!

 **说明:**
类目名称和类目英文名不支持纯数字组合命名。

8 运维调度

8.1 运维调度概述

运维调度基于策略对建模研发、编码研发、数据萃取生成的代码任务进行调度与运维管控。运维调度包括全局运维、逻辑表运维、萃取运维三部分。

进入运维调度

您可以通过以下任一方式进入运维调度：

- 在Dataphin产品首页，单击顶部研发菜单，默认进入开发模块，然后切换到运维页签，即可进入运维调度页面。
- 在Dataphin产品首页，通过快速入口进入运维调度页面：

单击运维调度，或者单击调度运维后的进入，进入运维调度页面。



在Dev-Prod模式中，Dev和Prod环境中的调度任务互相独立：

- Dev环境中调度任务包括全局运维、逻辑表运维。
- Prod环境中调度任务包括全局运维、逻辑表运维、监控报警，仅Prod环境下可配置监控报警任务。

您可以通过单击页面左上角的 **Prod/Dev** 切换进入到该项目对应环境的运维调度页面。例如，当前处于开发环境下项目test_dev的运维调度页面，单击 **Prod/Dev** 切换为Prod生产环境，就会进入生产环境下test项目的运维调度页面。

在Basic模式中，运维调度包括萃取运维、全局运维、监控报警模块。



说明:

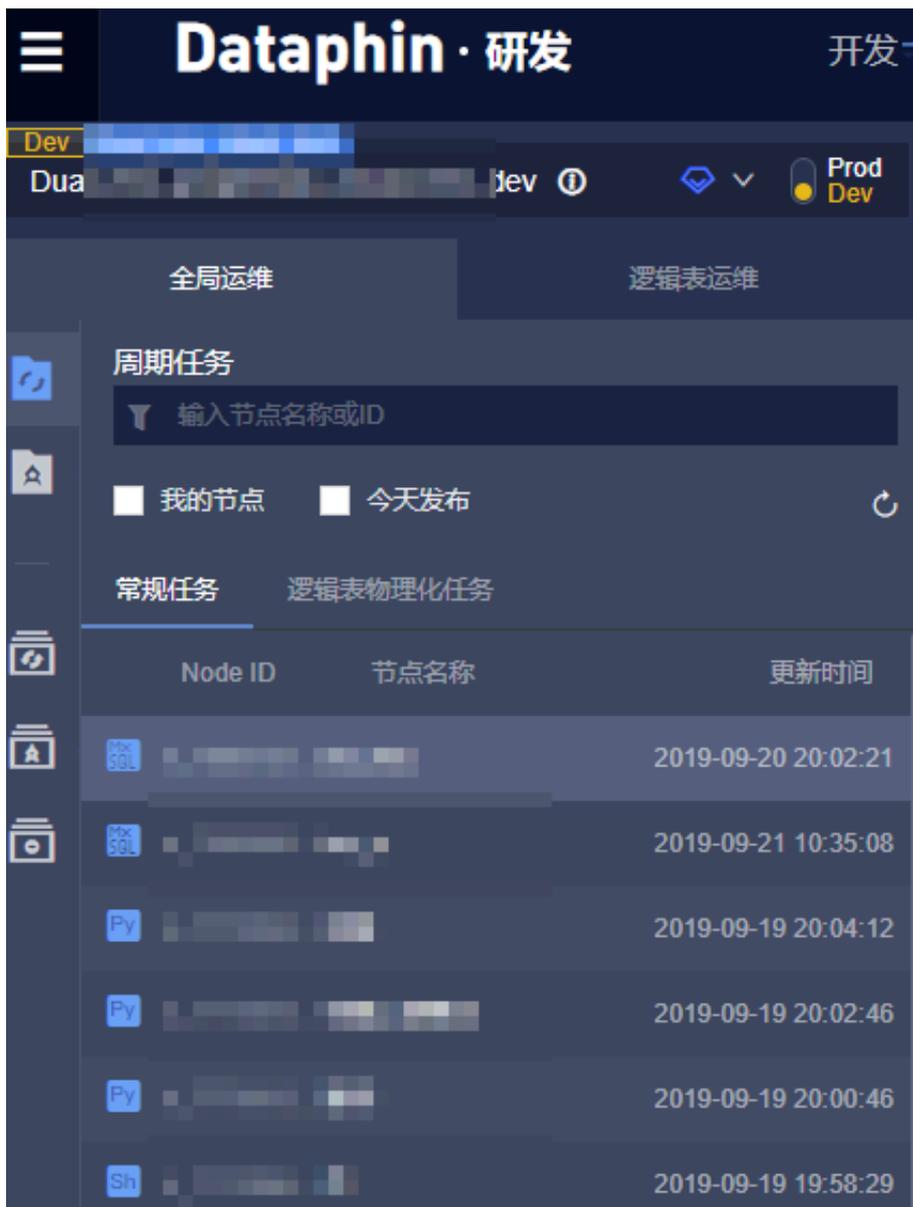
数据萃取项目下包含全局运维、萃取运维和监控报警模块，没有逻辑表运维模块。

Dev项目的调度任务不支持自动运行，您可以手动运行执行调度生成实例。Prod项目和Basic项目的调度任务将按照调度配置正常生成与运行。

Prod和Basic环境的调度实例正常生成与运行，可以配置任务监控规则并报警。



运维调度将任务、实例、逻辑表划分成了不同的类型，鼠标悬停于左侧图标时会展示分类名称。当您单击要查看的任务或实例类型时，会跳转至对应任务或实例的运维页面。进入运维调度后，默认展示周期任务 > 常规任务。



任务

任务是指所有代码脚本等提交后发布上线的、可周期/手动触发运行的对象，可以被调度，又称为节点。调度运维中将节点分为两种：

- 周期节点，这类节点会配置调度周期，Dataphin的作业调度系统会周期性地调度这类节点，每次节点被调度时都会生成对应的周期实例。
- 手动节点，这类节点不会被系统自动触发，用户可以根据需要自己手动运行对应的节点，每次运行手动节点会生成对应的手动实例。

调度执行顺序，会根据配置的调度周期、依赖关系、任务优先级三者共同决定：

- 每天晚上11点前更新节点的代码或配置，第二天生成的实例方会生效。
- 实例运行时，会读取最新的节点代码和调度配置。

- 调度支持的补数据时间范围是1970年-2099年，调度实例至多保留2周的实例信息。

实例

节点每运行一次生成一个实例。实例有3种不同的生成方式：

- 周期实例，这类实例是由周期节点自动被系统调度触发生成，具有周期性。
- 补数据实例，这类实例是通过用户对节点进行补数据操作时产生。
- 手动实例，这类实例通过运行手动节点生成。

逻辑表

逻辑表运维包含2个模块：

- 逻辑表任务，用于从逻辑表视角切入，为用户展现逻辑表内部任务关系。
- 逻辑表实例，用于查看已运行的逻辑表任务包含的节点实例及其状态。

萃取运维

萃取运维包含2个模块：任务和实例。萃取运维是从业务视角出发管理行为规则和标签相关任务和实例。

公测期说明

公测期间资源消耗说明如下：

- 每个租户分配最多40c160g的调度资源，当同步任务量大或者同步的数据量很大的情况下，可能同步速度较慢。
- 消耗资源的同步任务、调度等的调度资源费用，暂不做额外收费。



说明：

40c160g表示您可以使用的ECS资源为40个CPU Core + 160GB内存。ECS产品详情请参见[#unique_95](#)。

8.2 全局运维

8.2.1 周期任务

本文为您介绍周期任务功能模块。

进入周期任务页面

1. 进入运维调度，默认展示周期任务页面。



说明：

为区分您自己创建的任务和系统自动生成的任务，将周期任务分为常规任务和逻辑表物理化任务。例如，您自己创建的同步任务节点，可以在常规任务里查看、运维。

The screenshot shows the Dataphin interface with the '运维' (Operations) tab selected. On the left, there is a sidebar with '周期任务' (Periodic Tasks) and a search bar. Below it, a table lists tasks with columns for 'Node ID', '节点名称' (Node Name), and '更新时间' (Update Time). The table is sorted by update time in descending order. On the right, a DAG diagram is displayed, showing a flow from 'virtual_root_node_7...' to 'ins_hou'. A tooltip is visible over the 'ins_hou' node, providing details such as '节点id: n_780273271376510976', '节点: ins_hou', '优先级: 最高优先级', and '所在项目: Dual_env_proprietary_cloud'.

| Node ID | 节点名称 | 更新时间 |
|------------|------------|---------------------|
| [Redacted] | [Redacted] | 2019-09-20 20:02:21 |
| [Redacted] | [Redacted] | 2019-09-21 10:35:08 |
| [Redacted] | [Redacted] | 2019-09-19 20:04:12 |
| [Redacted] | [Redacted] | 2019-09-19 20:02:46 |
| [Redacted] | [Redacted] | 2019-09-19 20:00:46 |
| [Redacted] | [Redacted] | 2019-09-19 19:58:29 |
| [Redacted] | [Redacted] | 2019-09-19 19:57:24 |
| [Redacted] | [Redacted] | 2019-09-19 19:56:19 |
| [Redacted] | [Redacted] | 2019-09-19 15:55:40 |
| [Redacted] | [Redacted] | 2019-09-18 15:22:45 |
| [Redacted] | [Redacted] | 2019-09-18 15:21:25 |
| [Redacted] | [Redacted] | 2019-09-18 15:20:08 |
| [Redacted] | [Redacted] | 2019-09-18 15:19:02 |

2. 进入周期任务页面后，会按照更新时间倒排展示任务，并自动选择首个任务。

- 左侧功能区，主要用于搜索和展示任务列表。默认显示节点类型（以图标展示）、节点ID、节点名称及其更新时间，所有节点搜索的结果均按照更新时间倒排。
- 右侧的DAG图展示区，默认以当前选中节点为中心，展开其上下游各一层的节点DAG关系图。在DAG图的右上角有刷新按钮，单击可以查看最新的节点间的关系。同时，您可以根据节点ID或名称在当前DAG图中展开搜索，从而精准定位DAG图中的节点。在DAG图的下方有工具栏，您可以缩放整个DAG图的大小，调整默认展开的层数，也可以全屏展示DAG图。您还可以直接拖动节点，调整DAG图的位置。选中某个节点后，DAG图的右下方会展示该节点的详细信息。

搜索和过滤

· 搜索和过滤周期任务

在左侧功能区上方的搜索框内，可以输入节点名称或ID来搜索节点，同时可以框选我的节点和今天发布以快速过滤责任人为当前用户的节点和当天发布的节点。在搜索的时候，节点ID需要精准匹配，而节点名称则支持模糊匹配。

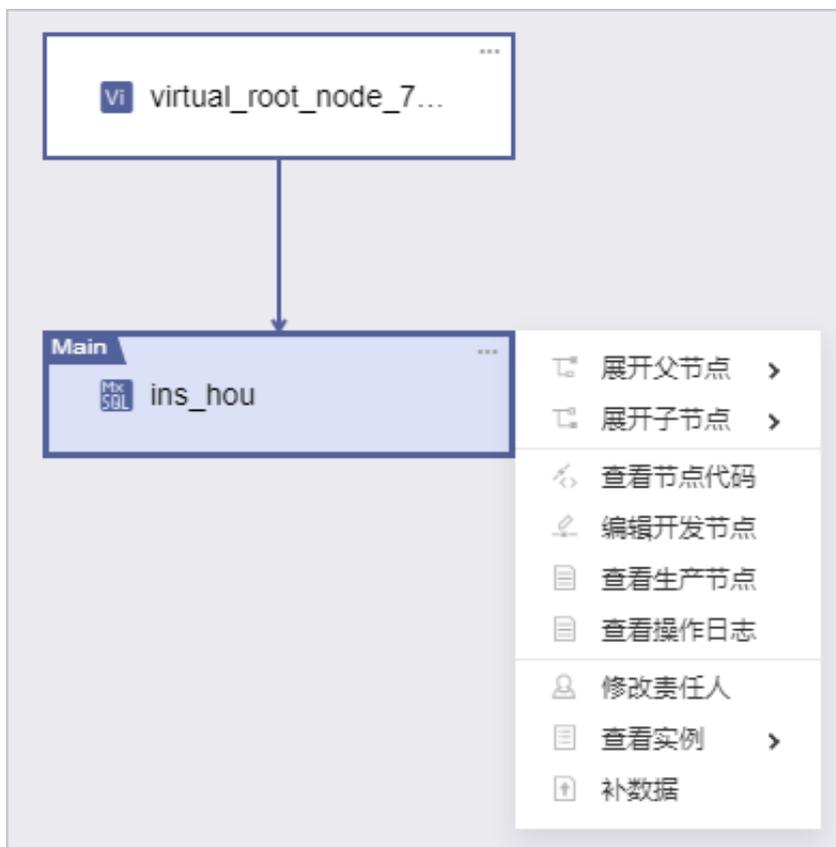
· 高级搜索过滤周期任务

如下图所示，您可以单击搜索框左侧“漏斗”图标进行高级搜索过滤。目前支持根据节点的责任人、任务类型进行过滤。



右键操作项

如下图所示，在DAG图上右键单击节点，展示该节点对应的可操作项。在有权限的情况下，可以进行：查看节点的上下游信息、查看节点代码、编辑节点、查看节点操作日志、修改节点责任人、查看节点实例、补数据等操作。



说明:

执行补数据操作时，业务日期的范围为1970~2099年，并且单次同步的最长时间周期为1年，不得超出该范围。

8.2.2 手动任务

本文为您介绍手动任务功能模块。

进入手动任务页面

1. **进入运维调度**，单击左侧菜单栏的**手动任务**，进入手动任务页面，如下图所示。



2. 当您进入手动任务页面后，会按照更新时间倒排展示任务。：

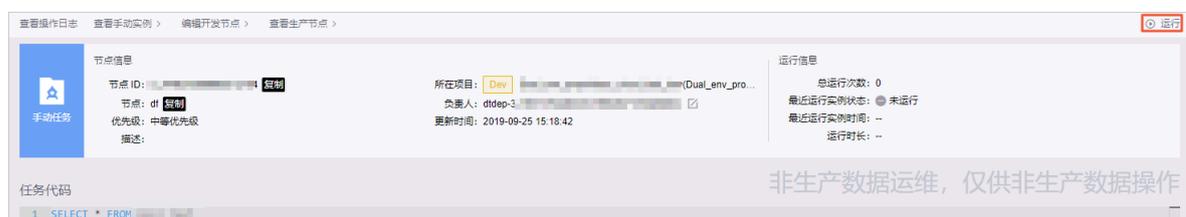
- 与周期任务页面类似，左侧功能区主要用于搜索和展示任务列表。默认显示节点类型（以图标展示）、节点ID、节点名称及其更新时间，所有节点搜索的结果均按照更新时间倒排。
- 手动任务的右侧页面则与周期任务有很大差别：右侧功能区主要用于展现手动节点的具体信息以及支持的操作项。右侧功能区上方有节点的信息展示栏，会展示当前节点的最近一次手动实例的运行状态，如果该手动节点尚未运行过，则会显示未运行。右侧功能区中间展示了可以对当前节点执行的各个操作项，包括运行、查看操作日志、查看实例，以及编辑节点等。右侧功能区的主体部分则用于展示当前节点的任务代码。

搜索和过滤

手动任务的搜索和过滤与周期任务的操作相同，请参见[周期任务](#)。

运行

1. 如下图所示，单击右侧功能区中间的**运行**按钮。



2. 进入手动节点运行编辑界面。

运行

实例名称: test_NJ01_2019-03-18_1552891721303

业务日期: 2019-03-17

修改运行实例名称，选择要执行的业务日期，单击确定，即可创建对应业务日期的手动实例。

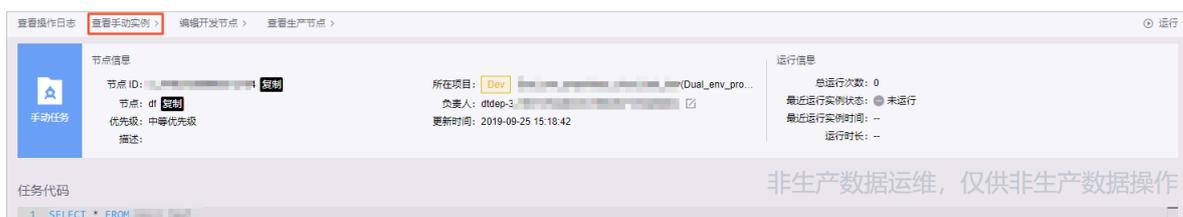


说明:

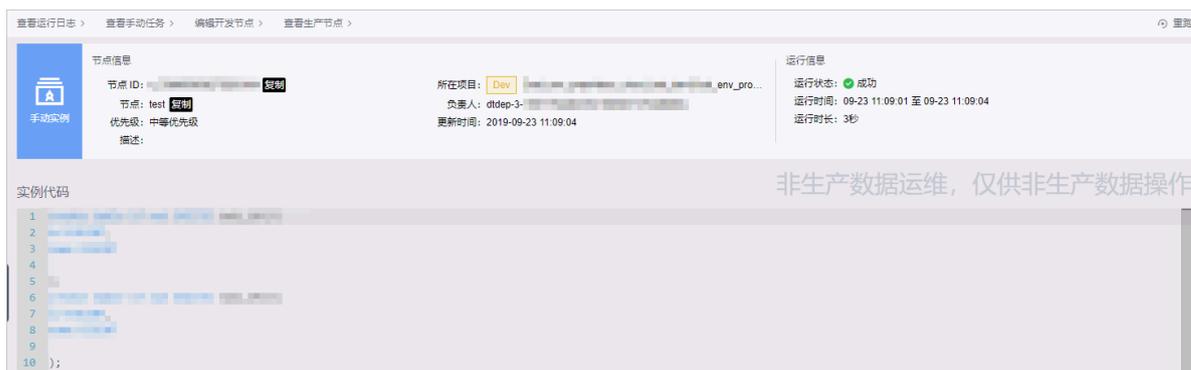
手动任务的实例名称，系统会自动生成，用户也可以自行修改。

查看实例

1. 如下图所示，单击右侧功能区中间的查看手动实例按钮。

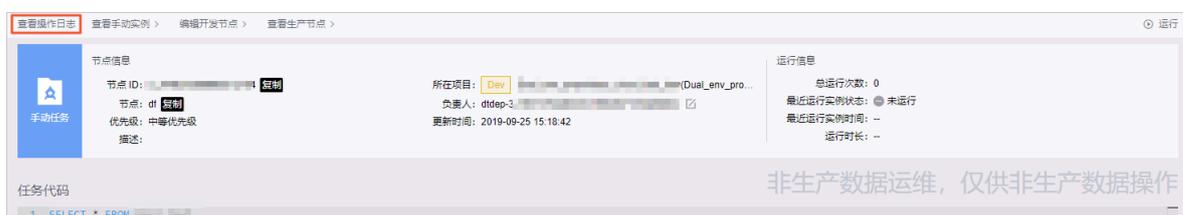


2. 当前页面刷新，跳转到手动实例页面，展示对应任务的运行实例，如下图所示。



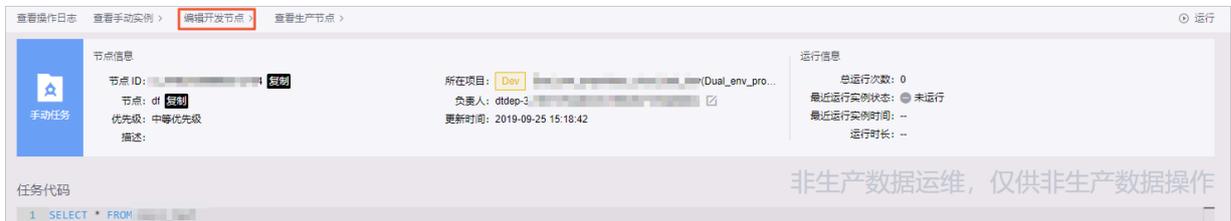
查看操作日志

1. 在手动任务页面，单击右侧功能区中间的查看操作日志按钮，进入操作日志页。



2. 用户可以看到对当前节点的操作历史，包括操作时间、操作人，以及执行的具体操作等信息。单击详情按钮，弹出操作详情页面，如下图所示。

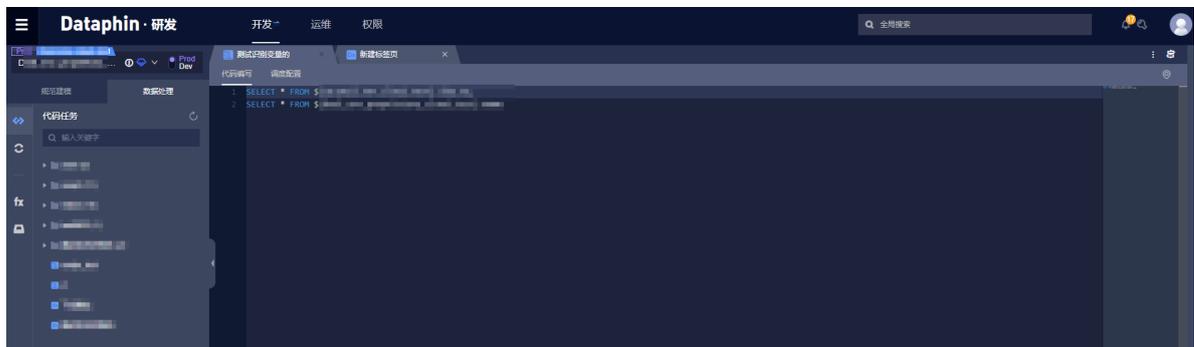
编辑节点



- 单击右侧功能区中间的编辑开发节点按钮，会进入开发环境下对应节点的编辑页面。



- 单击编辑生产节点按钮，会进入生产环境下对应节点的编辑界面。



8.2.3 周期实例

本文为您介绍周期实例功能模块。

进入周期实例页面

1. **进入运维调度**，单击左侧菜单栏的周期实例，进入周期实例页面，如下图所示。



2. 用户进入周期实例页面后，会按照实例任务的ID降序排列。

- 左侧功能区主要用于搜索和展示实例列表。
- 在右侧DAG图展示区，默认以当前选中实例为中心，展开其上下游各一层的实例DAG关系图。在DAG图的右上角有刷新按钮，单击可以查看最新的实例间的关系。同时您可以根据节点ID或名称搜索DAG图来精准定位当前DAG图中的实例。在DAG图的下方有工具栏，您可以缩放整个DAG图的大小，调整默认展开的层数，也可以全屏展示DAG图。您还可以直接拖动节点，调整DAG图的位置。选中某个节点后，DAG图的右下方会展示所选节点的详细信息。

搜索和过滤

- 搜索和过滤周期实例

在左侧功能区上方的搜索框内，您可以输入节点名称或ID来搜索实例，同时可以框选我的实例、出错实例和未完成实例以快速过滤责任人为当前用户的实例、出错的实例和未完成的实例。在搜索时，节点ID需要精准匹配，而节点名称则支持模糊匹配。默认搜索实例的业务日期是昨天，您可以选择前天、全部或单击日期选择对应的日期，对实例的搜索结果进行过滤。在下方则按照常规实例和逻辑表物理化实例对实例的搜索结果分页展示，默认是以实例的节点ID降序排列。

· 高级搜索过滤周期实例

如下图所示，单击搜索框左侧的“漏斗”图标，会出现高级搜索筛选的弹框。支持根据实例的负责人、任务类型，以及实例的状态来过滤实例的搜索结果。



右键操作项

如下图所示，右键单击DAG图中的周期实例，会展示该实例支持的所有可操作项，在有权限的情况下，可以进行：展开实例上下游、查看运行日志、查看节点代码、编辑节点、查看操作日志、终止实例运行、重跑并恢复调度、重跑下游、置成功继续调度、去除上游依赖、强制重跑、暂停等操作。

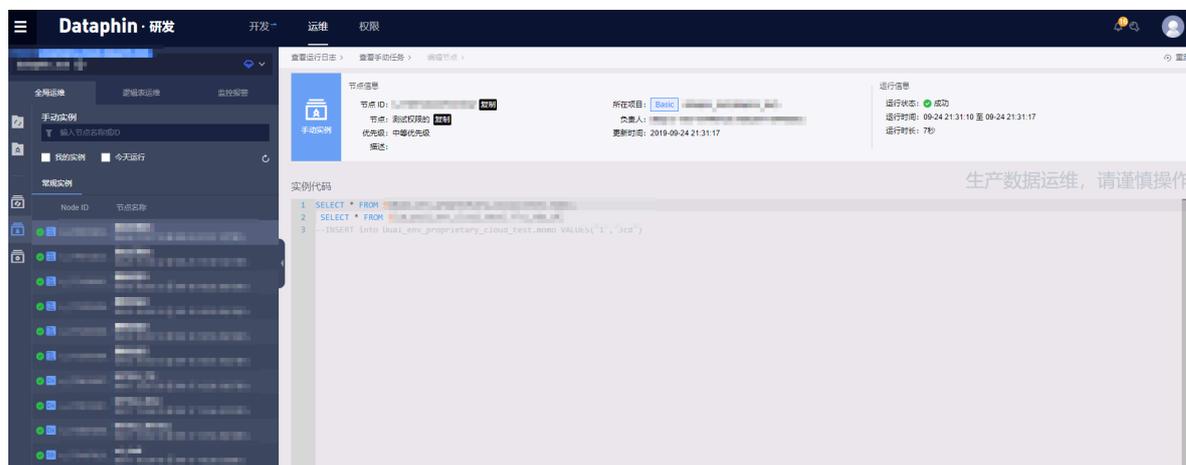


8.2.4 手动实例

本文为您介绍手动实例功能模块。

进入手动实例页面

1. **进入运维调度**，单击左侧菜单栏的手动实例，进入手动实例页面。

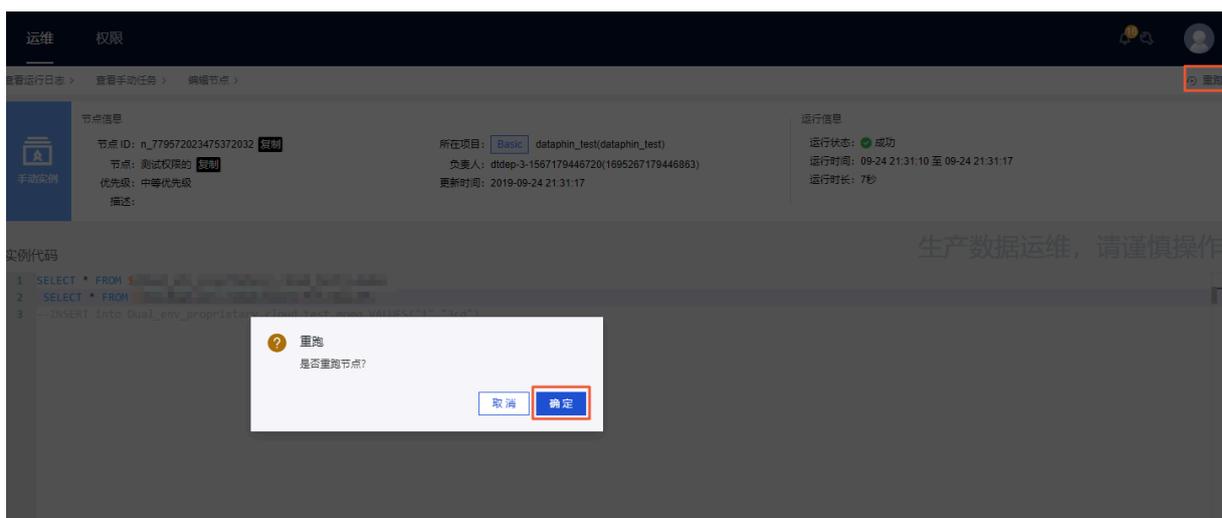


2. 进入手动实例页面后，默认按照运行时间倒排展示所有的手动实例。

- 左侧功能区主要用于搜索和展示实例列表。上方的搜索框内，用户可以输入节点名称或ID进行搜索，同时可以框选我的实例和今天运行以快速过滤责任人为当前用户的实例和当天创建的手动实例。下方按照手动实例的运行时间倒排显示所有符合搜索条件的实例，每一行展示了实例的任务类型（以图标展示）、节点id、节点名称，以及其开始运行时间，结束运行时间和实例运行耗时。
- 右侧功能区则展示了手动实例的代码、详细信息，以及可以对当前手动实例执行的操作项。详细信息在右侧功能区的上方，包括了当前实例的状态、实例ID、实例名称、执行时间、实例的任务类型及责任人等信息。手动实例的可执行操作项包括：重跑、终止、查看运行日志、查看操作日志、编辑节点。

重跑

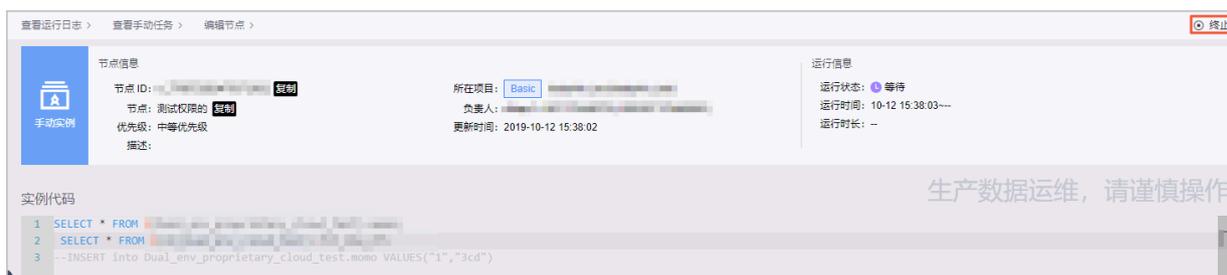
单击右侧功能区的操作项重跑，会出现弹框进行二次确认，单击确认即可重跑当前手动实例。



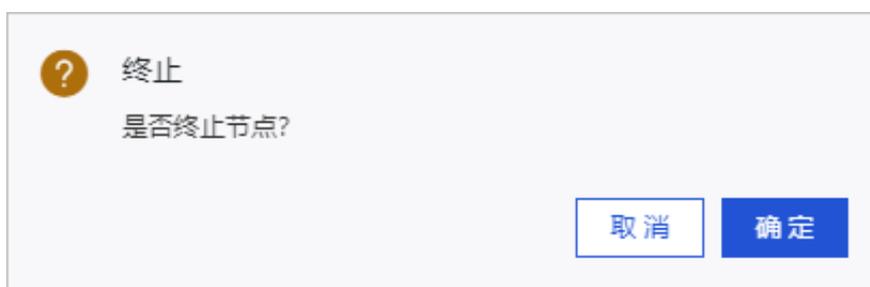
说明:

只有在当前手动实例的状态为已完成时（如成功、失败、未运行），才可以重跑实例，否则重跑按钮是不可见的。

终止运行中的实例



1. 如上图所示，如果当前手动实例的状态为非已完成状态（如初始状态、等待资源、运行中），可以在右侧功能区看到终止的按钮。
2. 单击终止，会出现弹窗进行二次确认，详情如下。

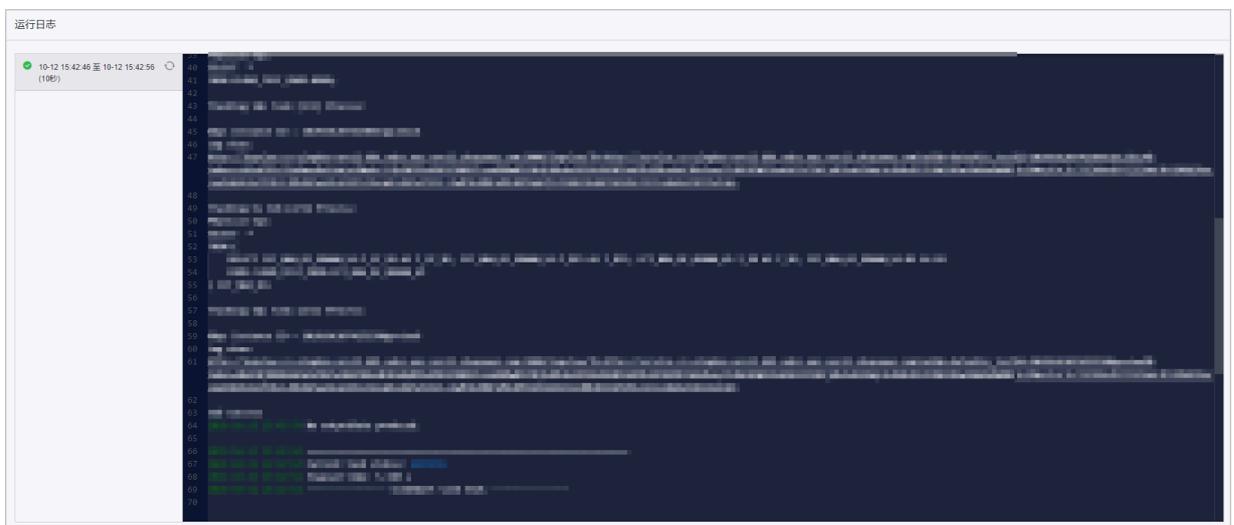


3. 单击确定，即可终止当前正在运行的实例。手动实例被终止后，其最终状态会被置为失败。

查看运行日志



如上图所示，单击右侧功能区中的查看运行日志按钮，新窗口打开运行日志页面，用户可以在此查看详细的运行过程。



查看操作日志

1. 如下图所示，单击右侧功能区中的查看操作日志按钮，会弹出操作日志页面。



2. 在操作日志页面，您可以看到历史操作信息。

| 操作日志 × | | | |
|---|------|----------------|---|
| 操作时间 | 操作人 | 操作 | 详情 |
| 2019-10-12 15:42:45 | [模糊] | 作为根节点 创建工作流 |  |
| 2019-10-12 15:42:13 | [模糊] | 手动重跑 |  |
| 2019-10-12 15:41:17 | [模糊] | 手动重跑 |  |
| 2019-10-12 15:38:02 | [模糊] | 手动重跑 |  |
| 2019-09-24 21:31:09 | [模糊] | 作为根节点 创建工作流 |  |
| 2019-09-20 15:36:41 | [模糊] | 作为根节点 创建工作流 |  |
| 2019-09-20 15:35:48 | [模糊] | 更新节点 |  |
| 2019-09-20 14:26:55 | [模糊] | 更新节点 |  |
| 2019-09-20 14:22:12 | [模糊] | 创建节点 |  |

3. 单击操作日志页面的详情按钮，可以查看历史操作的详情信息。

 操作详情

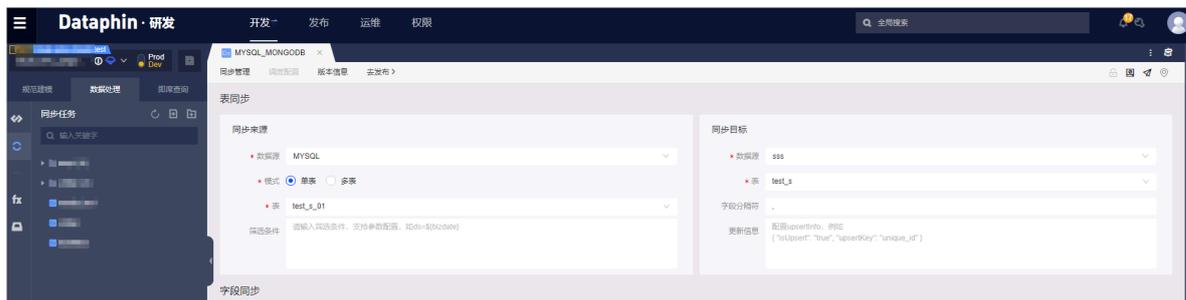
dtdep-3-[模糊]) 作为根节点创建
工作流 n_[模糊]

[知道了](#)

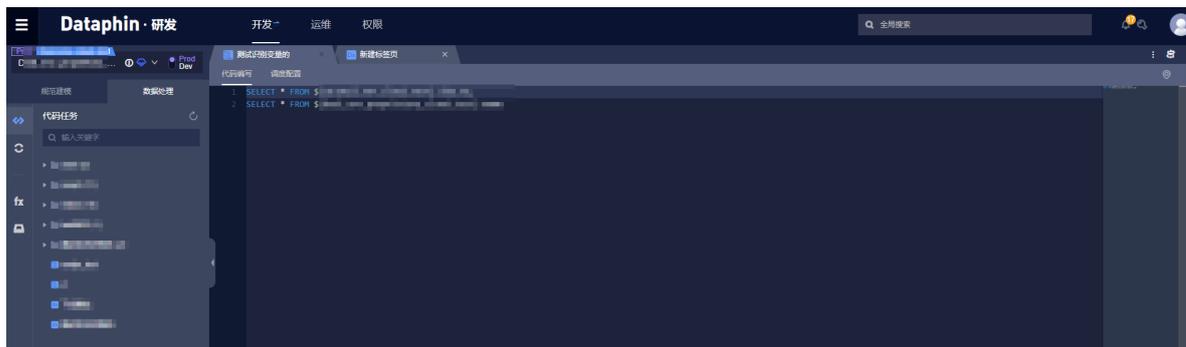
编辑节点



- 单击右侧功能区中的编辑开发节点按钮，会进入开发环境下对应节点的编辑页面。



- 单击编辑生产节点按钮，会进入生产环境下对应节点的编辑界面。



8.2.5 补数据实例

本文为您介绍补数据实例功能模块。

进入补数据实例页面

1. **进入运维调度**，单击左侧菜单栏的补数据实例，进入补数据实例页面。



2. 进入补数据实例页面后，默认展示运行日期为当天的补数据实例，并按照执行时间倒序排列，详情如下：

- 左侧功能区主要用于搜索和展示实例列表。
- 在右侧DAG图展示区，默认以当前选中实例为中心，展开其上下游各一层的实例DAG关系图。在DAG图的右上角有刷新按钮，单击可以查看最新的实例间的关系。同时用户可以根据节点id或名称搜索DAG图来精准定位DAG图中的实例。在DAG图的下方有工具栏，用户可以缩放整个DAG图的大小，调整默认展开的层数，还可以全屏展示DAG图。用户也可以直接拖动实例，调整DAG图的位置。在DAG图的右下方，会展示当前选中实例的详细信息。

搜索和过滤

· 搜索和过滤补数据实例

在左侧功能区上方的搜索框内，用户可以输入节点名称、节点id或补数据实例名称进行搜索，同时可以框选我的实例和我创建的以快速过滤责任人为当前用户的实例和当前用户自己创建的补数据实例。默认搜索补数据实例的运行日期是当天，您可以自行修改运行日期和业务日期的筛选条件以查看对应的补数据实例。

在下方的常规实例中，按照补数据实例生成的时间倒序排列。需要注意的是，补数据实例搜索结果的展示和周期实例有所区别，补数据实例分为三个层级进行展示，分别是：

- 第一层：补数据任务级别。
- 第二层：业务日期级别。
- 第三层：具体实例级别。

· 高级搜索过滤补数据实例

如下图所示，单击搜索框左侧的“漏斗”图标，会出现高级搜索筛选的弹框。支持根据实例的责任人、实例的状态来过滤实例的搜索结果。



右键操作项

补数据实例和周期实例的右键操作项没有区别，具体请参见[周期实例](#)。

8.3 逻辑表运维

8.3.1 逻辑表任务

本文为您介绍逻辑表任务功能模块。

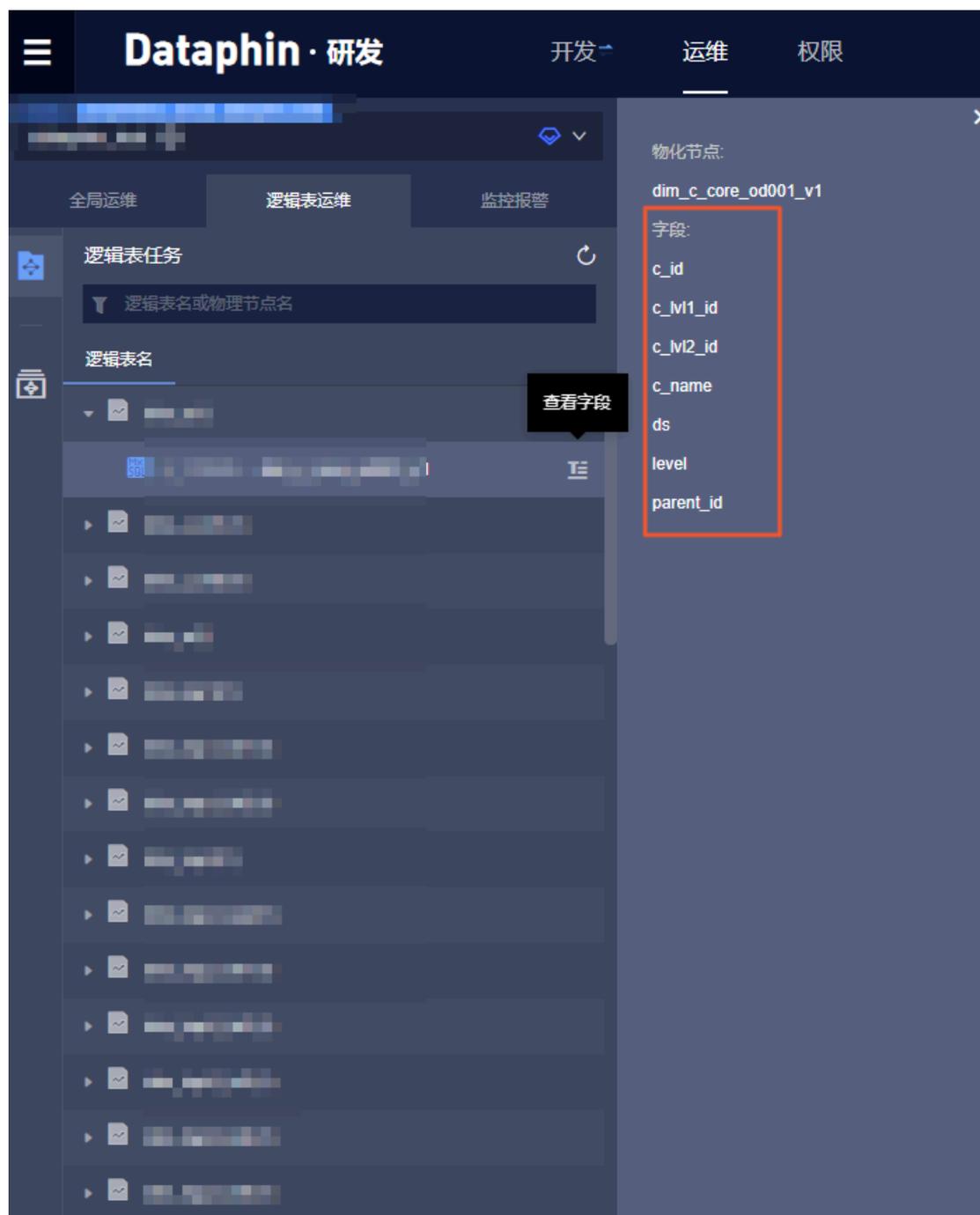
进入逻辑表任务页面

1. **进入运维调度**，单击左侧菜单栏中的逻辑表任务，进入逻辑表任务页面。



2. 在逻辑表任务页面：

- 左侧功能区主要用于搜索和展示逻辑表及其内部物理节点。
 - 在上方的搜索框内，用户可以输入逻辑表名、物理节点名称来搜索节点，单击左侧“漏斗”图标可以限定要搜索的逻辑表类型。搜索时，节点id是精准匹配，逻辑表名、物理节点名支持模糊匹配。
 - 下方的列表分两层显示，第一层为逻辑表，第二层为逻辑表包含的节点。默认显示节点类型（以图标展示）、节点id、节点名称，所有逻辑表按字母顺序排列。
 - 鼠标放置在逻辑表节点上，右侧会显示查看字段按钮，单击可以展开此节点内包含的详细字段。



- 右侧DAG图展示区，默认展示当前逻辑表内包含的所有节点和内部节点间的依赖关系，并在右下角动态展示缩略图。
 - 左上角显示当前逻辑表名称，单击旁边的“双箭头”图标，可以切换查看对应逻辑表实例或任务。
 - DAG图右上角有刷新按钮，单击可以查看最新的节点关系；在旁边的搜索框内，用户也可以输入节点名或节点产出的字段名称来精准定位DAG图中的节点。
 - DAG图下方有工具栏，用户可以缩放整个DAG图的大小，或全屏展示DAG图。
 - DAG图可拖动。单击DAG图上的节点，左下角会显示该节点的详细信息。

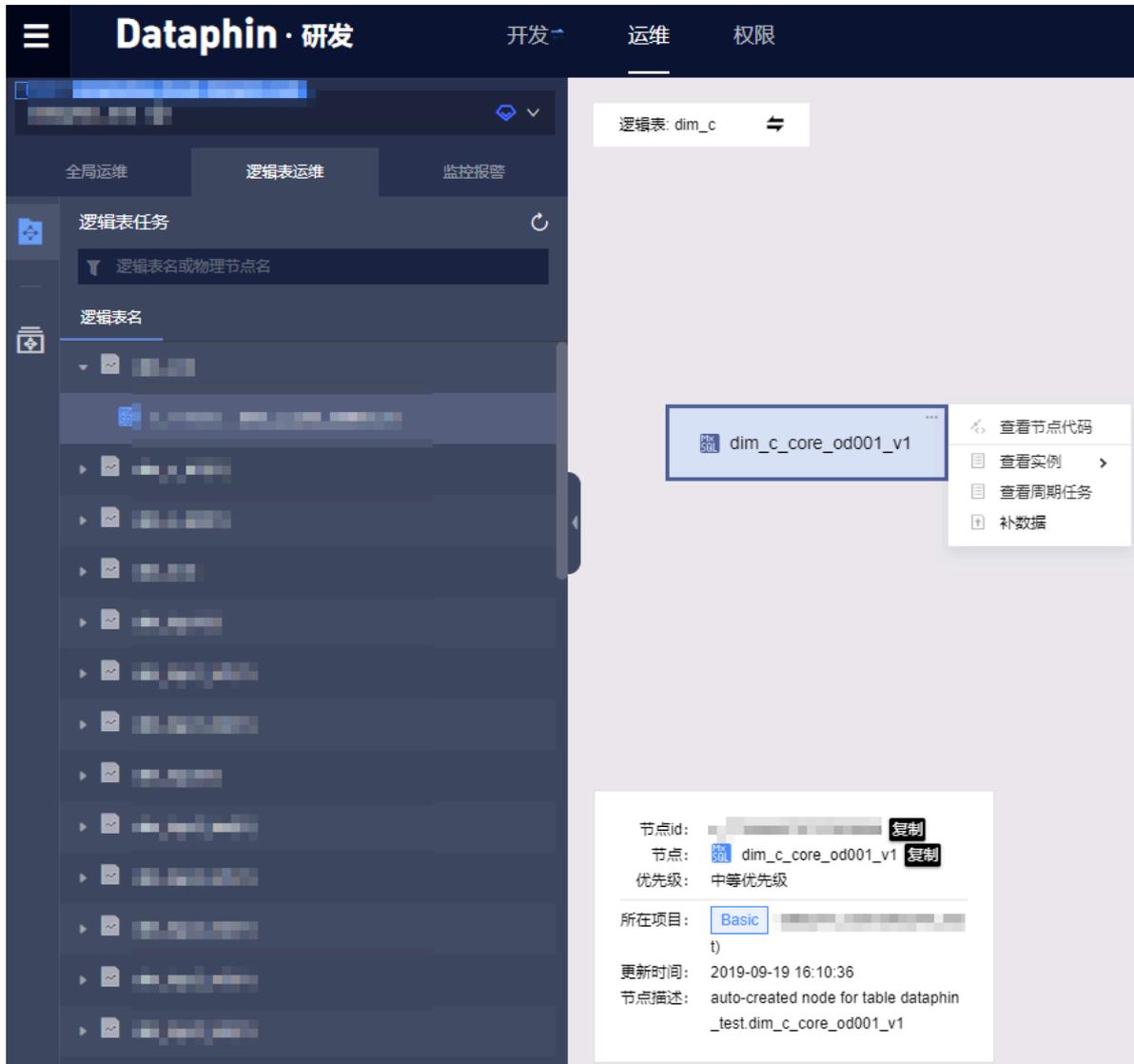
右键操作项

右键单击DAG图上的节点，会展示该节点对应的可操作项。在有权限的情况下，可以进行：**查看节点代码**、**查看实例**、**查看周期任务**、**补数据**等操作。



补数据

1. 右键单击DAG图上的节点，选择补数据。



2. 弹出补数据编辑页面。



说明:

补数据实例名称是系统自动生成的，也可以自行修改。

补数据 ×

实例名称: dim_bailu_level_1_leaf_core_od001_v1_2019-03-19_1552978178490

业务日期: 2019-03-18 ~ 2019-03-18

选择下游: 否 是

3. 选择补数据的业务日期、策略，单击确定即可生成补数据实例。选择下游默认为否，如果选择是，则会展开下游节点选择页面，可以勾选需要补数据的下游节点。



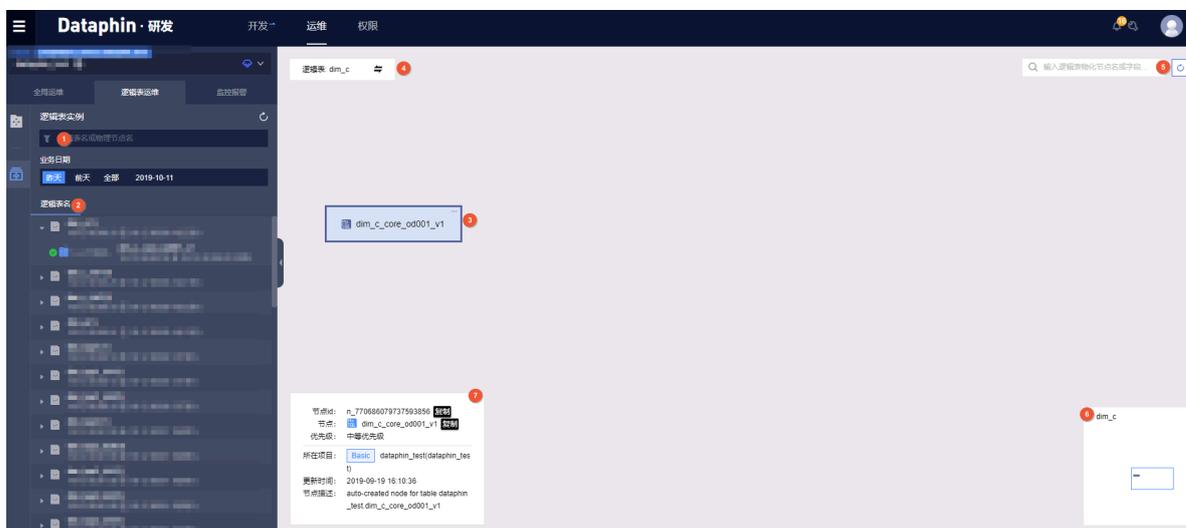
4. 补数据完成后，可以在补数据实例页面查看生成的实例。

8.3.2 逻辑表实例

本文为您介绍逻辑表实例功能模块。

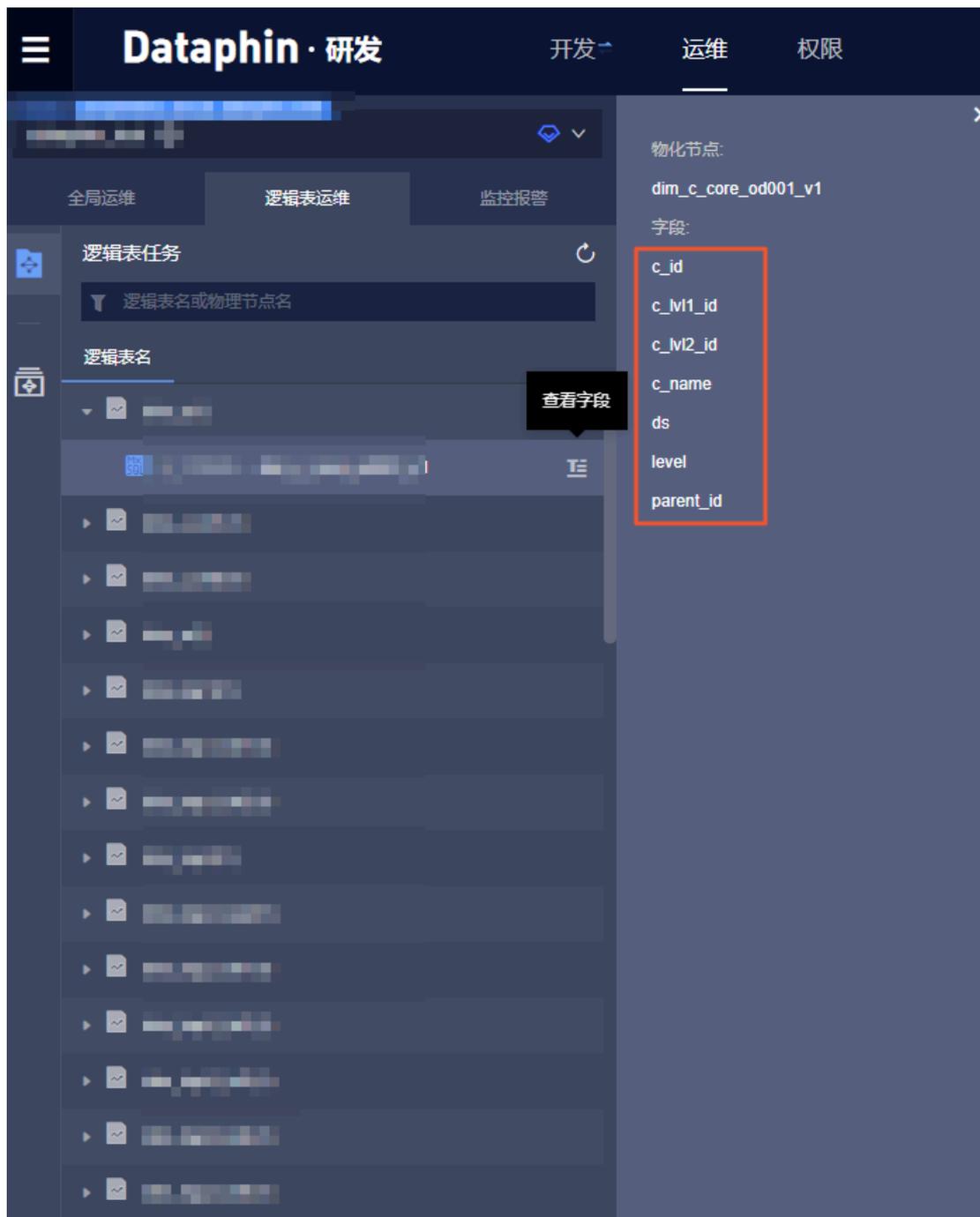
进入逻辑表页面

1. 进入运维调度，单击左侧菜单栏中的逻辑表实例，进入逻辑表实例页面。



2. 在逻辑表实例页面：

- 左侧功能区主要用于搜索和展示逻辑表及其内部物理节点实例，详情如下：
 - 在上方的搜索框内，用户可以输入逻辑表名、物理节点名称或ID来搜索对应实例。搜索时，节点id是精准匹配，节点名、逻辑表名则支持模糊匹配。默认搜索实例的业务日期是昨天，用户可以选择前天、全部或选择具体日期来过滤搜索结果。
 - 下方的搜索结果分两层展示，第一层为逻辑表，第二层为逻辑表包含的节点。默认显示节点类型（以图标展示）、节点id、节点名称，所有逻辑表按字母顺序排列。
 - 鼠标放置于逻辑表节点上，右侧显示查看字段按钮，单击可以展开此节点内包含的详细字段，如下图所示。



- 右侧DAG图展示区，默认展示当前逻辑表包含的所有节点实例及其状态（如运行中、运行成功、运行失败），并在左下角动态展示缩略图，详情如下：
 - 左上角显示当前逻辑表名称，单击旁边的“双箭头”图标可以切换查看对应实例或任务。
 - DAG图右上角有刷新按钮，单击可以查看最新的节点关系；在旁边的搜索框内，用户也可以输入节点名或字段来精准定位DAG图中的节点。
 - DAG图下方有工具栏，用户可以缩放整个DAG图的大小，或全屏展示DAG图。
 - DAG图可拖动。单击DAG上的节点，右下角会显示该节点的详细信息。

右键操作项

右键单击DAG图中的逻辑表实例，会展示所有支持的操作项。在有权限的情况下，可进行：查看运行日志、查看节点代码、重跑并恢复调度、查看周期任务、查看实例等操作。



8.4 萃取运维

8.4.1 行为规则任务

本文为您介绍行为规则任务功能模块，帮助您管理行为规则任务。

进入行为规则任务页面

1. **进入运维调度**，单击，选择data_distill项目，运维中心默认定位到萃取项目下的行为规则任务模块。



2. 进入萃取运维页面后，会按照最近更新时间降序展示，并自动选择首个任务。

- 左侧功能区主要用于搜索和展示行为规则及规则节点，您可以在左上方的搜索框中输入行为名称或行为规则ID来搜索行为规则任务。搜索时，按照行为规则精准匹配，行为名模糊匹配。

左侧下方列表分三层显示，第一层为行为，第二层为行为下的行为规则，第三层为该规则对应的节点。节点命名默认显示节点类型（图标展示）、节点ID、节点名称。鼠标悬浮在具体节点上，将展示该节点ID、节点名称、负责人和更新时间。

- 行为规则任务右侧DAG图，仅展示当前行为规则节点及其直接相关的上游萃取中心公用节点和下游标签相关节点。其中行为规则节点外增加行为规则框，展示行为规则相关信息。默认

除规则节点和标签节点外，其他节点收起在节点簇内。单击  展开节点簇查看内部的具体节点及整个DAG图的依赖关系，并在右下角动态展示缩略图。



管理行为规则

- 鼠标悬浮在行为名后的 ，可对该行为下行为规则进行补数据。



单击行为补数据，进入行为规则补数据页面，您可以对该行为下的所有行为规则进行补数据操作，详细内容请参见[补数据](#)。

- 鼠标悬浮在行为规则  按钮，可进行编辑行为规则、查看行为规则实例和对该条行为规则进行补数据操作。



- 单击编辑行为规则，编辑行为规则的详细内容请参见[编辑行为规则](#)。
- 单击查看行为规则实例，在行为规则实例页面，筛选出该规则实例且默认业务日期为昨天的行为规则实例。
- 单击行为规则补数据，进入行为规则补数据页面，详细内容请参见[补数据](#)。

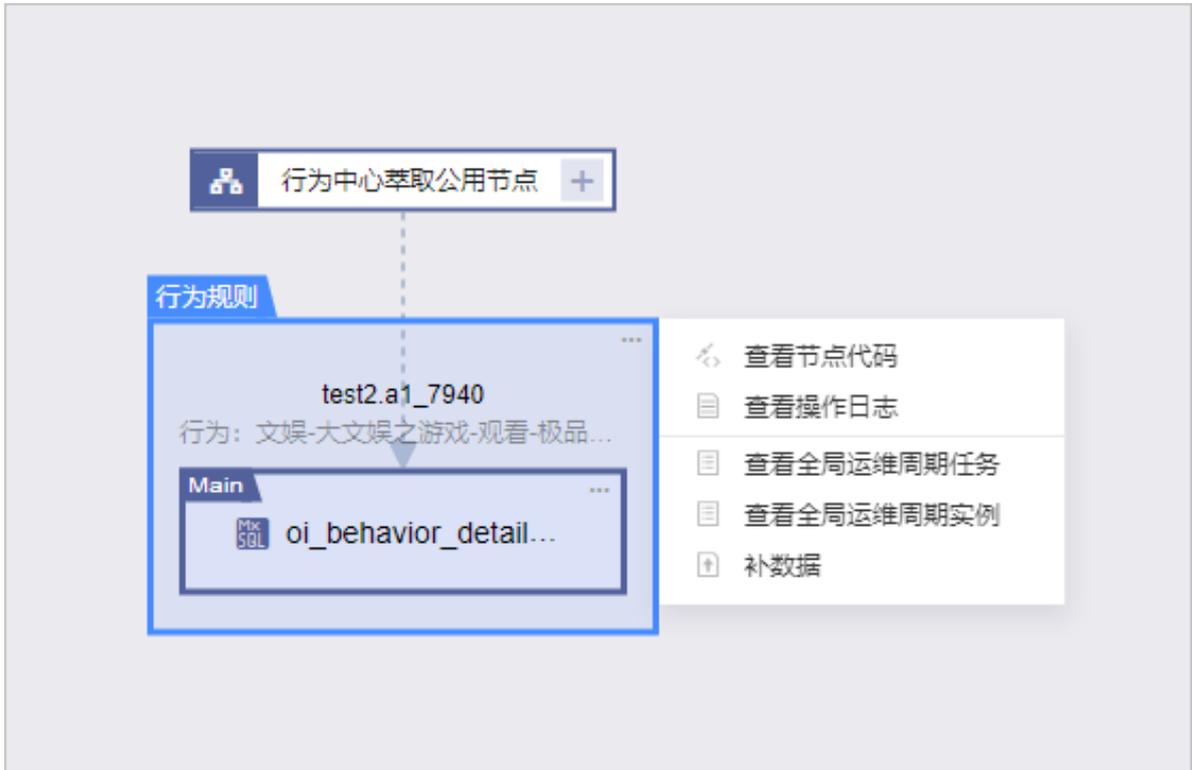
右键操作项

在DAG图右上角提供展开所有节点簇、收起所有节点簇、适应画布快捷按钮。单击DAG图右上角 ，可以查看最新的节点关系，您也可以通过搜索节点ID或节点产出字段名称来精确定位DAG图中的节点通过单击DAG图下方工具栏中的 、 和  可以缩放整个DAG图的大小，或单击  全屏展示DAG图。DAG图可拖动，单击选中节点，左下角会显示该节点的详细信息。



- 单击行为规则框的  按钮或右键，可执行行为规则补数据和编辑行为规则操作，同时您可以查看萃取运维实例中的行为规则实例 和行为规则补数据实例。
 - 单击行为规则补数据，进入行为规则补数据页面，详细内容请参见[补数据](#)。
 - 单击编辑行为规则，进入行为规则编辑页面，详细内容请参考[编辑行为规则](#)。
 - 单击查看萃取运维实例 > 行为规则实例，在行为规则实例页面，筛选出该规则实例且默认业务日期为昨天。
 - 单击查看萃取运维实例 > 行为规则补数据实例，在行为规则补数据实例页面，筛选出该规则补数据实例且默认补数据日期为今天。

- 单击某个节点（包括行为规则节点）的更多操作按钮或右键，可执行查看节点代码、查看操作日志、查看全局运维周期任务、查看全局运维周期实例和补数据操作。



- 单击查看节点代码，进入节点代码页面，展示该节点的代码内容，您可查看代码细节，但不能编辑。
- 单击查看操作日志，右侧弹出操作日志查看页面，您可看到该节点的历史操作记录。
- 单击查看全局运维周期任务，在全局运维周期任务界面，查看该节点的全局DAG图。
- 单击查看全局运维周期实例，在全局运维周期实例界面，查看该节点的全局DAG图。
- 单击补数据，进入节点补数据页面，根据页面提示完成参数配置，并单击确定。

补数据 ×

实例名称

业务日期 ~

选择下游 否 是

8.4.2 标签任务

本文为您介绍标签任务模块功能，助您管理标签相关的任务。

进入标签任务页面

1. 进入运维调度，单击 ，选择data_distill项目，运维中心展示萃取运维模块。



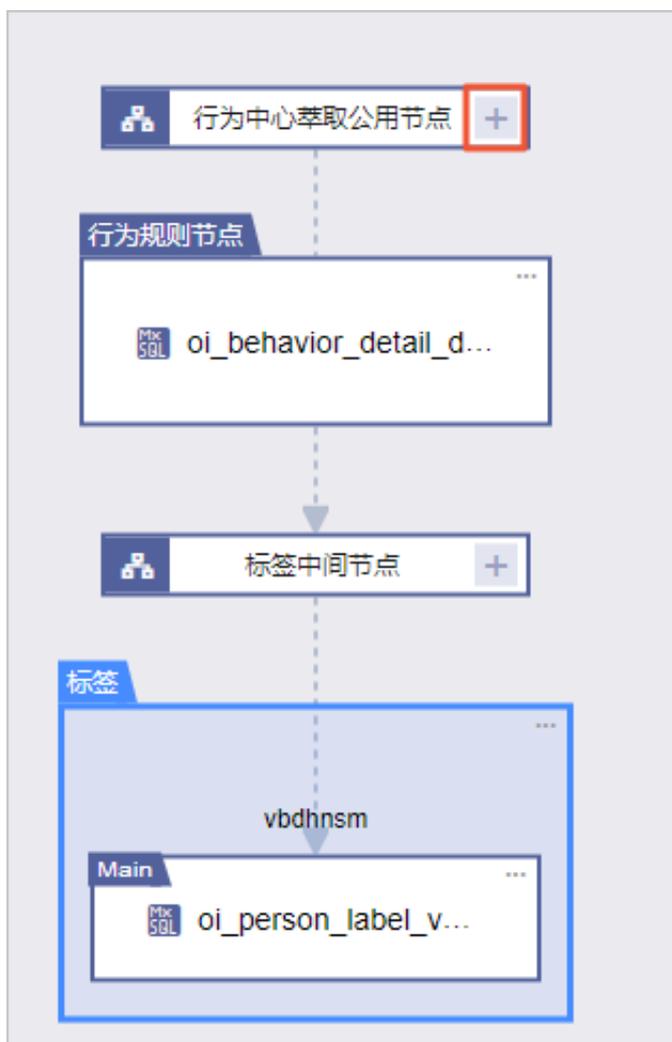
2. 单击左侧导航栏的  按钮，进入标签任务页面后，会按照更新时间倒排展示任务，并自动选择首个任务。

- 左侧功能区主要用于搜索和展示标签及其对应节点。您可以在左上方的搜索框中输入标签名称或标签ID来搜索标签任务。搜索时，按照标签ID精准匹配，标签名模糊匹配。

左侧下方列表分两层显示，第一层为标签，第二层为标签对应的节点。节点的名称默认显示节点类型（图标展示）、节点ID、节点名称命名。鼠标悬浮在具体节点，展示该节点ID、节点名称、负责人和更新时间。

- 标签任务右侧DAG图，仅展示当前标签节点及其相关标签中间节点、直接依赖的行为规则节点和规则相关的上游萃取中心公用节点。默认除规则节点和标签节点外，其他节点收起在节

点簇内。您可以单击  展开节点簇查看内部的具体节点及整个DAG图的依赖关系，并在右下角动态展示缩略图。



管理标签任务

鼠标悬浮左侧导航栏标签后的 ，您可以进行标签补数据、编辑标签和查看标签实例的操作。

- 单击标签补数据，进入标签补数据弹框。根据页面提示配置参数，详细内容请参见[补数据](#)。
- 单击编辑标签，进入编辑标签页面。根据页面提示完成参数配置，详细内容请参见[编辑工厂标签](#)。
- 单击查看标签实例，筛选出该标签实例且默认业务日期为昨天。



右键操作项

在DAG图中，页面右上角提供收起所有节点簇、适应画布、展开所有节点簇快捷按钮。您可以通过单击DAG图右上角的 ，查看最新的节点关系。您可以通过搜索节点ID或节点产出字段名称来精准定位DAG图中的节点。通过单击DAG图下方工具栏中的 、 和  可以缩放整个DAG图的大小，或单击  全屏展示DAG图。DAG图可拖动，单击节点右下角会显示该节点的详细信息。



- 单击某个节点（包括标签节点）的  按钮或右键，可执行查看节点代码、查看操作日志、查看全局运维周期任务、查看全局运维周期实例、补数据的操作。

生产数据运维



节点id: n_779773955666739200 [复制](#)

节点:  oi_person_label_value_7946 [复制](#)

优先级: 中等优先级

所在项目: [Distill](#) data_distill

负责人: SuperAdmin(SuperAdmin@dataphin)

更新时间: 2019-09-20 16:00:09

节点描述: auto-created node for oi_person_label_value_7946

- 单击查看节点代码，进入节点代码页面，展示该节点的代码内容，您可查看代码细节，但不能编辑。
- 单击查看操作日志，右侧弹出操作日志查看界面，可以看到该节点的历史操作记录。
- 单击查看全局运维周期任务，在全局运维周期任务界面，查看该节点的全局DAG图。
- 单击查看全局运维周期实例，在全局运维周期实例界面，查看该节点的全局DAG图。
- 单击补数据，进入节点补数据界面，根据页面提示完成参数的配置。并单击确定。

补数据 ×

实例名称

业务日期 ~

选择下游 否 是

- 单击标签框的  按钮或右键，可执行标签补数据和编辑标签操作、同时您可以查看萃取运维实例中的标签实例、标签补数据实例。



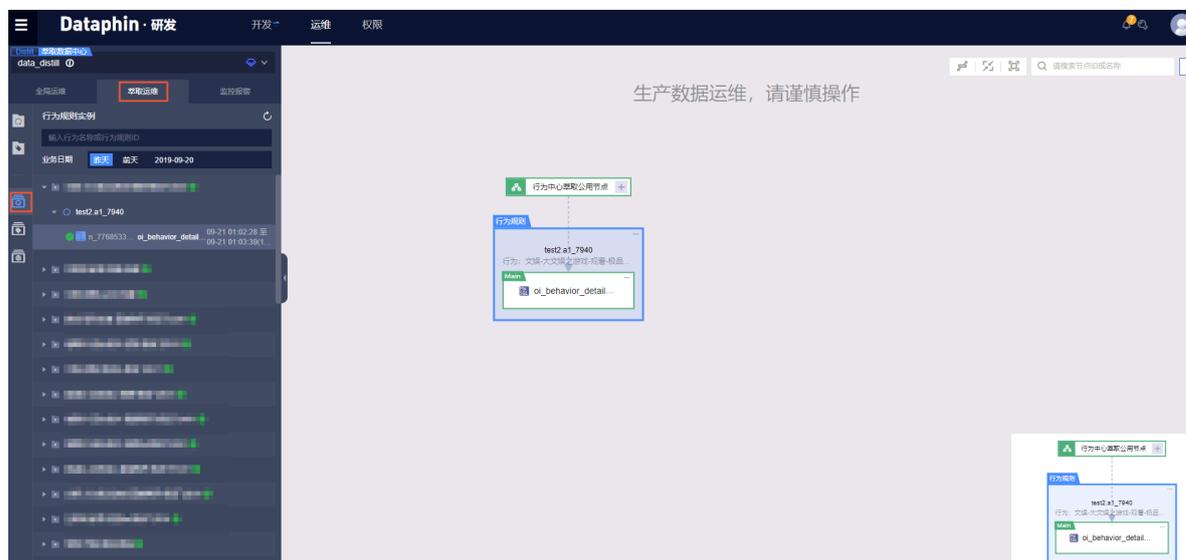
- 单击标签补数据，进入标签补数据页面。根据页面提示配置参数，详细内容请参见[补数据](#)。
- 单击编辑标签，在编辑标签页面，根页面提示完成参数配置，详细内容请参见[编辑工厂标签](#)。
- 单击标签实例，在标签实例页面，筛选出该标签任务对应的实例且默认业务日期为昨天。
- 单击标签补数据实例，在标签补数据实例页面，筛选出该标签对应的规则补数据实例且默认截止日期为今天。

8.4.3 行为规则实例

本文为您介绍行为规则实例模块功能。

进入行为规则实例页面

1. **进入运维调度**，默认展示萃取运维模块。如下图所示。



2. 单击左侧导航栏中的行为规则按钮，进入行为规则任务实例页面。

- 左侧功能区主要用于搜索和展示行为下的行为规则及规则节点。您可以在左上方的搜索框中输入行为名称或行为规则ID来搜索行为规则任务。搜索时，按照行为规则ID精准匹配，行为

名模糊匹配。默认搜索实例的业务日期是昨天，用户可以选择前天或选择具体日期来过滤搜索结果。

左侧下方列表分三层显示：

- 第一层为行为，行为名称后展示该行为下所有行为规则实例的整体运行状态；若所有规则实例都运行成功，则整体运行状态为成功；若有一个实例是失败，则整体运行状态为失败；若部分实例运行成功部分实例运行中，则整体运行状态为运行中。

- 第二层为行为下的行为规则，鼠标悬浮在行为规则的  按钮，可执行编辑行为规则、查看行为规则任务操作。

■ 单击编辑行为规则，进入编辑行为规则页面。页面参数的配置请参考[编辑行为规则](#)。

■ 单击查看行为规则任务，进入行为规则任务页面，查看该任务的详情。

- 第三层为该规则对应的实例节点，默认显示实例状态、节点类型、节点ID、节点名称、运行时间。鼠标hover具体节点，将展示该节点ID、节点名称、负责人和运行时间

- 节点框的颜色标识该节点运行状态，绿色表示运行成功，红色表示运行失败，灰色表示未运行、蓝色表示运行中、紫色表示等待运行。节点簇外框的颜色标识该节点簇整体运行状态。其中，内部全部节点运行成功节点簇为绿色，只要有一个节点运行失败即为红色。



- 右侧DAG图仅展示当前行为规则节点及其直接相关的上游萃取中心公用节点和下游标签相关节点；其中行为规则节点外增加行为规则框，展示行为规则相关信息。默认除规则节点和标签节点外，其他节点收起在节点簇内，可点击展开节点簇查看内部的具体节点及整个DAG图的依赖关系，并在右下角动态展示缩略图。

右键操作项

在DAG图中，页面右上角提供展开所有节点簇、收起所有节点簇、适应画布快捷按钮。DAG图右上角有刷新按钮，可以查看最新的节点关系，也可以通过搜索节点ID或节点产出字段名称来精确定位DAG图中的节点。DAG图下方有工具栏，可以缩放整个DAG图的大小，或全屏展示DAG图。DAG图可拖动，单击节点右下角会显示该节点的详细信息。

- 单击行为规则框的  按钮或右键，可执行以下操作：
 - 单击编辑行为规则，进入编辑行为规则页面。页面参数的配置请参考[编辑行为规则](#)。
 - 单击查看行为规则任务，进入行为规则任务页面，查看该任务的详情。



The screenshot displays a workflow node titled "行为中心萃取公用节点" (Behavior Center Extraction Common Node) with a plus sign. Below it, a blue box highlights a behavior rule node labeled "test2.a1_7940" with the description "行为: 文娱-大文娱之游戏-观看-极品...". A sub-node "Main" is visible, containing a "MySQL" node named "oi_behavior_detail...". To the right, a context menu offers "编辑行为规则" (Edit Behavior Rule) and "查看行为规则任务" (View Behavior Rule Task).

节点id: n_776853377905328128 [复制](#)

节点: [MySQL](#) oi_behavior_detail_d_7940 [复制](#)

优先级: 中等优先级

节点状态:  运行成功

所在项目: [Distill](#) data_distill

负责人: SuperAdmin(SuperAdmin@dataphin)

运行时间: 09-21 01:02:28 至 09-21 01:03:39

运行时长: 1分11秒

节点描述: auto-created node for oi_behavior_detail_d_7940

- 单击点击行为规则框的  按钮或右键，可执行以下操作：



- 单击查看运行日志，可查看该节点的运行日志。
- 单击查看节点代码，可查看应节点的代码内容，可以查看代码细节，但不能编辑。
- 单击查看操作日志，右侧弹出操作日志查看界面，可以看到该节点的历史操作记录。
- 单击查看全局运维周期任务，在全局运维周期任务界面，查看该节点的全局DAG图。
- 单击查看全局运维周期实例，在全局运维周期实例界面，查看该节点的全局DAG图。
- 单击终止，可终止节点处于等待运行和运行中的状态。
- 单击重跑并恢复调度，可重跑节点处于运行成功、失败或未运行当前及下游未跑数据。

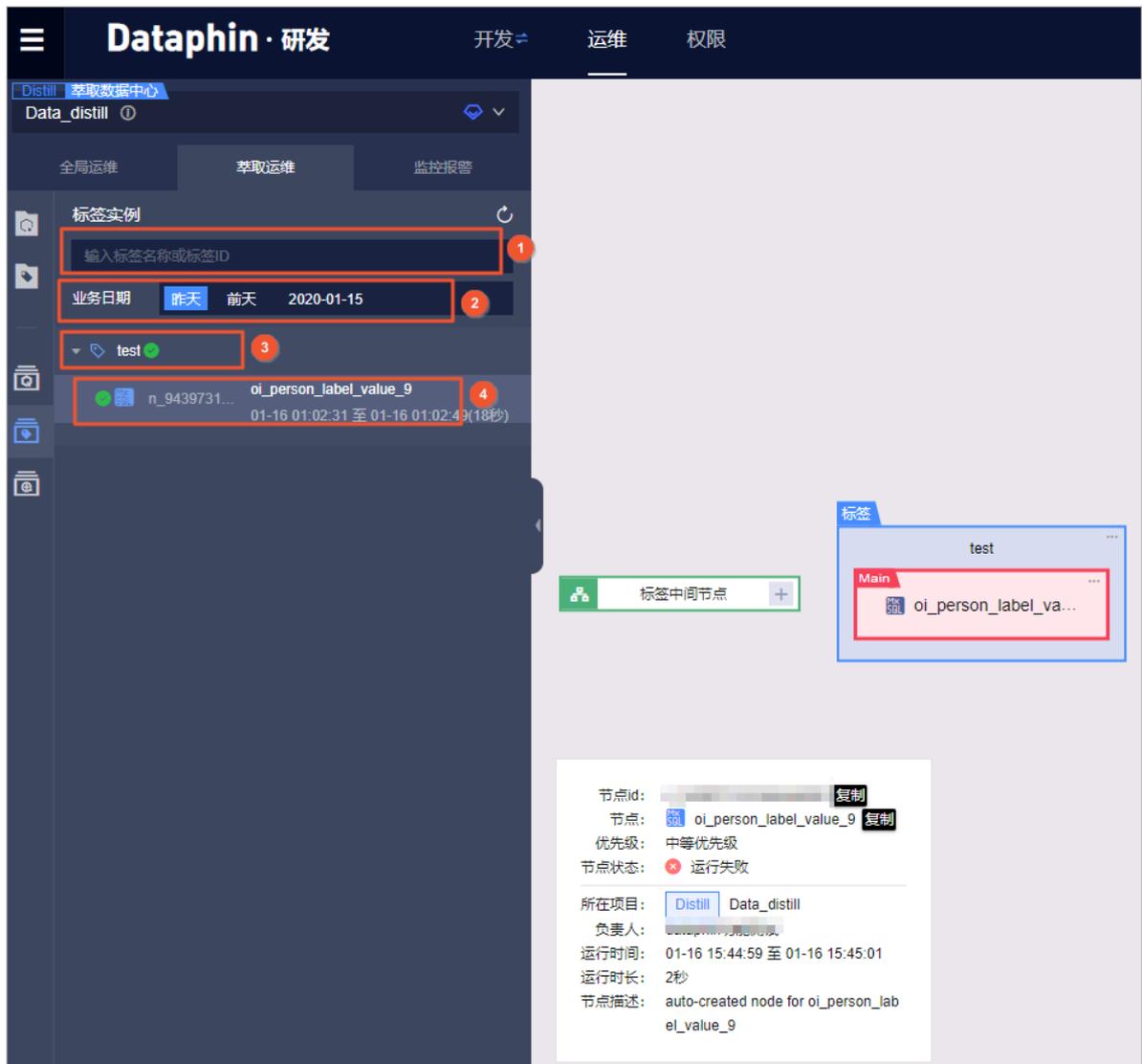
8.4.4 标签实例

标签实例为您展示标签任务的运行实例，本文为您介绍标签实例模块功能。

左侧功能区支持的

1. [进入运维调度](#)，默认展示萃取运维模块。

2. 单击左侧导航栏 ，进入标签实例页面。



| 序号 | 说明 |
|----|--|
| | 您可以在搜索框中输入标签名称或标签ID来搜索标签实例。搜索时，按照标签ID精准匹配，标签名称模糊匹配，默认搜索实例的业务日期是昨天。 |
| | 您可以选择前天、昨天或选择具体日期来过滤搜索标签实例。 |
| 3 | 表示标签的名称。 |

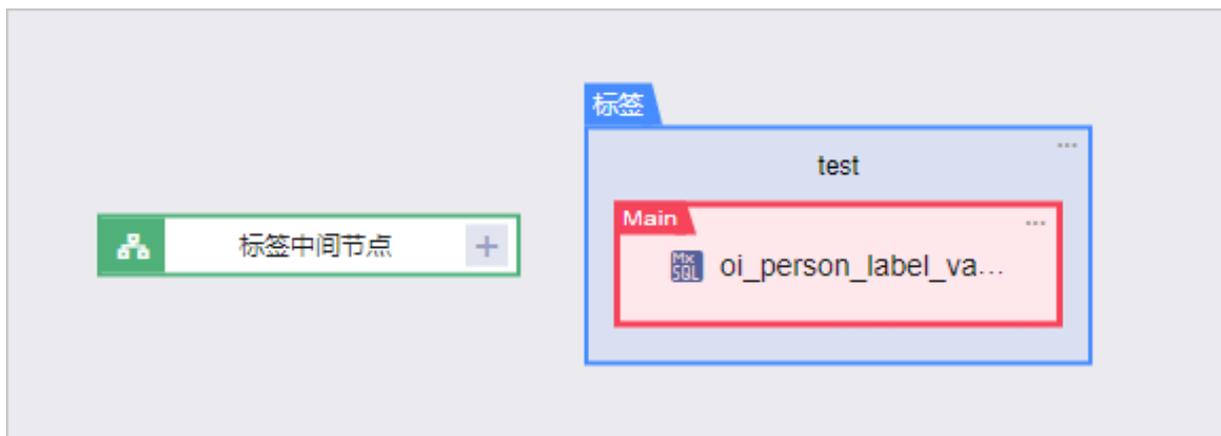
| 序号 | 说明 |
|----|---|
| 4 | 表示标签对应的标签实例。标签实例名称默认显示由实例状态、节点类型、节点ID、节点名称、运行时间组成。您可以将鼠标悬浮在具体节点，展示该节点ID、节点名称、负责人和运行时间。> |

鼠标悬浮左侧导航栏标签后的 ，可操作编辑标签和查看标签任务。



- 单击编辑标签，在编辑标签页面，根页面提示完成参数配置，详细内容请参见[编辑工厂标签](#)。
- 单击查看标签任务，进入行为规则任务页面，查看该任务的详情。

标签实例右侧DAG图仅展示当前标签节点及其相关标签中间节点。标签节点外增加标签框，展示标签相关信息。默认除规则节点和标签节点外，其他节点收起在节点簇内，您可单击 ，展开节点簇查看内部的具体节点及整个DAG图的依赖关系，并在右下角动态展示缩略图。



- 节点簇外框的颜色标识该节点簇整体运行状态。其中，内部全部节点运行成功节点簇为绿色，只要有一个节点运行失败即为红色。



- 节点框的颜色表示该节点运行状态：
 - 红色表示运行失败。
 - 灰色表示未运行。
 - 蓝色表示运行中。
 - 紫色表示等待运行。
 - 绿色表示运行成功。

单击标签框的  按钮或右键，可执行以下操作。



- 单击编辑标签，在编辑标签页面，根页面提示完成参数配置，详细内容请参见[编辑工厂标签](#)。
- 单击查看标签任务，进入行为规则任务页面，查看该任务的详情。

单击某个节点（包括标签节点）的  按钮或右键，可执行以下操作。



- 单击查看运行日志，可以查看该节点的运行日志。
- 单击查看节点代码，进入节点代码页面，展示该节点的代码内容，您可查看代码细节，但不能编辑。
- 单击查看操作日志，右侧弹出操作日志查看界面，可以看到该节点的历史操作记录。
- 单击查看全局运维周期任务，在全局运维周期任务界面，查看该节点的全局DAG图。
- 单击查看全局运维周期实例，在全局运维周期实例界面，查看该节点的全局DAG图。
- 单击终止，可以终止节点处于等待运行和运行中的状态。
- 单击重跑并恢复调度，可以重跑节点处于运行成功、失败或未运行的状态。

8.4.5 补数据实例

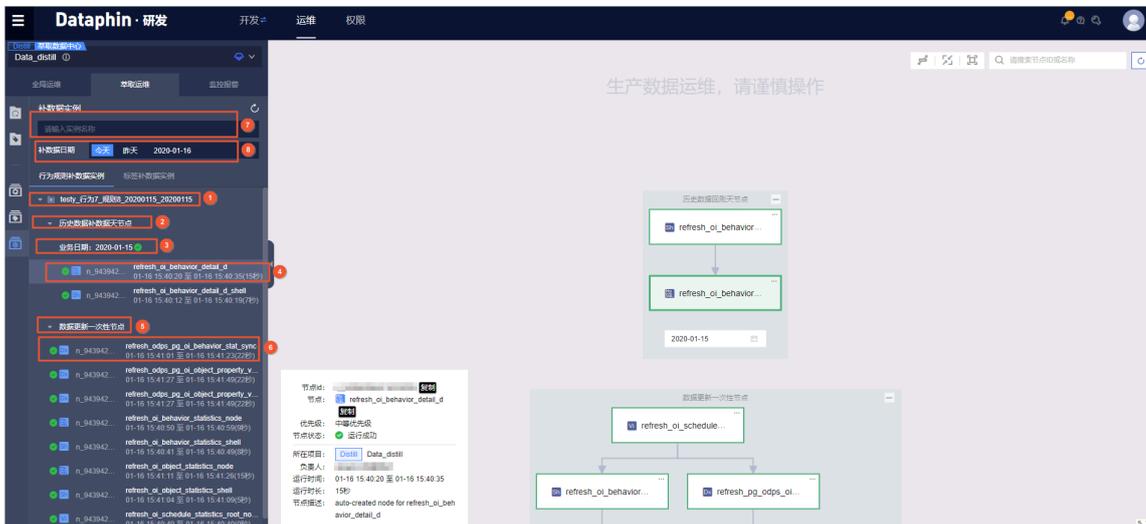
补数据实例为您展示补数据的结果，本文为您介绍补数据实例模块功能。

左侧功能区支持的操作项

1. **进入运维调度**，数据萃取项目下，默认展示萃取运维模块。

2. 单击左侧导航栏中的，进入补数据实例页面。萃取运维的补数据实例分为行为规则补数据实例和标签补数据实例：

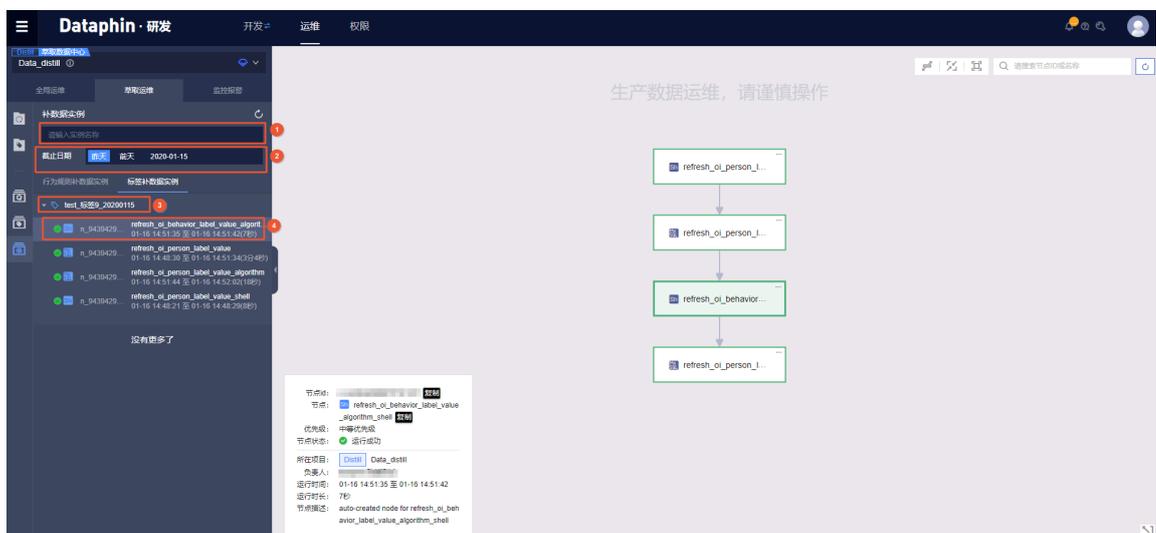
- 如果您选择行为规则补数据实例



| 序号 | 说明 |
|----|--|
| | 表示行为规则实例的名称，行为规则实例的命名规则为前缀_行为ID_规则ID_补数据开始时间_补数据结束时间 |
| 2 | 表示行为规则补数据实例的类型： <ul style="list-style-type: none"> - 历史数据回刷天节点支持在DAG图直接选择日期以查看对应实例，且每日运行。 - 数据更新一次性节点当且仅当所有业务日期的历史数据回刷天节点都运行成功后才会运行，且仅运行一次。 |
| 3 | 当行为规则补数据实例的类型选择历史数据回刷天节点，此处表示行为规则补数据实例的业务日期及该业务日期下所有实例的整体运行状态。 <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;"> <p> 说明： 当且仅当所有实例运行成功时显示为成功，如果有一个实例运行失败，则显示失败，如果部分实例运行成功，部分实例运行中时，显示运行中</p> </div> |
| 4 | 当行为规则补数据实例的类型选择历史数据回刷天节点，此处表示行为规则补数据实例名称，名称由节点状态、行为规则类型、节点ID、行为规则名称组成。 |
| 5 | 行为规则补数据实例的类型选择数据更新一次性节点 |

| 序号 | 说明 |
|----|--|
| 6 | 当行为规则补数据实例的类型选择历史数据回刷天节点，此处表示行为规则补数据实例名称，名称由节点状态、行为规则类型、节点ID、行为规则名称组成。 |
| 7 | 您可以在搜索框输入实例名称进行搜索，默认搜索补数据实例的发起补数据日期是当天，且采用模糊匹配方式。 |
| 8 | 您可以单击补数据日期后的昨天、今天、或日期选择补数据日期，进行筛选补数据实例，下方的补数据实例按照生成的时间倒排。 |

- 如果您选择标签补数据实例。



| 序号 | 说明 |
|----|--|
| 1 | 您可以在搜索框输入实例名称进行搜索，默认搜索补数据实例的截止日期是昨天，且采用模糊匹配方式。 |
| 2 | 您可以单击补数据日期后的昨天、今天、或日期选择补数据日期，进行筛选补数据实例。 |
| 3 | 表示标签实例的名称。标签补数据实例的命名规则为前缀_标签ID_补数据结束时间。 |
| 4 | 表示标签补数据实例的名称，名称由节点状态、行为规则类型、节点ID、行为规则名称组成。 |

右侧功能区支持的操作项

- 如果您选择行为规则补数据实例下的补数据实例。

右侧的DAG图展示区，分为历史数据回刷天节点数据更新一次性节点两个节点簇。您可单击 ，展开节点簇查看内部的具体节点及整个DAG图的依赖关系。

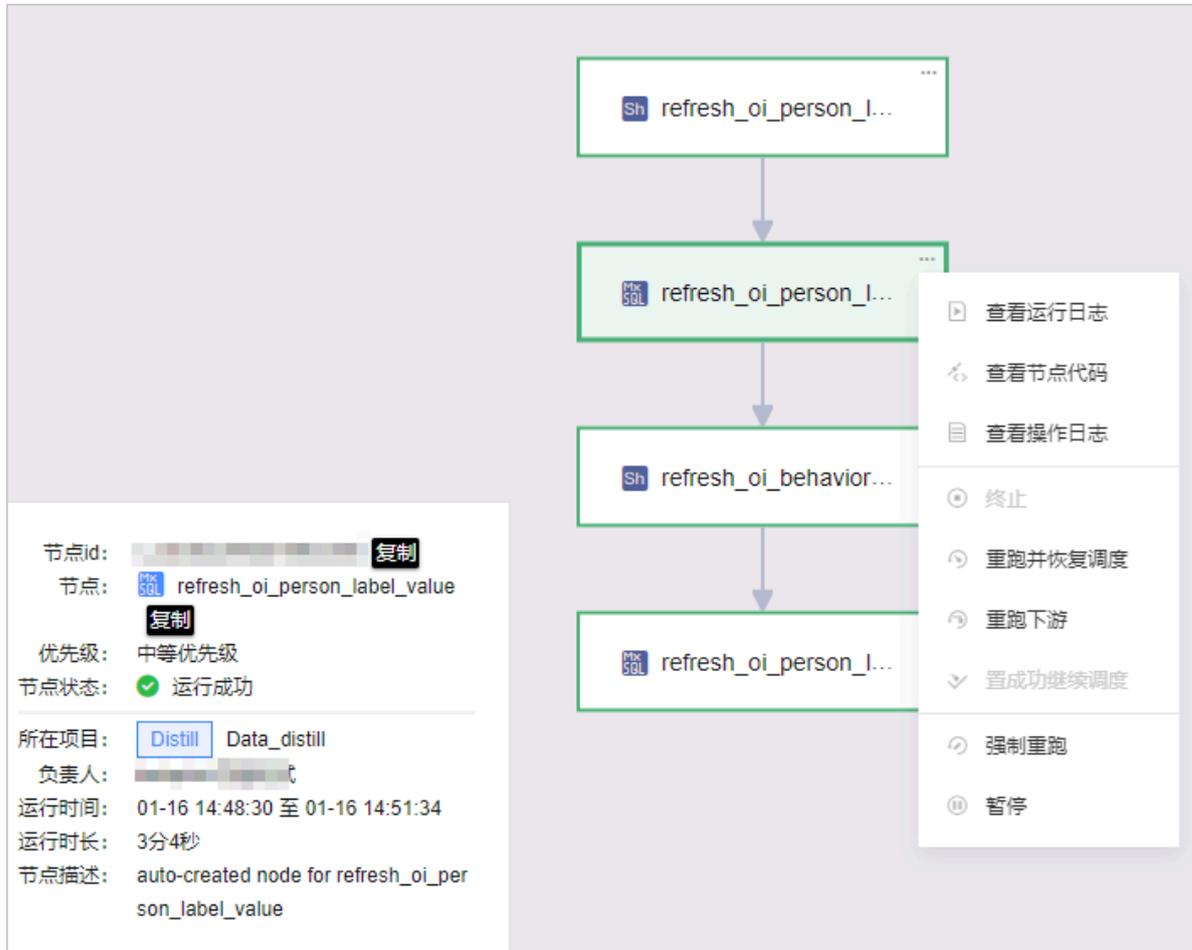
在任意一个节点上右击可以对该节点执行以下操作。



- 单击 ，选择查看补数据实例的日期。
- 单击查看运行日志，可以查看该节点的运行日志。
- 单击查看节点代码，进入节点代码页面，展示该节点的代码内容，您可以查看代码细节，但不能编辑。
- 单击查看操作日志，右侧弹出操作日志查看界面，可以看到该节点的历史操作记录。
- 单击重跑下游，重跑下游节点数据。
- 单击强制重跑，强制重跑该节点。
- 单击终止，可以终止节点处于等待运行和运行中的状态。
- 单击重跑并恢复调度，可以重跑节点处于运行成功、失败或未运行的状态。
- 单击暂停，只支持暂停当前周期该节点的运行。

- 单击置成功继续调度，可以将运行失败的节点，手动置于成功状态继续参与运维调度。
- 如果您选择行为规则补数据实例下的补数据实例。

标签补数据任务的DAG样式固定，由4个节点组成，在任意一个节点上右击可以对该节点执行查看运行日志、查看节点代码、查看操作日志等操作，详情请参见上文。



9 监控报警

9.1 任务监控配置

9.1.1 新建任务监控

配置任务监控报警，是对发布到生产环境的任务进行运行状态监测，方便您及时了解生产任务的情况。本章为您介绍如何新建任务监控。

1. 通过研发 > 运维 > 监控报警 > 任务监控配置，进入运维调度中心的任务监控配置页面。



说明:

如果是Dev-Prod模式，您需切换到生产环境项目的运维调度中心进行有关监控报警的所有操作。

2. 单击任务监控配置页面右上角新建任务监控按钮，进入新建任务监控页面。

新建任务监控

* 选择任务：

* 报警原因： 出错 完成 未完成

运行超过 分钟

* 接收人： 责任人 自定义

* 接收方式： 电话 邮件 短信 钉钉群机器人

3. 设置任务监控的各个配置项：

- **选择任务：**您可以选择一到三个任务进行监控。
- **报警原因：**可以选择出错、完成、截至某时刻未完成、运行时长超过设定值。
- **接收人：**可以直接指定该任务责任人为报警的接收人，也可以选择最多3位、至少1位的指定接收人。
- **接收方式：**当前支持电话、邮件、短信、钉钉群机器人方式。



说明：

- 系统支持每日使用短信、邮件、钉钉群机器人发送消息数量的上限分别是100、2000、2000条。
- 单账户每日可接收短信、邮件、钉钉群机器人消息数量的上限分别是20、400、400条。
- 系统支持每日使用电话语音服务发送消息的上限是100条（每条语音通知按1分钟计算）。

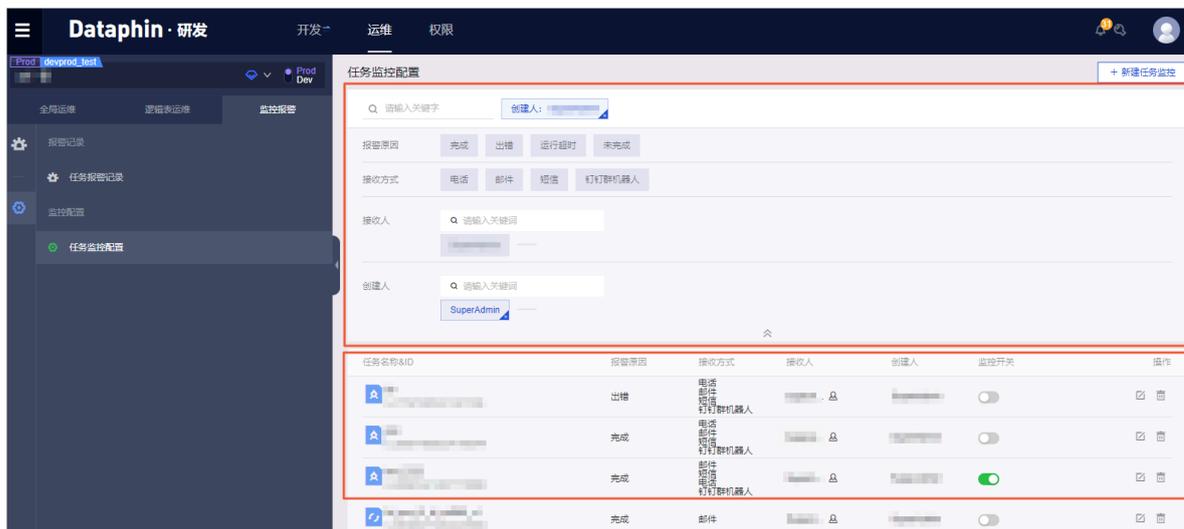
4. 上述设置完成之后，单击确定即可。

9.1.2 任务监控配置管理

本文为您介绍如何进行搜索、查看已有的任务监控，以及开启、关闭、编辑、删除任务监控等操作。

搜索、查看已有的任务监控

1. 通过研发 > 运维 > 监控报警 > 任务监控配置，进入运维调度中心的任务监控配置页面。



2. 如上图所示：

- 在任务监控配置页面。默认显示全部任务监控的会展示任务名称&ID、报警原因、接收方式、接收人、创建人、监控开关信息。
- 您可以通过在右侧面板上方的搜索框中，输入关键字搜索已有的任务监控配置。单击下拉箭头可以展开筛选项，支持通过选择报警原因、接收方式、接收人、创建人来过滤搜索结果。

下方显示搜索结果，会展示具体任务监控的信息。

开启/关闭任务监控

在下方的搜索结果展示区，单击指定任务后面的监控开关，可以开启或关闭该任务监控。

编辑、删除任务监控

在下方的搜索结果展示区，单击操作列的编辑按钮，弹出任务监控编辑面板，可以修改报警原因、接收人、接收方式。

编辑任务监控 ✕

任务：bb

* 报警原因： 出错 完成 未完成 00:00 🕒

运行超过 0 分钟

* 接收人： 责任人 自定义

* 接收方式： 电话 邮件 短信 钉钉群机器人

取消 确定

单击操作列的删除按钮，您可以在确认删除弹窗中单击删除，即可将该任务监控删除。

删除

确认删除 _v1 吗?

取消 确定

9.2 任务报警记录

当设定了监控配置的任务触发监控配置条件时，系统将通过您指定的渠道进行报警消息推送，同时会在任务报警记录页面生成报警记录。您可以在任务报警记录中心查看历史报警情况，便于报警的跟踪记录。

搜索、查看任务报警记录

1. 单击研发 > 运维 > 监控报警 > 任务报警记录，进入运维调度中心的任务报警记录模块。



| 任务名称&ID | 监控类型 | 报警原因 | 接收方式 | 发送状态 | 接收人 | 报警推送时间 | 操作 |
|-----------------------|------|------|--------|--------|----------|---------------------|--------|
| sbs_n_996914928224... | 任务监控 | 完成 | 邮件 | 发送失败() | [Avatar] | 2019-08-30 11:53:03 | [Icon] |
| sbs_n_996914928224... | 任务监控 | 完成 | 电话 | 发送失败() | [Avatar] | 2019-08-30 11:53:03 | [Icon] |
| sbs_n_996914928224... | 任务监控 | 完成 | 钉钉群机器人 | 发送成功 | [Avatar] | 2019-08-30 11:53:03 | [Icon] |



说明:

如果是Dev-Prod模式，您需切换到生产环境项目的运维调度中心查看任务报警记录。

2. 在右侧面板上方的搜索框内，您可以输入关键字搜索报警记录。展开搜索框下的筛选项，可以通过选择报警原因、接收方式、报警的发送状态等来过滤搜索结果。

发送状态包括成功、发送中、失败和已发送。

- 成功表示系统报警内容已按照报警接收方式发出，并获取到系统回执时所显示的状态。
- 发送中表示系统报警信息发送任务已提交，任务序列排队中。
- 失败表示出现如下场景时，系统回执时所显示的状态。

- 系统报警内容无法按照报警接收方式发出。

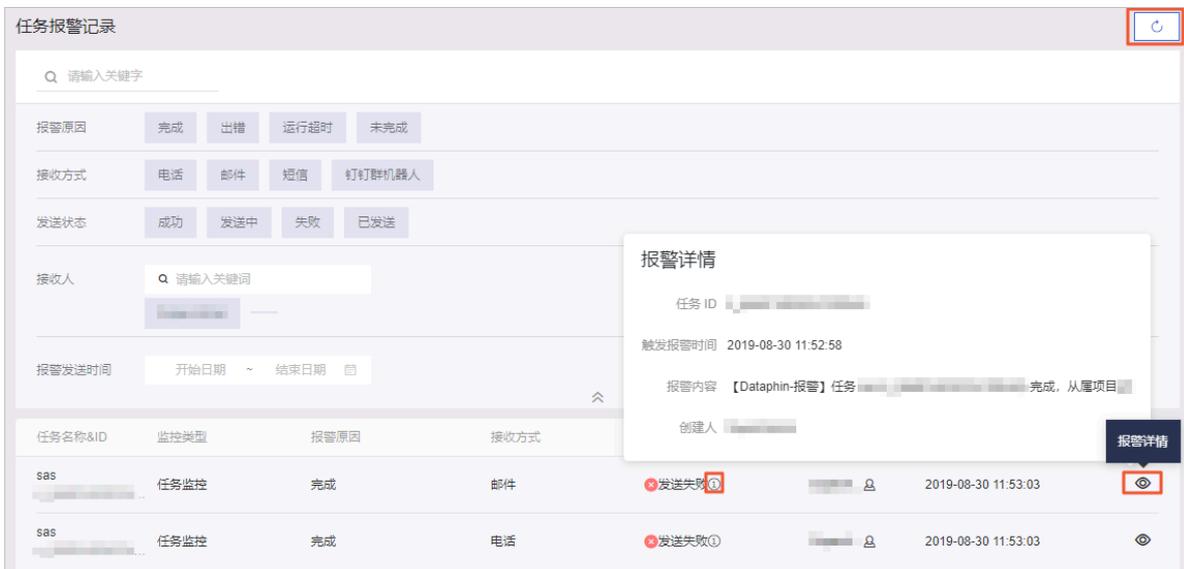
例如未填写手机号码、未填写邮箱地址、未填写钉钉群机器人地址。该租户本日可接收短信、邮件、钉钉群消息已达上限；系统本日可发送短信、邮件、钉钉群消息已达上限；租户列表中找不到指定接收人（已选接收人租户被删除情况）。

- 系统报警内容已按照报警接收方式发出，但系统获取到报错。

例如短信号码为空号、邮箱地址不存在。

- 系统错误，通过查看失败原因可获取具体的错误提示信息。
- 已发送表示系统报警内容已按照报警接收方式发出，但未获取到系统回执时所显示的状态。

3. 下方的报警记录展示区域，展示任务名称&ID、监控类型、报警原因、接收方式、发送状态、接收人、报警推送时间及操作信息。报警记录默认按报警推送时间倒序排列，通过单击可进行调序。



- 单击，可查看新增任务报警记录。
- 单击，可查看任务的报警详情。
- 鼠标悬浮在发送失败旁边的上，可查看失败原因。



说明：

如用户联系方式填写错误或未填写可能导致发送失败，您可以将鼠标悬停于旁边的信息图标可以查看具体原因。

使用限制

各种方式的报警消息推送数量有限。详情请参见 [新建任务监控](#)。