



# 数据智能中台-数据中台

## 用户操作手册

天翼云科技有限公司

# 目 录

---

<b>概述</b> .....	6
<b>业务术语</b> .....	6
<b>数据准备工作</b> .....	9
数据架构定义.....	9
数仓主题规划.....	10
新建主题.....	10
编辑主题.....	11
删除主题.....	12
数仓分层设计.....	12
新建分层.....	13
编辑分层.....	14
删除分层.....	15
公共目录管理.....	16
新增目录.....	16
移除目录.....	16
修改信息.....	17
编辑目录节点.....	17
查询目录.....	20
数据源准备.....	20
数据源接入.....	21
数据源配置操作说明.....	23
数据建模.....	27
数仓规划.....	28
新建模型.....	29
物理建表.....	30
物化物理表.....	37

上线物理表 .....	38
物理表管理 .....	40
<b>数据集成.....</b>	<b>42</b>
数据集成总览 .....	42
任务配置方式 .....	43
离线数据集成 .....	43
附录：支持的内置函数 .....	51
整库同步 .....	58
常见业务场景与最佳实践 .....	65
MySQL 单表离线周期同步至 Hive.....	65
<b>数据研发.....</b>	<b>78</b>
开发功能介绍 .....	79
离线开发 .....	80
脚本开发 .....	80
作业编排 .....	92
程序调试 .....	93
数据查询 .....	93
实时开发 .....	94
可视化开发 .....	94
<b>调度中心.....</b>	<b>97</b>
调度中心概述 .....	97
运维概览 .....	97
任务概览 .....	97
实例概览 .....	99
任务运维 .....	101
周期任务 .....	101
手动任务 .....	103
实时任务 .....	104
实例运维 .....	106
周期实例 .....	106
手动实例 .....	109
实时实例 .....	111

补数实例 .....	114
告警管理 .....	117
监控告警任务页面介绍 .....	117
全局监控管理 .....	119
<b>数据治理 .....</b>	<b>121</b>
元数据 .....	121
元数据总览 .....	122
元模型管理 .....	126
新建模型 .....	130
查看元模型 .....	131
查看元模型的元数据 .....	132
查询模型 .....	134
模型操作 .....	135
采集器管理 .....	137
元数据采集 .....	139
数据标准 .....	144
标准总览 .....	145
字段标准 .....	147
命名标准 .....	169
参考数据管理 .....	175
字段标准稽核 .....	191
稽核结果查询 .....	204
数据质量 .....	207
数据质量总览 .....	207
规则模板 .....	209
任务监控 .....	215
报告查询 .....	227
数据资源管理 .....	230
1.1.1.1.    检索配置 .....	231
1.1.1.2.    数据地图 .....	232
1.1.1.3.    我的数据 .....	237
1.1.1.3.1.    我负责的 .....	238
1.1.1.3.2.    我收藏的 .....	239

<b>数据服务</b> .....	<b>240</b>
服务开发.....	240
创建 API.....	240
向导模式：快速生成数据 API.....	240
脚本模式：自定义 SQL 查询.....	240
测试 API.....	241
使用 API.....	241
服务管理.....	241
API 集市.....	241
应用管理.....	242
<b>平台管理</b> .....	<b>242</b>
我的权限.....	242
成员管理.....	244

# 概述

星海·数据中台是面向数据全生命周期管理的下一代数据中台。通过 AI 大模型技术赋能，为用户一站式提供采、存、管、算、用的大数据能力，降低用户使用门槛，提高数据加工流转效率，帮助客户充分发掘数据价值。本文以星海·数据中台的部分核心功能为例，指导您使用星海·数据中台接入数据并进行业务处理和周期调度。

通过快速入门，您可以完成以下操作：

- **数据建模**：在星海·数据中台的架构规划模块中，完成数据模型的构建。
- **数据同步**：在星海·数据中台的数据集成模块中，创建离线同步任务，完成业务数据入仓入湖。
- **数据开发**：在星海·数据中台的数据开发模块中，对业务数据进行处理和挖掘。
- **任务调度**：为数据同步和数据加工流程配置周期性调度，使其定时执行。
- **数据治理**：在星海·数据中台的数据治理模块中，对业务数据配置质控规则。
- **服务开发**：在星海·数据中台的数据服务模块中，快速完成 API 开发。

# 业务术语

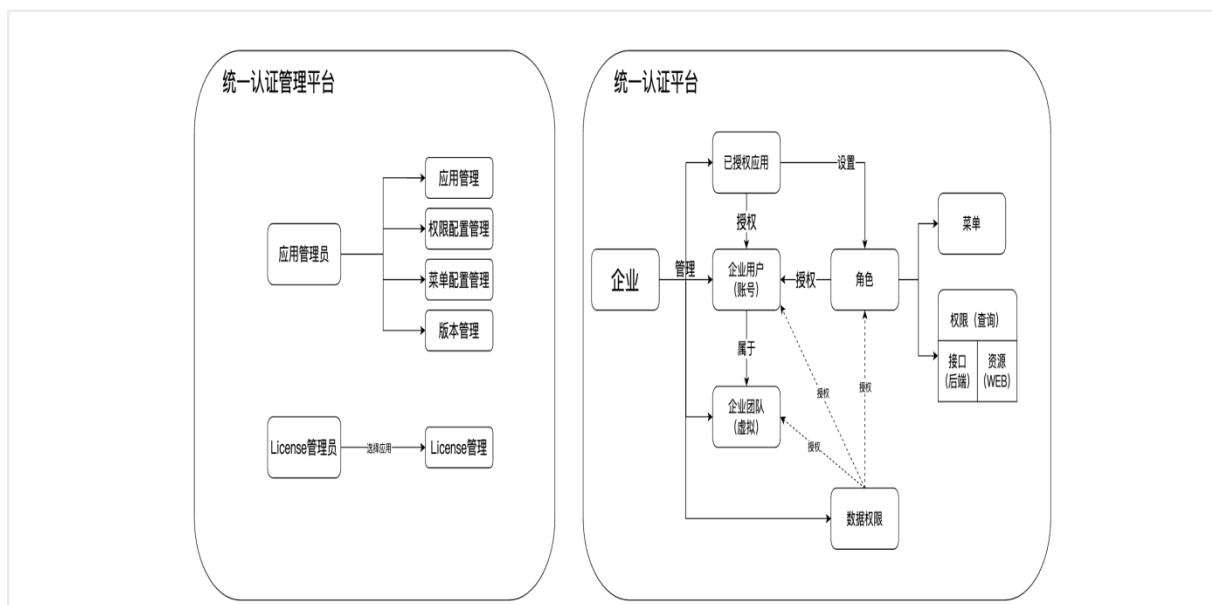
名词	解释
工作空间（团队）	<p>工作空间是从系统层面为管理者提供对使用星海数据中台的用户（成员）权限、资源的管理能力。</p> <p>工作空间作为成员管理、角色和权限分配的基本单元，每个团队都可具有独立的工作空间。</p> <p>您只有在加入工作空间并被分配权限后，才可具备数据架构、数据集成、数据开发、数据治理、数据服务等组件的操作权限。</p>
成员和角色	成员是被授予工作空间访问或使用权限的平台账号。在添加工作空间

	<p>成员时，您需要同时为添加的成员设置相应的角色。</p> <p>角色是一组操作权限的集合。不同的角色拥有不同的操作权限，把角色授予成员后，成员即具有了角色的所有权限。每位成员至少要拥有一个角色，并且可以同时拥有多种角色。</p>
数据源	<p>即数据的来源，本质是存储或处理数据的媒介，比如：关系型数据库、数据仓库、数据湖等。每一种数据源不同，其数据的存储、传输、处理和应用的模式、场景、技术和工具也不相同。</p>
数据连接	<p>定义访问数据实体存储（计算）空间所需的信息的集合，包括连接类型、名称和登录信息等。</p>
脏数据	<p>脏数据是对于业务没有意义或者格式非法的数据。例如，源端是 VARCHAR 类型的数据写到 INT 类型的目标列中，导致因为转换不合理而无法写入的数据。</p>
程序	<p>程序是一种批处理文件的延伸，是一种纯文本保存的程序，一般来说的计算机脚本程序是确定的一系列控制计算机进行运算操作动作的组合，在其中可以实现一定的逻辑分支等。</p>
作业	<p>在数据开发中，作业由一个或多个节点组成，共同执行以完成对数据的一系列操作。</p>
资源	<p>用户可以上传自定义的代码或文本文件作为资源，并在节点运行时调用。</p>
环境变量	<p>环境变量是在操作系统中一个具有特定名字的对象，它包含了一个或者多个应用程序所将使用到的信息。</p>
补数据	<p>手工触发周期方式调度的作业任务，生成过去某时间段内的实例。</p>
主题设计	<p>通过分层架构表达对数据的分类和定义，帮助厘清数据资产，明确业务领域和业务对象的关联关系。</p>
数据标准	<p>数据标准用于描述公司层面需共同遵守的数据含义和业务规则。其描述了公司层面对某个数据的共同理解，这些理解一旦确定下来，就应作为公司层面的标准在公司内被共同遵守。</p>
码表	<p>通常只包括一系列允许的值和附加文本描述，与数据标准关联用于生成值域校验质量监控。</p>
关系建模	<p>关系建模是用实体关系（Entity Relationship，ER）模型描述企业业务，它在范式理论上符合 3NF，出发点是整合数据，将各个系统中的数据以整个企业角度按主题进行相似性组合和合并，并进行一致性处理，为数据分析决策服务，但是并不能直接用于分析决策。</p>
维度建模	<p>维度建模是一种数据仓库设计方法，旨在通过将数据组织成事实表和</p>

	<p>维度表的形式，以支持高效的数据分析和查询。它以业务过程为核心，将数据分解为描述业务事件的<b>事实表</b>和描述业务属性的<b>维度表</b>。事实表存储与业务事件相关的量化数据，如销售额、订单数量等；而维度表则存储与业务相关的描述性信息，如时间、地点、客户等。这种结构使得数据易于理解、查询和分析，尤其适合于数据仓库和商业智能（BI）场景，能够高效支持复杂的多维分析和报告生成。例如，在零售行业，通过维度建模可以快速分析不同时间段、不同地区的销售数据，帮助企业管理者做出更精准的决策。</p> <p>在星海数据中台数据架构中，维度建模是以维度建模理论为基础，构建总线矩阵、抽象出事实和维度，构建维度模型和事实模型，同时对报表需求进行抽象整理出相关指标体系，构建出汇总模型。</p>
元数据	<p>元数据是关于数据的组织、数据域及其关系的信息，简言之，元数据就是关于数据的数据。元数据包括元数据实体和元数据元素。元数据元素是元数据的基本单元，若干个相关的元数据元素构成了元数据实体。</p> <p>在星海·数据中台中，元数据是数据的描述数据，可以为数据说明其属性（数据连接、类型、名称、大小等），或其相关数据（位于拥有者、标签、分类、密级等）。</p>
数据服务	<p>数据服务是基于数据分发、发布的框架，将数据作为一种服务产品提供，满足客户的实时数据需求，能复用并符合企业和工业标准，兼顾数据共享和安全。</p>
API	<p>API（应用程序编程接口）是软件和系统之间进行交互的桥梁，它允许不同的应用程序通过一组定义好的协议和工具来请求服务或交换数据。在数据服务环境中，API 主要用于“读取数据”的操作，即从数据库或数据表中检索信息。开发者可以通过 API 执行查询命令，以获取所需的数据。</p>
统一认证系统	<p>统一身份认证管理平台是一个多功能的在线服务系统，它通过互联网提供存储、计算、网络、数据库、应用托管、开发测试环境、安全措施、备份恢复以及自动化编排等一系列服务，旨在帮助用户灵活地管理和扩展 IT 资源，加速软件开发周期，提高业务效率，并支持各种规模和类型的组织实现数字化转型。</p>

# 数据准备工作

整个统一认证系统的产品架构如图：

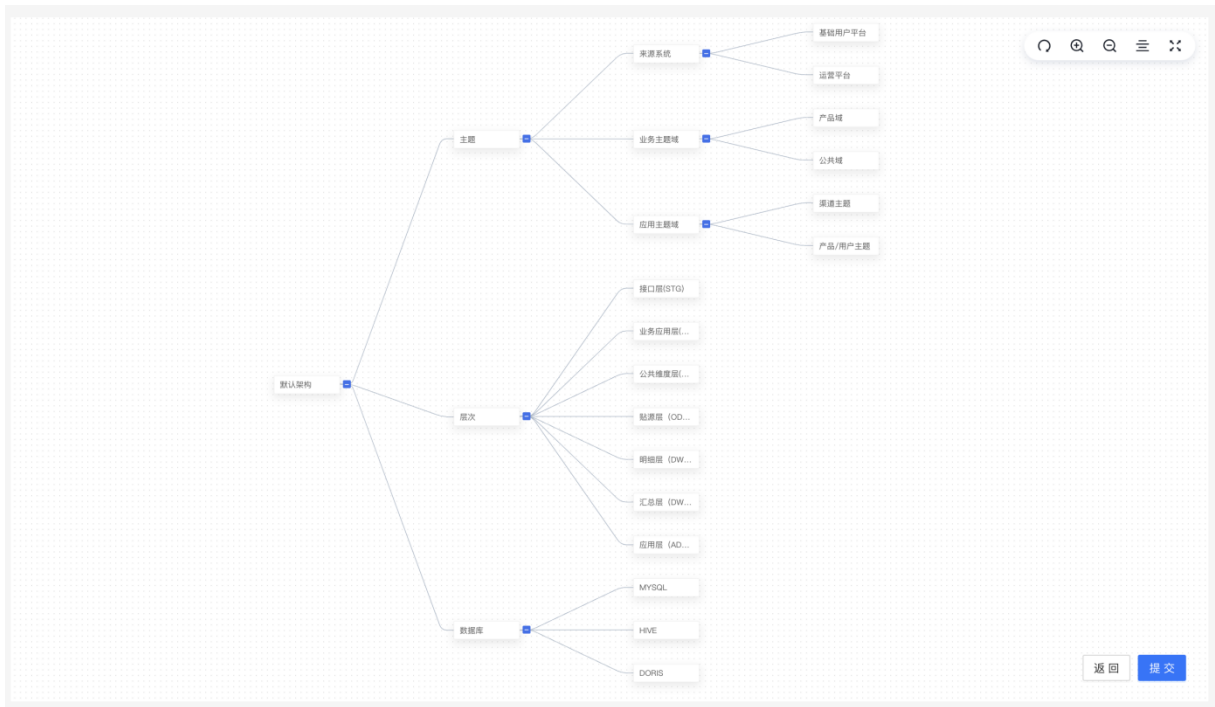


## 数据架构定义

使用星海·数据中台进行数据建模时，数仓架构师或者项目空间成员可以在数仓规划页面对数据分层和主题域进行设计。完成设计后，模型设计师在建模过程中可以依赖数仓规划中的数据分层和主题域对所建模型进行分层划域管理。

星海·数据中台的数据分层和分域管理在[架构规划> 数据架构> 数据架构管理](#)中进行。数据架构通过树的形式在画布中进行展示，数据架构管理中包含数据分层和数据主题，数据架构会默认提供主题和分层两个二级节点。在数据架构画布中您可以对画布进行以下操作：

- 刷新：刷新画布内容；
- 放大：放大画布中的元素显示比例；
- 缩小：缩小画布中的元素显示比例；
- 全屏：全屏显示画布内容。

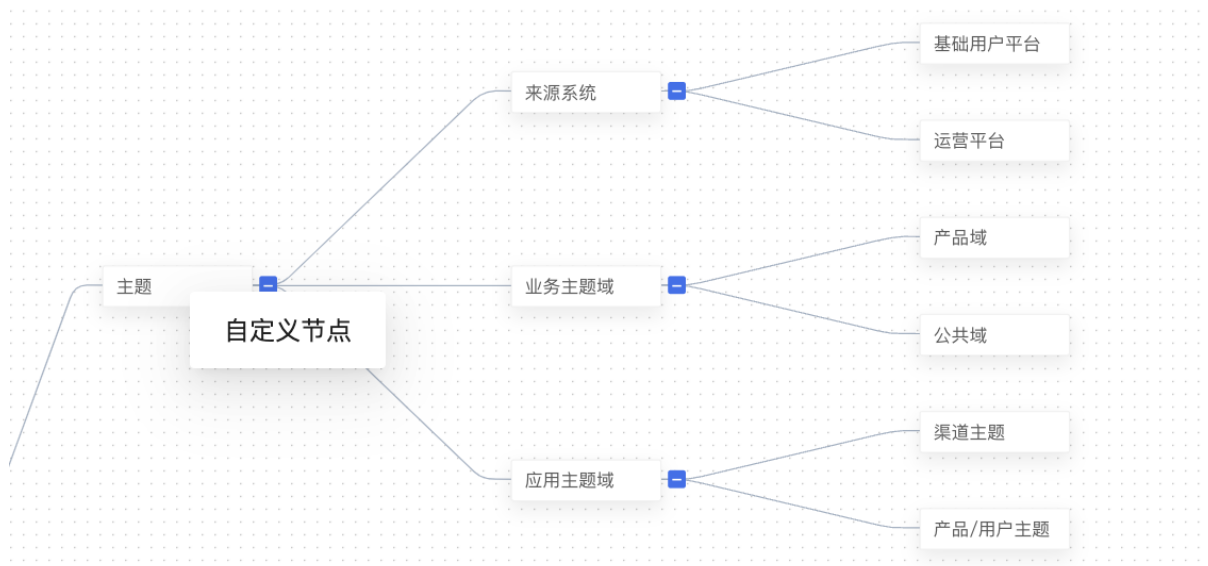


## 数仓主题规划

如您所在的集团或公司业务复杂且较为庞大时，可对数据按照主题进行划分，便于对业务数据进行更加方便的管理。数据主题用于将数据模型按照分析视角进行划分，通常是联系较为紧密的数据主题的集合。您可以根据业务的关注点，将这些数据主题划分至不同的主题域。例如，电商行业通常分为交易域、会员域、商品域等。

### 新建主题

1. 进入数据架构。
2. 登录星海·数据中台，单击顶部菜单架构规划，单击左侧导航栏的数据架构> 数据架构管理。
3. 在数据架构管理页面，选择主题，点击右键> 点击自定义节点，即可新建主题。



4. 在弹出的对话框中填写节点名称和节点描述。

### 新增自定义节点

\* 节点名称:

节点描述:

0 / 200

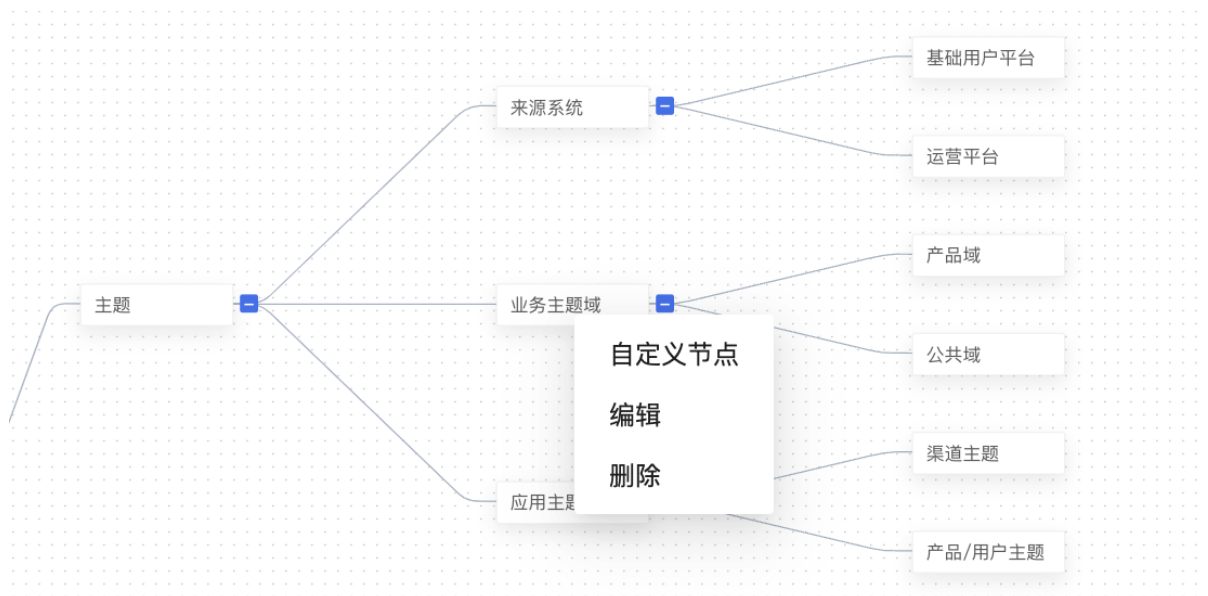
取消

确定

参数	说明
节点名称	主题的名称，20 个字符以内，必须填写。
节点描述	节点的描述信息，200 个字符以内。

### 编辑主题

1. 进入数据架构。
2. 登录星海·数据中台，单击顶部菜单架构规划，单击左侧导航栏的数据架构> 数据架构管理。系统会默认提供一套默认架构
3. 在数据架构管理页面，选择主题，点击右键> 点击编辑节点，即可编辑主题。



4. 在弹出的对话框配置分层的基本信息，配置信息内容同**新建主题**。

## 编辑

\* 节点名称:

节点描述:

0 / 200

## 删除主题

1. 进入**数据架构**。
2. 登录星海·数据中台，单击顶部菜单**架构规划**，单击左侧导航栏的**数据架构> 数据架构管理**。
3. 在**数据架构管理**页面，选择**层次**，单击**右键> 删除节点**，即可**删除主题**。删除主题时请确认数据建模没有引用相关的主题信息。

## 数仓分层设计

数仓分层是结合对业务场景、实际数据、使用系统的综合分析，对数据模型进行整体架构设计及层级划分。用于将不同用途的数据，归类划分至不同的分层，便于您更好地组织、管理、维护数据。

为了便于您快速启动数仓建设，系统按照行业通用规范为您提供了一套数仓默认分层方案。

星海·数据中台的数仓分层功能，默认为您创建了**数据引入层 ODS（Operational Data Store）**、**公共维度层 DIM（Dimension）**、**明细数据层 DWD（Data Warehouse Detail）**、**汇总数据层 DWS（Data Warehouse Summary）**及**应用数据层 ADS（Application Data Service）**等五层，各个分层的功能介绍如下：

- **数据引入层 ODS（Operational Data Store）**

ODS 层用于接收并处理需要存储至数据仓库系统的原始数据，其数据表的结构与原始数据所在的数据系统中的表结构一致，是数据仓库的数据准备区。ODS 层对原始数据的操作具体如下：

- 将原始的结构化数据增量或全量同步至数据仓库中。
- 将原始的非结构化数据（例如，日志信息）进行结构化处理，并存储至 Hive 或其他数据库。
- 根据实际业务需求，记录原始数据的历史变化或对原始数据进行简单的清洗。

- **明细数据层 DWD（Data Warehouse Detail）**

DWD 层通过企业的业务活动事件构建数据模型。基于具体业务事件的特点，构建最细粒度的明细数据表。您可以结合企业的数据使用特点，将明细数据表的某些重要维度属性字段适当冗余，即宽表化处理。同时，也可以减少明细数据表及维度表的关联，提高明细表的易用性。

- **汇总数据层 DWS（Data Warehouse Summary）**

DWS 层通过分析的主题对象构建数据模型。基于上层的应用和产品的指标需求，构建公共粒度的汇总指标事实表。

例如，从 ODS 层中对用户的行为做一个初步的归类汇总，抽象出来一些通用的维度，假设维度为时间、IP、ID，并根据这些维度统计出相关数据，比如用户每个时间段在不同登录 IP 购买的商品数。则在 DWS 层可以进一步添加一层轻度的汇总，可以让计算更加的高效。例如在此基础上计算仅 7 天、30 天、90 天的行为会节省很多时间。

- **应用数据层 ADS（Application Data Service）**

ADS 层用于存放数据产品个性化的统计指标数据，输出各种报表。例如，某电商企业，在 6 月 9 日至 6 月 19 日，杭州地域出售的各大球类商品的数量及排行情况。

- **公共维度层 DIM（Dimension）**

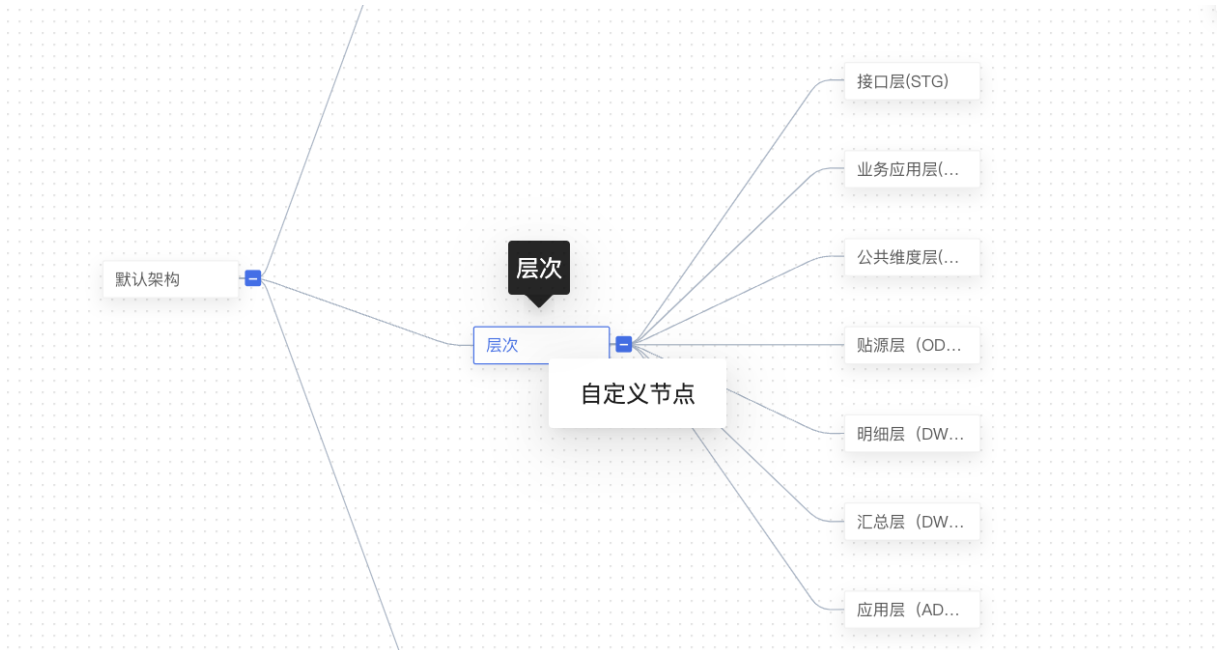
DIM 层使用维度构建数据模型。可基于实际业务，存放逻辑模型的维度表；或存放概念模型的维度定义，通过定义维度，确定维度主键，添加维度属性，关联不同维度等操作，构建整个企业的一致性数据分析维表，帮助您降低数据计算口径和算法不统一的风险。

数仓分层的详细操作参考：

## 新建分层

1. 进入**数据架构**。

2. 登录星海·数据中台，单击顶部菜单**架构规划**，单击左侧导航栏的**数据架构> 数据架构管理**。系统会默认提供一套默认架构
3. 在**数据架构管理**页面，选择**层次**，点击**右键> 点击自定义节点**，即可新建分层。



4. 输入节点（分层）名称和节点描述，点击**确定**，即可完成新增分层。

### 新增自定义节点

\* 节点名称:

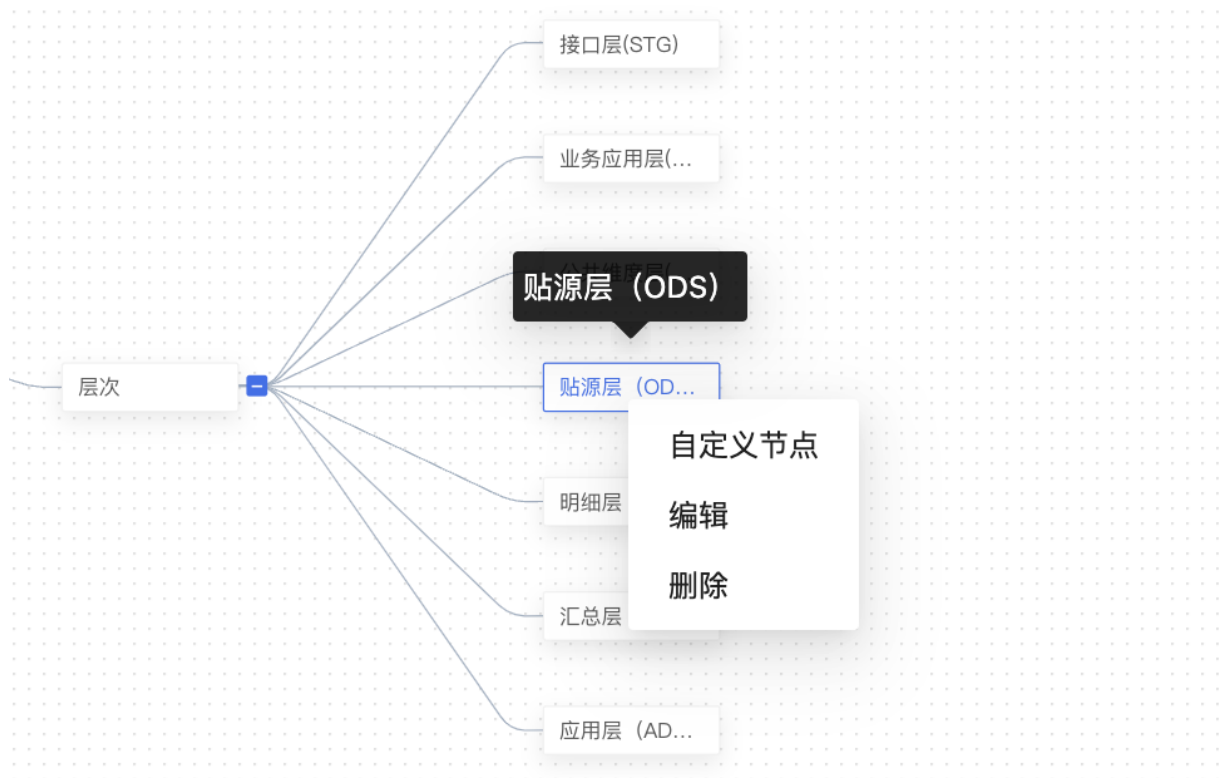
节点描述:

0 / 200

参数	说明
节点名称	节点（分层）的名称，20 个字符以内，必须填写。
节点描述	节点的描述信息，200 个字符以内。

### 编辑分层

1. 进入**数据架构**。
2. 登录星海·数据中台，单击顶部菜单**架构规划**，单击左侧导航栏的**数据架构> 数据架构管理**。系统会默认提供一套默认架构
3. 在**数据架构管理**页面，选择**层次**，点击**右键> 点击编辑节点**，即可编辑分层。



4. 在弹出的对话框配置分层的基本信息，配置信息内容同**新建分层**。

### 编辑

\* 节点名称：

节点描述：

0 / 200

取消

确定

### 删除分层

1. 进入**数据架构**。
2. 登录星海·数据中台，单击顶部菜单**架构规划**，单击左侧导航栏的**数据架构>数据架构管理**。系统会默认提供一套默认架构
3. 在**数据架构管理**页面，选择**层次**，点击**右键>点击删除节点**，即可**删除分层**。删除分层时请确认数据建模没有引用相关的分层信息。

## 公共目录管理

公共目录管理功能主要负责对客户的数据资源进行分类、编目、管理以及提供数据目录的查询服务。

### 新增目录

新增目录，可以点击“新增”按钮，进行新增目录。新增目录需要输入目录编码、选择目录类型和输入目录名称。目录编码和目录类型只能在新增时进行输入，后续将无法修改目录编码和目录类型。



目录类型可选择固定的物种类型，分别是：

模型开发目录：对数据模型进行系统的组织和分类管理；

程序开发目录：组织和管理与程序开发相关的各种资源和组件；

数据源目录：管理和组织所有接入数据中台的数据源；

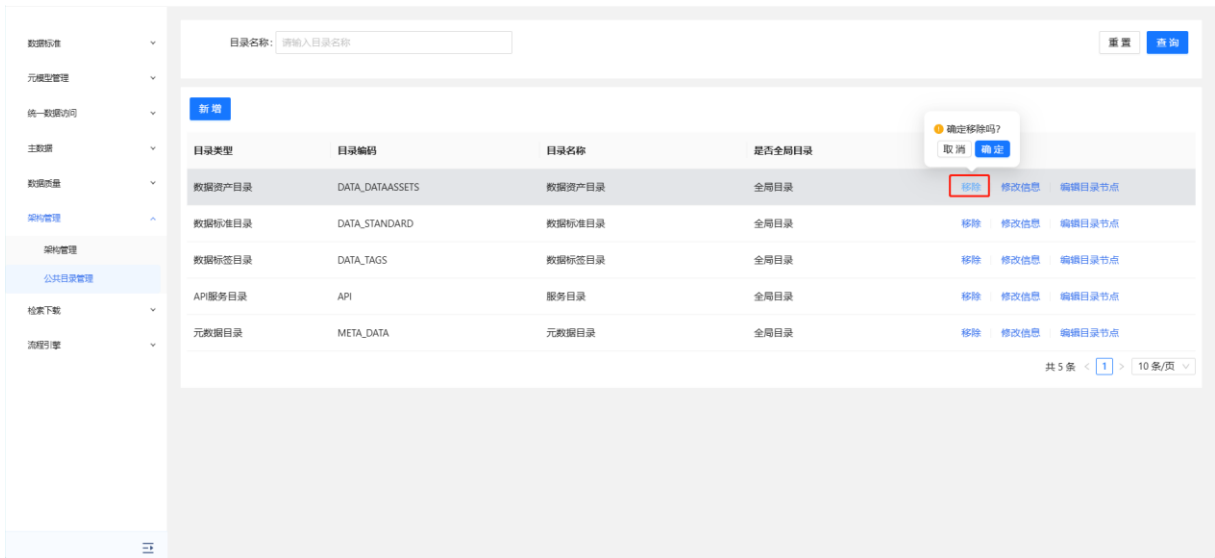
私有数据目录：内部管理和组织数据资源；

调度管理目录：管理和监控数据处理作业的调度和执行。

信息填充完成后，点击“确定”，新增的目录将呈现在目录列表里。

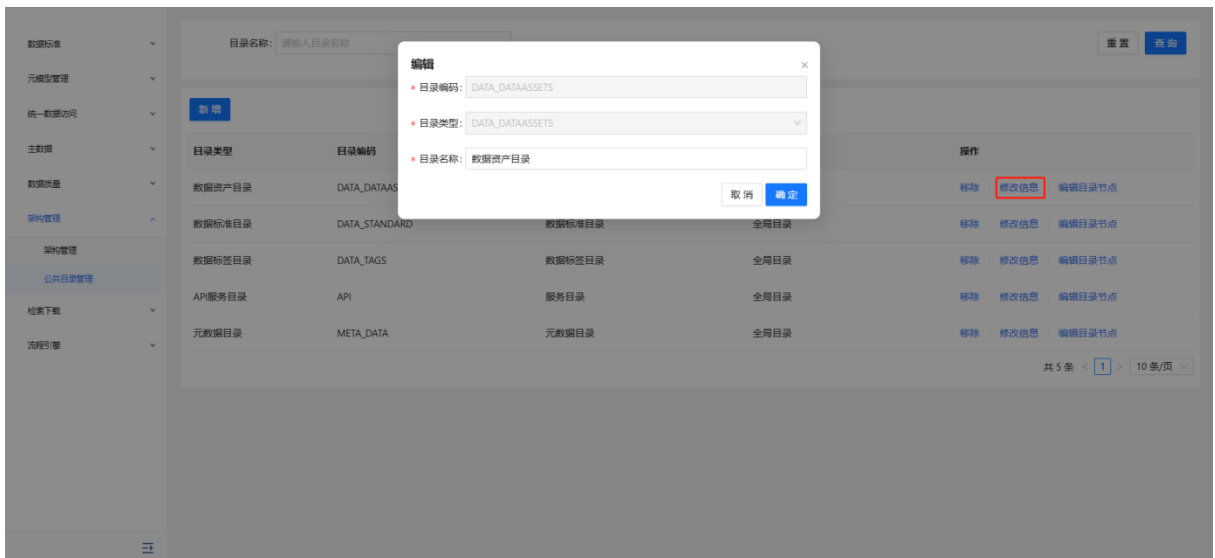
### 移除目录

移除当前目录，可点击目录列表右侧操作列“移除”按钮进行移除。移除需要二次确认。



## 修改信息

如需修改当前目录信息，可点击“修改信息”按钮，修改目录名称后，点击“确定”，保存修改信息。



## 编辑目录节点

修改当前目录中的目录节点，可点击“编辑目录节点”按钮，进入编辑画布界面进行编辑。

目录类型	目录编码	目录名称	是否全局目录	操作
数据资产目录	DATA_DATAASSETS	数据资产目录	全局目录	删除 修改信息 <b>编辑目录节点</b>
数据标准目录	DATA_STANDARD	数据标准目录	全局目录	删除 修改信息 编辑目录节点
数据标签目录	DATA_TAGS	数据标签目录	全局目录	删除 修改信息 编辑目录节点
API服务目录	API	服务目录	全局目录	删除 修改信息 编辑目录节点
元数据目录	META_DATA	元数据目录	全局目录	删除 修改信息 编辑目录节点

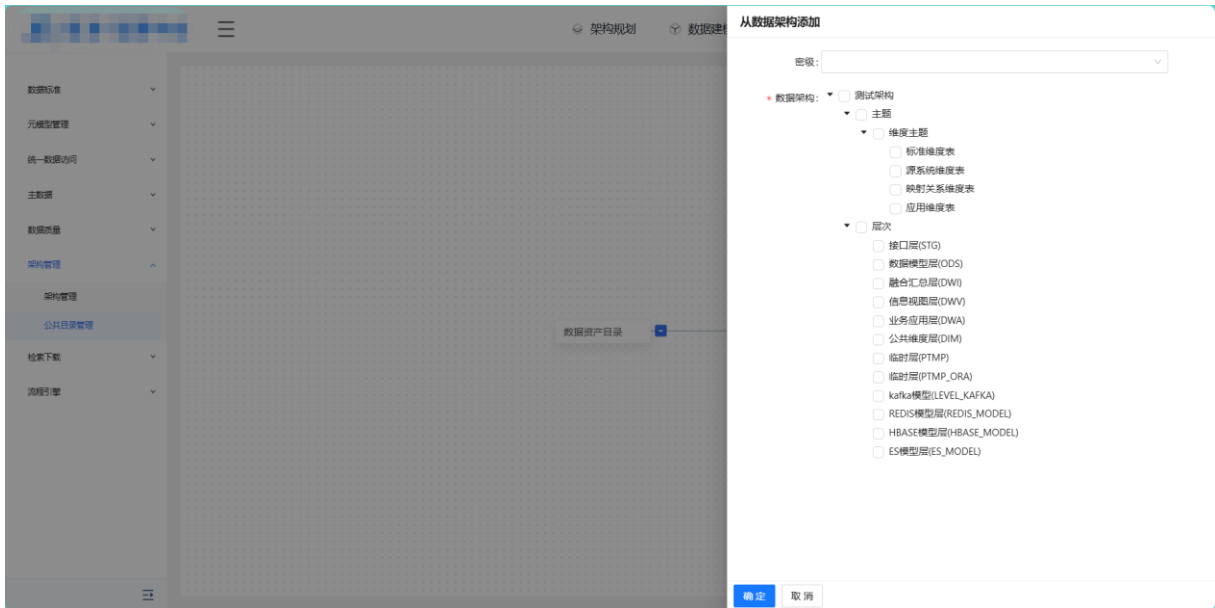
进入画布后，可对目录节点进行配置。右键目录节点，即可进行新增自定义节点、数据架构添加、移动、编辑和删除。值得注意的是，主节点无法进行删除。

新增自定义节点需要输入节点编码、节点名称，选择密级，输入节点描述，并且支持对该节点添加自定义属性，该属性会跟随该节点，并进行展示。

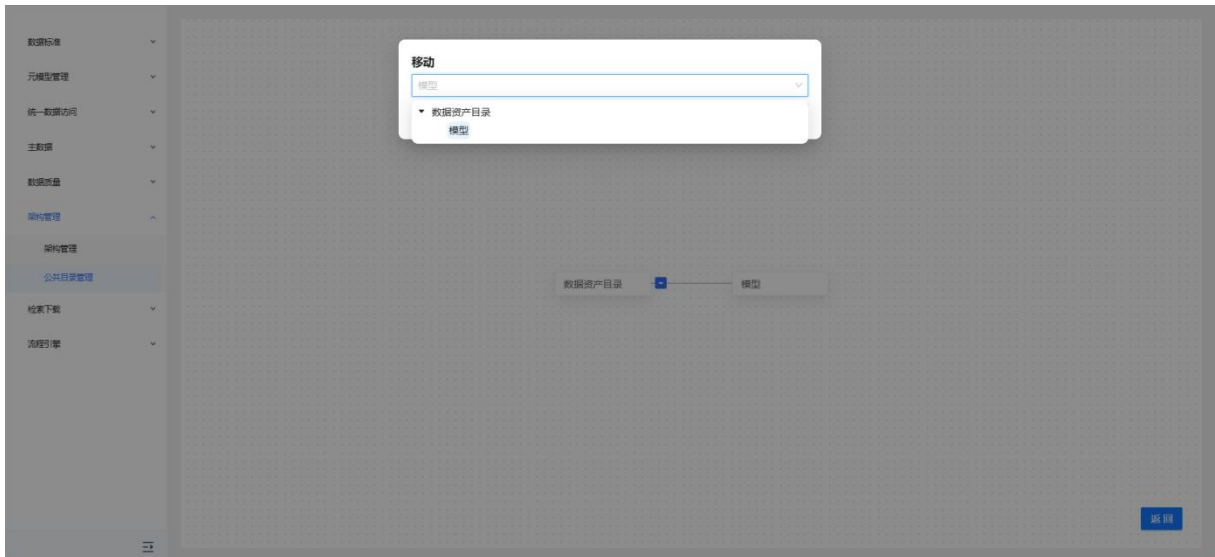
密级共有公开、内部、秘密、机密、绝密五个等级可供选择。

信息输入完成后，即可增加子节点。

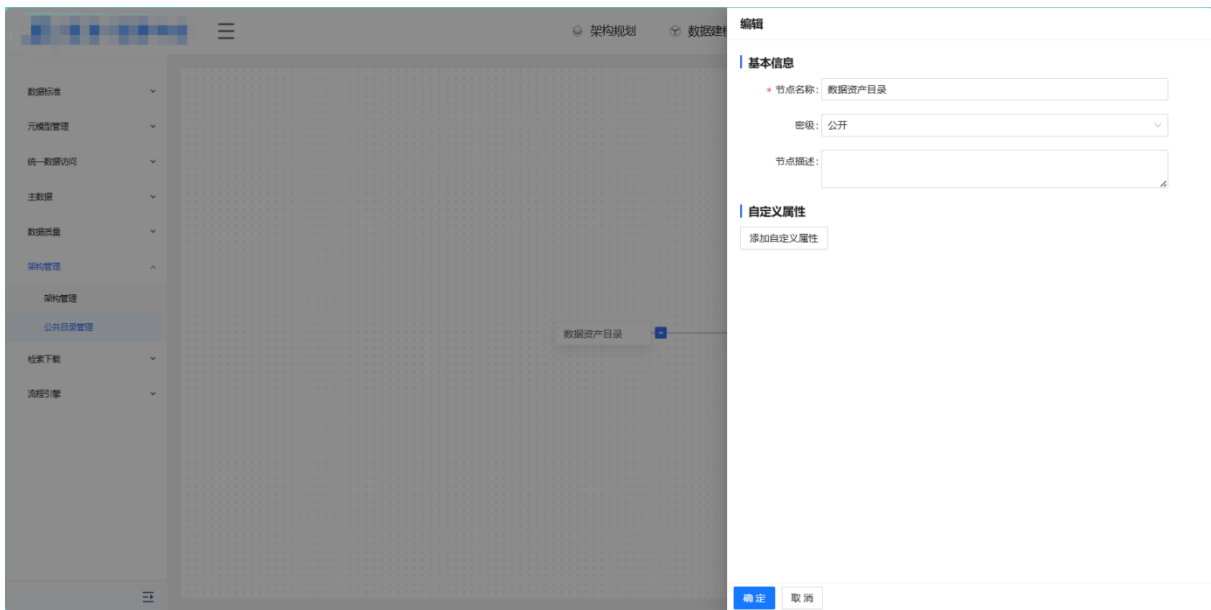
数据架构添加支持从数据架构中选择节点添加。数据架构为架构管理中已经启用的架构。支持多选，选择完成后即可增加所选的子节点。



移动功能支持将该节点移动至其他节点中。

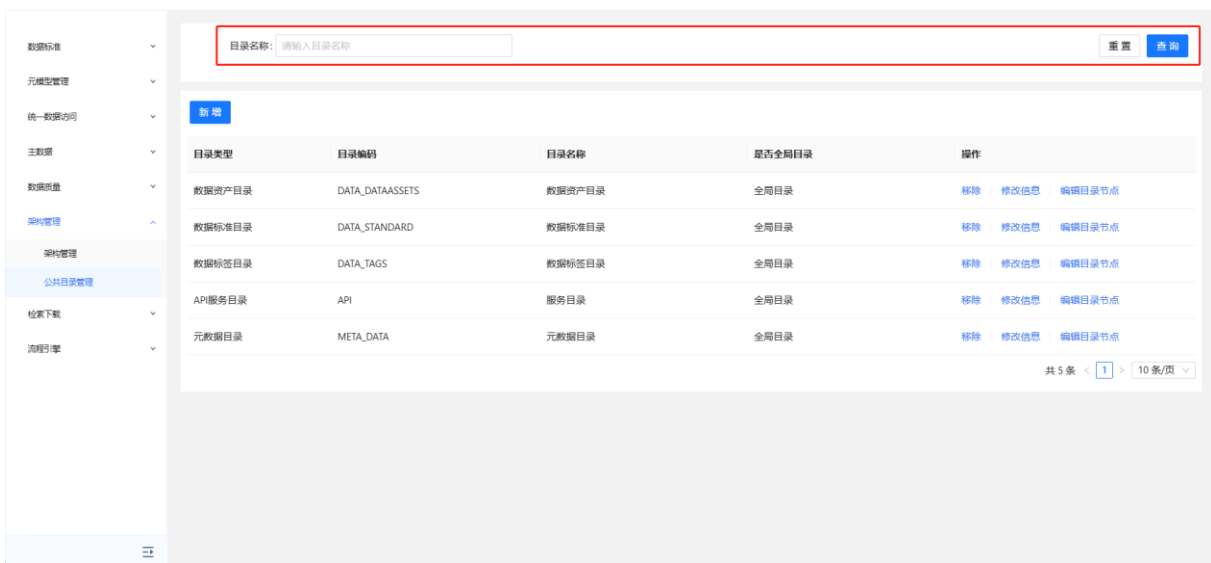


节点编辑功能可以修改节点名称、密级和描述，并支持添加自定义属性。



## 查询目录

通过目录名称，即可检索和定位目录。



## 数据源准备

若您想在星海·数据中台操作您的数据库或数据仓库，需先在星海·数据中台的数据源管理页面，将数据库或数据仓库创建为星海·数据中台的数据源，并将该计算资源绑定至星海·数据中台相应的功能模块使用。例如，同步 doris 数据前，需先创建 doris 数据源，创建完成后可在数据集成模块选择该数据源，控制同步任务读取或写入的数据库或数据仓库。

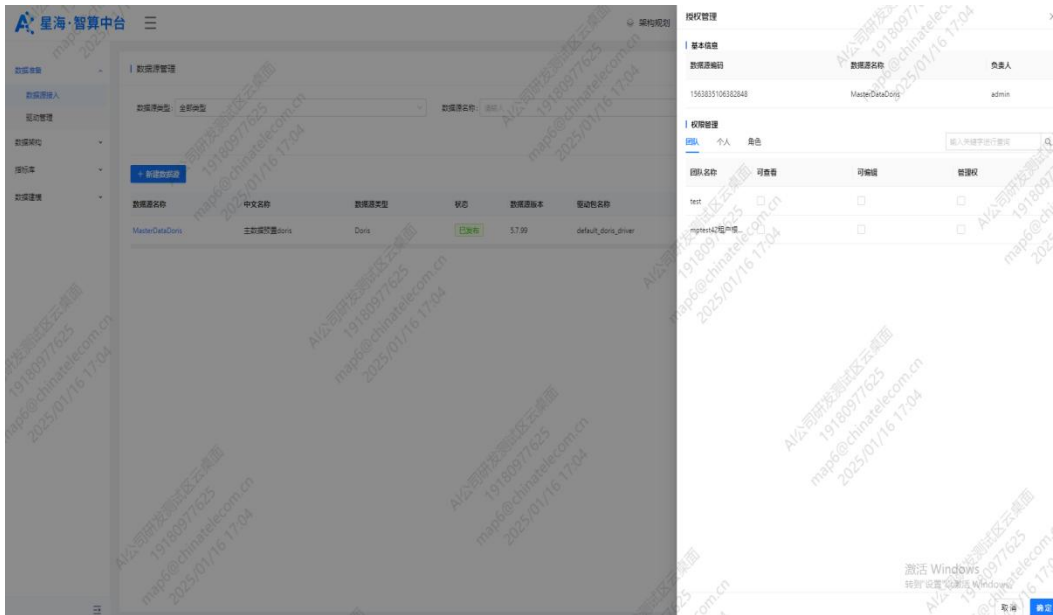
当您建立了自己的数据源之后，就拥有了当前数据源的管理和授权权限（不包括数据权限），对于数据源来说，权限主要有两个层面。

- 已公开的数据源

已公开的数据源，对于其他租户不可见，对于当前租户下的所有团队的所有成员可见。

- 未公开的数据源

未公开的数据源，需要 root 用户进行授权使用，授权管理页面如下：



可以分别给团队、个人以及角色进行授权，具体权限及说明如下：

权限	说明
可查看	可以查看并使用数据源。
可编辑	可以查看，编辑数据源的信息。
管理权	可以对数据源进行所有的操作，包括数据源的发布、公开、删除以及授权。

## 数据源接入

目前支持的数据源类型如下表

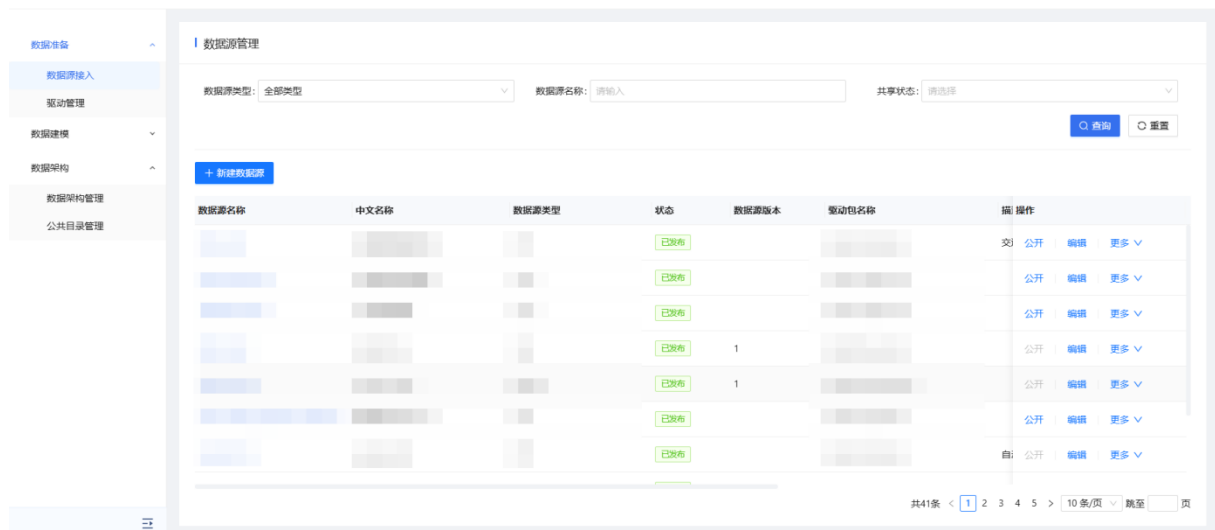
序号	数据源类型	是否提供内置驱动文件
1	MySQL	是
2	Oracle	是
3	DM	是
4	PostgreSQL	是
5	SQLSever	是
6	KingBase	是
7	DB2	是
8	Hana	是
9	OceanBase	是

10	GuassDB	是
11	ClickHouse	是
12	Doris	是
13	GBase	是
14	Vertica	是
15	Http	是
16	Webservice	是
17	Https	是
18	NebulaGraph	是
19	Neo4j	是
20	Ceoh	是
21	Kafka	是
22	FTP	是
23	SFTP	是
24	RTMP	是
25	RTSP	是
26	M3U8	是
27	Redis	是
28	HBase	是
29	MongoDB	是
30	Phoenix	是
31	OpenTSDB	是
32	InfluxDB	是
33	ES	是
34	Hive	是
35	HDFS	是
36	Hudi	是
37	Impala	是
38	SparkSQL	是
39	Presto	是
40	Paimon	是
41	Hadoop	是
42	JuiceFS	是
43	M3U8	是
44	Kyuubi	是
45	Milvus	是

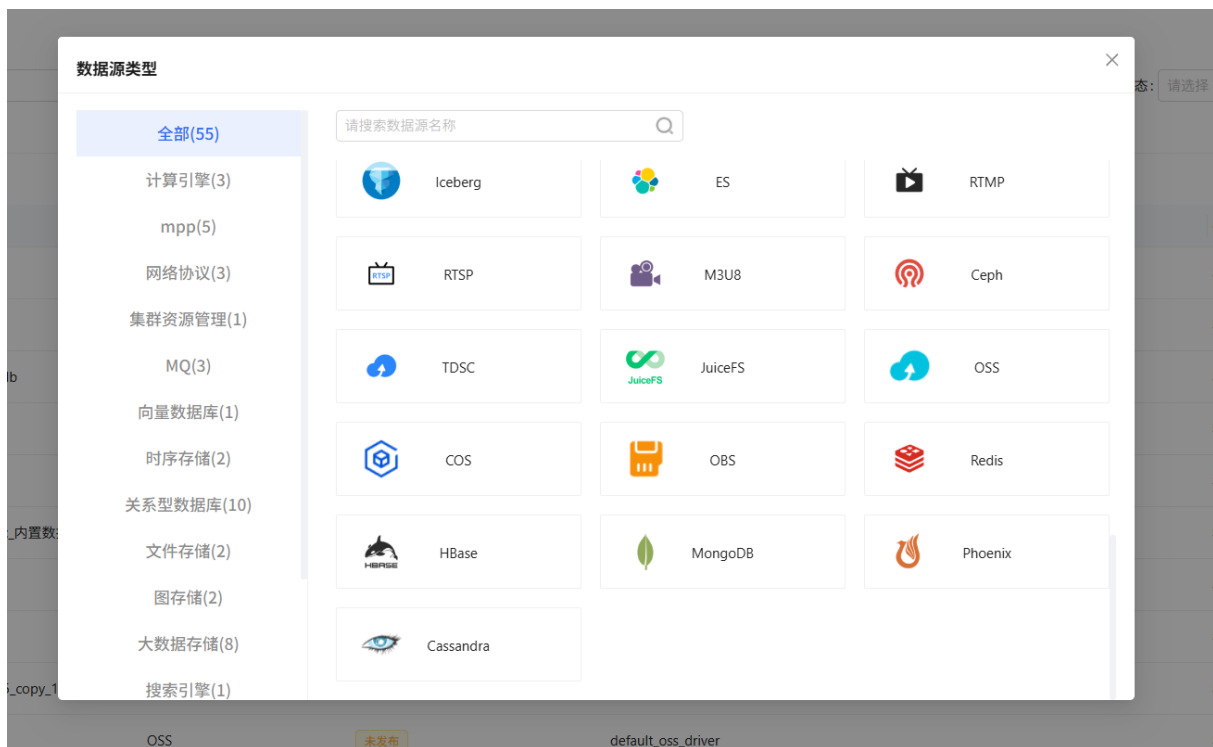
46	Cassandra	是
47	Neo4j	是
48	Ceph	是
49	OBS	是
50	COS	是
51	TDSC	是
52	OSS	是
53	RocketMQ	是
54	RabbitMQ	是
55	Iceberg	是

## 数据源配置操作说明

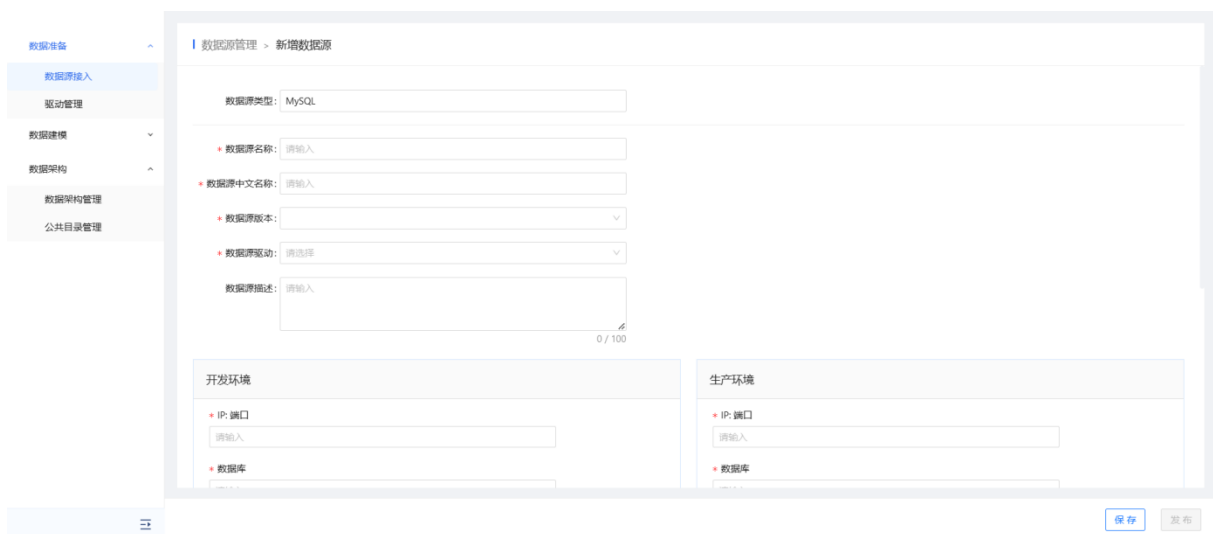
步骤一：在数据源接入管理页面，点击新建数据源，选择需要新建的数据源类型



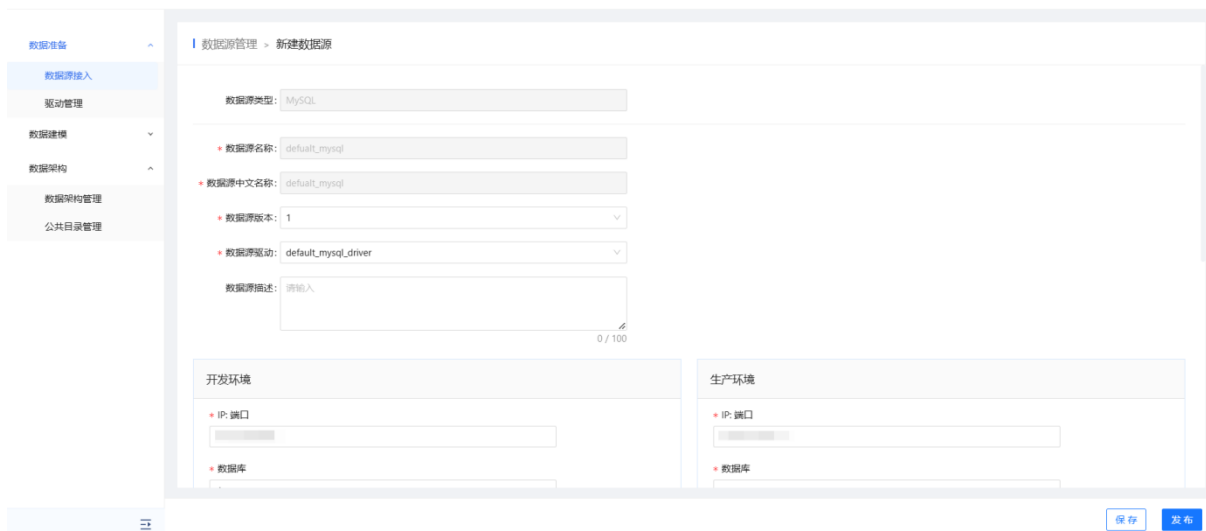
在数据准备-->数据源接入页面，点击【新建数据源】按钮，弹出数据源类型选择框。



如选择 MySQL 类型，将会跳转至数据源配置页面。不同的数据源类型，其配置项也不同。



## 步骤二：配置数据源基础信息



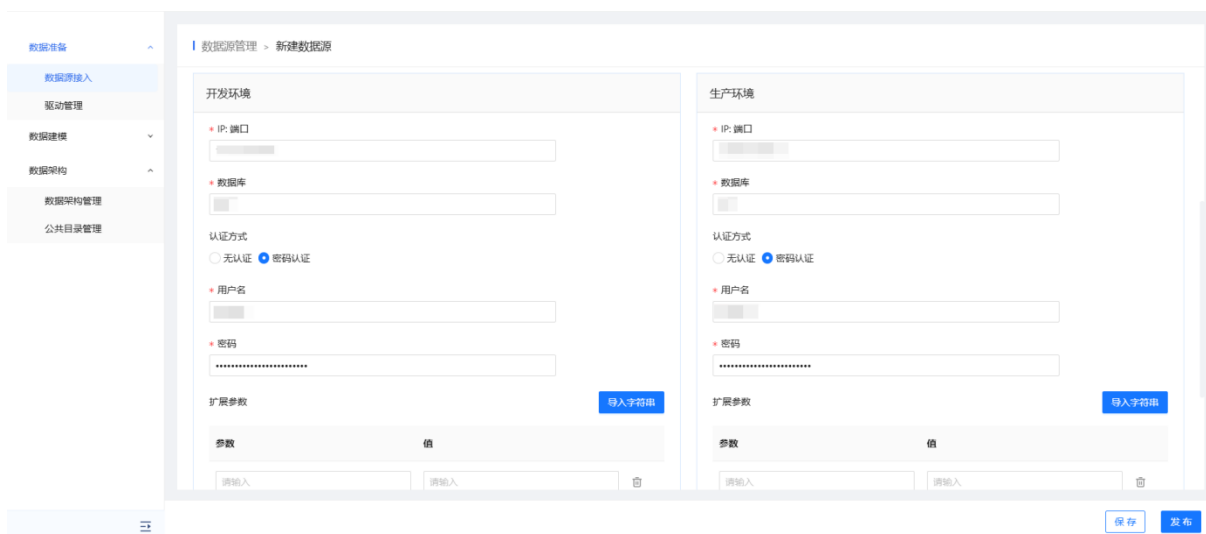
数据源基础信息包括以下内容：

配置项	配置说明
数据源名称	必填项，用户自定义的数据源英文名称，数据源英文名称有唯一性校验，数据源发布后不可更改。
数据源中文名称	必填项，用户自定义的数据源中文名称，数据源发布后不可更改。
数据源版本	数据源所属的版本信息，下拉选项，用于标识该数据源的版本。
数据源驱动	用于数据源连通的驱动文件包，驱动文件包的创建详见【2.3.5.1 驱动管理操作说明】
数据源描述	数据源的补充信息

### 步骤三：配置数据源环境信息

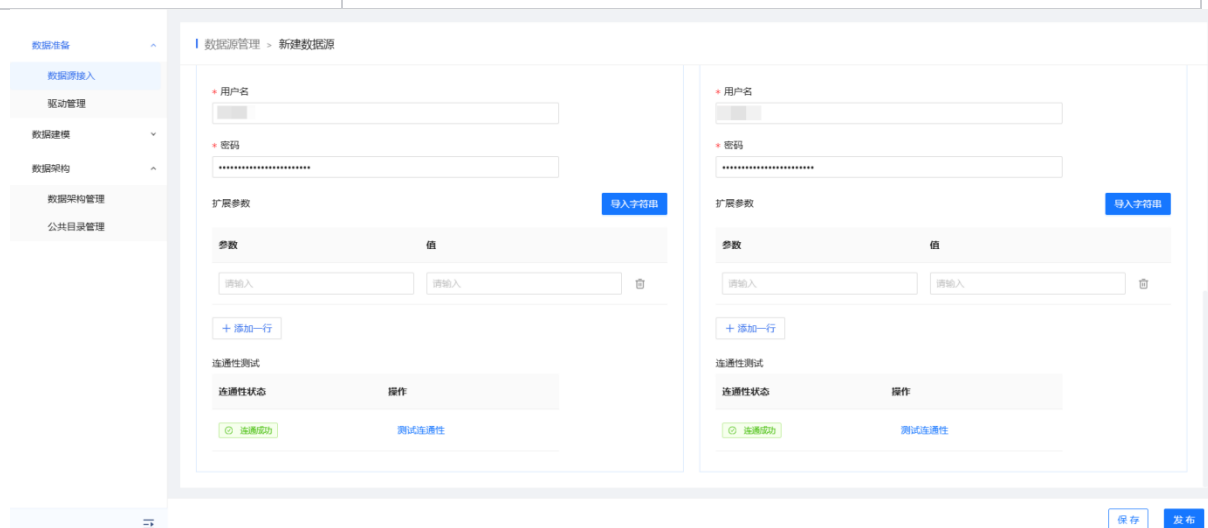
每个数据源分为开发环境与生产环境两个环境。每个类型的数据源的配置项也不同。

两个环境的配置项相同，配置内容可以相同，也可以不相同。



以开发环境配置为例：

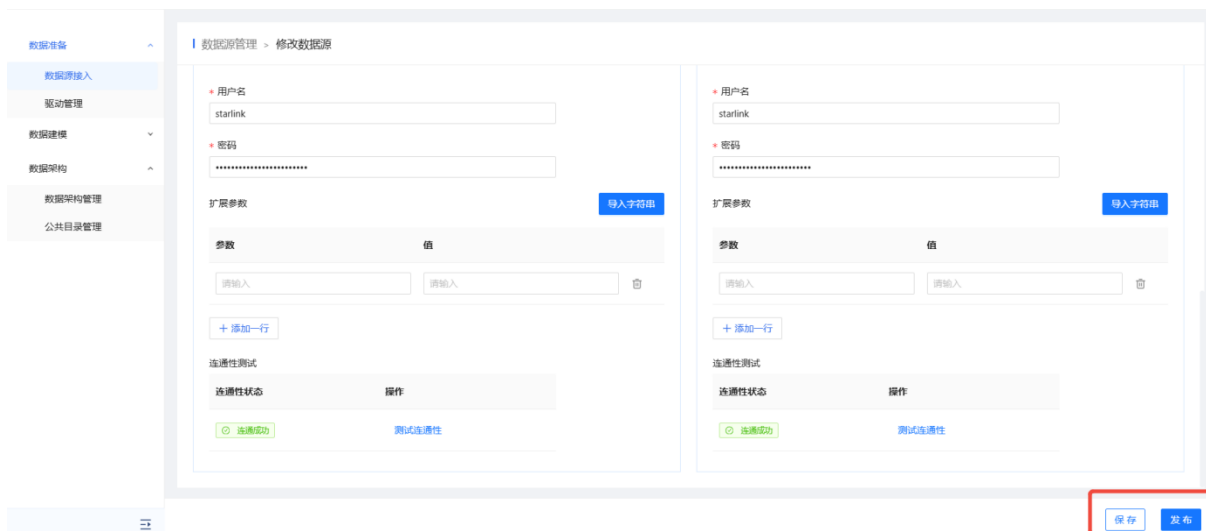
配置项	配置说明
IP: 端口	以 JDBC 连接方式进行连接的数据源类型，需要配置其 IP 地址及对应端口。
数据库	用于区分数据源开发环境与生产环境的数据库名
认证方式	数据源安全保障机制，根据不同的数据源类型提供账密认证、kerberos 认证等方式
用户名	若认证方式为账密认证，此处填写账户用户名
密码	若认证方式为账密认证，此处填写账户对应的密码
扩展参数	若连接数据源需要除上述配置项以外的配置项，在此处进行配置，以 K-V 键值对的方式进行配置的输入。



#### 步骤四：连通性测试

环境配置完成后，支持进行连通性测试，验证配置的正确性。

#### 步骤五：数据源发布



当数据源连通性测试完成后，可以进行数据源的发布，发布后的数据就可以在后续模块中进行使用了。

## 数据建模

随着业务的快速发展，企业数据呈几何倍增长，数据量庞大、复杂、各类数据间标准不一致，往往会出现数据难以管理的现象。星海·数据中台智能数据建模服务，将无序、杂乱、繁琐、庞大且难以管理的数据，进行结构化有序的管理。使企业中的数据产生更多的价值，将数据价值最大化。

### 前提条件

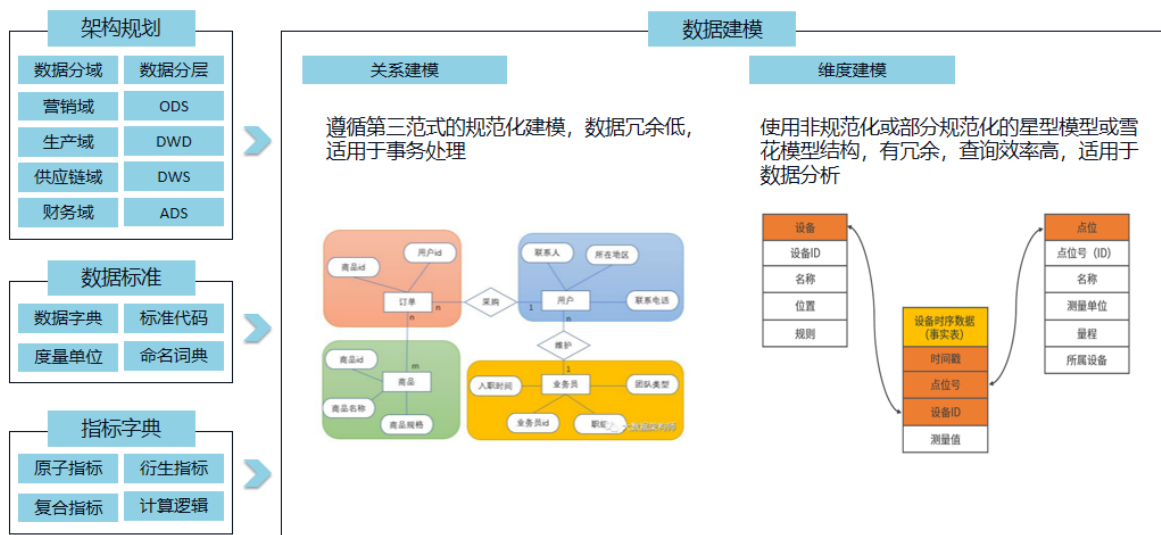
- 需要完成主题规划和数仓分层设计信息配置后，才能进行物理建模操作，主题划分和数仓分层设计请参见 2.3.1 数仓主题规划和 2.3.2 数仓分层设计。
- 需要完成数据源接入，才可进行物理建模操作。

### 使用限制

- 已拥有数据建模角色权限。
- 已拥有所需数据源使用权限。

### 概述

星海·数据中台数据建模支持数仓规划设计、制定并沉淀企业数据标准、关系建模、维度建模、数据指标定义，通过使用星海·数据中台数据建模，您可以将建模设计产出的维度表、明细表和汇总表物化到计算引擎中并进一步应用。



## 数仓规划

使用星海·数据中台进行数据建模时，您可以在数仓规划页面进行数据分层、业务分类、主题域的规划设计。

### ● 数据分层

您可以结合业务场景、数据场景综合考虑设计数仓的数据分层，星海·数据中台为您默认创建业界通用的五层数仓分层：

- 数据引入层 ODS（Operational Data Store）
- 明细数据层 DWD（Data Warehouse Detail）
- 汇总数据层 DWS（Data Warehouse Summary）
- 应用数据层 ADS（Application Data Service）
- 公共维度层 DIM（Dimension）

### ● 数据主题

当企业业务比较复杂，不同类型业务彼此间需要共享数据域，但是又希望能在模型设计和应用过程中快速定位本业务的数据时，您可结合真实业务情况，规划不同的数据分类，在后续建模的维度表和明细表中，将其关联到对应的数据主题中。

## 数仓规划

我们采用了分层式数据仓库架构，通过对各个数据层级中的模型信息进行系统化的管理和标准化处理，实现了数据模型的一致性和规范性。此外，借助于模块化设计理念，该架构不仅能够支持用户依据具体的业务需求灵活地扩展或调整模型，同时也保证了整个数据模型的全局一致性和易于维护性。这样的设计方式极大地提高了系统的适应性和灵活性，同时确保了数据模型在长期运行中的稳定性和高效性。

## 前提条件

- 需要完成主题规划和数仓分层设计信息配置后，才能进行物理建模操作，主题划分和数仓分层设计请参见 2.3.1 数仓主题规划和 2.3.2 数仓分层设计。
- 需要完成数据源接入，才可进行物理建模操作。

### 使用限制

- 已拥有数据建模角色权限。
- 已拥有所需数据源使用权限

### 新建模型

数仓分层的模型支持用户根据实际业务场景进行自定义。具体操作如下：

1. 进入数仓规划主页面。
2. 在数据数仓规划控制台，点击切换的“物理表 > 维度模型”。
3. 单击上面的“+新建”，进入“新建模型”页面。
4. 配置模型相关参数。
5. 维度模型创建

### 维度模型

\* 模型名称:  0 / 40

\* 数据库类型:

\* 层次:

描述:  0 / 300

参数

说明

*模型名称	定义模型名称。只能包含中文、英文字母、数字和下划线。
数据库类型	定义模型所对应的数据库类型。
层次	选择当前模型所处的层次。
描述	数仓模型描述信息。支持的长度 0~600 字符。

## 6. 逻辑模型创建

### 逻辑模型

\* 模型名称:

 0 / 40

描述:

 0 / 300

取消

确定

- 单击“确定”。模型新建完成。
- 更多操作如下：单击已新建的数仓模型右侧的“编辑”，可以修改模型的信息，
- 单击“删除”，可以删除数仓模型。删除操作无法恢复，请谨慎操作。如果模型包含实体，无法删除。
- 单击模型的“数据表”或“字段”可以跳转到对应的“物理建表”、“维度建表”页面。

## 物理建表

### 物理表

物理表是指按照一定规则和方法，将物理表中所定义的实体、属性、属性约束、关系等要素转换为数据库软件所能够识别的表关系图(Table Relationship Diagram)的一种物理描述。

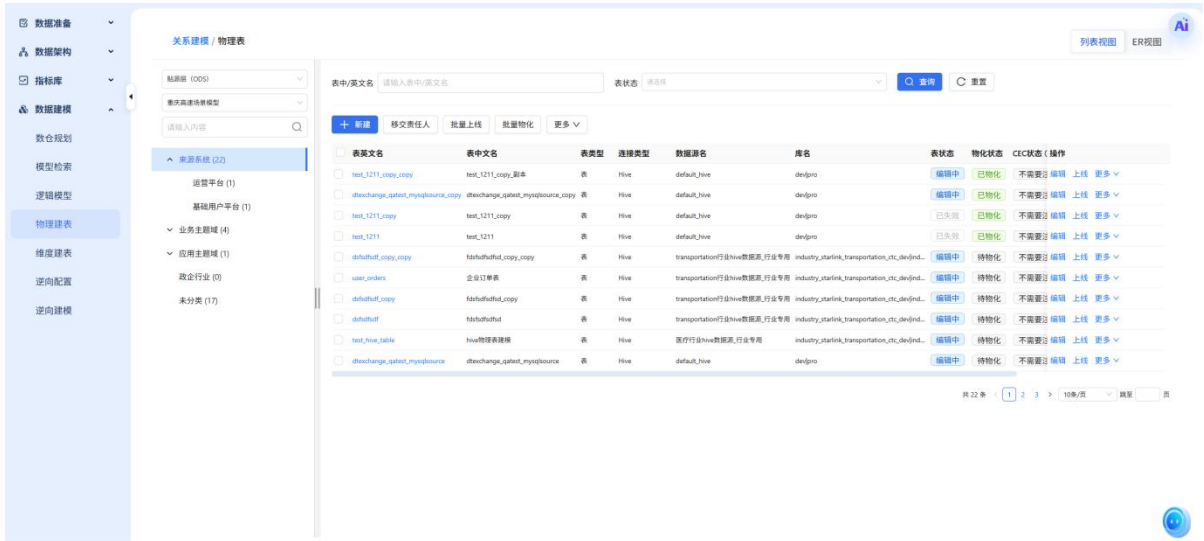
可视化建模：

## 前提条件

- 需要完成主题规划和数仓分层设计信息配置后，才能进行物理建模操作，主题划分和数仓分层设计请参见数仓主题规划与数仓分层设计。
- 需要完成数据源接入，才可进行物理建模操作。

## 使用流程

- 1、登录数据中台，单击上侧导航栏的**数据建模>物理建表**，进入物理表页面。



- 2、创建物理表

a.在物理表页面，鼠标点击 **+ 新建** 图标，创建物理表。

- 3、在“新建物理表”页面，根据页面提示完成相关配置。

a.填写基本配置参数。



参数名称	说明
------	----

● 所属主题	选择所属的主题信息
● 数据源	选择处于已发布状态的数据源
● 表中文名	以中文、英文、数字、下划线，且以中文或英文开头,最少 4 字符,最多 64 字符
● 表英文名	以英文、数字、下划线组成，且以英文开头,最少 4 字符,最多 64 字符表英文名
表类型	<p>物理表支持以下表类型：</p> <p>表：表是数据库中用于存储数据的物理结构，它由行和列组成，每一行代表一个记录，每一列代表一个字段。当需要长期存储数据，且数据需要频繁更新和查询时，选择表。</p> <p>视图：视图是基于一个或多个表的虚拟表，它不存储数据，而是通过 SQL 查询语句动态生成数据，需要提供数据的抽象层，简化用户查询时，选择视图。</p>
关联逻辑模型	选择物理表所需关联的逻辑模型
关联逻辑实体	关联当前【关联逻辑模型】中生效的逻辑实体
分区类型	<p>当数据库类型是 DM、可选如下分区类型：</p> <ul style="list-style-type: none"> <li>○ 范围分区（Range Partition）：是根据表中某个列或多个列的值的范围来创建分区。应用场景为时间序列数据和数值范围数据。</li> <li>○ 列表分区（List Partition）：列表分区是通过指定表中某个列的离散值集来创建分区，应用场景为地区或分类数据、状态或类型数据</li> </ul> <p>当数据库类型是 GaussDB、可选如下分区类型</p> <ul style="list-style-type: none"> <li>○ Range：根据分区键的值范围将数据映射到分区，适用于时间序列数据，如按日期范围分区。</li> <li>○ list：通过指定分区键的离散值列表来控制数据映射到分区，适用于已知离散值的场景，如按特定分类或状态分区。</li> <li>○ hash：基于分区键使用哈希算法将数据均匀映射到分区，适用于数据分布均匀且无明显分区键的场景。</li> </ul> <p>当数据库类型是 Doris、可选如下分区类型</p> <p>Range（范围分区）：根据指定列的值范围进行分区，适用于数据具有连续数值范围的场景。</p> <p>List（列表分区）：根据指定列的离散值列表进行分区，适用于数据具有离散值集合的场景。</p>
● 是否外部表	<p>当数据库类型是 Hive、需选是否外部表信息</p> <p>用于指定 Hive 表是否为外部表。外部表允许用户访问存储在 Hive 外部的文件系统中的数据，而不必将其移动到 Hive 的默认存储位置。</p>
● 文件格式	<p>当数据库类型是 Hive、需选文件格式信息</p> <ul style="list-style-type: none"> <li>○ RCFILE：列式存储格式，支持压缩，适用于 Hive 分析。</li> </ul>

	<ul style="list-style-type: none"> <li>○ TEXTFILE: 行式存储, 默认编码 UTF-8, 简单通用。</li> <li>○ GBK_TEXTFILE: 行式存储, GBK 编码, 适合中文数据。</li> <li>○ SEQUENCEFILE: 二进制存储, 支持压缩, 效率高。</li> <li>○ PARQUET: 列式存储, 支持复杂数据类型, 查询优化。</li> <li>○ ORC: 列式存储, 高效压缩, 适合大规模数据分析。</li> </ul>
● 列分隔符	<p>当数据库类型是 Hive、需定义列分隔符信息</p> <p>指定 Hive 表中字段值之间的分隔方式, 以便正确解析行式存储文件中的数据</p>
压缩类型	<p>当数据库类型是 Hive、需定义压缩类型信息</p> <p>压缩类型配置用于指定数据存储时采用的压缩算法, 以减少存储空间和提高 I/O 效率</p> <p>当数据库类型是 KingBase、需选压缩类型信息。</p> <ul style="list-style-type: none"> <li>○ (5, 5)压缩算法: 性能与压缩比相对均衡的压缩模式</li> <li>○ (3, 1)压缩算法: 压缩率最大的压缩模式。</li> </ul>
表类型	<p>当数据库类型是 KingBase、需选择如下表类型。</p> <ul style="list-style-type: none"> <li>○ 复制表: 数据完整存放在各个主机上, 适用于配置表、字典表等数据量小的表。</li> <li>○ 哈希表: 数据按 hash 策略分布存储与不同主机上, 以实现分布式存储和计算</li> </ul> <p>当数据库类型是 Doris、需选择如下表类型。</p> <p><b>Aggregate:</b> 聚合表, 用于存储聚合后的数据, 适用于快速查询统计结果。</p> <p><b>Unique:</b> 唯一表, 保证表中每一行数据的唯一性, 适用于需要去重的数据存储。</p> <p><b>Duplicate:</b> 重复表, 允许表中存在重复数据, 适用于存储原始数据, 不进行去重处理。</p>
Nocopies 表	<p>当数据库类型是 KingBase、需选 Nocopies 表信息。</p> <p><b>0:</b> 表示该表是复制表, 即在数据库集群中, 该表的数据会复制到所有的节点上, 以保证数据的高可用性。</p> <p><b>1:</b> 表示该表是非复制表, 即在数据库集群中, 该表的数据只存储在创建表的节点上, 不会复制到其他节点, 适用于存储不经常访问或不需要高可用性的数据。</p>
行列模式	<p>当数据库类型是 KingBase、需选行列模式信息。</p> <ul style="list-style-type: none"> <li>○ 行存储: 数据整行存储, 适用于经常需要关注整行数据的 OLTP(在线联机事务处理系统)。</li> <li>○ 哈希表: 数据按 hash 策略分布存储与不同主机上, 以实现分布式存储和计算)</li> </ul>
TTL	<p>当数据库类型是 kafka、需填 TTL 信息。</p> <p>TTL 用于设置 Kafka 主题中消息的存活时间, 超过此时间的消息将被自动删除。</p>

	单位是天 (day)、小时 (hours) 或分钟 (minutes)
● 分区个数	<p>当数据库类型是 kafka、需填分区个数信息。</p> <p>分区个数的作用是提高并行处理能力和消息吞吐量，分区数越多，可同时处理的消费者和生产者请求就越多。对于小型集群 (少于 6 个 broker)，可以选择是 broker 数量的 3 倍；对于大型集群 (超过 12 个 broker)，可以选择是 broker 数量的 2 倍。</p>
● 当前分区	<p>当数据库类型是 kafka、需填当前分区信息。</p> <p>指定创建主题 (topic) 时的初始分区数量，如果没有指定，则使用集群配置的默认值，默认的当前分区数值取决于 Kafka 集群的配置</p>
● 备份数	<p>当数据库类型是 kafka、需填备份数信息。</p> <p>备份数 (Replication Factor)：是指定每个分区的副本数量，以提高数据的可靠性和容错能力</p>
● 数据类型	<p>当数据库类型是 kafka、需填数据类型信息。</p> <p>json：用于指定消息体为 JSON 格式，这是一种轻量级的数据交换格式。</p> <p>json 数组：用于指定消息体为 JSON 数组格式，允许单个消息包含多个 JSON 对象，适用于批量处理数据的场景。</p> <p>Csv：用于指定消息体为 CSV 格式，其中的数据通常由纯文本组成，字段由逗号分隔，适用于存储表格数据的场景。</p>
● 字符集	<p>当数据库类型是 kafka、需填写字符集信息。</p> <p>f-8：指定使用 UTF-8 字符集，是一种广泛使用的字符编码，能够表示任何 Unicode 字符，适用于国际化场景</p> <p>gbk：指定使用 GBK 字符集，是一种用于简体中文的字符编码，适用于涉及简体中文字符的场景</p>
结构类型	<p>当数据库类型是 kafka、需选择结构类型。</p> <p>单一模式：消息体只包含单一结构的数据，如单个 JSON 对象或 CSV 记录，适用于简单的数据模型。</p> <p>复合模式：消息体包含复合结构的数据，如 JSON 数组或嵌套 JSON 对象，适用于复杂的数据模型，可以包含多个记录或字段。</p>
● 表引擎	<ul style="list-style-type: none"> <li>○ 当数据库类型是 ClickHouse、需选择如下表引擎信息。</li> <li>○ MergeTree：一种高效的列式存储引擎，适用于处理大量数据的写入和快速查询，支持数据自动合并以优化存储和查询性能。</li> <li>○ ReplacingMergeTree：类似于 MergeTree，但具有去重功能，自动删除具有相同主键的旧记录，保留最新的记录。</li> <li>○ TinyLog：轻量级的日志引擎，适用于小规模数据的写入，不支持索引，查询</li> </ul>

	<p>性能不如 MergeTree。</p> <ul style="list-style-type: none"> <li>○ Log: 简单的行式存储引擎, 适用于快速写入和简单查询, 不支持复杂的查询操作, 数据存储为纯文本格式。</li> </ul> <p>当数据库类型是 Doris、需选择如下表引擎信息。</p> <ul style="list-style-type: none"> <li>○ OLAP: 由 Doris 负责数据管理和存储, 适用于分析型查询。</li> </ul>
分桶类型	<p>当数据库类型是 Doris、需填分桶类型信息。</p> <p>"hash"分桶: 即根据哈希值将数据均匀分配到不同的桶中, 以实现负载均衡</p>
分桶数量	<p>当数据库类型是 Doris、需填分桶数量信息。</p> <p>自动分桶: 默认选择"auto"让系统根据数据量和查询模式自动决定最优的分桶数量。</p> <p>手动分桶: 支持用户手动设置分桶数, 分桶数量的最小值为 1-5000。</p>
拓展属性	<p>当数据库类型是 Doris、需填写拓展属性。</p> <p>拓展属性: 允许用户为表设置额外的配置选项, 如副本数量、存储介质和数据冷却时间等, 以满足特定的性能和存储需求。</p>

b.在“物理表属性”页面添加所需要的物理表字段属性, 物理表字段属性参数说明参考。



参数名称	说明
● 英文名称	以英文、数字、下划线组成, 且以英文开头,最少 4 字符,最多 64 字符。
● 属性名	以中文、英文、数字、下划线, 且以中文或英文开头, 最少 4 字符,最多 64 字符。
● 数据类型	设置属性的数据类型。
数据标准	选择一个已发布数据标准与逻辑实体属性关联
主键	表中唯一标识每条记录的字段或字段组合
分区	选中表示基于该字段值进行数据分区
不为空	选中表示该字段不可为空
索引信息	<p>当数据库类型是 DM、需填写索引信息。</p> <ul style="list-style-type: none"> <li>○ 无索引: 不配置索引。</li> <li>○ 唯一索引: 要求索引列的值在表中唯一, 用于确保数据的唯一性, 提高查询效率。</li> <li>○ 组合索引: 在表的多个字段上建立的索引, 适用于需要同时在多个字段上进行查询过滤的场景, 可以提高多字段联合查询的效率。</li> </ul>

加密方式	<p>当数据库类型是 Hive、需填写加密方式信息。</p> <ul style="list-style-type: none"> <li>○ AES_ENC: 通过 AES_ENCRYPT 函数实现, 适用于快速加密大量敏感数据, 如用户信息和财务数据。</li> <li>○ SM4_ENC: 是一种对称加密算法, 适用于需要符合国家标准和法规要求的场景, 如政府数据、金融数据等。</li> </ul>
日期表达式	<p>当数据库类型是 kafka、需填写日期表达式信息。</p> <p>用于定义 Kafka 消息中日期字段的格式, 例如: yyyy-MM-dd</p>
选择器表达式	<p>当数据库类型是 kafka、需填写选择器表达式信息。</p> <p>jsonpath: 使用 JSONPath 表达式从 JSON 格式的消息中提取特定字段, 适用于处理 JSON 消息。</p> <p>xpath: 使用 XPath 表达式从 XML 格式的消息中提取数据, 适用于处理 XML 消息。</p> <p>列表下标: 通过指定列表的下标来提取数组中的特定元素, 适用于处理包含数组的数据结构。</p>
分布排序键	<p>当数据库类型是 ClickHouse、需设置分布排序键。</p> <p>分布排序键” (Distribution Key): 用于指定数据在分布式表中的分布方式, 通常选择一个或多个列作为分布键, 以确保数据按照这些列的值分布到不同的节点上, 从而优化查询性能和负载均衡。</p>
分桶字段	<p>当数据库类型是 Dorise、需设置分桶字段。</p> <p>设置该字段为分桶字段, 这个字段的值将影响数据在不同桶中的分布, 从而对查询性能和负载均衡产生影响。</p>
聚合 key	<p>当数据库类型是 Dorise、需设置聚合 key 字段。</p> <p>设置改字段为聚合字段, 这些字段决定了数据的聚合方式, 如 SUM、MAX、MIN 等, 适用于数据预聚合场景</p>
聚类类型	<p>当数据库类型是 Dorise、需选择聚合类型。</p> <p>SUM: 对数值列进行求和, 适用于需要计算总量或累计值的场景。</p> <p>REPLACE: 用于替换 Value 列的值。当新数据的 Key 列与已有数据的 Key 列相同时, 新数据的 Value 列会替换掉旧数据的 Value 列</p> <p>MAX: 找出数值列的最大值, 适用于需要确定最大值的场景。</p> <p>MIN: 找出数值列的最小值, 适用于需要确定最小值的场景。</p>
分区 key	<p>当数据库类型是 Dorise、需设置分区 key。</p> <p>“分区 key”参数用于指定表的分区键</p>

c.在“属性信息”页面, 单击“创建关系”进入创建数据表之间关系。

关系模型 > 物理表 > 编辑

实体基本信息 属性信息 1 关系

+ 新建

序号	当前表	当前字段(FK)	关系	关联表	关联字段(PK)	操作
1	endek_info	单击	单击	不可与当前物理表重复选择	单击	删除

参数名称	说明
当前表	默认选择上一步骤创建的物理表
当前字段	默认选择上一步骤创建的物理表的属性信息
关系	<p>1:1 表示此物理表数据在关联物理表中只有一条数据与之对应。</p> <p>1:N 表示此物理表数据在关联物理表中至少有一条数据与之对应。</p> <p>N:1 表示此物理表数据有多条数据与关联物理表中一条数据与之对应。</p>
关联表	与同模型下已生效物理表进行关联
关联字段	与同模型下已生效物理表字段进行关联

6、完成物理表基本信息后，单击“确定”，完成物理表信息创建。

## 后续步骤

您可进行物理表物化、上线转为已生效状态，已生效的物理表可在其他模块应用。

## 物化物理表

物理表完成创建完成后可进行物理表物化，完成开发环境物理表新建。

## 前提条件

- 已创建所需物理表，数据中台支持可视化建模方式。
- 需要完成主题规划和数仓分层设计信息配置后，才能进行物理表操作。
- 需要完成数据源接入，才可进行物理建模操作。

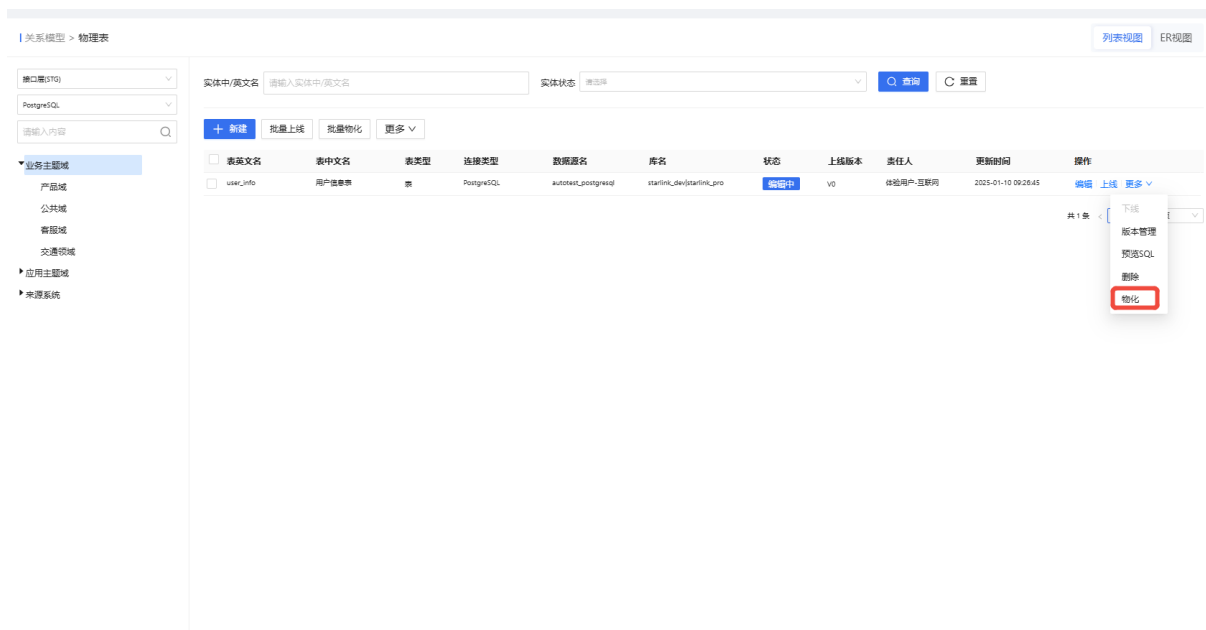
## 使用限制

- 已拥有数据建模角色权限。
- 已拥有所需数据源使用权限。

## 使用流程

物理表创建完成后可通过“物化”操作使其物理表发布至开发环境。

- 1、物理表创建完成后，您可点击物理表操作项中的“更多—物化”按钮，完成物理表物化。



2、在弹出的上线提示框中确定如下信息。



3、点击物化后，物化成功物理表将发布至开发环境。

说明：只有编辑中的物理表才可进行物理表物化，切不支持重复物化。

## 后续步骤

物理表物化后，可进行物理表上线发布至生产环境

## 上线物理表

物理表完成物化后可直接进行物理表上线的发布转为已生效状态。

## 前提条件

- 已创建所需物理表，且已完成物理表物化。
- 需要完成主题规划和数仓分层设计信息配置后，才能进行物理表操作。
- 需要完成数据源接入，才可进行物理建模操作。

## 使用限制

- 已拥有数据建模角色权限。
- 已拥有所需数据源使用权限。

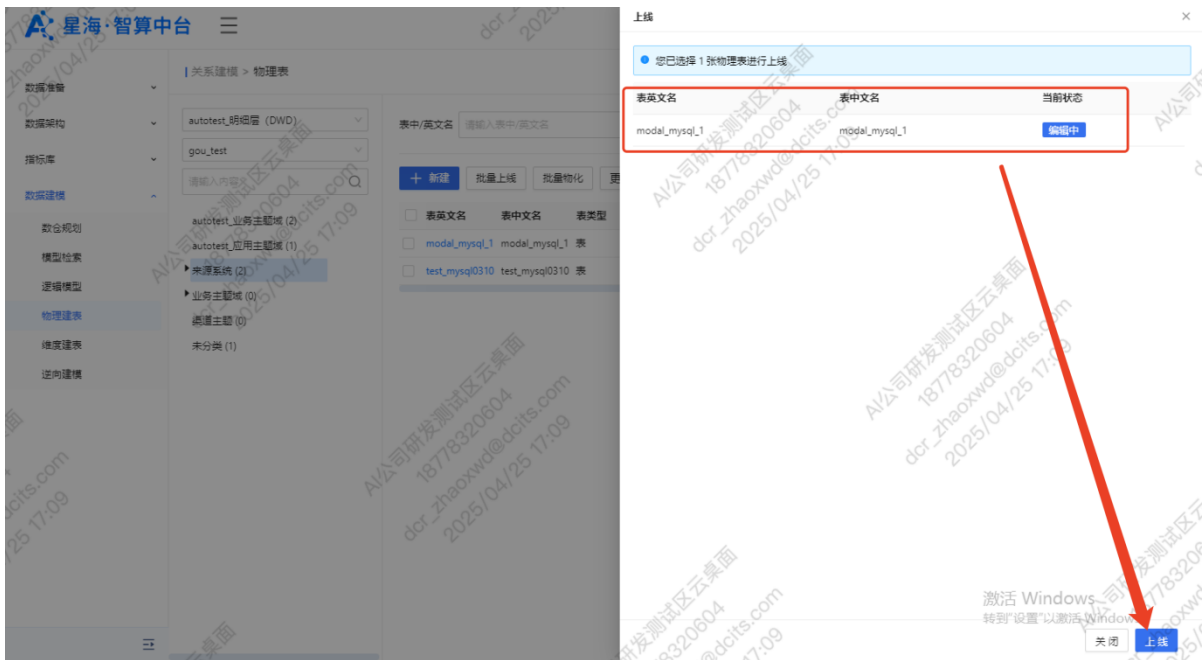
## 使用流程

物理表物化完成后支持上线转为已生效状态，已生效状态的物理表支持其他模块进行应用。

1、物理表创建完成后，您可点击物理表操作项中的“上线”按钮。



2、在弹出的上线提示框中确定如下信息。



3、点击上线后，物理表转为“待上线”状态，等待审核人员审核通过后，返回物理表管理页面，在列表中可以查看已转为“已生效”状态的物理表。

说明：只有处于编辑中、变更中状态的物理表才可进行上线操作。

## 物理表管理

物理表列表帮助您统一管理已创建的模型。在列表页，您可查看指定条件的模型，同时，也可对多个目标模型执行批量操作。本文为您介绍如何查看物理表列表并批量修改模型信息。

### 使用限制

- 已拥有数据建模角色权限。
- 已拥有所需数据源使用权限。

### 使用流程

1、进入物理表页面。

a.进入物理表页面

点击数据建模>关系建模，进入关系模型管理页面，关系模型包含逻辑模型和物理表。

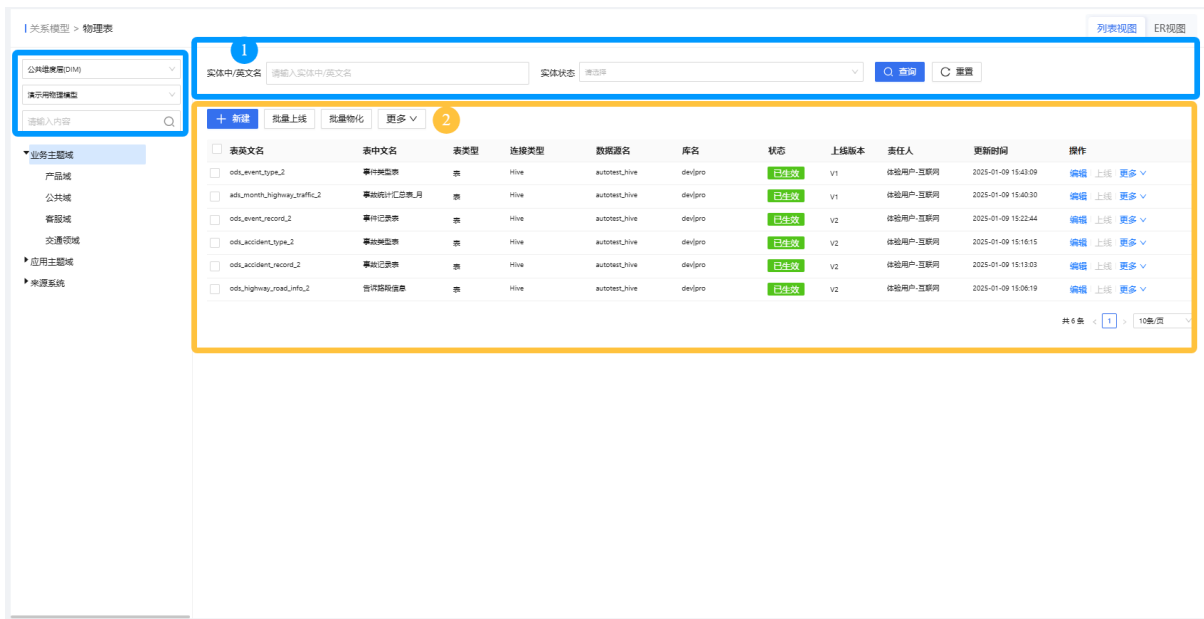
b.进入物理表页面

点击物理表中选择所需要的物理表，单击该模型进入物理表管理页面。

2、查看物理表及批量修改模型。

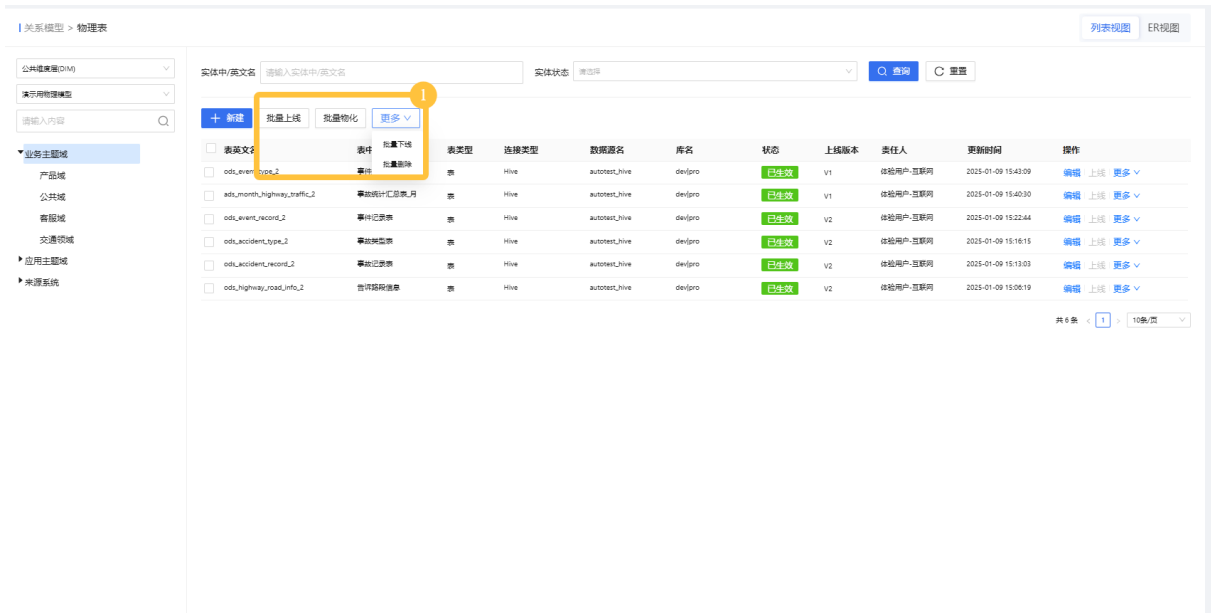
- 查看物理表列表

查看物理表的基本信息或执行新建、上线等单个物理表操作。



区域	描述
1	<p>在该区域，您可通过切换模型、切换主题和通过物理表名称、物理表状态查看指定条件的物理表</p> <p><b>说明：</b>通过名称搜索支持模糊匹配，即输入关键字，即可搜索出所有名称包含关键词的物理表</p>
2	<p>在该区域，您可查看物理表的基本信息，或执行单个物理表的相关操作。</p> <p><b>新建：</b>快速创建全部物理表。</p> <p><b>查看：</b>双击物理表英文名，进入目标物理表详情页</p> <p><b>编辑：</b>支持物理表基本信息、属性信息、关系的配置编辑</p> <p><b>物化：</b>支持单个物理表物化操作</p> <p><b>上线：</b>支持物理表上线，转为已生效状态。</p> <p><b>下线：</b>支持物理表下线，下线后支持重新上线。</p> <p><b>版本管理：</b>包含物理表版本历史、版本对比、发布日志。</p> <p><b>预览 SQL：</b>支持查看物理表创建所需的 DDL 语句。</p> <p><b>删除：</b>支持物理表信息删除</p> <p><b>说明：</b>物理表删除后不可恢复，请您谨慎操作。</p>

● 批量修改物理表



区域	描述
1	<p>在该区域，你可批量进行物理表操作。</p> <p><b>批量物化：</b>支持选择多个物理表物化操作。</p> <p><b>批量上线：</b>支持选择多个物理表上线操作。</p> <p><b>批量下线：</b>支持选择多个物理表下线操作。</p> <p><b>批量删除：</b>支持选择多个物理表一次性删除。</p>

## 数据集成

### 数据集成总览

数据集成总览将数据集成模块相关指标进行可视化监控。支持查看接入数据源的总数、连通正常的的数据源数量，连通异常的数据源数量、各数据类型任务情况、资源占用情况、采集总数据量、今日源端库表数据情况、今日数据集成实例情况、采集速率、今日数仓贴源层数据情况、昨日实时采集数据、昨日离线采集数据量、昨日多模态采集数据量等指标。

## 任务配置方式

### 离线数据集成

#### 前提条件

- 已拥有数据集成角色权限
- 完成来源数据源以及目标源数据源的接入工作。
- 完成对来源数据源以及目标数据源的元数据采集。

#### 使用限制

来源数据源以及目标数据源完成授权。

来源数据表以及目标数据表完成授权。

来源数据表以及目标数据表完成模型创建。

#### 使用说明

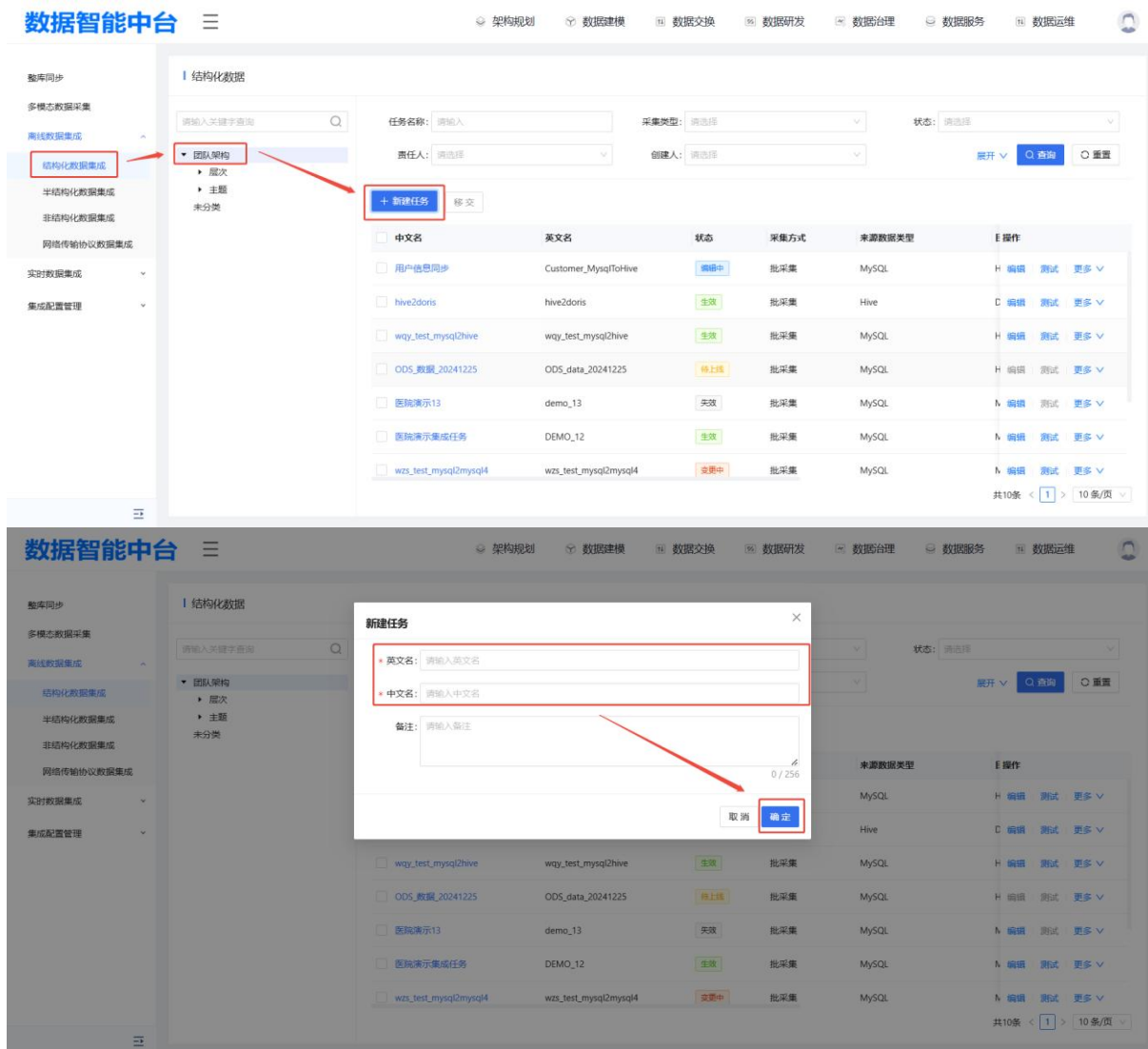
离线数据集成任务采用引导式配置方式，一共分为以下 6 个步骤：

- 1、新建离线数据集成任务
- 2、定义来源端与目标端
- 3、配置资源队列
- 4、字段映射
- 5、数据预处理与数据过滤
- 6、调度信息配置

#### 配置过程

##### 第一步：新建离线数据集成任务

进入到离线数据集成-结构化数据集成目录，选择架构目录，进行“新建任务”的操作，填写任务的中文名、英文名信息，点击“确定”。



## 第二步：定义来源端与目标端

在新增任务页面中，源与目标配置中，来源配置中左侧为 source 端，右侧为 sink 端。左侧 source 端选择 Mysql，选择相应数据源、数据库、数据表信息；右侧 sink 端选择 Hive，选择相应的目标数据源、目标模型，压缩格式。

交换模式  增量同步  全量同步

### 第三步：配置资源队列

资源配置中选择对应的 yarn 资源、生产队列、测试队列信息，点击“下一步”，进入到字段映射配置中。

资源队列详情介绍详见运维中心章节。

### 第四步：字段映射

在字段映射配置中，数据源会获取到相应所填写的表的字段信息，你可以通过 3 种方式进行对应字段关系的设置：

- 手动选择下拉框将对应字段进行匹配；
- 同名映射：当数据源中的字段与目标数据源字段有相同名称时，可以通过该方式直接进行映射；
- 同行映射：当数据源中的字段数量大于目标数据源中字段的数量时，可以通过该种方式直接

进行映射关系对应。

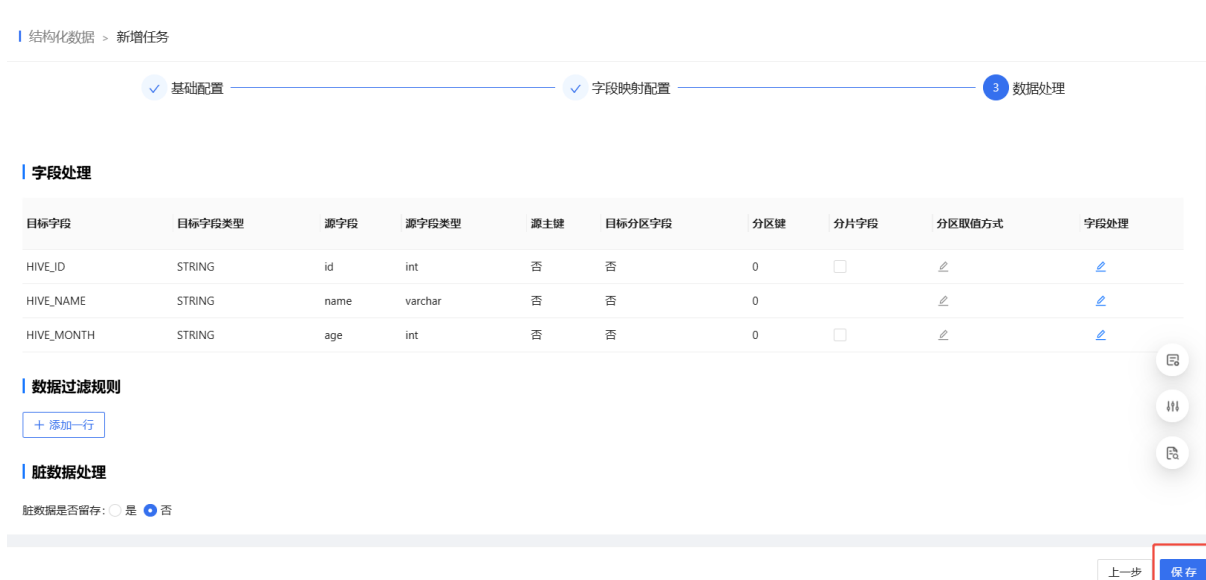
确认数据源与目标数据源字段映射关系无误后，点击“下一步”进入到数据处理中。



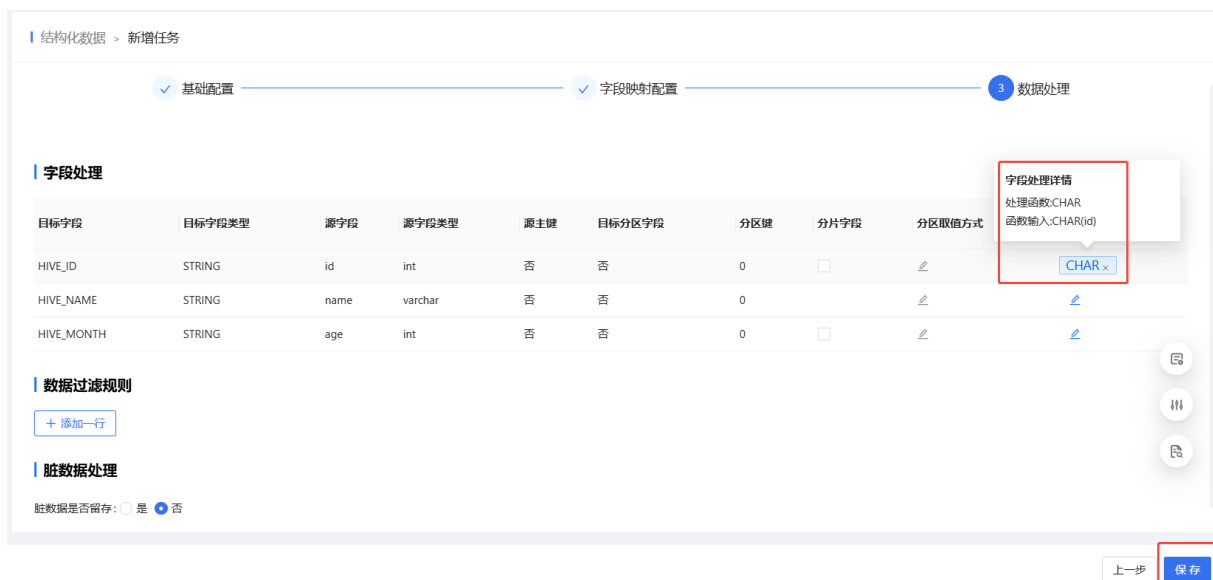
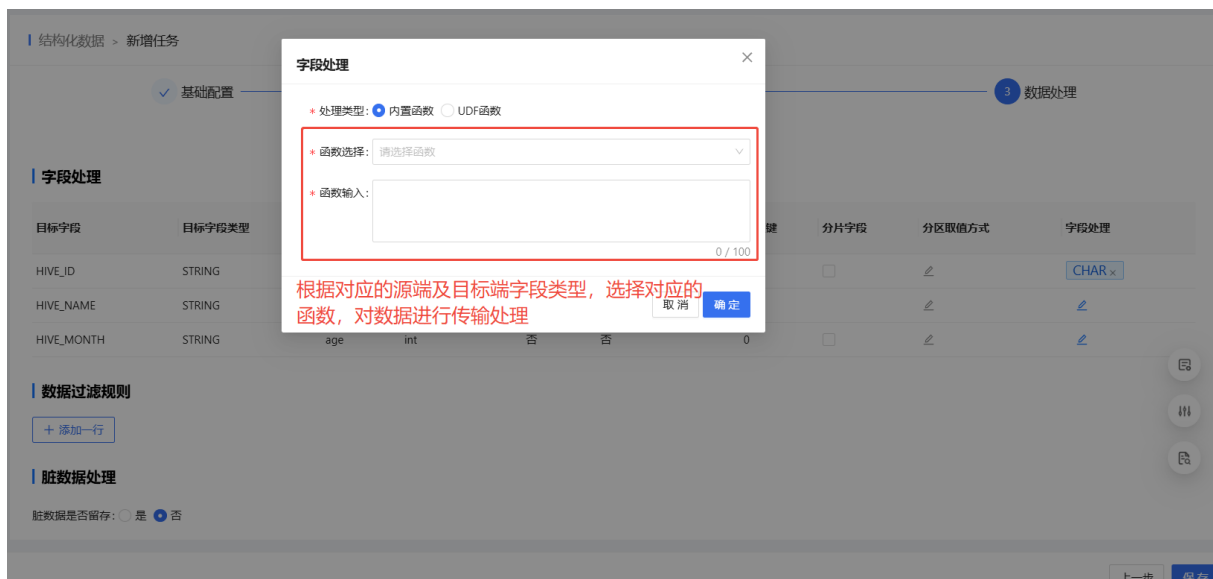
## 第五步：数据预处理与数据过滤

在数据处理中，可以对数据进行以下的操作：

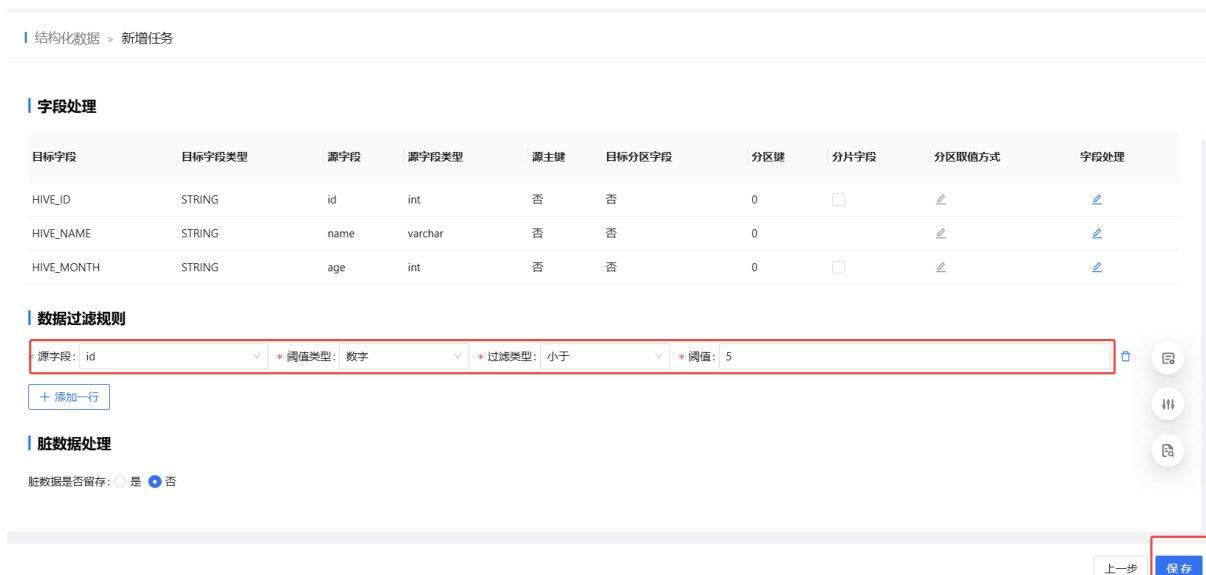
a.若对数据进行直接传输，从 source 端复制传输到 sink 端，不进行任何数据的加工，则可以直接点击“保存”即可；



b.若要对数据进行传输的数据进行处理，则可以在字段处理中选择相应的函数，对传输中的数据进行处理；（如：source 端的字段类型为 int，目标端为 string，则可以使用 CHAR()函数，将 source 端的字段类型转换为 string 类型再传入到 sink 端中）

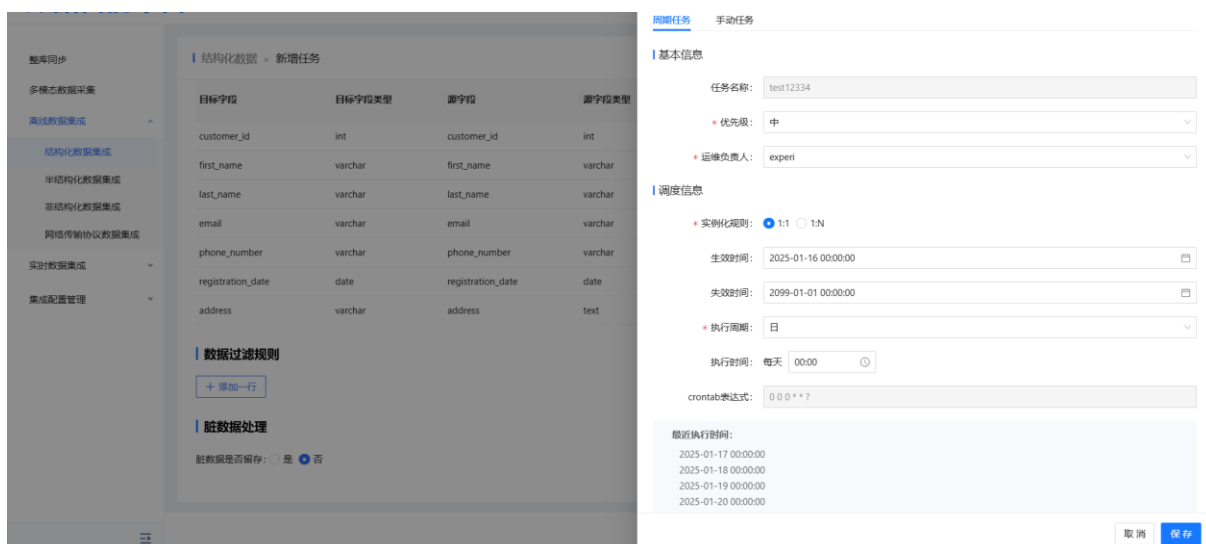


c.若要对数据进行筛选的操作，则可以添加过滤规则来对数据传输过程中进行筛选。（如：将 source 端中的 id 大于 5 的数据全部传输到 sink 端中，则可以设置源字段 id，类型为数字，过滤类型为小于，阈值为 5，设置完成后，点“保存”）



## 第六步：调度信息配置

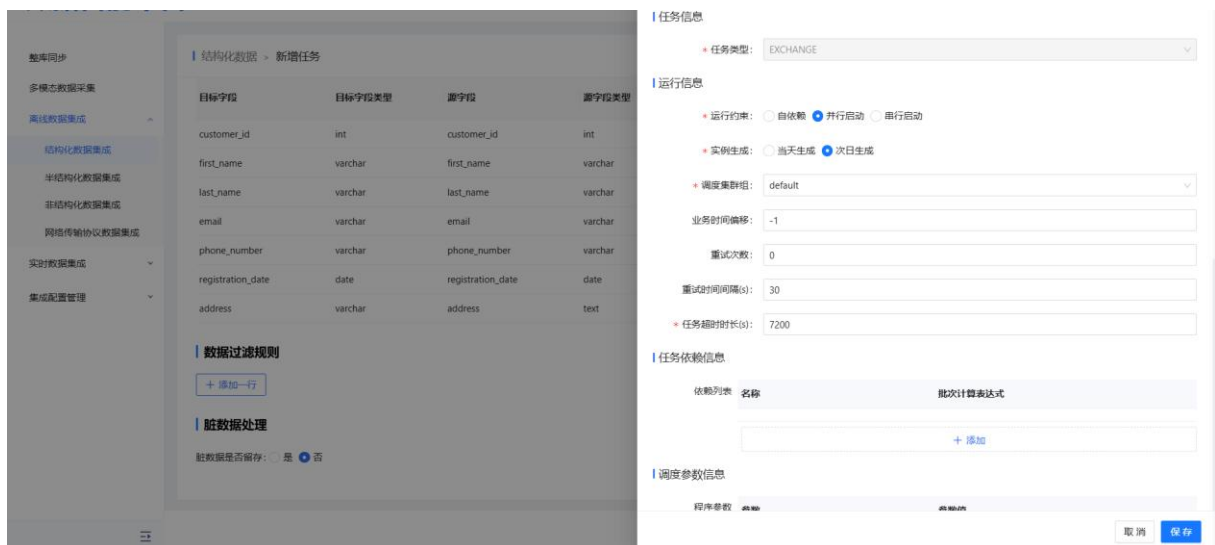
在任务页面右侧悬浮球中，可以进行调度配置。



调度配置分为以下几个部分：

- 基本信息
  - 任务名称：默认与任务英文名保持一致，不可更改。
  - 优先级：分为极高、高、中、极低等 5 个优先级，可根据实际情况配置，默认为中。
  - 运维负责人：可指定该任务的运维负责人。
- 调度信息
  - 实例化规则：分为 1:1 及 1:N 两种情况：

- 1:1：指一个任务生成 1 个运行实例。
- 1:N：指一个任务生成 N 个运行实例。
- 生效时间：指当前任务开始生效的时间，默认时间为当天 00:00。
- 失效时间：指该周期任务执行到什么时间失效，默认为 2999-12-31。
- 执行周期：指该周期任务的周期类型，分为月、日、小时、分钟，本案例中配置为日周期任务。
- 执行时间：指任务实例开始运行的具体时间。根据不同的执行周期，具有不同的含义及填写内容：
  - 若执行周期为月：执行时间配置项为每月 x 号 x 时 x 分，如每月 1 日 12 时 00 分运行实例。
  - 若执行周期为日：执行时间配置项为每日 x 时 x 分，如每日 00 时 00 分运行实例。
  - 若执行周期为小时：执行时间配置项为每小时 x 分起，如每小时 5 分起运行实例，注意此处的范围为 0~59。
  - 若执行周期为分钟：执行时间配置项为每分钟 x 秒起，如每分钟 5 秒起运行实例，注意此处的范围为 0~59。
- crontab 表达式：由上述执行周期与执行时间配置组合出的表达式，不可编辑。
- 最近执行时间：指预计该任务上线后，各实例运行的时间，不可编辑。



- 任务信息

- 标识该任务属于何种业务，每个业务模块默认，不可编辑，用户无需关心此项配置。

- 运行信息

- 运行约束：用于定义同一任务中不同实例的依赖情况，默认为并行启动。

- 自依赖：当前的周期实例的执行，必须要等到上个周期实例执行成功后才能开始执行。
- 并行启动：任务的多个周期实例，可以一起运行，相互之间没有依赖关系。
- 串行启动：当前的周期实例的执行，只需要上一个周期实例是运行结束状态即可（不管成功与否）。

- 实例生成：用于定义任务实例的生成时间，默认为次日生成。

- 当天生成：任务上线后，实例当天就开始生成。
- 次日生成：任务上线后，实例第二天才开始生成。

- 调度集群组：需选择用于调度的调度集群组，默认为 default 集群组。若有其他集群组也可选择。

- 业务时间偏移：对于数据的账期偏移量。偏移量的值为 0 至负数。如当前日期为 2025 年 1 月 15 日，以下对一些数值进行说明：

- 业务时间偏移填写 0：表示该任务采集的是数据日期为当日的的数据，即任务实例运行时，采集数据时间为 2025 年 1 月 15 日的数据。
- 业务时间偏移填写-1：表示该任务采集的是数据日期为前一日的的数据，即任务实例运行时，采集数据时间为 2025 年 1 月 14 日的数据。
- 业务时间偏移填写-2：表示该任务采集的是数据日期为前两日的的数据，即任务实例运行时，采集数据时间为 2025 年 1 月 13 日的数据。
- 依此类推。

此处需注意，增量同步中，若增量条件也配置了时间相关的变量，此时增量条件中的时间变量受到业务偏移量的影响。

以本案例为例，增量条件中配置了时间变量`batchNo?calDate(0,'D')`，若业务偏移填写为-1，则增量条件中表示为采集前一天的数据。若业务偏移填写为 0，则增量条件中表示为采集当天的数据。

- 重试次数：指当任务实例失败后，需要自动重试的次数。
- 重试时间间隔：指当任务实例失败后，每次自动重试的时间间隔。
- 任务超时时长：可设置定义任务超时的时长。默认为 7200 秒。
- 任务依赖信息
  - 依赖列表：指不同任务间的依赖关系配置列表。
- 调度参数信息
  - 程序参数：用于扩展额外的配置项，以 k-v 的形式进行录入，如有特殊情况需要可以添加配置项，无额外需求可不配置此项。

## 附录：支持的内置函数

函数名	中文名	函数规范	函数说明
ASCII	获取 ASCII 编码	ASCII('Hi')	返回字符串中第一个字符的 ASCII 值。此方法返回一个整数。
BIT_LENGTH	位长度	BIT_LENGTH(NAME)	返回二进制字符串中的位数。该方法返回一个长整型
CHAR_LENGTH	字符长度	CHAR_LENGTH(NAME)	这个方法返回一个字符串中字符的数量，返回类型为 long。
LENGTH	文本长度	LENGTH(NAME)	这个方法返回一个字符串中字符的数量，返回类型为 long。
OCTET_LENGTH	二进制字符长度	OCTET_LENGTH(NAME)	返回二进制字符串中字节的数量。此方法返回一个 long 类型的值。
CHAR	ASCII 转字符	CHAR(65)	返回表示 ASCII 值的字符。该方法返回一个字符串。
CHR	ASCII 编码转字符	CHR(65)	返回表示 ASCII 值的字符。该方法返回一个字符串。
CONCAT	连接字符串	CONCAT(NAME, '_')	组合字符串。与运算符    不同，NULL 参数会被忽略，不会导致结果变为 NULL。如果所有参数都是 NULL，则结果是一个空字符串。该方法返回一个字符串。

CONCAT_WS	拼接字符串	CONCAT_WS(' ', NAME, '')	使用分隔符组合字符串。如果分隔符为 NULL，则会被视为空字符串。其他 NULL 参数会被忽略。剩余的非 NULL 参数（如果有）将用指定的分隔符连接起来。如果没有剩余参数，则结果是一个空字符串。该方法返回一个字符串。
HEXTORAW	十六进制转原始字符串	HEXTORAW(DATA)	将字符串的十六进制表示转换为字符串。每个字符串字符使用 4 个十六进制字符。
RAWTOHEX	原始字符串转十六进制	RAWTOHEX(DATA)	将字符串或字节转换为十六进制表示。每个字符串字符使用 4 个十六进制字符。该方法返回一个字符串。
INSERT	指定位置插入字符串	INSERT(NAME, 1, 1, '')	在原始字符串的指定起始位置插入额外的字符串。长度参数指定在原始字符串的起始位置删除的字符数。该方法返回一个字符串。
LOWER	转小写	LOWER(NAME)	将字符串转换为小写形式。
LCASE	转小写（同 LOWER）	LCASE(NAME)	将字符串转换为小写形式。
UPPER	转大写	UPPER(NAME)	将字符串转换为大写形式。
UCASE	转大写（同 UPPER）	UCASE(NAME)	将字符串转换为大写形式。
LEFT	左截取	LEFT(NAME, 3)	返回最左边的一定数量的字符。
RIGHT	右截取	RIGHT(NAME, 3)	返回最右边的一定数量的字符。
LOCATE	查找字符位置	LOCATE('.', NAME)	返回字符串中搜索字符串的位置。如果使用了起始位置参数，则忽略它之前的字符。如果位置参数是负数，则返回最右边的位置。如果未找到搜索字符串，则返回 0。请注意，即使参数不区分大小写，此函数也区分大小写。
INSTR	查询字符位置	INSTR(NAME, '.')	返回字符串中搜索字符串的位置。如果使用了起始位置参数，则忽略它之前的字符。如果位置参数是负数，则返回最右边的位置。如果未找到搜索字符串，则返回 0。请注意，即使参数不区分大小写，此函数也区分大小写。
POSITION	寻找字符位置	POSITION(NAME, '.')	返回字符串中搜索字符串的位置。如果使用了起始位置参数，则忽略它之前的字符。如果位置参数是负数，则

			返回最右边的位置。如果未找到搜索字符串，则返回 0。请注意，即使参数不区分大小写，此函数也区分大小写。
LPAD	左限制填充	LPAD(AMOUNT, 10, '*')	将字符串左侧填充到指定的长度。如果长度比字符串短，则字符串将在末尾被截断。如果未设置填充字符串，则使用空格填充。
RPAD	右限制填充	RPAD(TEXT, 10, '-')	将字符串右侧填充到指定的长度。如果长度比字符串短，则字符串将被截断。如果未设置填充字符串，则使用空格填充。
LTRIM	去除左空格	LTRIM(NAME)	移除字符串中所有前导空格或其他指定的字符。此函数已被弃用，请使用 TRIM 替代。
RTRIM	去除右空格	RTRIM(NAME)	移除字符串中所有尾随空格或其他指定的字符。此函数已被弃用，请使用 TRIM 替代。
TRIM	去除空格	TRIM(NAME)	移除字符串中所有前后导空格或其他指定的字符。
REGEXP_REPLACE	正则替换字符串	REGEXP_REPLACE('Hello World', '+', '')	替换与正则表达式匹配的每个子字符串。详情请参阅 <a href="#">Java String.replaceAll()</a> 方法。如果任何参数为 null（除了可选的 flagsString 参数），则结果为 null。
REGEXP_LIKE	正则匹配字符串	REGEXP_LIKE('Hello World', '[A-Z]*', 'i')	将字符串与正则表达式匹配。详情请参阅 <a href="#">Java Matcher.find()</a> 方法。如果任何参数为 null（除了可选的 flagsString 参数），则结果为 null。
REGEXP_SUBSTR	正则子串	REGEXP_SUBSTR('2020-10-01', '\d{4}')	将字符串与正则表达式匹配，并返回匹配的子字符串。
REPEAT	重复字符串	REPEAT(NAME, 10)	返回重复字符串
REPLACE	替换字符串	REPLACE(NAME, '')	在文本中替换所有出现的搜索字符串为另一个字符串。如果没有指定替换字符串，则从原始字符串中移除搜索字符串。如果任何参数为 null，则结果为 null。
SOUNDEX	声音编码转换	SOUNDEX(NAME)	表示字符串发音。此方法返回一个字符串，如果参数为 null，则返回 null。
SPACE	空格字符串	SPACE(80)	返回由一定数量的空格组成的字符

			串。
SUBSTRING	截取子串	SUBSTRING('[Hello]', 2)	返回从指定位置开始的字符串的子串。
SUBSTR	截取字符串	SUBSTR('[Hello]', 2)	返回从指定位置开始的字符串的子串。
TO_CHAR	转换为指定格式字符串	TO_CHAR(SYS_TIME, 'yyyy-MM-dd HH:mm:ss')	Oracle 兼容的 TO_CHAR 函数可用于格式化时间戳、数字或文本。
TRANSLATE	转换替换	TRANSLATE('Hello world', 'eo', 'EO')	Oracle 兼容的 TRANSLATE 函数用于将字符串中的一系列字符替换为另一组字符。
ABS	绝对值	ABS(I)	返回指定值的绝对值。返回的值与参数的数据类型相同。注意，TINYINT、SMALLINT、INT 和 BIGINT 数据类型无法表示它们的最小负值的绝对值，这可能导致异常。
ACOS	反余弦	ACOS(D)	计算反余弦值。该方法返回一个双精度浮点数。
ASIN	反正弦	ASIN(D)	计算反正弦值。该方法返回一个双精度浮点数。
ATAN	反正切	ATAN(D)	计算反正切值。该方法返回一个双精度浮点数。
COS	余弦	COS(ANGLE)	计算三角余弦值。该方法返回一个双精度浮点数。
COSH	双曲余弦	COSH(X)	计算双曲余弦值。该方法返回一个双精度浮点数。
COT	余切	COT(ANGLE)	计算三角余切值 (1/TAN(角度))。该方法返回一个双精度浮点数。
SIN	正弦	SIN(ANGLE)	计算三角正弦值。该方法返回一个双精度浮点数。
SINH	双曲正弦	SINH(ANGLE)	计算双曲正弦值。该方法返回一个双精度浮点数。
TAN	正切	TAN(ANGLE)	计算三角正切值。该方法返回一个双精度浮点数。
TANH	双曲正切	TANH(X)	计算双曲正切值。该方法返回一个双精度浮点数。
MOD	模运算	MOD(A, B)	取模运算表达式。结果与除数的类型相同。结果与被除数的符号相同，或者等于 0。

CEIL	向上取整	CEIL(A)	返回大于或等于参数的最小整数值。
CEILING	向上取整函数	CEILING(A)	返回大于或等于参数的最小整数值。
EXP	指数	EXP(A)	计算参数的指数值，该方法返回一个双精度浮点数。
FLOOR	地板函数	FLOOR(A)	返回小于或等于参数的最大整数值。
LN	自然对数	LN(A)	计算自然对数（以 e 为底）。参数必须是一个正数值。
LOG	对数	LOG(2, A)	计算以指定底数的对数。底数和参数必须是正数值。
LOG10	以 10 为底的对数	LOG10(A)	计算以 10 为底的对数。参数必须是一个正数值。
RADIANS	弧度	RADIANS(A)	将角度转换为弧度。
SQRT	平方根	SQRT(A)	计算参数的平方根，返回一个双精度浮点数。
PI	圆周率	PI()	返回圆周率常数 PI。
POWER	幂运算	POWER(A, B)	计算底数的幂。返回一个双精度浮点数。
RAND	随机数	RAND()	生成一个介于 0 和 1 之间的伪随机数。如果带有参数调用，则将会给该会话的随机数生成器设定种子。
RANDOM	随机数字	RANDOM()	生成一个介于 0 和 1 之间的伪随机数。如果带有参数调用，则将会给该会话的随机数生成器设定种子。
ROUND	四舍五入	ROUND(N, 2)	将数值四舍五入到指定的小数位数，只支持 Double 类型。
SIGN	数值符号	SIGN(N)	返回数值的符号。如果值小于 0，则返回 -1；如果值为 0，则返回 0；如果值大于 0，则返回 1。
TRUNC	截断小数	TRUNC(N, 2)	将数值截断到指定的小数位数。
TRUNCATE	截断小数部分	TRUNCATE(N, 2)	将数值截断到指定的小数位数。
CURRENT_DATE	当前日期	CURRENT_DATE()	返回当前日期。这些函数在事务（默认）或命令内部返回相同的值，具体取决于数据库模式。
CURRENT_TIME	当前时间	CURRENT_TIME()	返回带有系统时区的当前时间。实际可用的最大精度取决于操作系统和 JVM，可以是 3（毫秒）或更高。在 Java 9 之前不支持更高的精度。

CURRENT_TIMESTAMP	当前时间戳	CURRENT_TIMESTAMP()	返回带有系统时区的当前时间戳。实际可用的最大精度取决于操作系统和 JVM，可以是 3（毫秒）或更高。在 Java 9 之前不支持更高的精度。
NOW	当前系统时间	NOW()	返回带有系统时区的当前时间戳。实际可用的最大精度取决于操作系统和 JVM，可以是 3（毫秒）或更高。在 Java 9 之前不支持更高的精度。
DATEADD	日期加法	DATEADD(CREATED, 1, 'MONTH')	将单位添加到日期时间值中。 datetimeFieldString 表示单位。使用负值来减去单位。当操作毫秒、微秒或纳秒时，addIntLong 可能是一个 long 值，否则其范围被限制为 int。
TIMESTAMPADD	时间戳加法	TIMESTAMPADD(CREATED, 1, 'MONTH')	将单位添加到日期时间值中。 datetimeFieldString 表示单位。使用负值来减去单位。当操作毫秒、微秒或纳秒时，addIntLong 可能是一个 long 值，否则其范围被限制为 int。
DATEDIFF	日期差	DATEDIFF(T1.CREATED, T2.CREATED, 'MONTH')	返回两个日期时间值之间跨越的单位边界数。此方法返回一个 long 值。 datetimeField 表示单位。
DATE_TRUNC	日期截断	DATE_TRUNC(CREATED, 'DAY')	将指定的日期时间值截断到指定的字段。
DAYNAME	星期名称	DAYNAME(CREATED)	返回星期几的名称（英文）。
DAY_OF_MONTH	月份中的天	DAY_OF_MONTH(CREATED)	返回月份中的日期（1-31）。
DAY_OF_WEEK	周中的天	DAY_OF_WEEK(CREATED)	返回星期几的数值（1-7）（星期一至星期日），根据本地化设置。
DAY_OF_YEAR	年份中的天	DAY_OF_YEAR(CREATED)	返回一年中的日期（1-366）。
EXTRACT	提取时间	EXTRACT(SECOND FROM CURRENT_TIMESTAMP)	从日期/时间值中返回特定时间单位的值。该方法对于 EPOCH 字段返回一个数值，对于其他字段返回一个整数。
FORMATDATETIME	格式化日期时间	FORMATDATETIME(CREATED, 'yyyy-MM-dd HH:mm:ss')	将日期、时间或时间戳格式化为字符串。最重要的格式字符包括：y（年）、M（月）、d（日）、H（时）、m（分）、s（秒）。
HOUR	小时	HOUR(CREATED)	从日期/时间值中返回小时（0-23）。
MINUTE	分钟	MINUTE(CREATED)	从日期/时间值中返回分钟（0-59）。 该函数已经被弃用，请使用 EXTRACT

			替代。
MONTH	月份	MONTH(CREATED)	从日期/时间值中返回月份（1-12）。 该函数已经被弃用，请使用 EXTRACT 替代。
MONTHNAME	月份名称	MONTHNAME(CREATED)	返回月份的名称（英文）。
PARSEDATETIME	解析日期时间	PARSEDATETIME('2021-04-08 13:34:45','yyyy-MM-dd HH:mm:ss')	解析一个字符串并返回一个 TIMESTAMP WITH TIME ZONE 值。
TO_DATE	转换为日期	TO_DATE('2021-04-08 13:34:45','yyyy-MM-dd HH:mm:ss')	解析一个字符串并返回一个 TIMESTAMP WITH TIME ZONE 值。
QUARTER	季度	QUARTER(CREATED)	从日期/时间值中返回季度（1-4）。
SECOND	秒	SECOND(CREATED)	从日期/时间值中返回秒数（0-59）。 该函数已经被弃用，请使用 EXTRACT 替代。
WEEK	周	WEEK(CREATED)	返回日期/时间值中的周数（1-53）。 该函数使用当前系统的区域设置。
YEAR	年	YEAR(CREATED)	返回日期/时间值中的年份。
FROM_UNIXTIME	时间戳转时间	CALL FROM_UNIXTIME(1672502400, 'yyyy-MM-dd HH:mm:ss')	将从 UNIX 纪元（1970-01-01 00:00:00 UTC）开始的秒数转换为表示该时刻时间戳的字符串。
CAST	强转换类型	CAST(NAME AS INT)	将一个值强转为另一个数据类型，如不支持会报错。支持的数据类型有： STRING   VARCHAR, INT   INTEGER, LONG   BIGINT, BYTE, FLOAT, DOUBLE, DECIMAL(p,s), TIMESTAMP, DATE, TIME。
COALESCE	返回第一个非空值	COALESCE(A, B, C)	返回第一个非空值。
IFNULL	返回两者的第一个非空值	IFNULL(A, B)	返回第一个非空值。
NULLIF	相等则为空	NULLIF(A, B)	如果 'a' 等于 'b'，则返回 NULL，否则返回 'a'。
MASK_MIDDLE	掩码屏蔽	MASK_MIDDLE(fieldName, 1, 4, '*')	对 fieldName 一定长度内进行屏蔽
ENCRYPT	加密	ENCRYPT(NAME, 'AES')	将输入字段进行加密，字段名称,字段加密方式支持 ASE
DECRYPT	解密	DECRYPT(NAME, 'AES')	将输入字段进行解密，字段名称,,字段

## 整库同步

### 前提条件

- 已拥有数据集成角色权限
- 完成来源数据源以及目标源数据源的接入工作。
- 完成对来源数据源以及目标数据源的元数据采集。

### 使用限制

- 来源数据源以及目标数据源完成授权。
- 来源数据表以及目标数据表完成授权。
- 来源数据表以及目标数据表完成建模。

### 使用说明

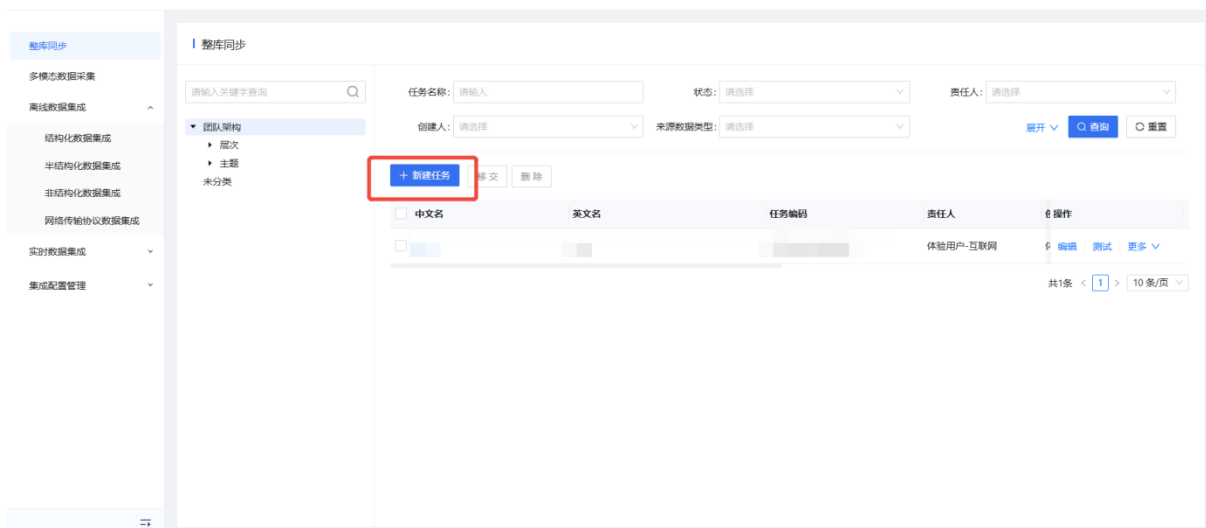
整库同步任务采用引导式配置方式，一共分为以下 5 个步骤：

- 1、新建整库同步任务
- 2、定义来源端与目标端
- 3、配置资源队列
- 4、配置目标表映射
- 5、调度信息配置

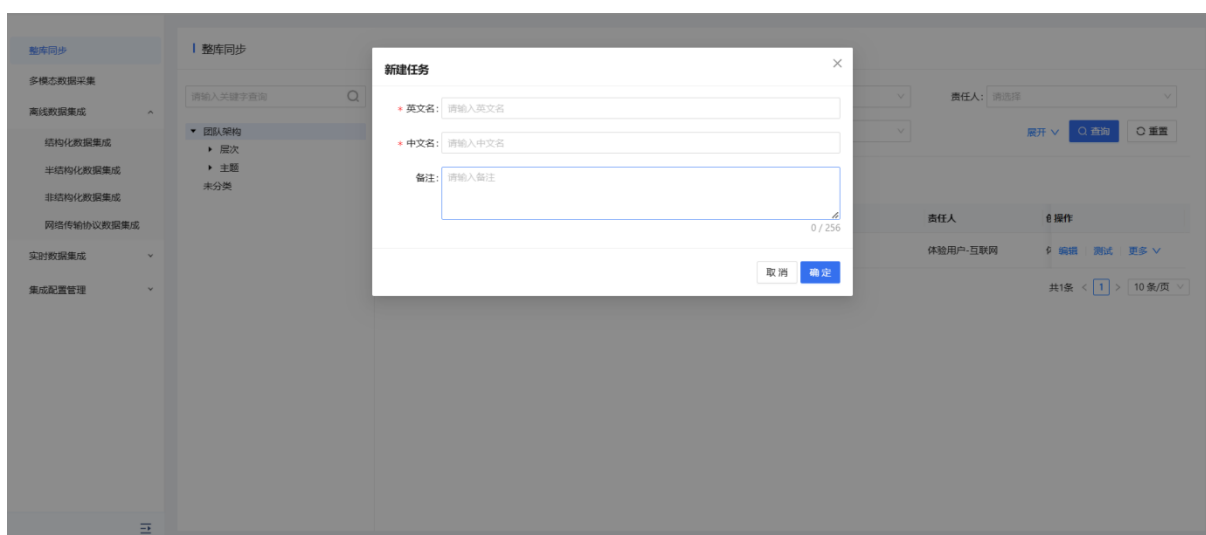
### 配置过程

#### 第一步：新建整库同步任务

在整库同步任务管理页面，点击新建任务按钮，新建任务。



在弹出的弹窗中，输入任务的基础信息



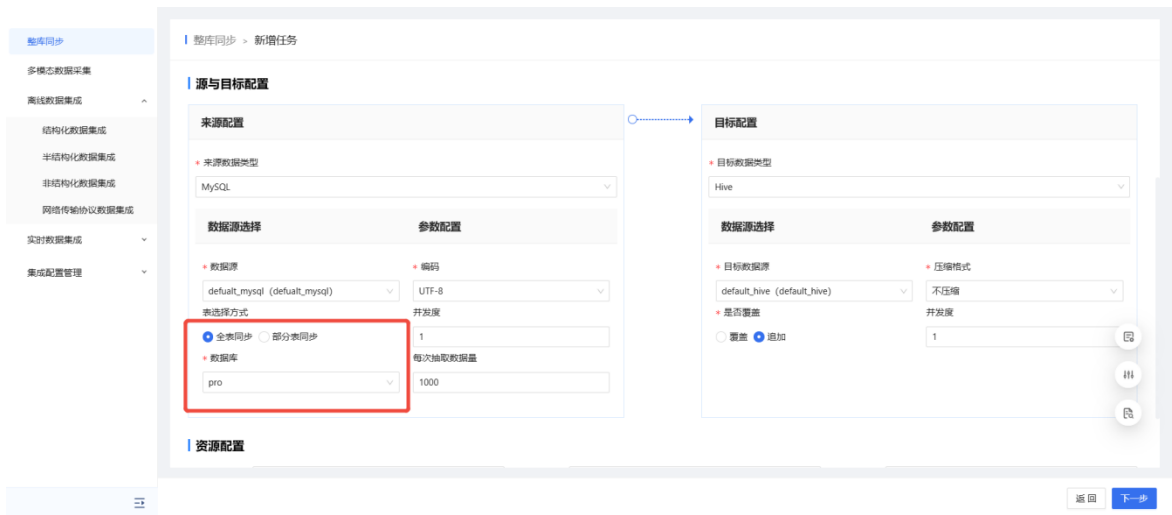
## 第二步：定义来源端与目标端

### 来源端配置

整库同步来源端有两种数据表同步方式

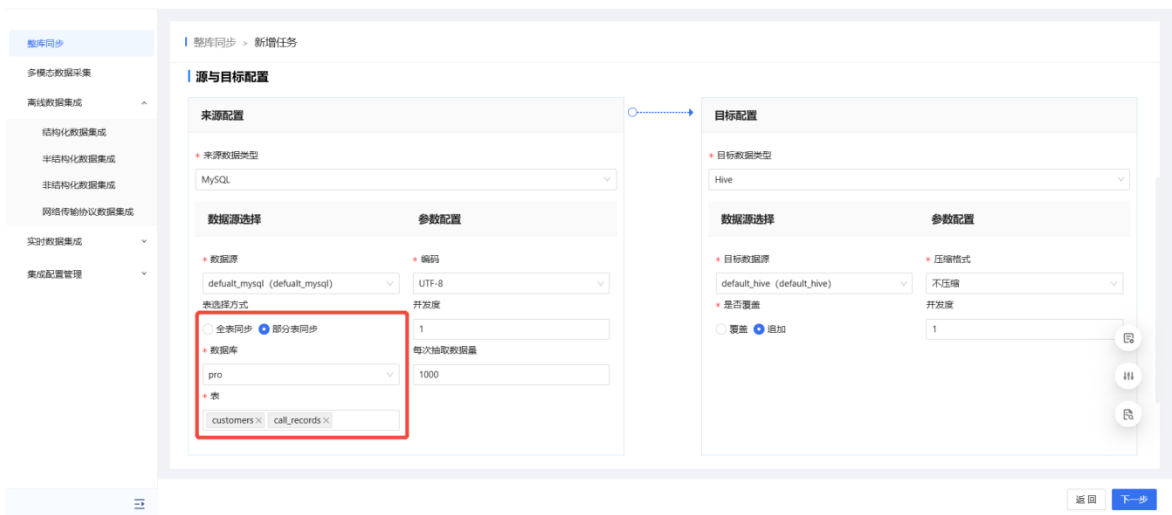
- 全表同步

全表同步表示将所选的来源库中的所有表同步至目标库中。



## ● 部分表同步

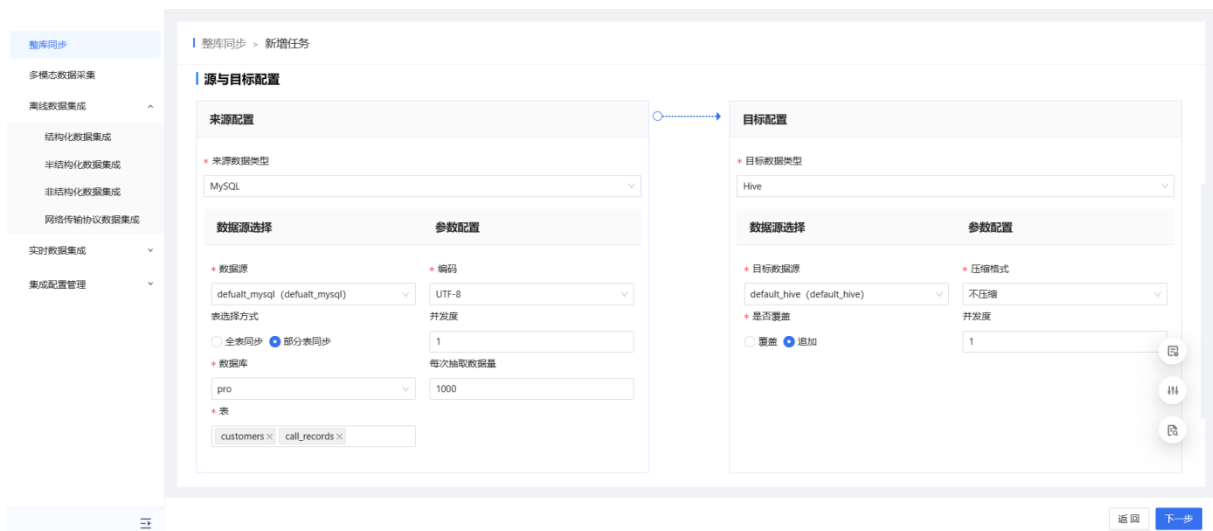
部分表同步表示将来源库中的部分表同步至目标库中，用户可自行选择需要同步的数据表。



来源配置还包括编码、并发度、每次抽取数据量等。

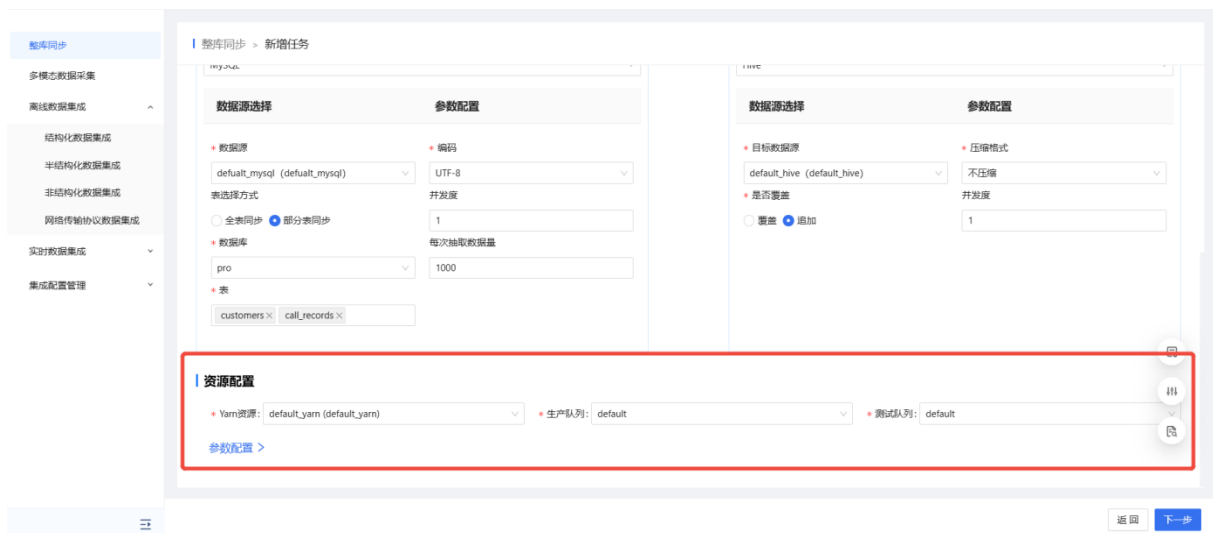
## 目标端配置

目标端选择需要存放数据表的数据源，并支持配置写入方式为覆盖或追加。支持压缩格式配置与并发度配置。

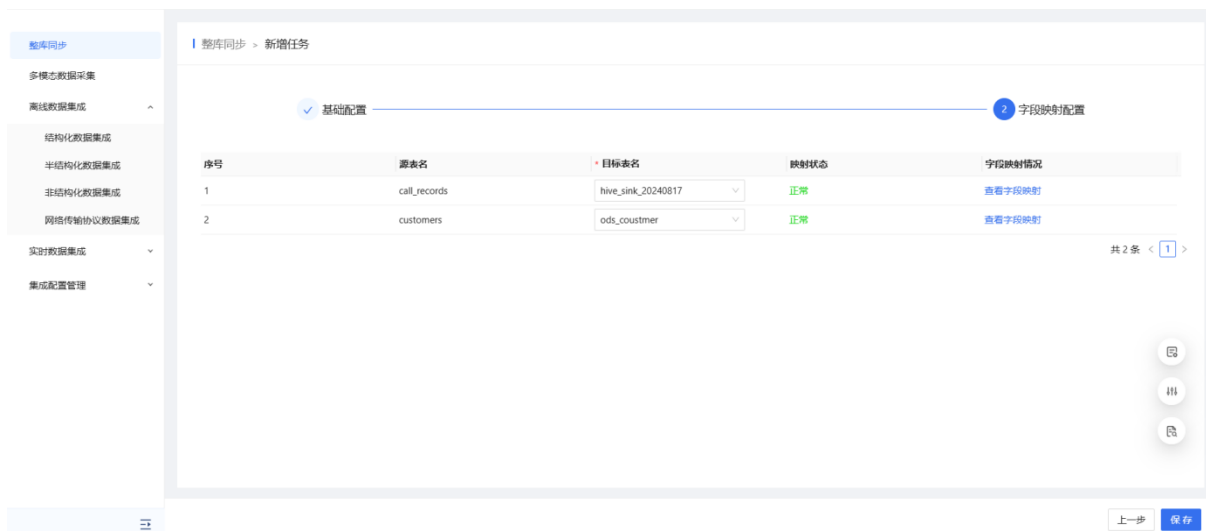


### 第三步：配置资源队列

资源配置中选择对应的 yarn 资源、生产队列、测试队列信息，点击“下一步”，进入到字段映射配置中。



### 第五步：配置目标表映射



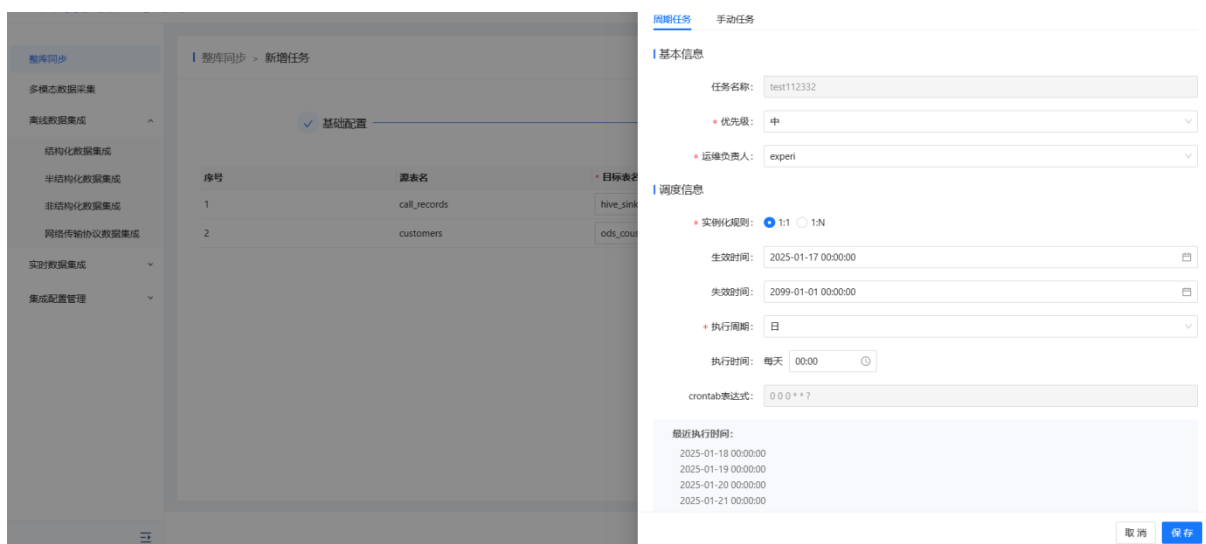
当目标库中存在与所选数据表名称相同的物理表时，系统会自动匹配并建立字段映射关系。

若目标库中不存在与所选数据表名称相同的物理表，用户可以进行异名数据表映射并手动建立字段映射关系。

若目标库中不存在数据表，则需先通过数据建模功能建立数据模型与物化物理表。

## 第六步：调度信息配置

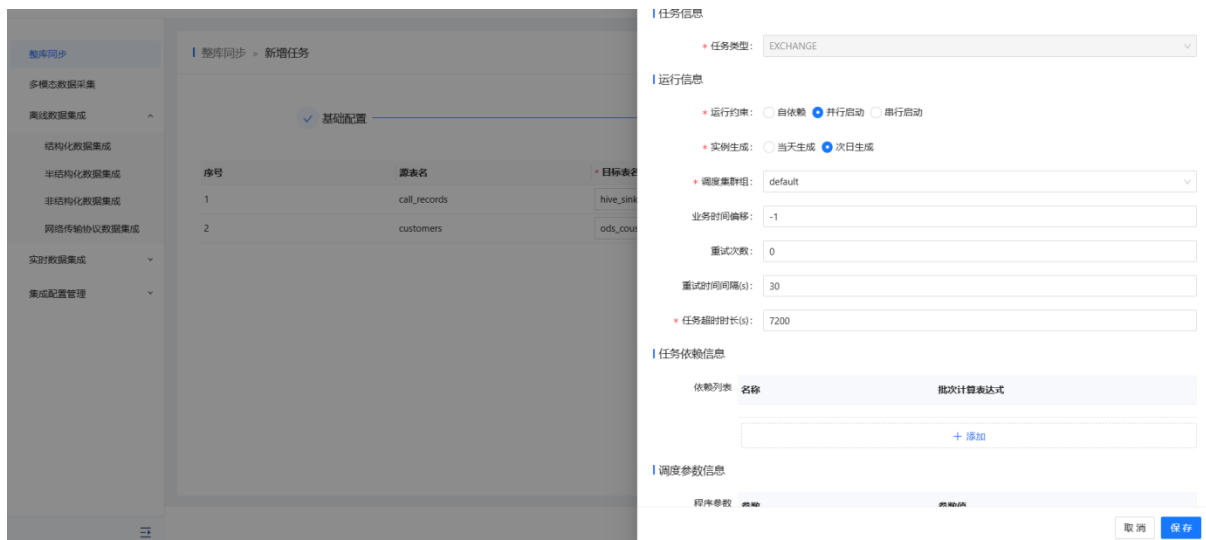
在任务页面右侧悬浮球中，可以进行调度配置。



调度配置分为以下几个部分：

- 基本信息
  - 任务名称：默认与任务英文名保持一致，不可更改。
  - 优先级：分为极高、高、中、极低等 5 个优先级，可根据实际情况配置，默认为中。

- 运维负责人：可指定该任务的运维负责人。
- 调度信息
  - 实例化规则：分为 1:1 及 1:N 两种情况：
    - 1:1：指一个任务生成 1 个运行实例。
    - 1:N：指一个任务生成 N 个运行实例。
  - 生效时间：指当前任务开始生效的时间，默认时间为当天 00:00。
  - 失效时间：指该周期任务执行到什么时间失效，默认为 2999-12-31。
  - 执行周期：指该周期任务的周期类型，分为月、日、小时、分钟，本案例中配置为日周期任务。
  - 执行时间：指任务实例开始运行的具体时间。根据不同的执行周期，具有不同的含义及填写内容：
    - 若执行周期为月：执行时间配置项为每月 X 号 X 时 X 分，如每月 1 日 12 时 00 分运行实例。
    - 若执行周期为日：执行时间配置项为每日 X 时 X 分，如每日 00 时 00 分运行实例。
    - 若执行周期为小时：执行时间配置项为每小时 X 分起，如每小时 5 分起运行实例，注意此处的范围为 0~59。
    - 若执行周期为分钟：执行时间配置项为每分钟 X 秒起，如每分钟 5 秒起运行实例，注意此处的范围为 0~59。
  - crontab 表达式：由上述执行周期与执行时间配置组合出的表达式，不可编辑。
  - 最近执行时间：指预计该任务上线后，各实例运行的时间，不可编辑。



- 任务信息

- 标识该任务属于何种业务，每个业务模块默认，不可编辑，用户无需关心此项配置。

- 运行信息

- 运行约束：用于定义同一任务中不同实例的依赖情况，默认为并行启动。
  - 自依赖：当前的周期实例的执行，必须要等到上个周期实例执行成功后才能开始执行。
  - 并行启动：任务的多个周期实例，可以一起运行，相互之间没有依赖关系。
  - 串行启动：当前的周期实例的执行，只需要上一个周期实例是运行结束状态即可（不管成功与否）。
- 实例生成：用于定义任务实例的生成时间，默认为次日生成。
  - 当天生成：任务上线后，实例当天就开始生成。
  - 次日生成：任务上线后，实例第二天才开始生成。
- 调度集群组：需选择用于调度的调度集群组，默认为 default 集群组。若有其他集群组也可选择。
- 业务时间偏移：对于数据的账期偏移量。偏移量的值为 0 至负数。如当前日期为 2025 年 1 月 15 日，以下对一些数值进行说明：
  - 业务时间偏移填写 0：表示该任务采集的是数据日期为当日的的数据，即任务实例运行时，采集数据时间为 2025 年 1 月 15 日的数据。

- 业务时间偏移填写-1：表示该任务采集的是数据日期为前一日的的数据，即任务实例运行时，采集数据时间为 2025 年 1 月 14 日的数据。
- 业务时间偏移填写-2：表示该任务采集的是数据日期为前两日的数据，即任务实例运行时，采集数据时间为 2025 年 1 月 13 日的数据。
- 依此类推。

此处需注意，增量同步中，若增量条件也配置了时间相关的变量，此时增量条件中的时间变量受到业务偏移量的影响。

以本案例为例，增量条件中配置了时间变量`batchNo?calDate(0,'D')`，若业务偏移填写为-1，则增量条件中表示为采集前一天的数据。若业务偏移填写为 0，则增量条件中表示为采集当天的数据。

- 重试次数：指当任务实例失败后，需要自动重试的次数。
- 重试时间间隔：指当任务实例失败后，每次自动重试的时间间隔。
- 任务超时时长：可设置定义任务超时的时长。默认为 7200 秒。
- 任务依赖信息
  - 依赖列表：指不同任务间的依赖关系配置列表。
- 调度参数信息
  - 程序参数：用于扩展额外的配置项，以 k-v 的形式进行录入，如有特殊情况需要可以添加配置项，无额外需求可不配置此项。

## 常见业务场景与最佳实践

### MySQL 单表离线周期同步至 Hive

#### 使用限制

无

#### 前提条件

- 1、已完成来源数据源、目标数据源的配置

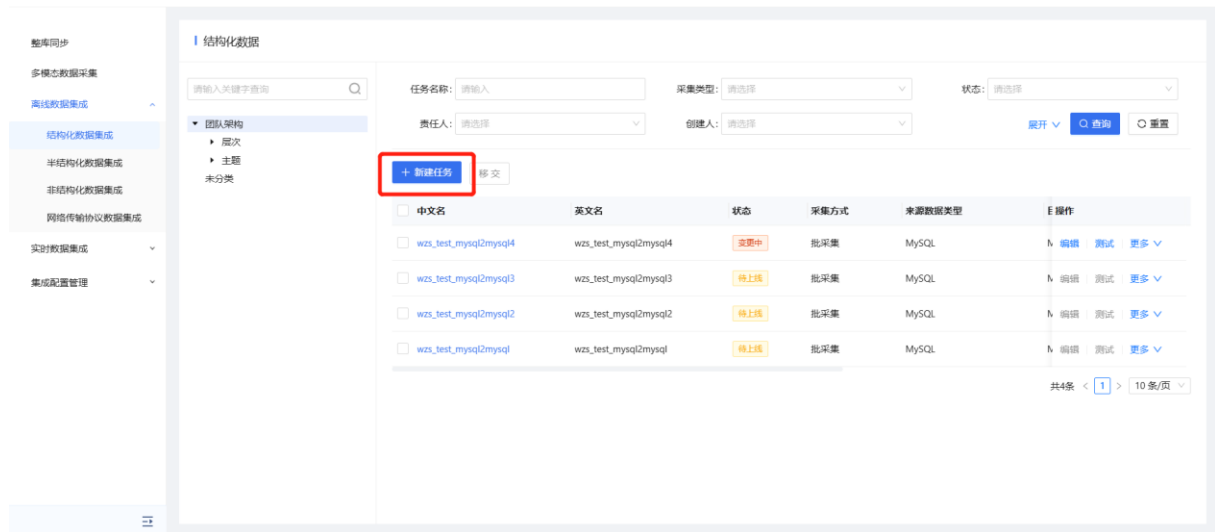
本场景以 MySQL 作为来源数据源，以 Hive 作为目标数据源。因此需要在数据源管理中新增

MySQL 数据源和 Hive 数据源。

- 2、完成目标数据源的数据建模。
- 3、完成来源数据源与目标数据源的元数据采集，并将元数据权限授予当前登录账号。
- 4、如在字段处理部分需要使用 UDF 函数，则需在 UDF 函数管理部分完成 UDF 函数的创建。

## 创建同步任务

- 1、登录并进入数据集成模块页面，点击进入【离线数据集成】-->【结构化数据集成】任务管理页面。点击【新建任务】按钮。



- 2、填写任务的基础信息，包括【任务英文名】、【任务中文名】、【描述】等，点击确定。

配置项	配置说明
任务英文名	任务英文名团队内唯一，必须以英文字母开头，支持字母、数字、下划线，如：ods_data_20241225.
任务中文名	任务中文名不做唯一限制，支持汉字、字母、数字、下划线，如 ODS_数据_20241225.
描述	任务的相关描述。

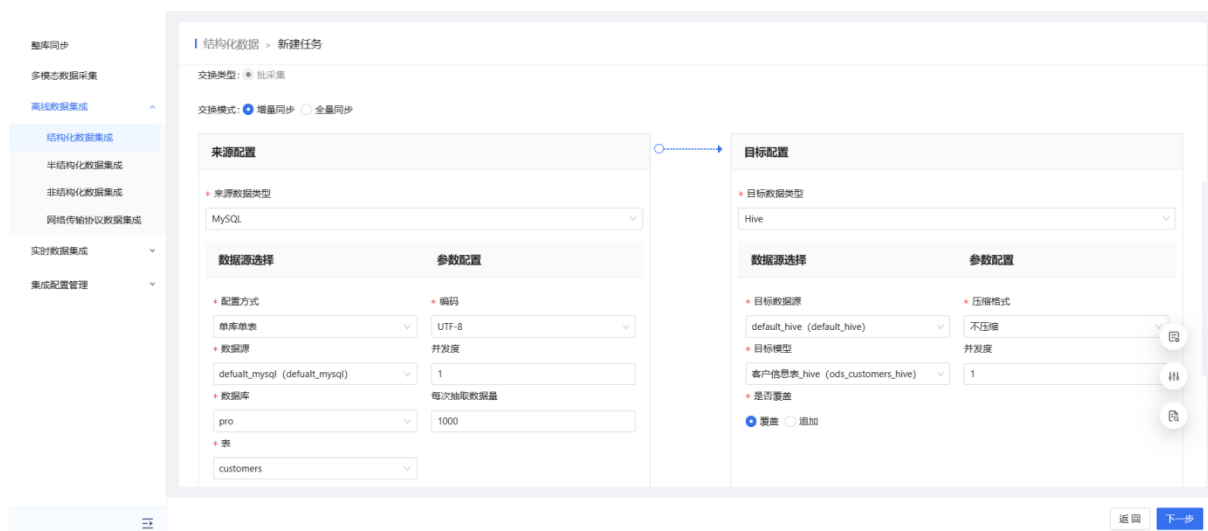


3、确认后进入数据集成任务配置页面，配置基本的任务信息。

- **交换类型：** Mysql 数据源仅支持批采集，即批次采集。
- **交换模式：**
  - 若您需要进行一次性全量或周期性全量同步，此处请选择**全量同步**。
  - 若您需要进行一次性增量或周期性增量同步，此处请勾选**增量同步**。
- **来源配置：**
  - 选择来源数据类型为：**MySQL**
  - 配置方式默认选择：**单库单表**
  - 数据源：选择需要采集数据的 MySQL 数据源，如 **default\_mysql**
  - 数据库：选择需要采集数据的 MySQL 数据库，如 **pro**
  - 数据表：选择需要采集数据的数据表，如 **customers**
  - 编码：指数据采用的编码格式，默认为 **UTF-8**
  - 并发度：指同时能够进行读取操作的数量，当并发度大于 1 时，step3 数据处理必须选择分片字段，分片字段建议为主键或索引字段；若您资源较少，过高的并发会影响您其他任务的执行，建议您使用默认并发度，**默认并发度为 1**。如大于 1，则需要指定分片字段，分片字段的指定在后续步骤中介绍。
  - 每次抽取数据量：指每次执行 SQL 时抽取的数据条数，建议填写范围为 [100,100000]。如填写 **1000**，则表示每次执行 SQL 时抽取 1000 条数据。

## ● 目标配置：

- 选择目标数据类型为：**Hive**
- 目标数据源：选择需要存放数据的 Hive 数据源，如 **default\_hive**
- 目标模型：选择需要存放数据的 Hive 模型，如**客户信息表\_hive**  
(ods\_customers\_hive)
- 是否覆盖：指写入数据时，对于目标表中的存量数据，写入时的策略
  - 若您不需要保留目标表的存量数据，此处可以选择覆盖，即新数据将覆盖老数据进行写入。
  - 若您需要保留目标表的存量数据，此处可以选择追加，即新数据将追加在老数据后进行写入。
  - 在周期性集成中，为保证数据的完整性，一般建议选择**追加**。
- 压缩格式：指数据是否需要压缩以节省数据存储空间，目前提供 **LZO** 和**不压缩**两种选择，这里选择**不压缩**。
- 并发度：指同时能够进行写入操作的数量，若您资源较少，过高的并发会影响您其他任务的执行，建议您使用默认并发度，默认并发度为 1。



## ● 资源配置

- Yarn 资源：集成任务执行时的 yarn 队列，下拉选择需要的队列，如 **default\_yarn**。
- 生产队列：所选 Yarn 队列中配置的生产队列，下拉选择，如 **default**。
- 测试队列：所选 Yarn 队列中配置的测试队列，下拉选择，如 **default**。

- 参数配置：此项为高级配置项
  - jobManager 内存：可设置 jobManager 内存，单位为 MB
  - taskManager 并发：taskManager 并发默认值为 1，范围为[1,64]
  - taskManager Vcore：可设置 taskManager Vcore 的数量
  - taskManager 内存：可设置 taskManager 内存，单位为 MB
- 流量控制（选填）
  - 若您需要按带宽限制数据流量，此处选择**按字节控制**。
  - 若您需要按数据量限制数据流量，此处选择**按条数控制**。

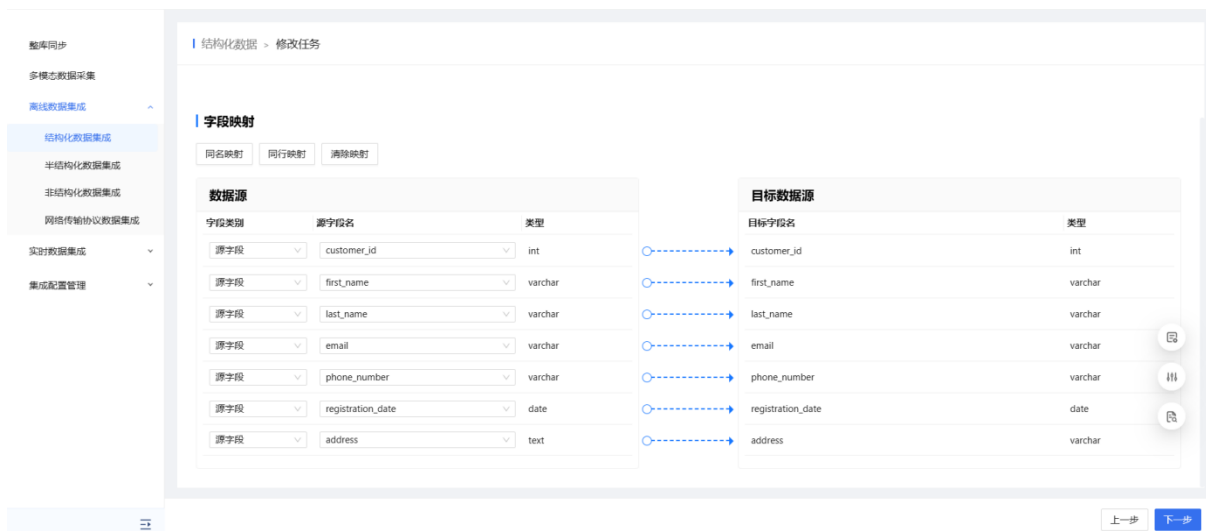
The screenshot shows a configuration page for '结构化数据' (Structured Data) with the following sections:

- 结构化数据 > 修改任务**: Includes a dropdown menu with 'pro' and a text input field containing '1000'. Below it is a table with a '+' icon and a dropdown menu showing 'customers'.
- 资源配置**: Includes dropdown menus for 'Yarn资源: default\_yarn (default\_yarn)', '生产队列: default', and '测试队列: default'.
- 参数配置**: Includes input fields for 'jobManager内存: 请输入...' (with 'MB' unit), 'taskManager并发: 1', 'taskManager Vcore: 请输入taskMnager Vcore数量', and 'taskManager内存: 请输入...' (with 'MB' unit).
- 流量控制**: Includes a dropdown menu for '流量控制方式'.

Navigation buttons '返回' and '下一步' are located at the bottom right of the configuration area.

4、基本信息配置完成后，点击右下角【下一步】，进入字段映射配置。

- 字段映射提供两种映射方式（默认为同名映射）：
  - 若您来源表与目标表的字段名一致，此处选择**同名映射**
  - 若您只需要将来源表的字段按照顺序映射到目标表，此处选择**同行映射**



- 字段映射提供目标字段赋值功能，字段类别选择**自定义字段**，支持为目标字段赋予常量、变量等。
  - 若您需要对目标字段赋予固定不变的值，如同步的一批数据都属于一个地区四川省成都市，来源表中没有地区字段，而目标表中有地区字段，则可以使用常量进行目标表地区字段的赋值，可在字段映射时，配置**自定义字段**，选择**常量**，填写：**四川省成都市**，此时同步后的数据都会带有该地区值。
  - 若您需要对目标字段赋予跟随调度账期而变化的值，可以选择变量赋值，常见的场景如周期为天的集成任务，来源表中未记录数据所属日期，而目标表中要求记录数据所属日期，此时可以通过选择**自定义字段**，选择**变量**，填写：**\${batchNo}**，此时将会填充调度账期至目标表的时间字段中。

5、字段映射配置完成后，点击右下角【下一步】，进入数据处理配置。

- 数据处理分为三个部分：字段处理、数据过滤、脏数据处理。
- 分片字段：分片字段定义主要用于数据分片，若在基础信息配置中，读取并发度配置大于 1，则在此处必须指定分片字段。
- 分区取值方式：若目标表设置了分区字段，则需进行分区取值方式的配置。分区取值方式分为两种：枚举取值、时间取值。
  - 枚举取值：枚举值是根据源端字段的实际值做动态分区值处理，一般建议源端字段此值个数有限，属于可枚举的状态。使用枚举值作为分区取值时，需要注意分区数量每天不能超过 1000 个，超过后交换程序会报错，无法正常执行。

- 时间取值：时间值主要用于源字段取值为时间值的情况，系统基于配置的规则对字段取值进行识别和转化后作为动态分区所使用的分区值。选择时间取值方式时，需要选择来源格式和目标格式。

- 来源格式可选 13 位时间戳、10 位时间戳、时间字符串。

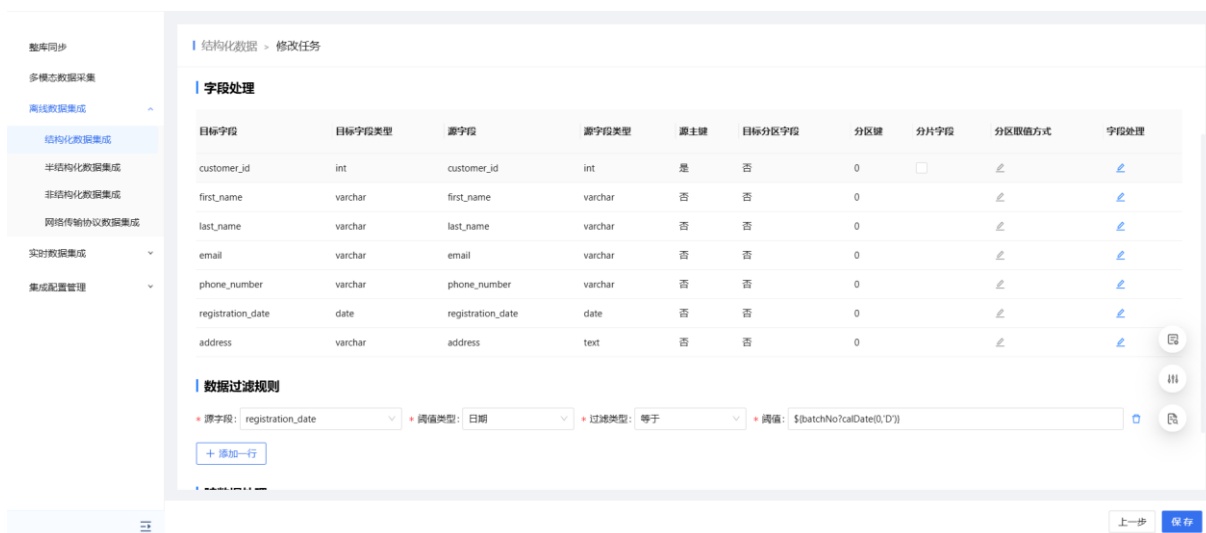
- 10 位时间戳：秒级别时间戳，指定后系统直接将源字段取值当做 10 位时间戳来识别。
- 13 位时间戳：毫秒级别时间戳，指定后系统直接将源字段取值当做 13 位时间戳来识别。
- 时间字符串：本身是字符串，但是使用一定序列化方式代表时间，如“2023-08-01 08:14:50”，此类来源需要单独指定来源字符串格式。

- 目标格式：系统根据指定的规则将对应字符串写入目标端，需要保证填写的目标格式能够从序列化的来源字符串中获取，若系统无法获取，会将整条数据作为脏数据处理。。

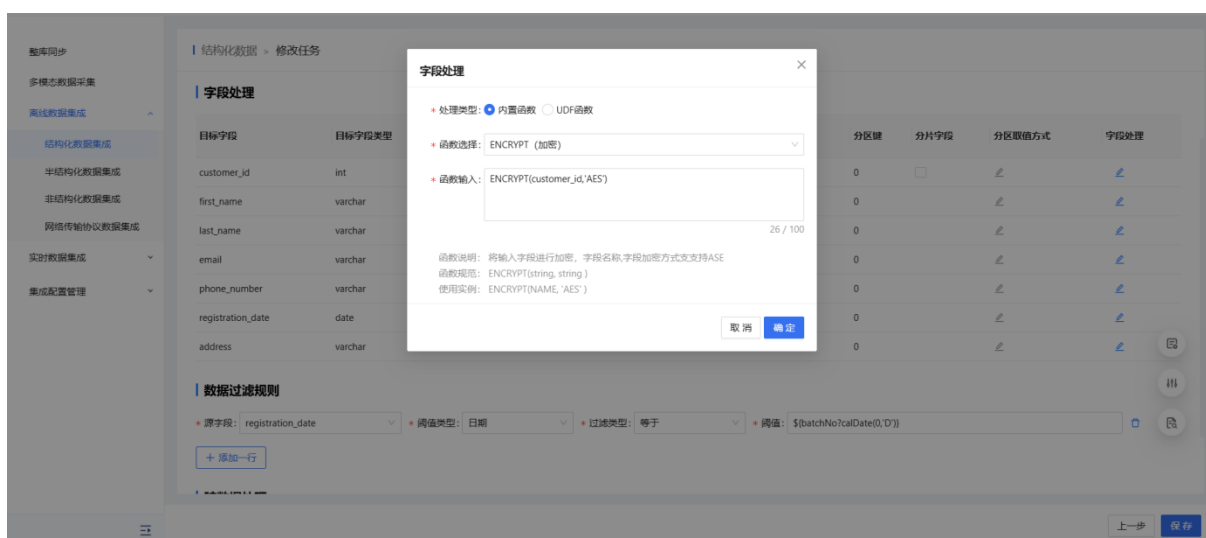
- yy: 2 位纪年
- yyyy: 4 位纪年
- MM: 2 位月份
- dd: 2 位日期
- HH: 2 位小时，24 小时制
- hh: 2 位小时，12 小时制
- 若需要基于原表字段取值的小时进行分区，则输入 yyyyMMddHH

- **字段处理**

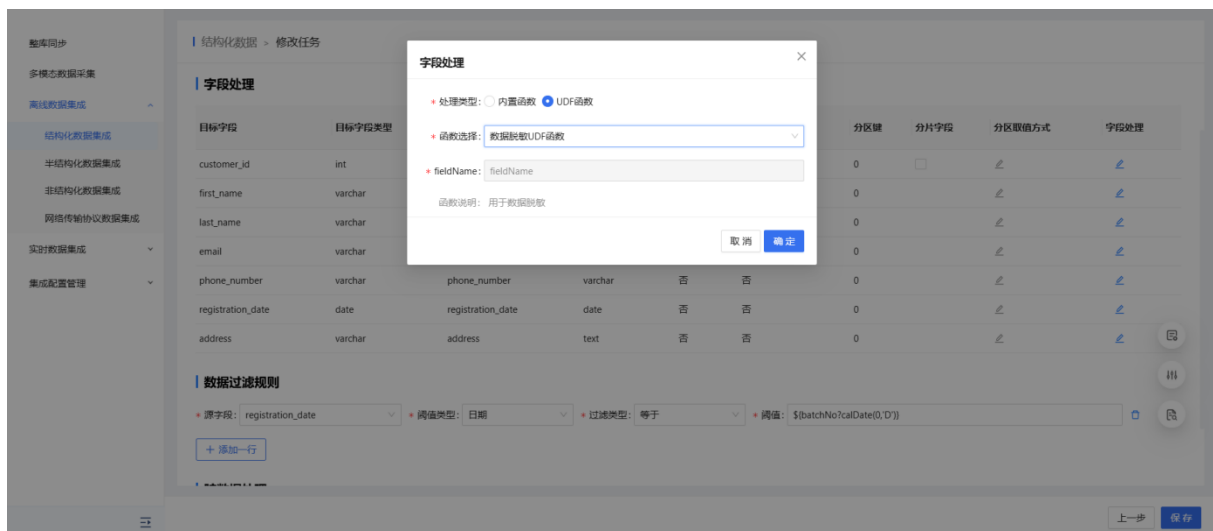
- 您可以根据自身业务需要，采用内置函数进行数据处理或 UDF 函数进行数据处理。
- 内置函数：内置为 seatunnel 原生支持的字段处理函数，包含数值型函数、字符串函数、日期函数等 90+函数类型。如需对 mysql\_id 字段进行加密，则可点击 mysql\_id 字段后的【编辑】图标，设置加密函数。



- 点击【编辑】按钮后，弹出函数编辑框，选择**内置函数**，选择加密函数 **ENCRYPT**，根据提示的函数规范输入：**ENCRYPT(customer\_id,'AES')**后，点击确定，完成内置函数的配置。



- 若您需要选择 UDF 函数进行字段数据处理，则选择 UDF 函数。
- 如需要对 customer\_id 使用 UDF 函数进行数据脱敏，则在处理类型处选择 UDF 函数，选择之前创建好的 UDF 函数：**数据脱敏 UDF 函数**，进行数据脱敏。



- 数据过滤（增量条件定义）

- 若您在之前选择的是增量同步，则在此处可配置增量条件。

- 如需要增量采集前一天的数据，则可在此处配置：

- 源字段：此处以 registration 字段作为增量字段，所以选择 **registration**。

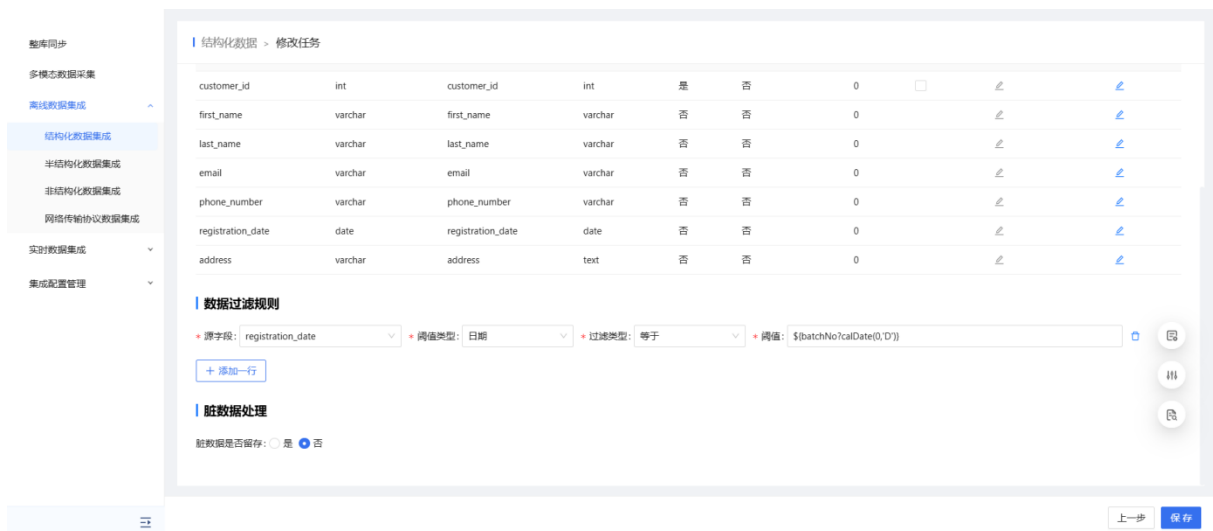
- 阈值类型：此处选择日期

- 若所选字段的字段类型为 **date**，阈值类型需选择**日期**；

- 若所选字段的字段类型为 **datetime 或 timestamp**，阈值类型需选择**日期时间**

- 过滤类型：根据实际情况选择，如本案例中需要采集前一天的数据，这里可以选择**等于**。

- 阈值：指判断条件，如需要引用调度账期，可以查看页面右侧第二个悬浮按钮，进入全局变量进行查看。本案例中需同步前一天的数据，结合调度业务偏移量来选择，此处填写`${batchNo?calDate(0,'D')}`



注意：当增量条件涉及到日期时间，需要结合调度配置中的业务偏移量来选择对应的变量值。

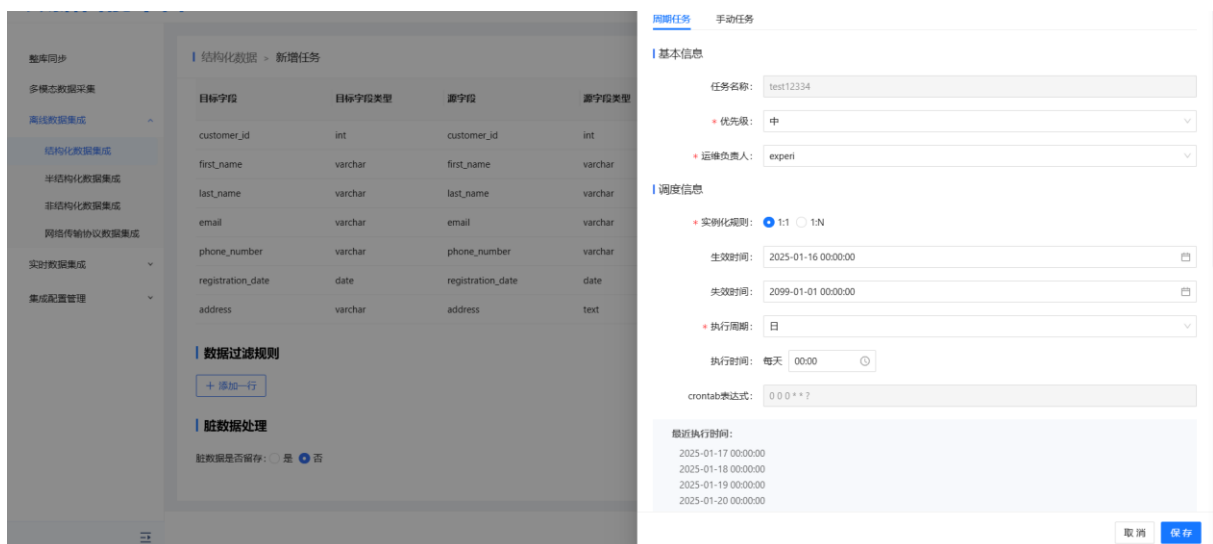
- 脏数据过滤

- 当前版本定义的脏数据，为数据过滤规则过滤后的数据，可以选择是否留存这些数据。

- 用户可以指定该批被过滤数据的留存路径，同时系统提供默认的留存路径：

hdfs://dirty\_data/\${job\_id}。

6、整体任务配置完成后，需要进行调度配置。



调度配置分为以下几个部分：

- 基本信息

- 任务名称：默认与任务英文名保持一致，不可更改。

- 优先级：分为极高、高、中、极低等 5 个优先级，可根据实际情况配置，默认为

中。

○ 运维负责人：可指定该任务的运维负责人。

● 调度信息

○ 实例化规则：分为 1:1 及 1:N 两种情况：

■ 1:1：指一个任务生成 1 个运行实例。

■ 1:N：指一个任务生成 N 个运行实例。

○ 生效时间：指当前任务开始生效的时间，默认时间为当天 00:00。

○ 失效时间：指该周期任务执行到什么时间失效，默认为 2999-12-31。

○ 执行周期：指该周期任务的周期类型，分为月、日、小时、分钟，本案例中配置为日周期任务。

○ 执行时间：指任务实例开始运行的具体时间。根据不同的执行周期，具有不同的含义及填写内容：

■ 若执行周期为月：执行时间配置项为每月 X 号 X 时 X 分，如每月 1 日 12 时 00 分运行实例。

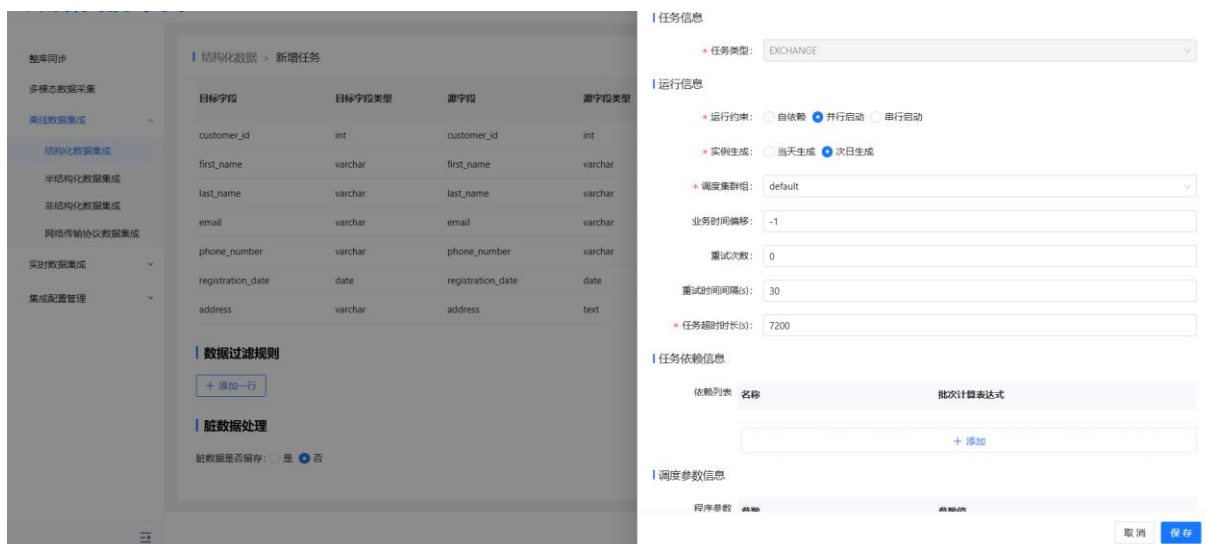
■ 若执行周期为日：执行时间配置项为每日 X 时 X 分，如每日 00 时 00 分运行实例。

■ 若执行周期为小时：执行时间配置项为每小时 X 分起，如每小时 5 分起运行实例，注意此处的范围为 0~59。

■ 若执行周期为分钟：执行时间配置项为每分钟 X 秒起，如每分钟 5 秒起运行实例，注意此处的范围为 0~59。

○ crontab 表达式：由上述执行周期与执行时间配置组合出的表达式，不可编辑。

○ 最近执行时间：指预计该任务上线后，各实例运行的时间，不可编辑。



## ● 任务信息

- 标识该任务属于何种业务，每个业务模块默认，不可编辑，用户无需关心此项配置。

## ● 运行信息

- 运行约束：用于定义同一任务中不同实例的依赖情况，默认为并行启动。
  - 自依赖：当前的周期实例的执行，必须要等到上个周期实例执行成功后才能开始执行。
  - 并行启动：任务的多个周期实例，可以一起运行，相互之间没有依赖关系。
  - 串行启动：当前的周期实例的执行，只需要上一个周期实例是运行结束状态即可（不管成功与否）。
- 实例生成：用于定义任务实例的生成时间，默认为次日生成。
  - 当天生成：任务上线后，实例当天就开始生成。
  - 次日生成：任务上线后，实例第二天才开始生成。
- 调度集群组：需选择用于调度的调度集群组，默认为 default 集群组。若有其他集群组也可选择。
- 业务时间偏移：对于数据的账期偏移量。偏移量的值为 0 至负数。如当前日期为 2025 年 1 月 15 日，以下对一些数值进行说明：
  - 业务时间偏移填写 0：表示该任务采集的是数据日期为当日的的数据，即任务

实例运行时，采集数据时间为 2025 年 1 月 15 日的数据。

- 业务时间偏移填写-1：表示该任务采集的是数据日期为前一日的的数据，即任务实例运行时，采集数据时间为 2025 年 1 月 14 日的的数据。
- 业务时间偏移填写-2：表示该任务采集的是数据日期为前两日的的数据，即任务实例运行时，采集数据时间为 2025 年 1 月 13 日的的数据。
- 依此类推。

此处需注意，增量同步中，若增量条件也配置了时间相关的变量，此时增量条件中的时间变量受到业务偏移量的影响。

以本案例为例，增量条件中配置了时间变量 $\${batchNo?calDate(0,'D')}$ ，若业务偏移填写为-1，则增量条件中表示为采集前一天的数据。若业务偏移填写为 0，则增量条件中表示为采集当天的数据。

- 重试次数：指当任务实例失败后，需要自动重试的次数。
- 重试时间间隔：指当任务实例失败后，每次自动重试的时间间隔。
- 任务超时时长：可设置定义任务超时的时长。默认为 7200 秒。
- 任务依赖信息
  - 依赖列表：指不同任务间的依赖关系配置列表。
- 调度参数信息
  - 程序参数：用于扩展额外的配置项，以 k-v 的形式进行录入，如有特殊情况需要可以添加配置项，无额外需求可不配置此项。

# 数据研发

星海·数据中台是智能湖仓一体数据开发平台，沉淀电信多年大数据研发经验，深度适配 Hadoop 生态的各类大数据组件，例如 Spark、Flink、Ceph 和 Doris 等数十种大数据计算服务，为数据仓库、数据湖、湖仓一体数据架构提供智能化 ETL 开发服务

## 数据开发相关概念：

概念	描述
程序	程序是一种批处理文件的延伸，是一种纯文本保存的程序，一般所说的计算机脚本程序是确定的一系列控制计算机进行运算操作动作的组合，在其中可以实现一定的逻辑分支等。
任务	任务由一个或多个程序组成，执行任务可以完成对数据的一系列操作。
资源	用户可以上传自定义的代码或文本文件作为资源，以便在节点运行时调用。
环境变量	环境变量是在操作系统中一个具有特定名字的对象，它包含了一个或者多个应用程序所使用到的信息。
workflow	以业务视角出发的 DAG 可视化开发界面，轻松管理复杂的任务工。Workflow 支持编排数十种不同类型的节点，提供 Python、Shell、MySQL、Doris、Hive、Presto、DM（达梦）、Spark、PySpark、PostgreSql、GaussDB、SqlServer、Kingbase、Clickhouse、Oracle 等不同类型的节点；同时支持 Workflow 级的调度配置。
手动 workflow	面向某一特定业务需求的任务、表、资源、函数的集合。 手动 workflow 与周期 workflow 的区别为：手动 workflow 中的任务需手动触发运行，而周期 workflow 中的任务是按计划来定时触发运行。
任务节点	任务节点是星海·数据开发平台的基本执行单元。星海·数据开发平台提供多种类型的节点，包括用于数据处理的引擎计算节点（例如，Python、Shell、MySQL、Doris），以及可对引擎计算节点进行复杂逻辑处理的通用节点，多种节点配合使用，满足不同的数据处理需求。
依赖关系	任务间通过依赖关系定义任务的运行顺序。如果节点 A 运行后，节点 B 才能运行，我们称 A 是 B 的上游依赖，或者 B 依赖 A。在 DAG 中，依赖关系用节点间的箭头表示。
输出表名	输出表名建议配置为当前任务的产出表，正确填写输出表名可以方便下游设置依赖时确认数据是否来自期望的上游表。自动解析生成输出表名时不建议手动修改，输出表名仅作为标识，修改输出表名不会影响 SQL 脚本实际产出的表

	<p>名，实际产出表名以 SQL 逻辑为准。</p> <p><b>说明</b></p> <p>节点的<b>输出名</b>需要全局唯一，而<b>输出表名</b>无此限制。</p>
<b>调度资源组</b>	指用于任务调度的资源组。资源组介绍详情请参见。
<b>调度参数</b>	调度参数是代码中用于调度运行时动态取值的变量。代码在重复运行时若希望获取到运行环境的一些信息，例如日期、时间等，可根据星海·数据中台调度系统的调度参数定义，动态为代码中的变量赋值。
<b>业务日期</b>	通常指的是与业务活动直接相关的日期，这个日期反映了业务数据的实际发生时间。这个概念在离线计算场景中尤为重要，例如在零售业务中，您需要统计 20241010 日的营业额，往往会在 20241011 日凌晨再开始计算，这时所计算出来的数据实际是 20241010 日的营业额，这个 20241010，就是业务日期。
<b>定时时间</b>	<p>用户为周期任务设置预期执行的时间点，可精确到分钟级别。</p> <p><b>重要</b></p> <p>影响任务运行的因素较多，并不意味着定时时间到了，任务就会立即执行。在任务执行前，星海·数据中台会检测上游任务是否运行成功、定时时间是否已达到、调度资源是否充足，当上述条件均已满足后，才开始正式触发任务的运行。</p>

## 开发功能介绍

数据开发的主要功能介绍如下：

功能	说明
离线开发	<ul style="list-style-type: none"> <li>● <b>脚本开发</b>：支持多种脚本程序开发（Presto、Mysql、DM、Doris、Hive、Spark、Shell、Python、PySpark、Jar、PostgreSql、GaussDB、SqlServer、KingBase、Clickhouse、Oracle）等十几种数据源，能够满足多种数据处理和分析需求。</li> <li>● <b>低代码开发</b>：通过可视化组件拖拉拽的方式进行开发，大大降低开发门槛，提升开发效率，使非专业开发人员也能快速构建数据处理和应用开发解决方案。</li> </ul>
实时开发	<ul style="list-style-type: none"> <li>● <b>脚本开发</b>：支持通过编写 FlinkSQL 实现实时数据处理脚本开发。</li> <li>● <b>低代码开发</b>：通过可视化组件拖拉拽的方式进行开发，大大降低开发门槛，提升开发效率，使非专业开发人员也能快速构建数据处理和应用开发解决方案。</li> <li>● <b>实时集群管理</b>：实时调度集群是负责管理和协调分布式任务调度</li> </ul>

	<p>的核心组件。用于在集群环境中分发任务、管理资源、监控任务状态以及处理故障恢复。实时调度集群 <b>Broker</b> 与多种大数据组件（<b>Flink</b>、<b>Spark</b>、<b>Hadoop</b>）集成，确保任务的高效执行和资源的最优利用。</p>
算法开发	<ul style="list-style-type: none"> <li>● 可视化建模：基于云原生架构 <b>Pipeline Service -- PAIFlow</b> 的可视化建模工具，提供端到端的机器学习全链路开发环境，帮助您实现端到端的机器学习开发流程。</li> </ul>
数据预处理	<ul style="list-style-type: none"> <li>● 文本数据处理：内置多种自然语言处理算子；</li> <li>● 图片数据处理：内置多种图片数据处理算子；</li> <li>● 音频数据处理：内置多种音频数据处理算子；</li> <li>● 视频数据处理：内置多种视频数据处理算子；</li> </ul>
作业开发	<ul style="list-style-type: none"> <li>● 提供图形化设计器，支持拖拉拽方式快速构建数据处理 workflow。</li> <li>● 预设数据集成、SQL、Shell 等多种任务类型，通过任务间依赖完成复杂数据分析处理。</li> </ul> <p>支持导入和导出作业。</p>
资源管理	<ul style="list-style-type: none"> <li>● 支持统一管理在脚本开发和作业开发使用到的 <b>file</b>、<b>jar</b>、<b>archive</b> 类型的资源。</li> </ul>
作业调度	<ul style="list-style-type: none"> <li>● 支持单次调度、周期调度和事件驱动调度，周期调度支持分钟、小时、天、周、月多种调度周期。调度周期配置为小时，系统支持按间隔小时和离散小时配置调度周期。</li> </ul>
运维监控	<ul style="list-style-type: none"> <li>● 支持对作业进行运行、暂停、恢复、终止等多种操作。</li> <li>● 支持查看作业和其内各任务节点的运行详情。</li> </ul> <p>支持配置多种方式报警，作业和任务发生错误时可及时通知相关人，保证业务正常运行。</p> <ul style="list-style-type: none"> <li>● 调度运维：实时开发的调度运维是确保数据处理实时任务高效、稳定运行的关键环节。调度运维的核心目标包括任务的自动化调度、监控、故障恢复和性能优化。</li> </ul>

## 离线开发

### 脚本开发

星海·数据中台将不同类型引擎任务封装为不同脚本任务，通过创建脚本来生成数据开发任务。

同时，数据开发支持使用资源和函数开发任务。

脚本开发功能提供如下能力：

- 提供在线脚本编辑器，支持 Python 脚本、Shell 脚本、MySQL 脚本、DM（达梦数据库）脚本、Hive 脚本、Spark 脚本、PySpark 脚本、Postgre SQL 脚本、Gauss DB 脚本、SQLserver 脚本、Kingbase 脚本、Clickhouse 脚本、Oracle 脚本等在线代码开发和调测。
- 支持导入和导出脚本。
- 支持使用变量和函数。
- 支持脚本的版本管理能力，支持生成保存版本和提交版本。
- 脚本开发支持细粒度权限管控，对数据开发脚本目录权限管控策略进行配置。

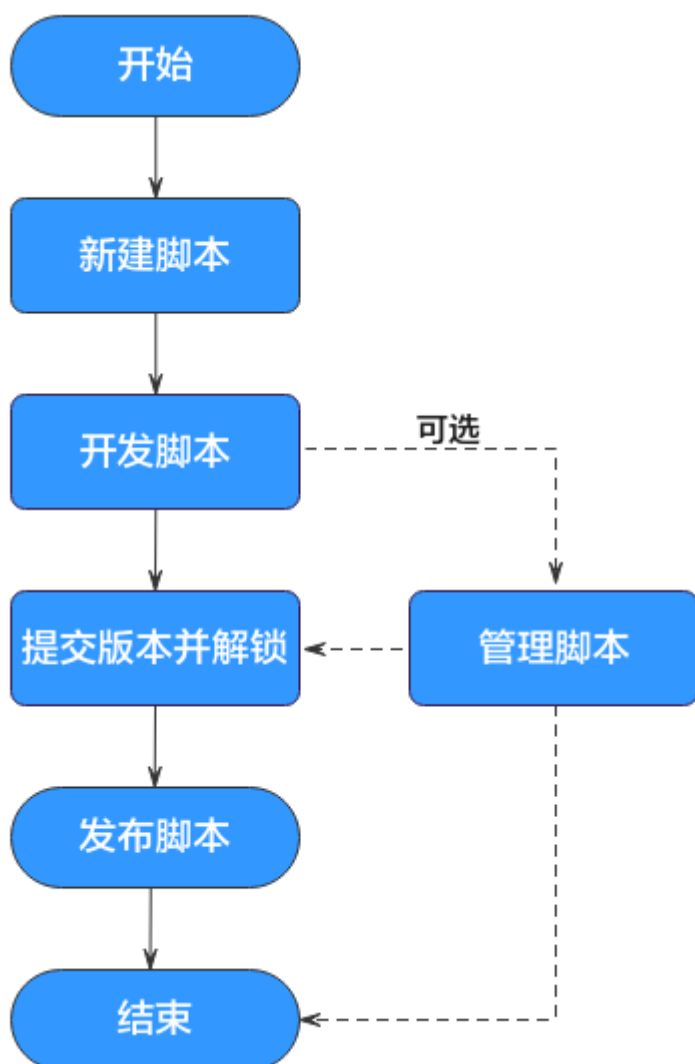
## 前提条件

需要完成数据源接入和数据建模，才能进行数据开发操作。

## 使用限制

- 已拥有开发角色权限。
- 已拥有所需数据源和数据表权限。

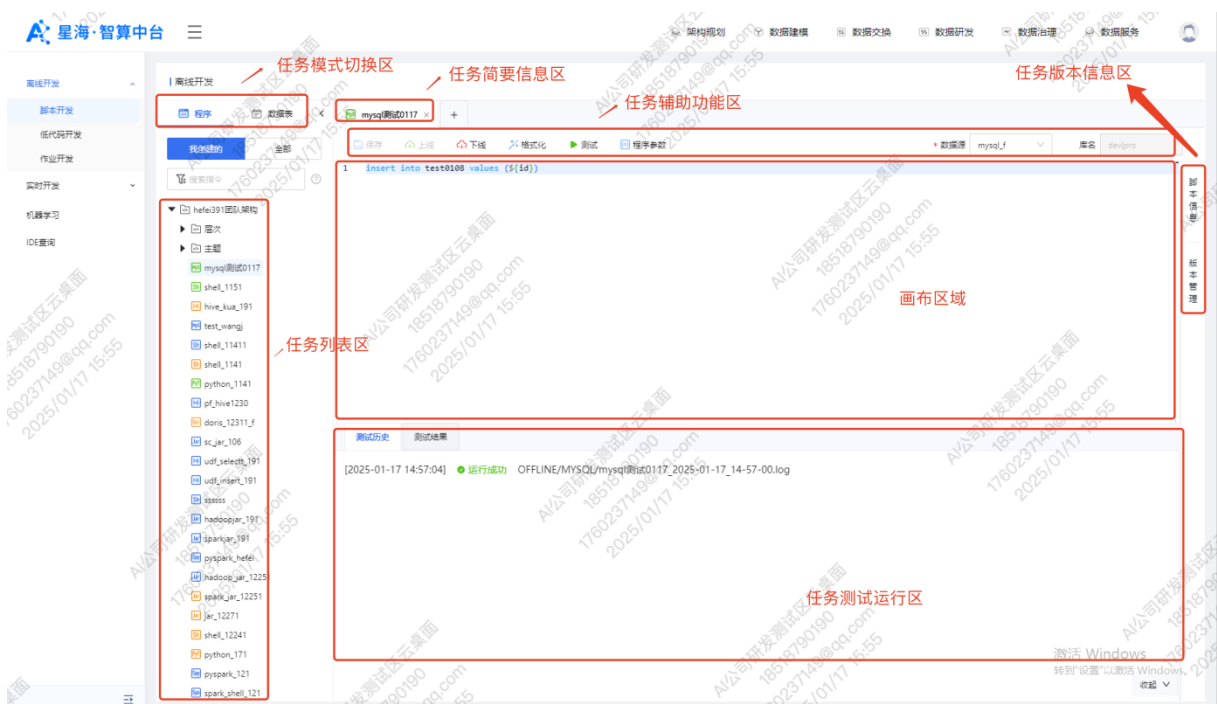
## 开发流程



- 新建脚本：新建相应类型的脚本。
- 开发脚本：基于新建的脚本，进行脚本的在线开发、调试和执行。。
- （可选）管理脚本：脚本开发完成后，您可以根据需要，进行脚本管理。
- 编排脚本：将脚本按照依赖关系进行编排，同时配置任务调度参数。
- 发布脚本：将测试通过的脚本发布上线到生产环境。

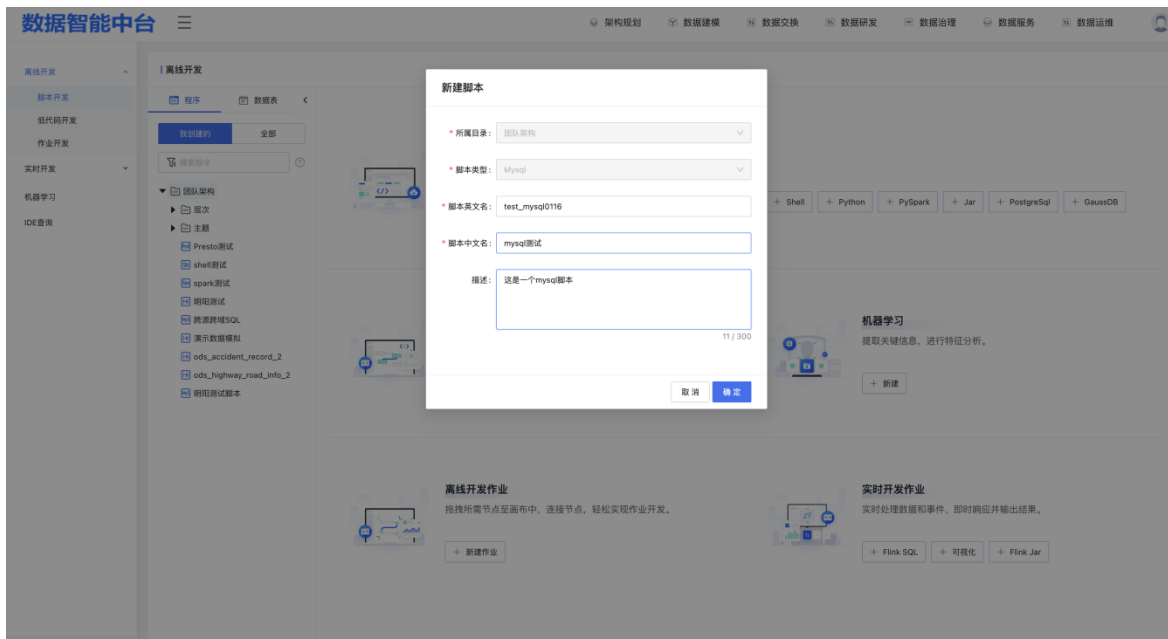
## 页面介绍

脚本开发共包含七个个区块：任务模式切换区、任务列表区、任务简要信息区、任务测试运行区、任务按本信息区、画布区以及任务辅助功能区。

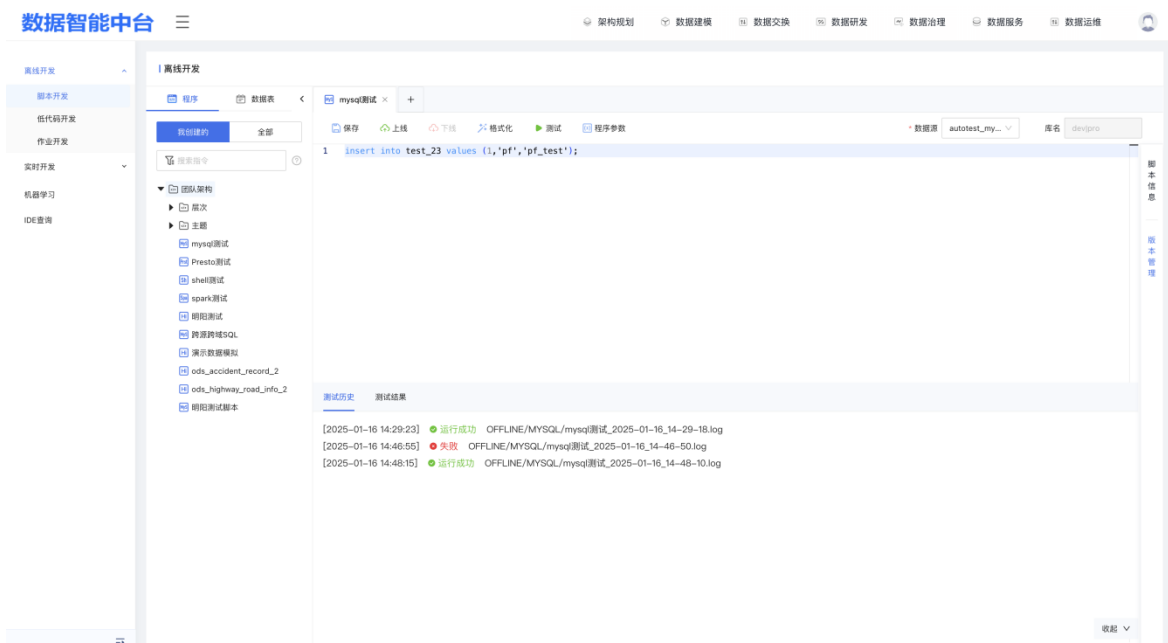
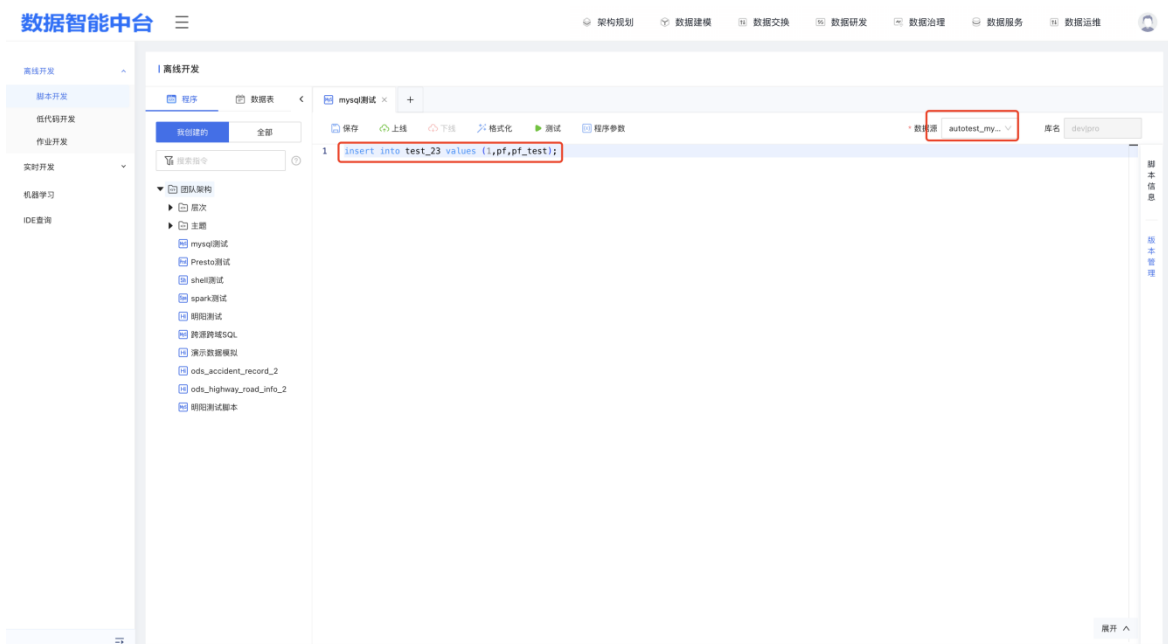


## 脚本开发

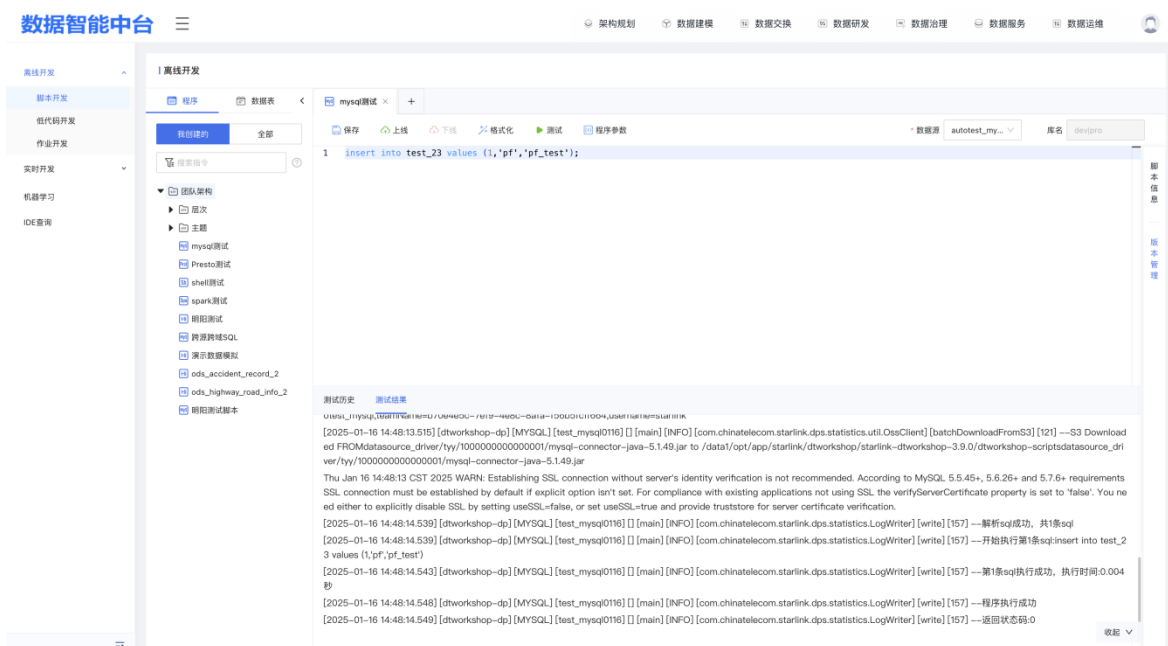
1. 新建 sql 类的脚本程序 (MySQL、PostgreSQL、Oracle、SQL Server、GaussDB、KingBase、DM)



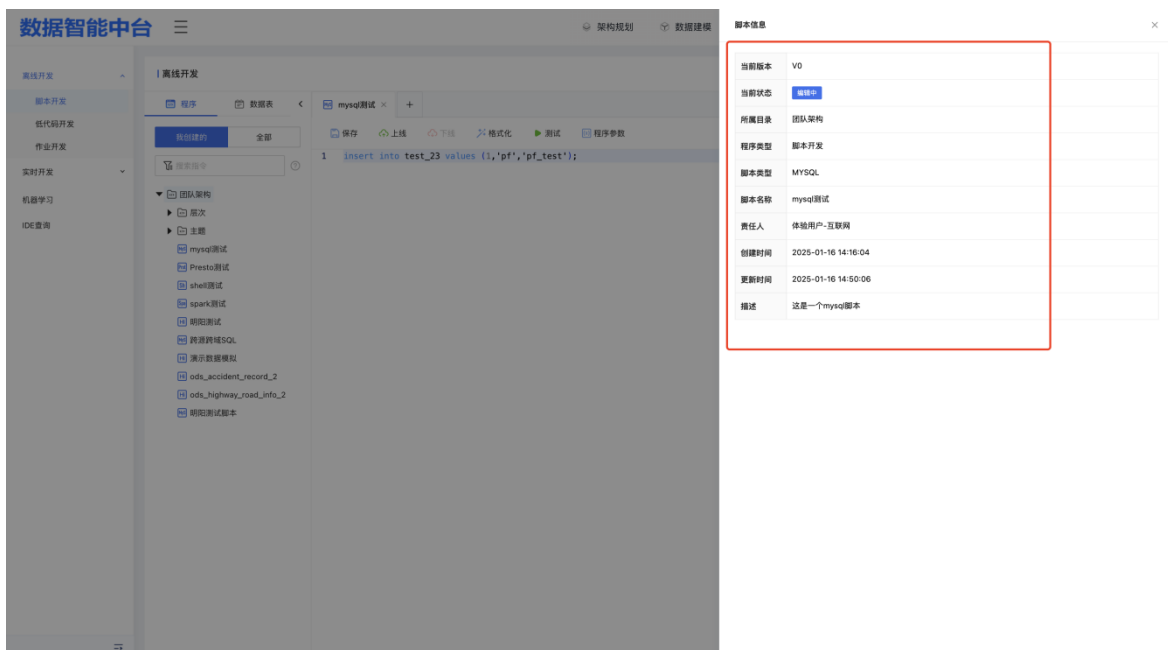
2. 在新建的脚本程序里写 sql，可以进行保存、测试



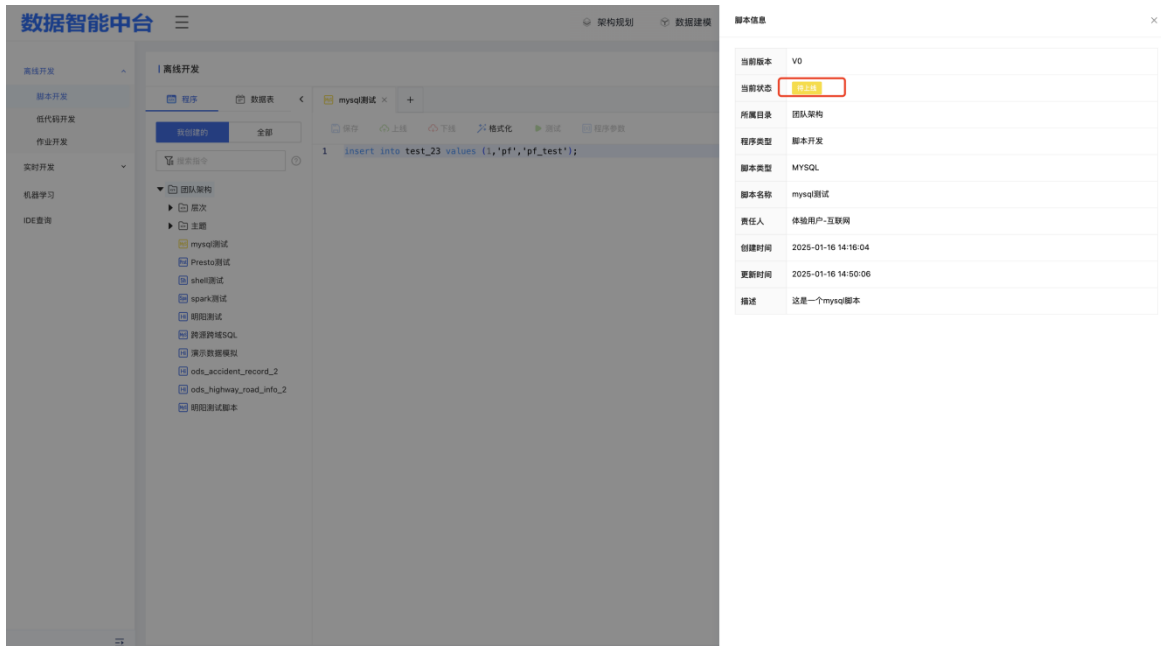
也可以点击测试结果查看具体的测试日志



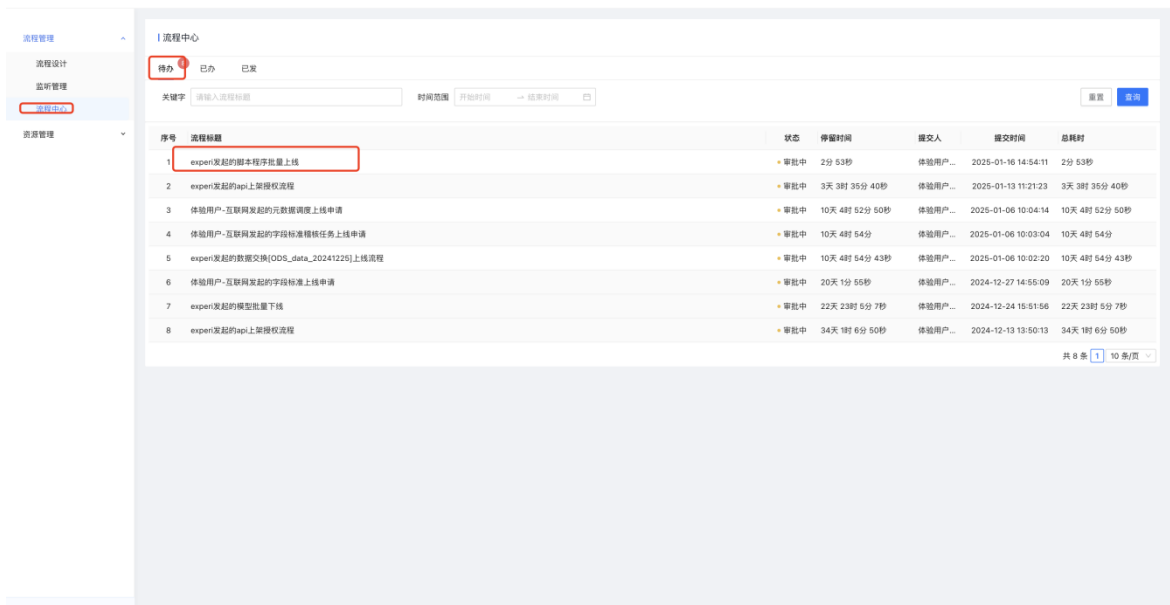
可以点击脚本信息查看脚本详情



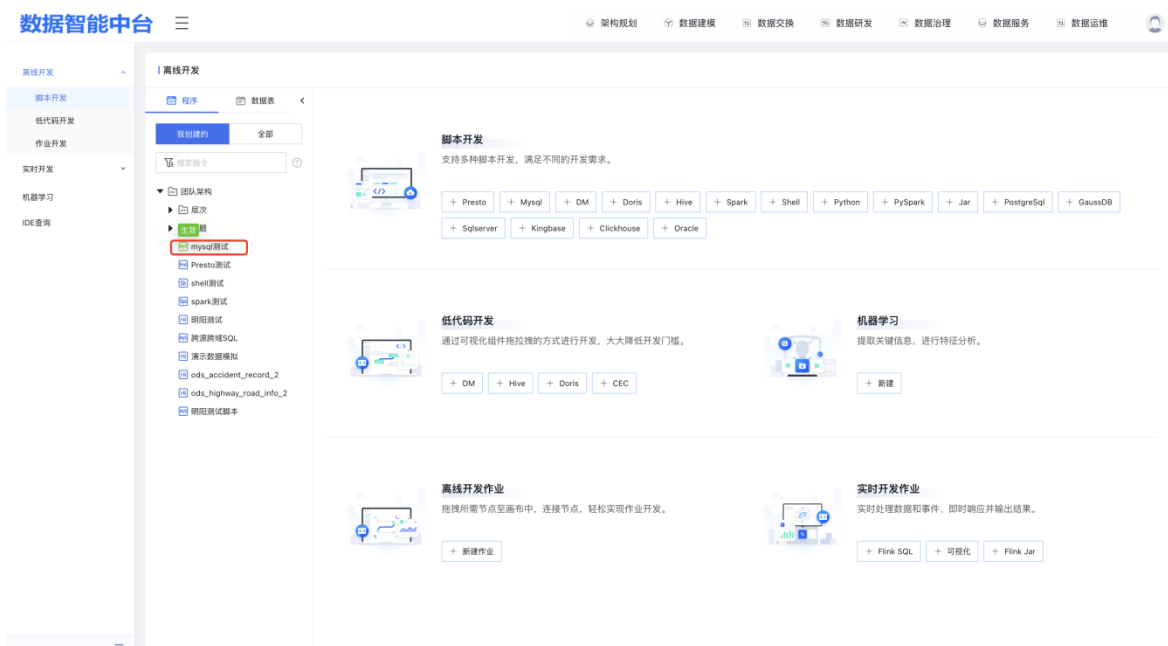
3. 上线脚本程序 (对于新创建的脚本, 都是编辑中状态, 只有点击上线, 会变更为待上线)



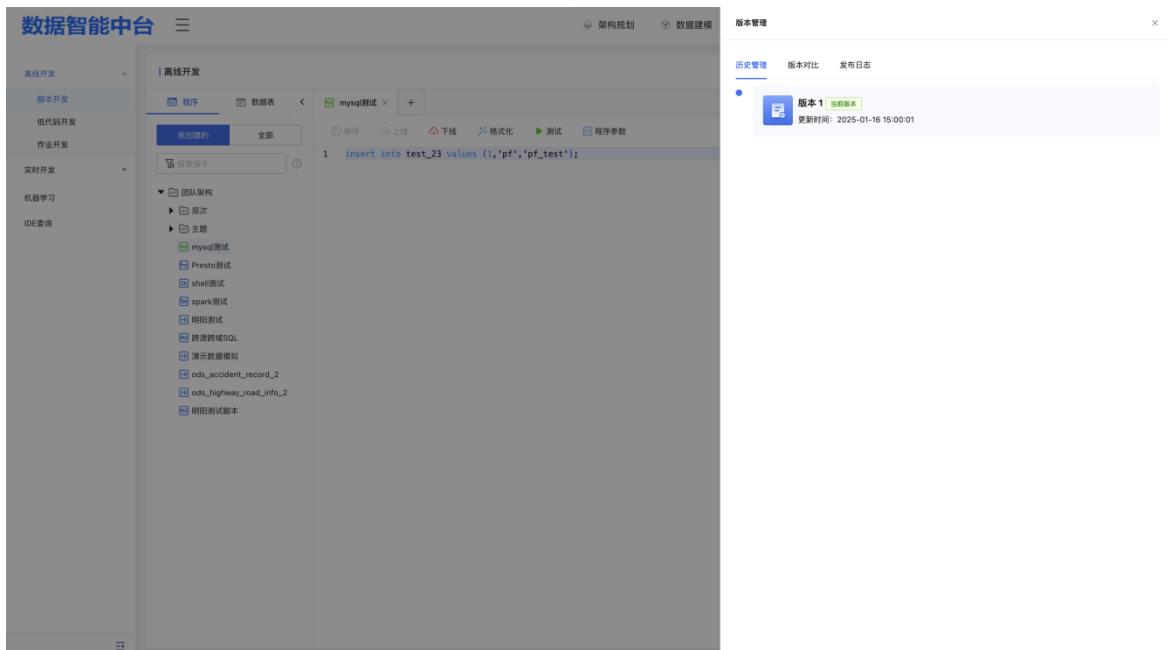
4. 当脚本变更为待上线状态时，说明该脚本已经走入到流程审批中，需要流程审批完，才可以变更为生效状态，上线流程会流转到 admin 管理员用户，需要管理员去流程中心通过审批

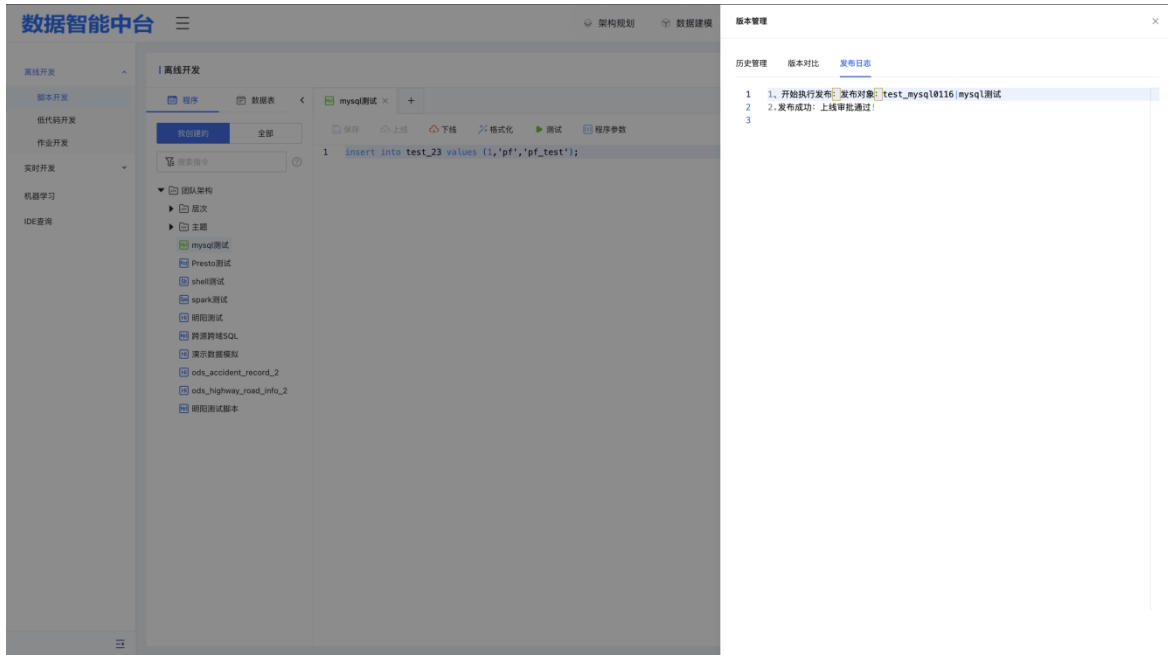
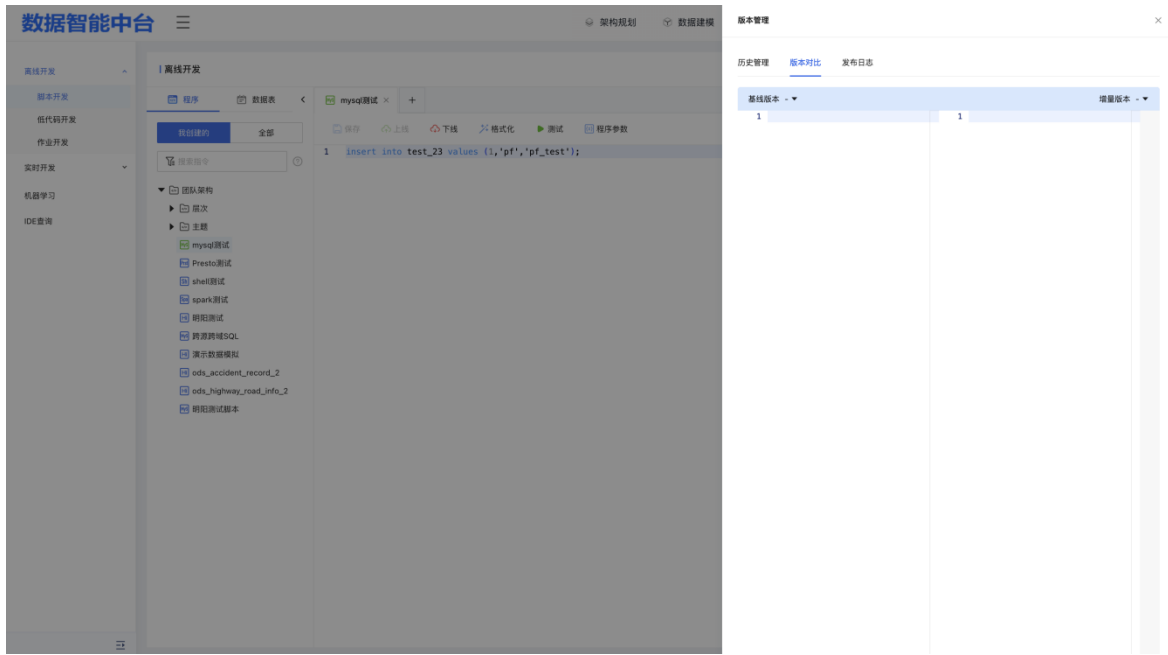


5. 通过审批后，脚本状态会变更为生效

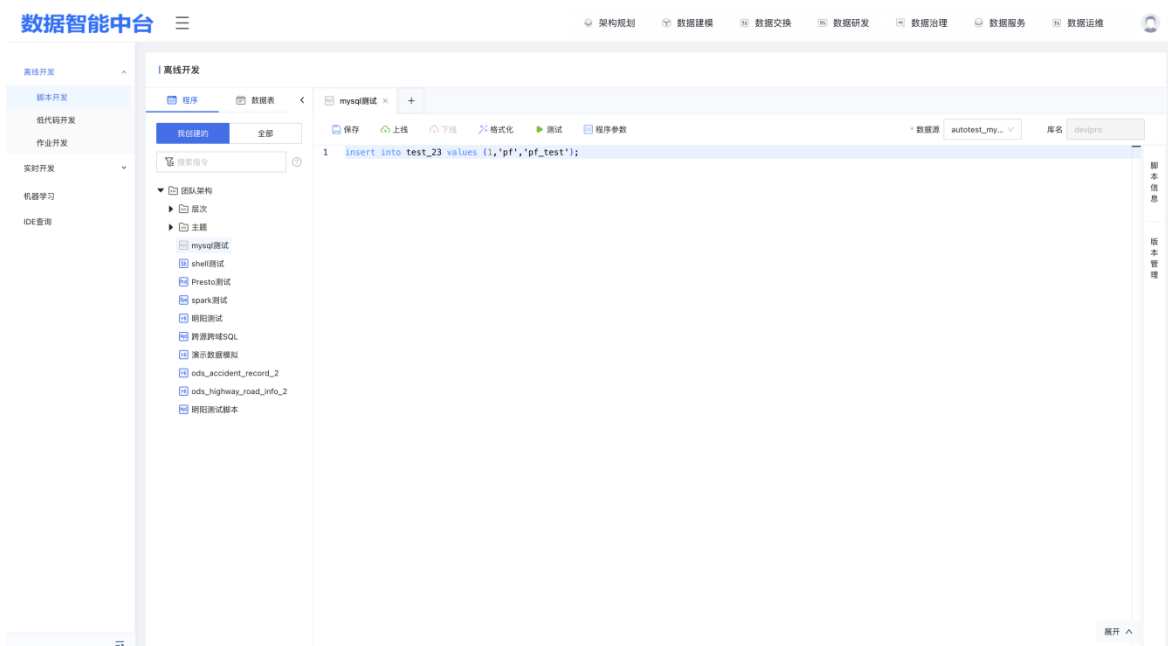


6. 只有生效的脚本才可以被后面的作业开发所引用，生效后可以看到历史管理，版本对比，发布日志

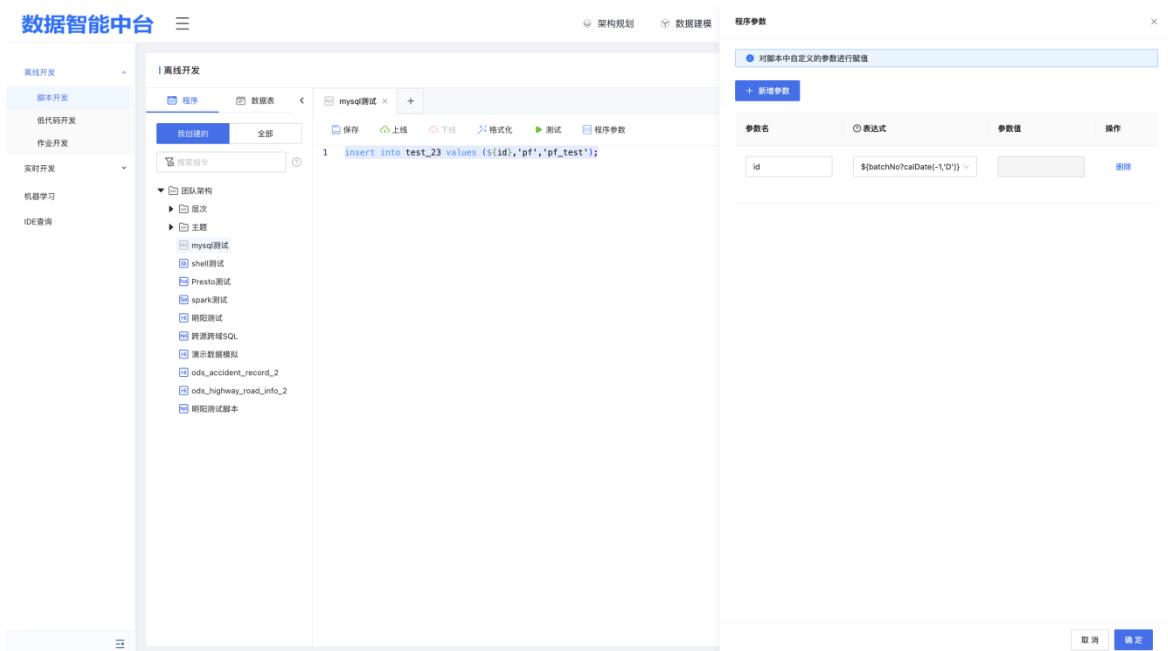


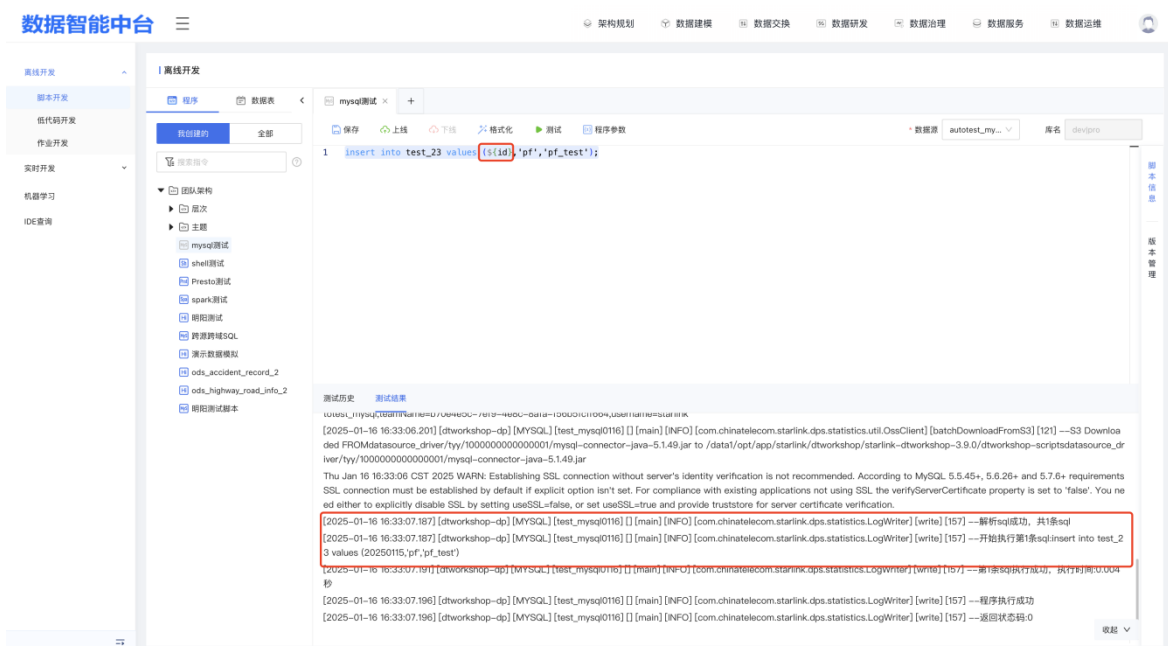


7. 如果想多该脚本进行变更，必须使其下线，生效状态不能变更，下线同上线流程一样，审批过后才可以变更。



8. 对脚本程序 sql 进行变更，添加程序参数(内置)，可以对其变量替换，执行测试





### 内置参数列表说明:

$\${batchNo?calDate(-1,'M','yyyy-MM-dd')}$ : 日批次对应上月同期日 (yyyy-mm-dd)

$\${batchNo?calDate(-6)}$ : 批次前 6 天 (yyyymmdd)

$\${batchNo?calDate(-1,'Y','yyyyMM')}$ : 去年同月 (yyyymm)

$\${batchNo?calDate(+2)}$ : 批次对应后 2 天 (yyyymmdd)

$\${batchNo?calDate(+3)}$ : 批次对应后 3 天 (yyyymmdd)

$\${batchNo?calDate(-1,'M','yyyyMM')}01$ : 上月 1 号 (yyyymm01)

$\${batchNo?calDate(-1,'M')}$ : 批次对应上月同期日 (yyyymmdd)

$\${batchNo?calDate(+1,'M','yyyyMM')}01$ : 下月一号 (yyyymm01)

$\${batchNo?calDate(+3,'M')}$ : 3 个月后批次 (yyyymmdd)

$\${batchNo?calDate(-2,'D','yyyyMMdd')}$ : 日批次前两天 (yyyymmdd)

$\${batchNo?calDate(-3,'D','yyyyMMdd')}$ : 日批次前三天 (yyyymmdd)

$\${batchNo?calDate(-6,'M')}$ : 上 6 月批次日期 (yyyymmdd)

$\${batchNo?calDate(-1,'D')}$ : 批次对应前一天 (yyyymmdd)

$\${batchNo?calDate(-1,'M','yyyyMMdd')?calDate(0,'L')}$ : 上月最后一天 (yyyymmdd)

$\${batchNo?calDate(-1,'M')?substring(4,6)}$ : 上月 (mm)

$\${batchNo?calDate(-7,'D')}$ : 批次对应上周同期日 (yyyymmdd)

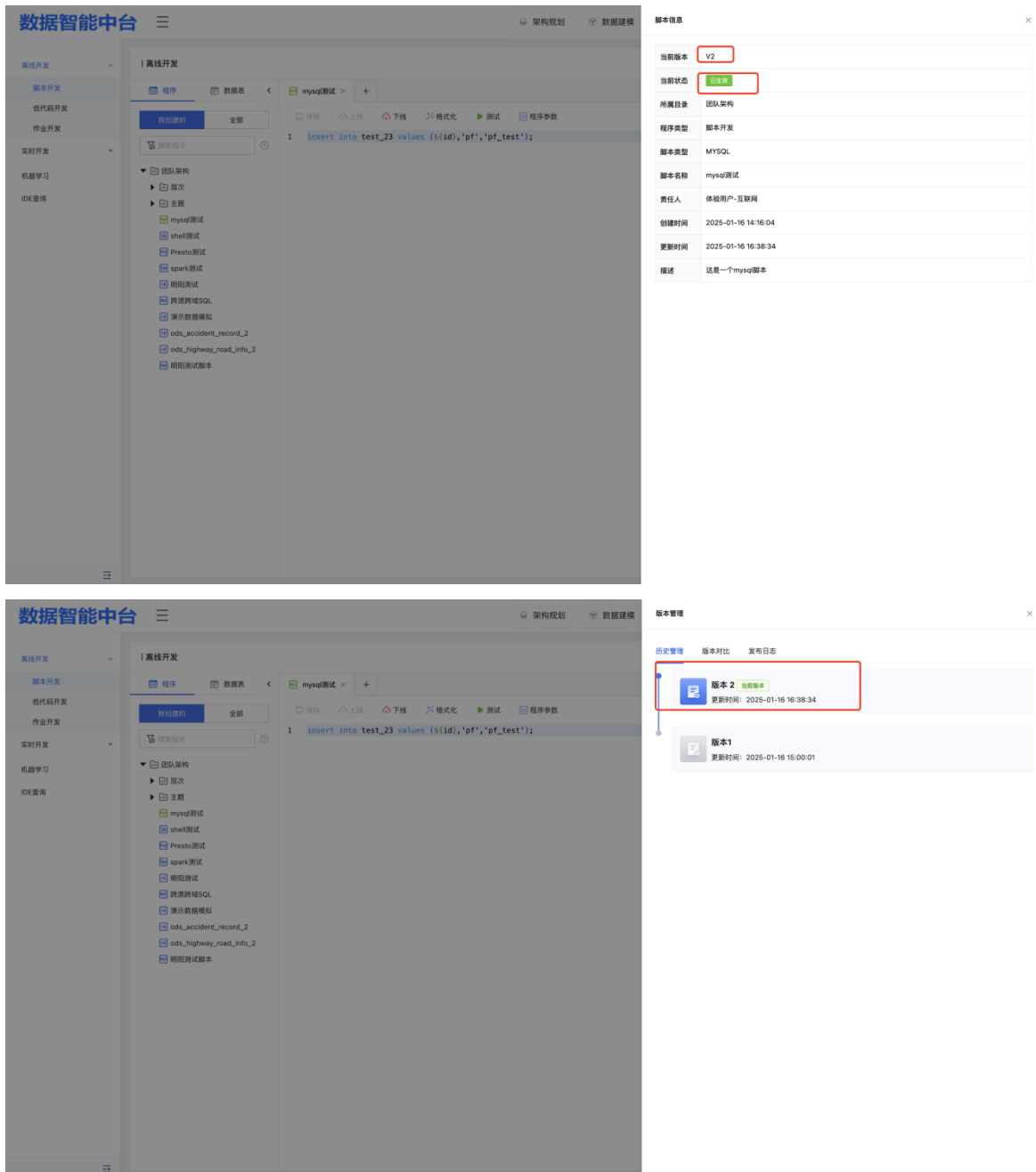
$\${batchNo?substring(0,6)}$ : 月

$\${batchNo?calDate(0,'L')?substring(6,8)}$ : 当月天数

$\${batchNo?calDate(0,'H','yyyyMMddHH')}$ : 小时

$\${batchNo?calDate(0,'MI','yyyyMMddHHmm')}$ : 分钟

## 9. 可以再次上线该脚本程序，上线后当前版本会自增



## 后续

程序代码开发完成后，您可根据需要执行如下操作：

- **调试代码：**根据业务需求调试运行单个程序，或调试运行整个业务流程，并在调试完成后查看运行结果。
- **配置调度：**进行节点的调度配置，后续节点将按照该配置周期性调度执行。
- **提交发布任务：**任务开发完成后，需提交至对应环境调度执行。

- 运维任务：任务发布完成后将默认展示在生产环境运维中心，您可前往生产环境运维中心查看生产环境任务的运行情况，并对生产任务执行相关运维操作。

## 作业编排

星海-数据中台将不同类型引擎任务封装为不同脚本程序，通过创建脚本来生成数据开发任作业。

同时，数据开发支持使用资源和函数开发任务。

- 提供图形化设计器，支持拖拉拽方式快速构建数据处理作业。
- 预设数据集成、计算&分析、资源管理、数据监控、其他等多种任务类型，通过任务间依赖完成复杂数据分析处理。
- 支持多种作业调度方式。
- 支持导入和导出作业。
- 支持作业状态运维监控和作业结果通知。
- 提供编辑锁定能力，支持多人协同开发场景。
- 支持作业的版本管理能力，支持生成保存版本和提交版本。

## 前提条件

需要完成脚本开发和低代码开发，才能进行 workflow 编排操作。

## 使用限制

- 已拥有开发角色权限。
- 已拥有所需任务权限。

## 使用流程

- 新建作业：当前支持离线开发程序的作业编排，支持脚本开发、低代码开发和机器学习三种模式。
- 作业编排：基于新建的作业，您可以进行新建、编排和配置节点程序，进行作业开发。
- 调度配置：配置作业调度信息。针对批处理作业，您可以配置任务级别的调度任务，即以工作流为一个整体进行调度，支持单次调度、周期调度、事件驱动调度三种调度方式。
- 上线：作业调度配置完成后，您需要提交任务并通过审批上线后才能用于调度运行。
- 管理作业：作业开发完成后，您可以根据需要，进行作业管理。

## 配置过程

[数据中台 workflow 编排操作手册](#)。

## 程序调试

程序代码开发完成后，您可根据需要，通过运行、带参运行、快捷运行等功能调试完整代码或代码片段的正确性，并在调试完成后查看运行结果。本文为您介绍程序调试的相关内容。

### 前提条件

需要完成程序开发，才能进行程序调试。

### 使用限制

- 已拥有开发角色权限。
- 已完成任务开发。

### 使用说明

在左侧导航栏的脚本开发或低代码开发模块，双击程序目录下的程序名称，进入该节点的编辑页面。

调试运行程序：在低代码程序编辑页面，脚本编辑页面，工具栏中点击测试运行，调试代码逻辑。

## 数据查询

若您仅需在开发环境，查询数据及相关 SQL 代码，测试代码的实际情况与期望值是否相符或验证代码的正确性；而无需将数据或 SQL 代码发布至生产环境操作生产环境引擎，则可通过数据查询实现。

### 前提条件

需要完成数据源的接入，同时当前用户具有数据源的查询权限。

### 使用限制

- 已拥有开发角色权限。
- 已拥有数据源的查询权限。

### 使用流程

- 进入[数据开发](#)页面。

- 在左侧导航栏，单击 IDE **查询**。
- 支持 Hive、Doris、MySQL、DM、GaussDB、PostgreSQL、Clickhouse、Oracle、SQLserver、Kingbase 等数据源的数据查询。

## 实时开发

实时开发是数据开发的一种模式，它处理的是连续不断产生的数据流（。开发人员通过编写实时处理程序（如 Flink 任务），对这些流动的数据进行即时计算、分析，并在极短的时间内输出结果。

星海·数据中台为您提供了便捷的实时开发页面辅助您进行实时任务开发，本文将会为您介绍开发页面及其元素构成，方便您快速上手产品。

### 前提条件

需要完成数据源的接入。

### 使用限制

- 引擎版本：支持多引擎版本兼容，目前已支持 Flink 1.17.2。
- 角色权限：已拥有**开发**角色权限，已拥有数据源的读写权限。

## 可视化开发

### 使用流程

#### 1. 新建实时任务

登录星海·数据中台。

快捷导航栏 选择数据开发-实时开发。

切换至目标工作空间。

单击 新建开发任务，在弹出窗口中填写任务配置信息。

单击 确定。

### 新建可视化开发

\* 所属目录:

\* 可视化英文名:

\* 可视化中文名:

描述:

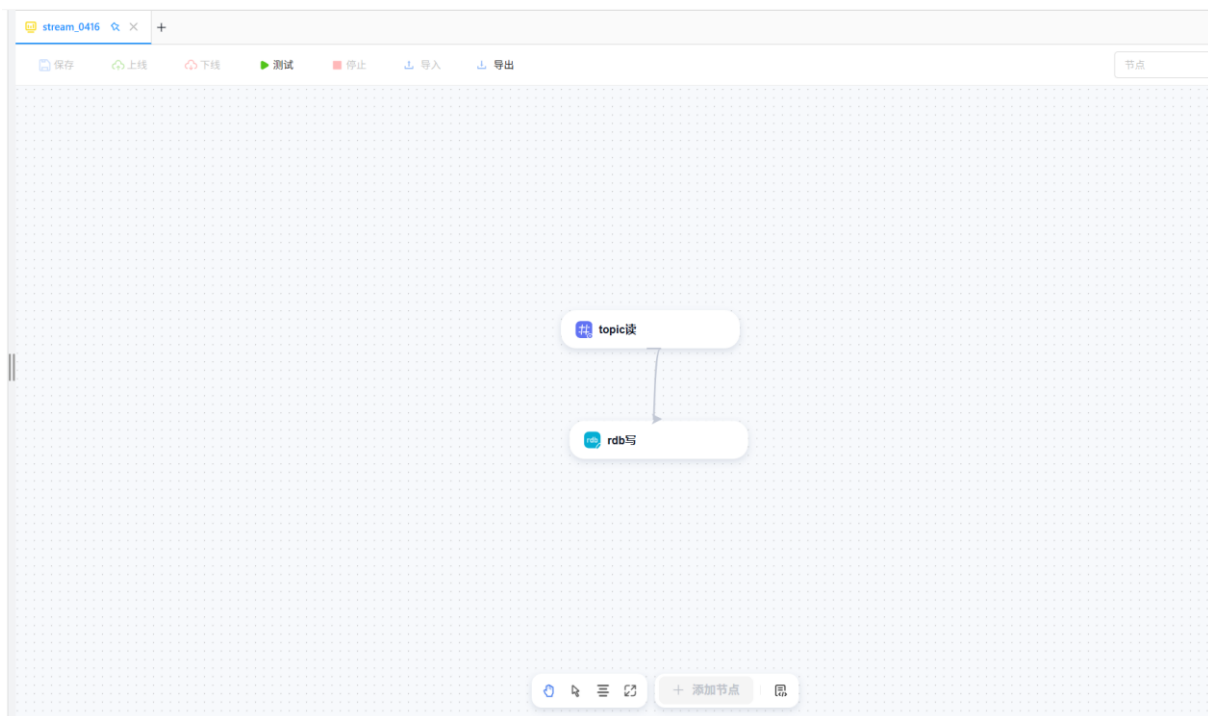
0 / 300

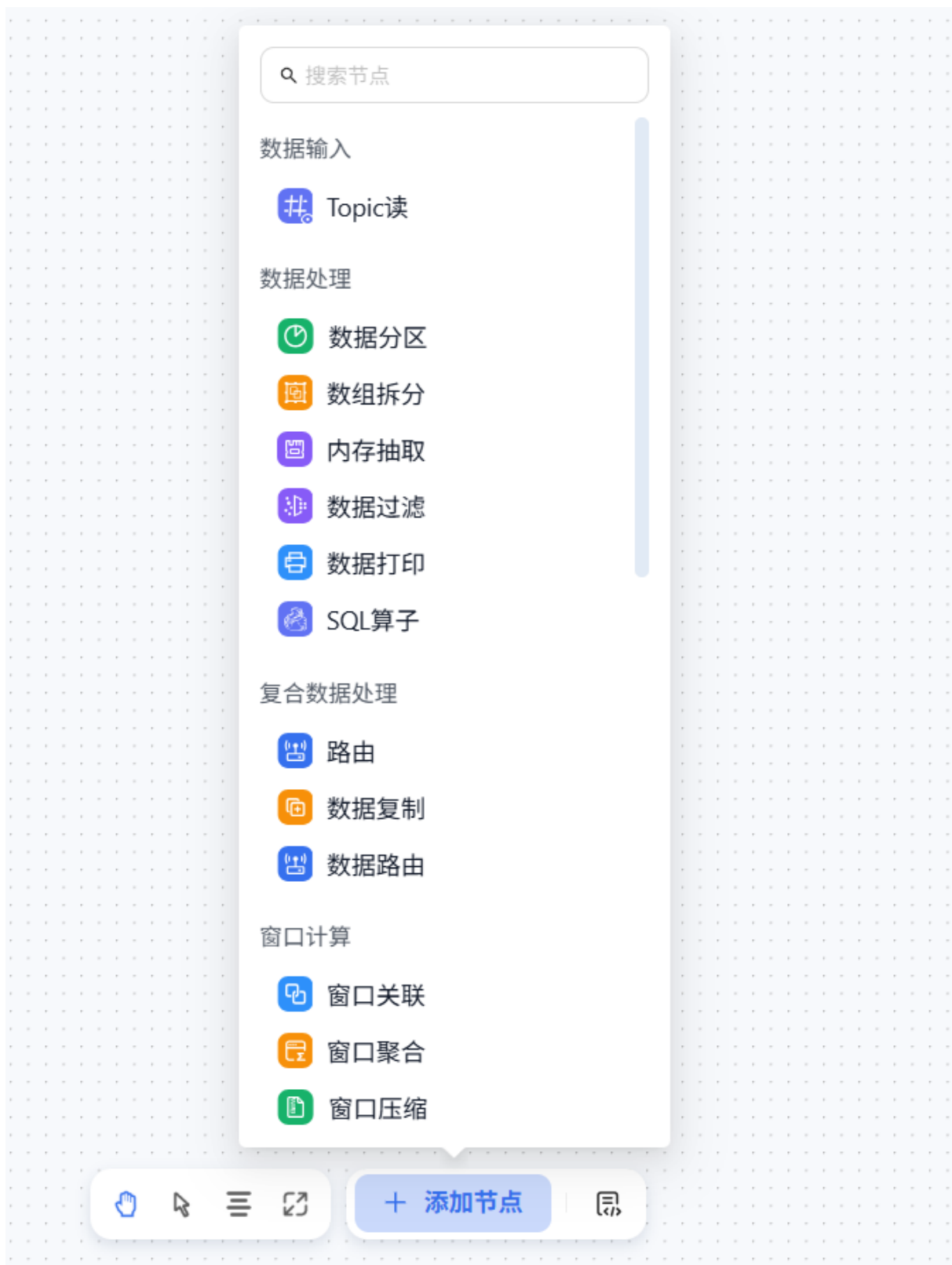
## 2. 配置实时任务

选择数据源节点;

选择数据处理节点, 配置数据处理节点以内;

选择写入节点。





### 3. 任务提交前验证和上线

完成低代码任务配置；

单击 运行配置，在右侧抽屉窗口中根据所需配置任务与资源配置。

单击 提交上线，在弹出窗口中确认相关信息并填写描述。

单击 提交，即可将实时任务提交上线，将任务发布到生产环境。

# 调度中心

---

## 调度中心概述

在星海·数据中台中，数据交换、数据开发、数据治理等相关模块的任务配置调度保存上线后，您可以在运维中心进行任务运维。运维中心为您提供任务总览、实例总览、任务运维、实例运维、基线运维、告警管理、配置管理 7 大功能模块，您可针对提交的任务和生成的实例进行多方位的运维管理；还支持实时查看任务运行状态，提供智能诊断、重跑等运维操作，帮助您对异常任务进行基础运维；提供智能基线，帮助您解决重要任务产出时间不可控，海量任务监控难的问题，保障任务产出的时效性；提供引擎、资源、调度等多方位的运维能力。

## 运维概览

### 任务概览

统计系统内各个业务模块产生的任务总数，也提供任务状态详情查看以及每个业务模块上线任务和未上线任务的数量查看；最后提供任务告警排行信息。

### 任务总览入口

在中台首页，单击顶部横向菜单栏的数据运维。默认进入数据运维页面。

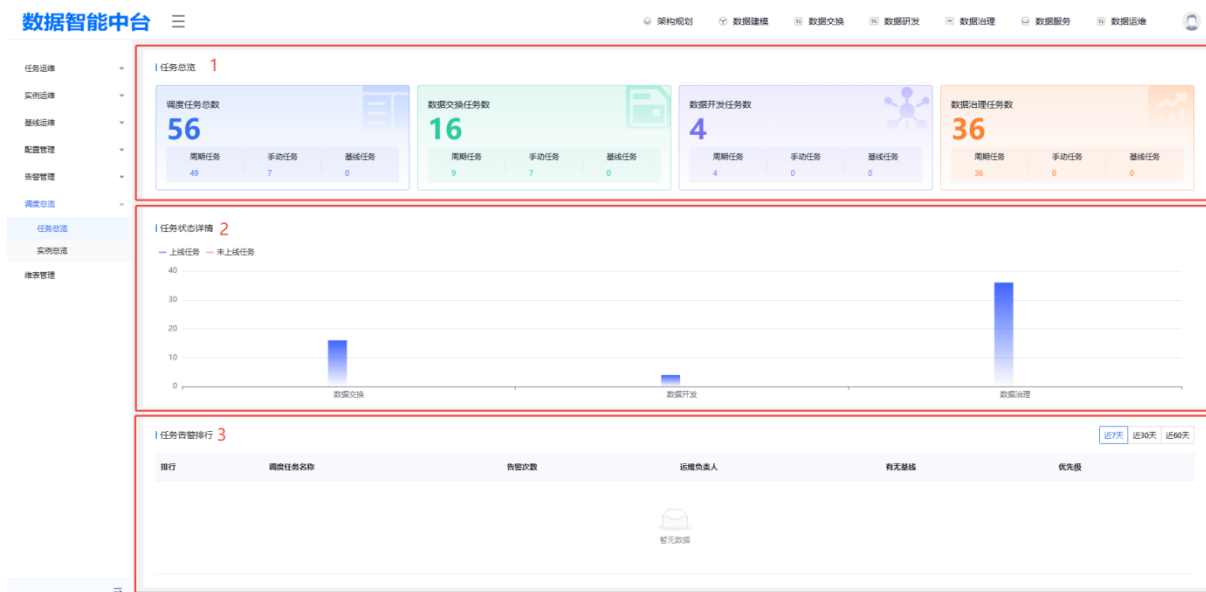
按照以下操作指引，进入任务总览页面。

依次单击侧边菜单的调度总览->任务总览。



## 任务总览页面介绍

任务总览页面主要由任务总数统计、任务状态详情、任务告警排行三部分组成。



编号区域	描述
①任务数量统计	展示系统存在的调度任务总量，包含的数据集成、数据开发、数据治理各模块的任务总量查看；再往下钻到任务类型维度，分别查看其中周期任务、手动任务、基线任务数量是多少。
②任务状态详情	展示数据集成、数据开发、数据治理任务上下线状态情况；各个业务模块上线和下线任务数量通过柱状图展示，鼠标悬停至柱状图上展示详细数值。 <div style="border: 1px solid gray; padding: 5px; margin: 10px auto; width: fit-content;"> <p><b>数据开发</b></p> <ul style="list-style-type: none"> <li><span style="color: blue;">●</span> 上线任务 4</li> <li><span style="color: orange;">●</span> 未上线任务 0</li> </ul> </div>

### ③任务告警排行

展示近 7 天、近 30 天、近 60 天的任务告警次数前 10 的排行列表。

## 实例概览

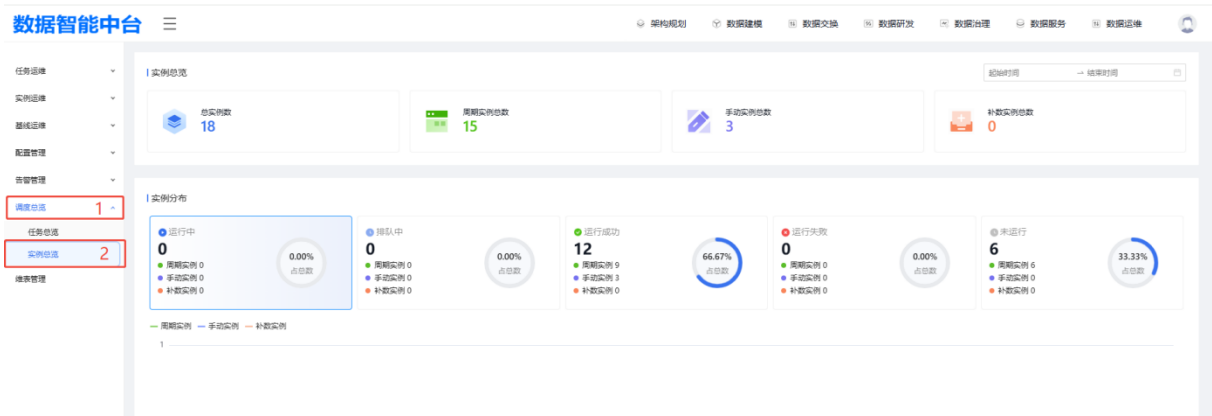
实例总览功能默认提供当天的总实例数查看，也可以进行时间筛选查看；总实例下的周期实例、手动实例、补数实例总数查看。也提供实例分布查看，可查看数据集成、数据开发、数据治理模块运行中、排队中、运行成功、运行失败、未运行这 5 种以状态维度的周期实例、手动实例、补数实例的数量。最后提供实例运行趋势和运行失败实例数排行的展示，且这两个展示内容都能通过近 7 天、近 30 天、近 60 天的时间维度做切换。

## 实例总览入口

在中台首页，单击顶部横向菜单栏的数据运维。默认进入数据运维页面。

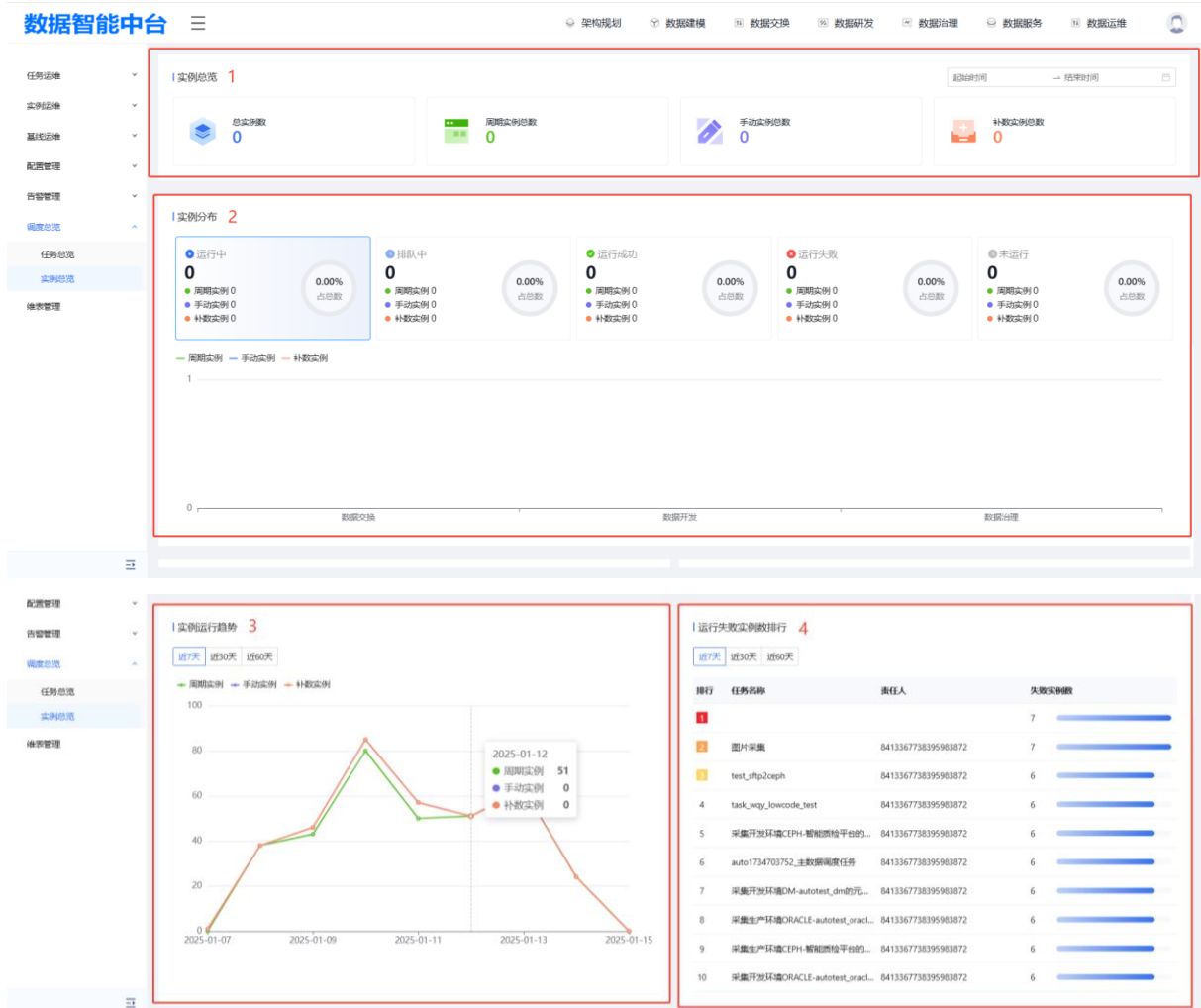
按照以下操作指引，进入实例总览页面。

依次单击侧边菜单的调度总览->实例总览。



## 实例总览页面介绍

实例统计页面主要由实例总览、实例分布、趋势与排行组成。您可以在实例统计页面查看实例运行情况。



编号区域	描述
①实例总览	展示系统存在的实例总量，包含的周期实例、手动实例、补数实例等各类型的实例总量查看。
②实例分布	展示的实例分布情况，包括各状态实例数量的占比情况。同时，分别展示各状态周期实例、手动实例和补数据实例的实例总数量。并且各个状态的实例数量通过柱状图展示，鼠标点击至柱状图上展示状态实例来联动下方柱状图，进行详细的周期实例数量、补数据实例数量和手动实例数查看。
③实例运行趋势	展示最近 7 天、30 天、60 天的周期实例、手动实例和补数据实例，各状态的运行数据及趋势。鼠标悬停至某一运行日期上，支持查看当日的周期实例、手动实例和补数实例运行情况。
④运行失败实例排行	展示最近 7 天、30 天、60 天的任务所产生的运行失败实例数排行。

# 任务运维

## 周期任务

### 周期任务通用运维概述

周期任务是指调度系统按照调度配置自动定时运行的任务。本文为您介绍周期的通用功能，包括使用说明，支持的调度类型和相关功能说明。

### 周期任务页面入口

在数据中台首页，单击顶部菜单栏的数据运维。默认进入数据运维页面。

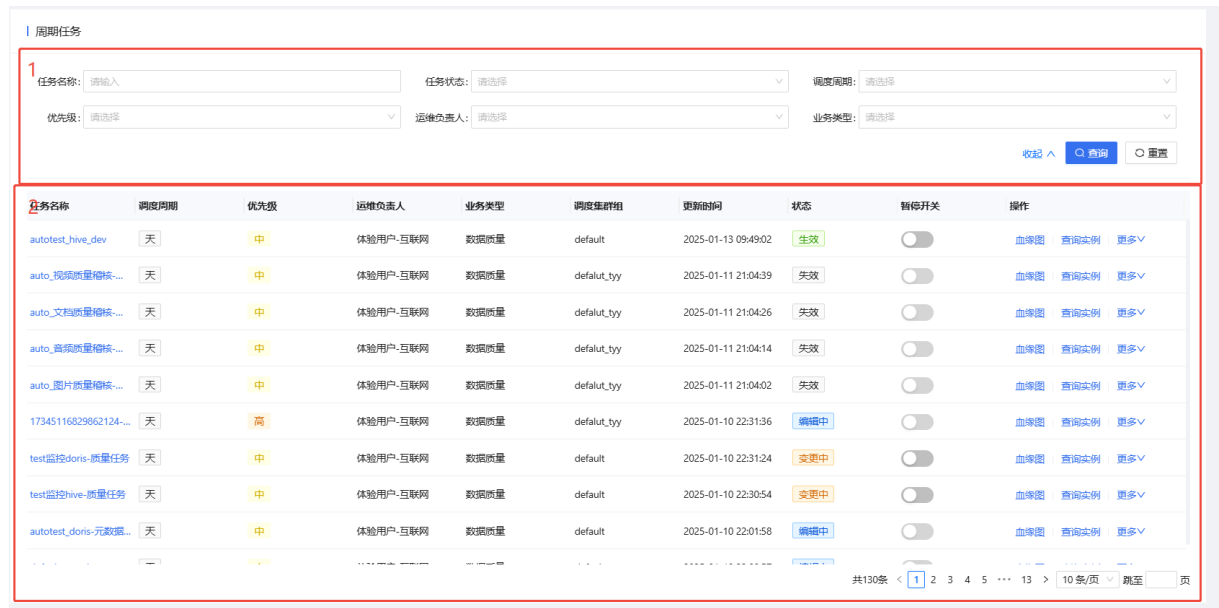
按照以下操作指引，进入周期任务运维页面。

依次单击数据运维->任务运维->周期任务。



### 周期任务页面介绍

周期任务页面主要由筛选区、周期任务列表组成。您可以在周期任务页面对提交的周期任务进行相关的运维操作。



编号区域	描述
------	----

<p>① 筛选区</p>	<p>快捷筛选支持通过输入任务名称来筛选周期任务，适用于知晓任务名称情况下来进行快捷筛选出所需的周期任务。</p> <p>同时支持任务状态、调度周期、优先级、运维负责人、业务类型这几种维度下选择来进行列表内容的快捷筛选。</p> <ul style="list-style-type: none"> <li>● 任务状态：支持编辑中、变更中、生效、待上线、失效、待下线、暂停。</li> <li>● 调度周期：支持分钟、小时、天、月共 4 种。</li> <li>● 优先级：支持最高、高、中等、低、最低实例的运行优先级进行筛选。</li> <li>● 运维负责人：可根据周期任务的运维负责人进行筛选。</li> <li>● 业务类型：支持数据交换、数据治理、数据开发这三种业务模块的筛选。</li> </ul>
<p>② 周期任务列表</p>	<p>周期任务页面以列表形式为您展示提交到运维中心的周期任务。列表操作栏为您展示周期任务支持的运维管理操作。</p>

## 周期任务血缘图

单击周期任务操作栏的血缘图按钮，会弹出新的画布页面为您展示以当前任务为中心节点的血缘图。血缘图支持可视化动态呈现该节点的上下游依赖关系，默认展示当前任务（选中节点）及上下游第一层节点。如果周期任务没有上下游节点，则血缘图仅展示当前任务一个节点。同时，您也可以调整周期任务血缘图页面范围；以下以某个周期任务为例：



编号区域	描述
------	----

①任务信息	展示当前选中任务的概要信息。如：任务名称、调度周期、优先级、运维负责人等内容展示，你还可以点击查看实例，跳转到这个任务的实例列表页。
②画布调整区	<p>可以快速调整画布的内容展示层级、大小、全屏等</p> <ul style="list-style-type: none"> <li>● 展示层级：默认为 1 层，展示的是以当前节点的上下游第一层节点。</li> <li>● 缩小：缩小画布中的元素显示比例。</li> <li>● 放大：放大画布中的元素显示比例。</li> <li>● 居中：画布中的元素基于画布进行居中展示。</li> <li>● 全屏：全屏显示画布内容。</li> </ul>
③任务依赖关系图	展示实例的调度依赖关系图，支持向上向下展开更多节点；如果周期任务没有任务依赖关系，则图展示只有一个此任务的节点。

## 手动任务

### 手动任务通用运维概述

手动任务是已提交至运维中心，且调度类型为手动节点的任务。手动任务不会自动运行，只有手动触发运行才会生成实例。本文为您介绍手动任务运维功能。

### 手动任务页面入口

在数据中台首页，单击顶部菜单栏的数据运维。默认进入数据运维页面

按照以下操作指引，进入手动任务运维页面。

依次单击数据运维->任务运维->手动任务。



### 手动任务页面介绍

手动任务页面主要由筛选区、手动任务列表组成。您可以在手动任务页面对提交的手动任务进行相关的运维操作。

任务名称	优先级	运维负责人	业务类型	状态	调度集群组	更新时间	操作
Customer_MysqlToHive	中	丁书琳	数据交换	失效	default	2025-01-09 21:27:25	运行 <a href="#">查询实例</a> <a href="#">更多</a>
demo_13	中	周书剑	数据交换	失效	default	2024-12-27 14:38:46	运行 <a href="#">查询实例</a> <a href="#">更多</a>
demo	中	管理员	数据交换	编辑中	defalut_tyy	2024-10-23 09:27:23	运行 <a href="#">查询实例</a> <a href="#">更多</a>
LDH_CALL_RECORDS_MYSQL...	中	管理员	数据交换	生效	defalut_tyy	2024-10-16 16:33:18	运行 <a href="#">查询实例</a> <a href="#">更多</a>
LDH_CALL_RECORDS_MYSQL...	中	管理员	数据交换	生效	defalut_tyy	2024-10-16 14:50:47	运行 <a href="#">查询实例</a> <a href="#">更多</a>
LDH_BILLING_MYSQL_2_HIVE	中	管理员	数据交换	生效	defalut_tyy	2024-10-15 13:55:51	运行 <a href="#">查询实例</a> <a href="#">更多</a>
LDH_CALL_RECORDS_MYSQL...	中	管理员	数据交换	生效	defalut_tyy	2024-10-15 13:55:44	运行 <a href="#">查询实例</a> <a href="#">更多</a>
LDH_DEMO_MYSQL_2_HIVE	中	管理员	数据交换	变更中	defalut_tyy	2024-10-15 11:09:34	运行 <a href="#">查询实例</a> <a href="#">更多</a>
ytest_mysql	中	管理员	数据交换	编辑中	defalut_tyy	2024-10-09 17:36:42	运行 <a href="#">查询实例</a> <a href="#">更多</a>
lx_mysql_to_hive	中	体验用户-互联网	数据交换	生效	defalut_tyy	2024-09-30 10:07:13	运行 <a href="#">查询实例</a> <a href="#">更多</a>

编号区域	描述
① 筛选区	<p>快捷筛选支持通过输入任务名称来筛选周期任务，适用于知晓任务名称情况下来进行快捷筛选出所需的手动任务。</p> <p>同时支持任务状态、优先级、运维负责人、业务类型、更新时间这几种维度下选择来进行列表内容的快捷筛选。</p> <ul style="list-style-type: none"> <li>● 任务状态：支持编辑中、变更中、生效、待上线、失效、待下线、暂停。</li> <li>● 优先级：支持最高、高、中等、低、最低实例的运行优先级进行筛选。</li> <li>● 运维负责人：可根据手动任务的运维负责人进行筛选。</li> <li>● 业务类型：支持数据交换、数据治理、数据开发这三种业务模块的筛选。</li> <li>● 更新时间：支持快捷选中今天或设置任务更新的开始时间和结束时间范围进行筛选。</li> </ul>
② 手动任务列表	<p>手动任务页面以列表形式为您展示提交到运维中心的手动任务。列表操作栏为您展示手动任务支持的运维管理操作。</p>

## 实时任务

实时任务是指在中台中创建的 FLINK\_SQL、FLINK\_TEMPLATE\_SQL 等类型的计算任务。在中台的实时任务页面为您展示了任务实例的基本信息及运维管理选项，方便您查看实例任务并进行管理。本文为您介绍实时任务支持的运维管理功能。

## 实时任务页面入口

在数据中台首页，单击顶部菜单栏的数据运维。默认进入数据运维页面

按照以下操作指引，进入手动任务运维页面。

依次单击数据运维->任务运维->手动任务。



## 实时任务页面介绍

实时任务页面主要由筛选区、实时任务列表组成。您可以在实时任务页面对提交的实时任务进行相关的运维操作。



编号区域	描述
① 筛选区	<p>快捷筛选支持通过输入任务名称来筛选实时任务，适用于知晓任务名称情况下来进行快捷筛选出所需的实时任务。</p> <p>同时支持调度资源组、任务状态、运维负责人、业务类型这几种维度下选择来进行列表内容的快捷筛选。</p> <ul style="list-style-type: none"><li>● 任务状态：支持编辑中、变更中、生效、待上线、失效、待下线、暂停。</li><li>● 调度集群组：可根据支持任务运行的资源队列进行筛选。</li><li>● 运维负责人：可根据实时任务的运维负责人进行筛选。</li><li>● 业务类型：支持数据交换、数据治理、数据开发这三种业务模块的筛选。</li></ul>
② 实时任务列表	<p>实时任务页面以列表形式为您展示系统已生成的实时任务。实时任务列表信息，列表操作栏为您展示实时任务支持的运维管理操作。</p>

# 实例运维

## 周期实例

周期实例是周期任务按照调度配置自动定时运行生成的实例。周期任务每调度一次，便执行一个实例。同时，可以对已生成的实例进行运维管理，例如查看运行状态、重跑和查看节点代码等操作。本文将介绍周期实例的通用功能。

## 周期实例生成时间及运行规则

中台调度模块每晚 23 点生成次日需要运行的周期实例，生成的实例默认是未运行状态。

周期实例依赖的上游实例全部运行成功后，周期实例的状态从未运行变成等待调度时间。

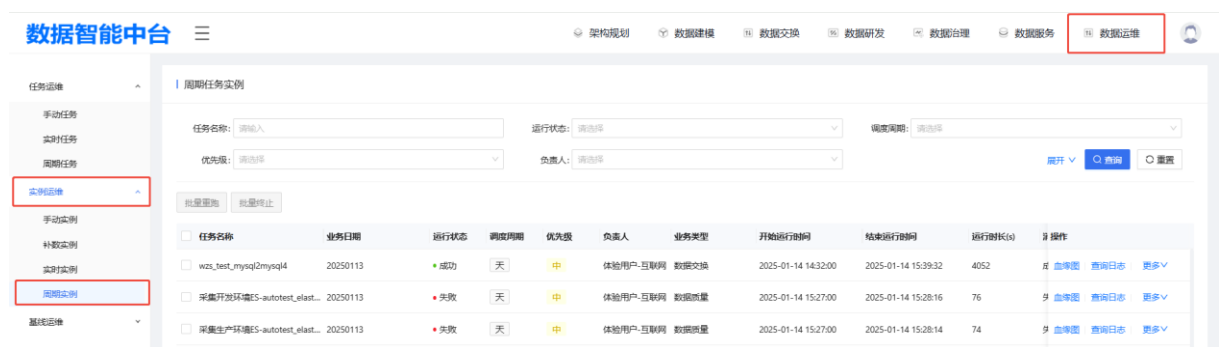
当到达当前实例的定时运行时间，且剩余可用调度资源充足、并通过所有命中的限流规则校验后，周期实例的状态会从等待调度时间变成运行中。

## 周期实例入口

在数据中台首页，单击顶部菜单栏的数据运维。默认进入数据运维页面

按照以下操作指引，进入周期实例运维页面。

依次单击数据运维->实例运维->周期实例。



## 周期实例页面介绍

周期实例页面主要由筛选区、周期实例列表、批量操作区组成。可以在周期实例页面对提交的周期实例进行相关的运维操作。

周期任务实例

1 任务名称: 请输入 运行状态: 请选择 调度周期: 请选择

优先级: 请选择 负责人: 请选择 展开 查询 重置

批量重跑 批量终止 3

任务名称	业务日期	运行状态	调度周期	优先级	负责人	业务类型	开始运行时间	结束运行时间	运行时长(s)	操作
wzs_test_mysql2mysql4	20250113	成功	天	中	体验用户-互联网	数据交换	2025-01-14 14:32:00	2025-01-14 15:39:32	4052	详情 血缘图 查询日志 更多
采集开发环境ES-autotest_elast...	20250113	失败	天	中	体验用户-互联网	数据质量	2025-01-14 15:27:00	2025-01-14 15:28:16	76	详情 血缘图 查询日志 更多
采集生产环境ES-autotest_elast...	20250113	失败	天	中	体验用户-互联网	数据质量	2025-01-14 15:27:00	2025-01-14 15:28:14	74	详情 血缘图 查询日志 更多
采集开发环境DM-autotest_dm...	20250113	失败	天	中	体验用户-互联网	数据质量	2025-01-14 15:26:00	2025-01-14 15:26:12	12	详情 血缘图 查询日志 更多
采集生产环境DM-autotest_dm...	20250113	失败	天	中	体验用户-互联网	数据质量	2025-01-14 15:23:00	2025-01-14 15:24:12	72	详情 血缘图 查询日志 更多
dd-稽核任务	20250114	成功	天	高	体验用户-互联网	数据质量	2025-01-14 15:22:00	2025-01-14 15:23:36	96	详情 血缘图 查询日志 更多
账单数据质量稽核-质量任务	20250113	成功	天	中	体验用户-互联网	数据质量	2025-01-14 15:21:00	2025-01-14 15:22:32	92	详情 血缘图 查询日志 更多
采集生产环境ORACLE-autotest...	20250113	失败	天	中	体验用户-互联网	数据质量	2025-01-14 15:19:00	2025-01-14 15:20:10	70	详情 血缘图 查询日志 更多
采集开发环境ORACLE-autotest...	20250113	失败	天	中	体验用户-互联网	数据质量	2025-01-14 15:18:00	2025-01-14 15:19:11	71	详情 血缘图 查询日志 更多

共16569条 < 1 2 3 4 5 ... 1657 > 10条/页 跳至 页

编号区域	描述
① 筛选区	<p>快捷筛选支持通过输入任务名称来筛选此周期任务产生的周期实例，适用于知晓任务名称情况下来进行快捷筛选出当前任务产生的所有实例信息。</p> <p>同时支持运行状态、调度周期、优先级、负责人、业务类型、业务日期、开始运行时间、结束运行时间这几种维度下选择来进行列表内容的快捷筛选。</p> <ul style="list-style-type: none"> <li>● 运行状态：支持新建、等待中、队列中、启动中、运行中、成功、失败、处理中、暂停。</li> <li>● 调度周期：支持分钟、小时、天、月共 4 种。</li> <li>● 优先级：支持最高、高、中等、低、最低实例的运行优先级进行筛选。</li> <li>● 负责人：可根据周期实例的运维负责人进行筛选。</li> <li>● 业务类型：支持数据交换、数据治理、数据开发这三种业务模块的筛选。</li> <li>● 业务日期：可根据周期实例的业务日期进行筛选。具体到某一天或某一业务时间区间（精确到秒）。</li> <li>● 开始运行时间：可根据周期实例的开始运行时间进行筛选。具体到某一天或某一业务时间区间（精确到秒）。</li> <li>● 结束运行时间：可根据实例结束运行时间进行筛选，具体到某一天或某一业务时间区间（精确到秒）。</li> </ul> <p>如果需要重新筛选，则可单击重置，即可快速清空所有筛选条件，恢</p>

	复系统默认。
②周期实例列表	周期实例页面以列表形式展示系统已生成的周期实例。列表操作栏展示周期实例支持的运维管理操作。
③批量操作区	可通过批量操作区域提供的重跑、终止功能，对周期实例进行批量处理，提高操作效率。

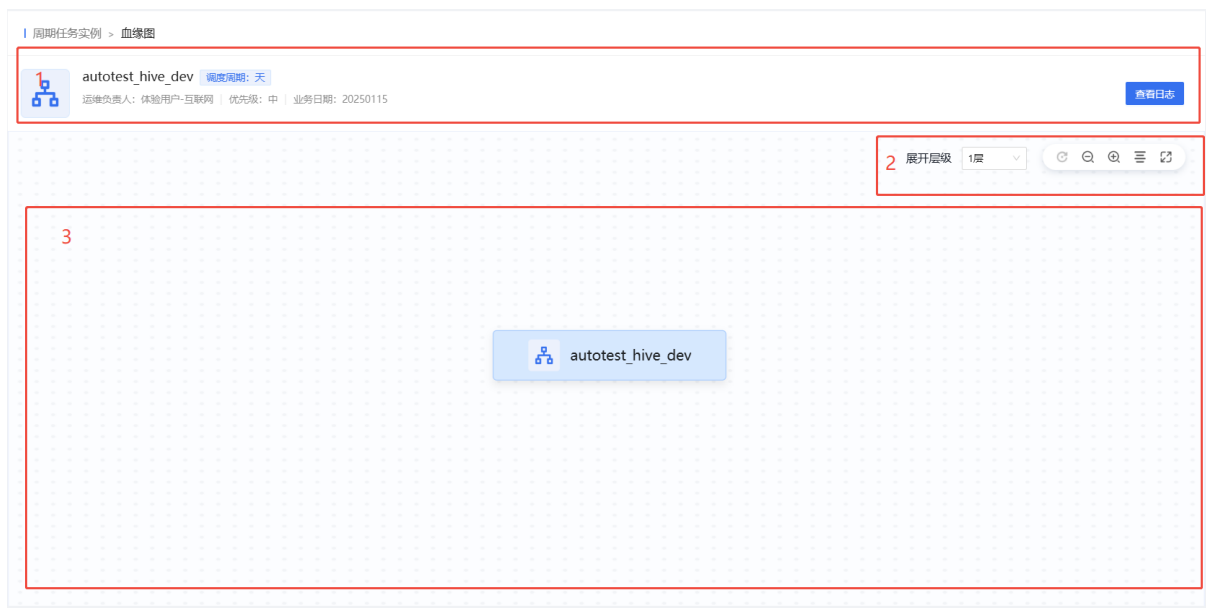
### 周期实例运行状态说明

- **新建**：实例初始化状态，代表实例刚刚生成，且还没有被调度机进行调度处理。
- **等待中**：任务等待执行，即执行时间还没到，或者任务的依赖未满足。
- **队列中**：任务被加入到执行队列中，即将提交到 worker 开始执行，在系统的任务负载不高的情况下，这个状态也是一个短暂存在的状态。
- **启动中**：当任务准备提交到 yarn/k8s 等计算资源上面时候，任务启动会持续一段时间，启动中即是代表任务提交到计算资源，但任务还没启动成功这一段时间的状态。
- **运行中**：正在运行中的实例。
- **成功**：运行成功的实例。强制成功的实例系统会直接标记为运行成功。
- **失败**：运行失败的实例。如果终止等待中或运行中状态的实例，则实例状态变更为失败。
- **处理中**：中间状态，代表后台正在处理实例相关指令，如重做、强制成功等，该状态不会长时间存在
- **暂停**：只有任务暂停后，相关的任务实例会变成暂停状态，是一个较为独立的实例状态。

### 周期实例血缘图

单击周期实例操作列表的实例对象名称，页面右侧将展示以当前实例为中心节点的血缘图。血缘图支持可视化动态呈现该节点的上下游依赖关系，默认展示当前实例（选中节点）及上下游第一层节点。如果实例没有上下游节点，则血缘图仅展示当前实例节点。

同时，您也可以调整实例血缘图页面范围；以下以某个周期实例为例：



编号区域	描述
①实例信息	展示当前选中实例的概要信息。如：所属任务名称、调度周期、优先级、业务日期等内容展示，你还可以点击查看日志，跳转到这个实例的启动、运行日志页进行详细日志信息查看。
②画布调整区	可以快捷调整画布的内容展示层级、大小、全屏等 <ul style="list-style-type: none"> <li>● 展示层级：默认为 1 层，展示的是以当前节点的上下游第一层节点。</li> <li>● 缩小：缩小画布中的元素显示比例。</li> <li>● 放大：放大画布中的元素显示比例。</li> <li>● 居中：画布中的元素基于画布进行居中展示。</li> <li>● 全屏：全屏显示画布内容。</li> </ul>
③实例依赖关系图	展示实例的调度依赖关系图，支持向上向下展开更多节点；如果周期实例没有相互的依赖关系，则图展示只有一个此实例的节点。

## 手动实例

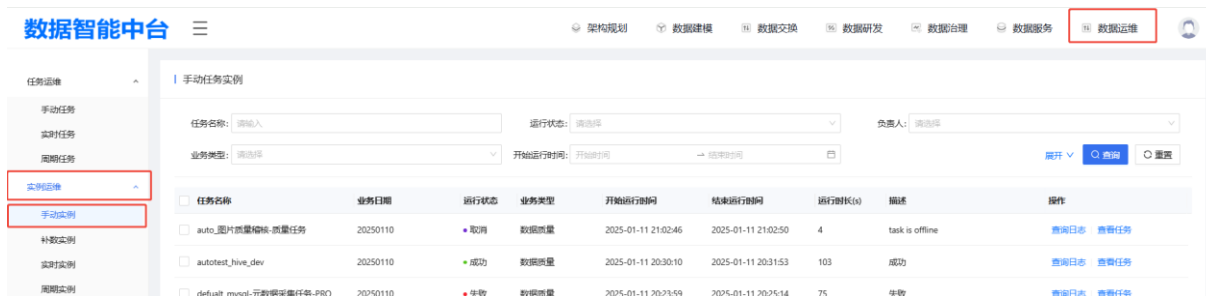
手动实例是手动任务运行后生成的实例数据。系统支持对手动实例进行运维管理，包括查寻日志、查看手动任务等。

### 手动实例入口

在数据中台首页，单击顶部菜单栏的数据运维。默认进入数据运维页面

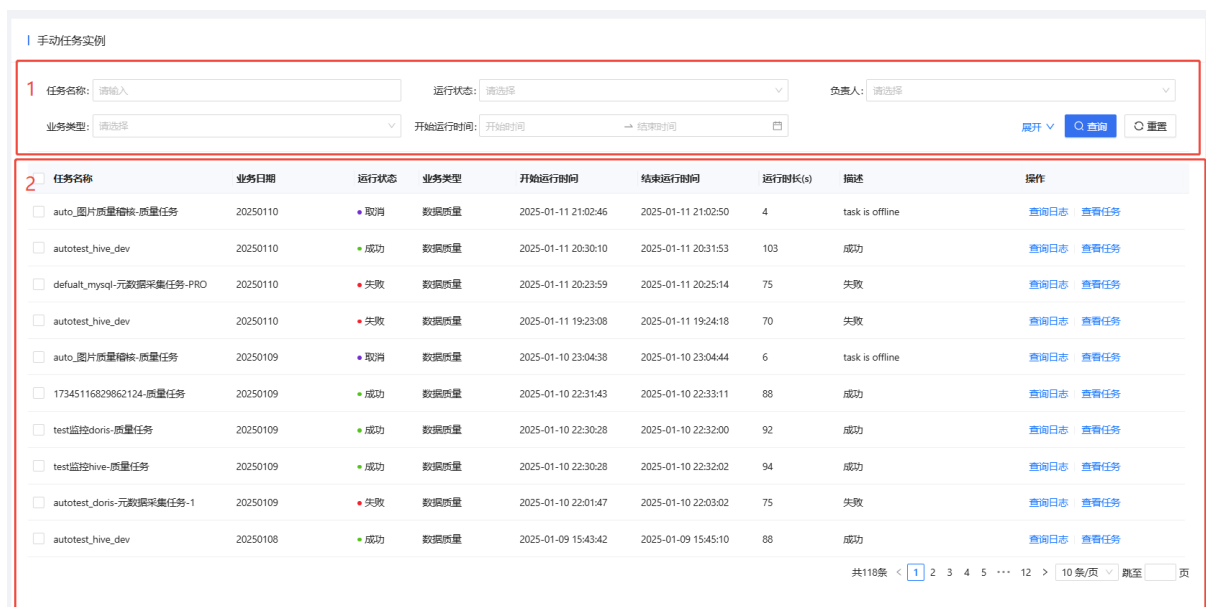
按照以下操作指引，进入手动实例运维页面。

依次单击数据运维->实例运维->手动实例。



## 手动实例页面介绍

手动实例页面以列表方式，展示系统已生成的手动实例。手动实例支持的运维管理操作包括查看手动任务、查看日志等。



编号区域	描述
①筛选区	<p>快捷筛选支持通过输入任务名称来筛选此手动任务产生的手动实例，适用于知晓任务名称情况下来进行快捷筛选出当前任务产生的所有手动实例信息。</p> <p>同时支持运行状态、优先级、负责人、业务类型、业务日期、开始运行时间、结束运行时间这几种维度下选择来进行列表内容的快捷筛选。</p> <ul style="list-style-type: none"> <li>● 运行状态：支持新建、等待中、队列中、启动中、运行中、成功、失败、处理中。</li> <li>● 负责人：可根据手动实例的运维负责人进行筛选。</li> <li>● 业务类型：支持数据交换、数据治理、数据开发这三种业务模块的筛选。</li> <li>● 业务日期：可根据手动实例的业务日期进行筛选。具体到某一天</li> </ul>

	<p>或某一业务时间区间（精确到秒）。</p> <ul style="list-style-type: none"> <li>● 开始运行时间：可根据手动实例的开始运行时间进行筛选。具体到某一天或某一业务时间区间（精确到秒）。</li> <li>● 结束运行时间：可根据手动实例结束运行时间进行筛选，具体到某一天或某一业务时间区间（精确到秒）。</li> </ul> <p>如果需要重新筛选，则可单击重置，即可快速清空所有筛选条件，恢复系统默认。</p>
②手动实例列表	<p>手动实例页面以列表形式为您展示系统已生成的手动实例。列表操作栏为您展示手动实例支持的运维管理操作。</p>

### 手动实例运行状态说明

- 新建：实例初始化状态，代表实例刚刚生成，且还没有被调度机进行调度处理。
- 等待中：任务等待执行，即执行时间还没到，或者任务的依赖未满足。
- 队列中：任务被加入到执行队列中，即将提交到 worker 开始执行，在系统的任务负载不高的情况下，这个状态也是一个短暂存在的状态。
- 启动中：当任务准备提交到 yarn/k8s 等计算资源上面时候，任务启动会持续一段时间，启动中即是代表任务提交到计算资源，但任务还没启动成功这一段时间的状态。
- 运行中：正在运行中的实例。
- 成功：运行成功的实例。强制成功的实例系统会直接标记为运行成功。
- 失败：运行失败的实例。如果终止等待中或运行中状态的实例，则实例状态变更为失败。
- 处理中：中间状态，代表后台正在处理实例相关指令，如重做、强制成功等，该状态不会长时间存在。

## 实时实例

实时实例是实时任务提交到运维中心后生成的实例数据，**一个实时任务只产生一个实时实例**。支持查看实时任务生成的实例数据，同时支持对实时实例进行运维管理，包括查看任务、查看日志等。本文为您介绍实时实例的状态及如何运维管理实时实例。

### 实时实例入口

在数据中台首页，单击顶部菜单栏的数据运维。默认进入数据运维页面

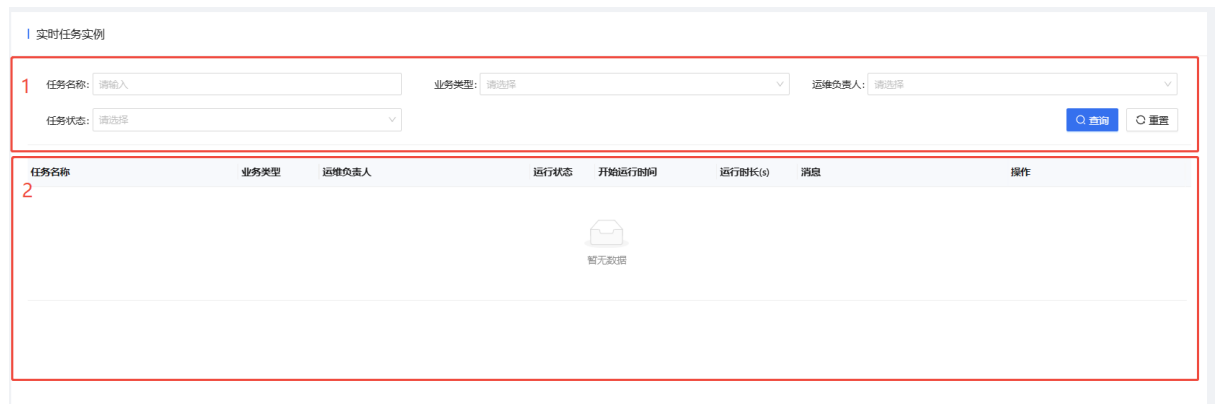
按照以下操作指引，进入实时实例运维页面。

依次单击数据运维->实例运维->实时实例。



### 实时实例页面介绍

实时实例页面以列表方式，为您展示系统已生成的实时实例。实时实例页面主要由搜索及筛选区、实时实例列表组成。您可以在实时实例页面进行相关的运维操作。



编号区域	描述
①筛选区	<p>快捷筛选支持通过输入任务名称来筛选此实时任务产生的实时实例，适用于知晓任务名称情况下来进行快捷筛选出当前任务产生的所有实时实例信息。</p> <p>同时支持运行状态、负责人、业务类型这几种维度下选择来进行列表内容的快捷筛选。</p> <ul style="list-style-type: none"> <li>● 运行状态：支持新建、启动中、运行中、停止中、停止、失败。</li> <li>● 负责人：可根据实时实例的运维负责人进行筛选。</li> <li>● 业务类型：支持数据交换、数据治理、数据开发这三种业务模块的筛选。</li> </ul> <p>如果需要重新筛选，则可单击重置，即可快速清空所有筛选条件，恢复系统默认。</p>
②实时实例列表	<p>实时实例页面以列表形式为您展示系统已生成的实时实例。列表操作栏为您展示实时实例支持的运维管理操作</p>

### 实时实例运行状态说明

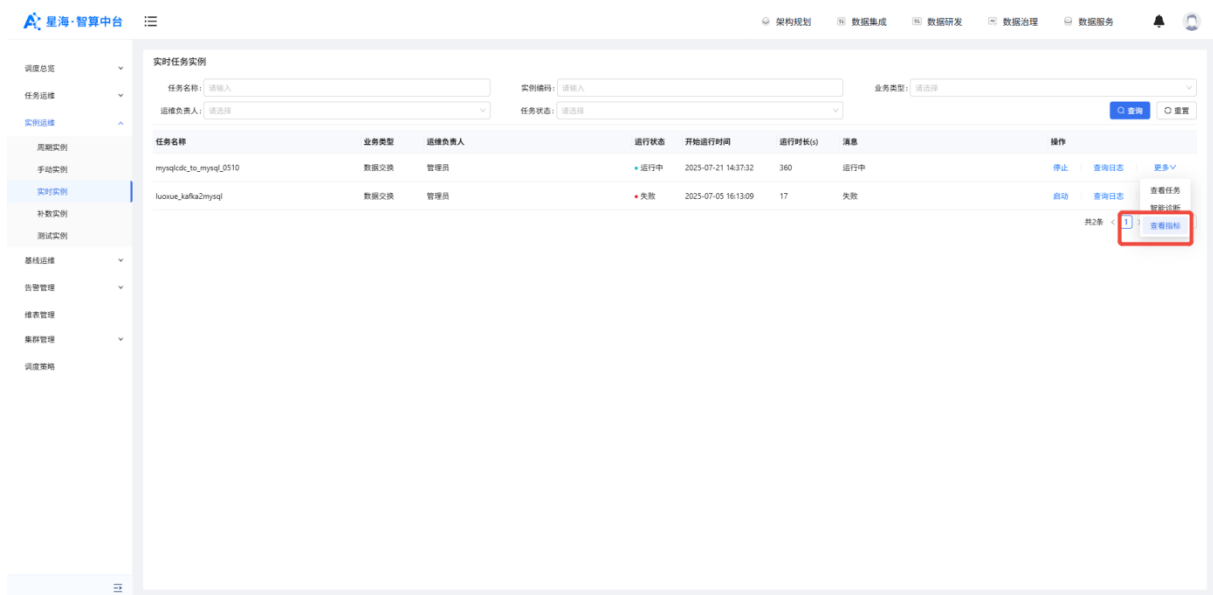
- 新建：实例初始化状态，代表实例刚刚生成，且还没有被调度机进行调度处理。

- 启动中：当任务准备提交到 yarn/k8s 等计算资源上面时候，任务启动会持续一段时间，启动中即是代表任务提交到计算资源，但任务还没启动成功这一段时间的状态。
- 运行中：正在运行中的实例。
- 停止中：已触发停止实例的操作，系统正在执行停止实例。
- 停止：停止正在运行的实例。
- 失败：启动没有成功导致实例失败。

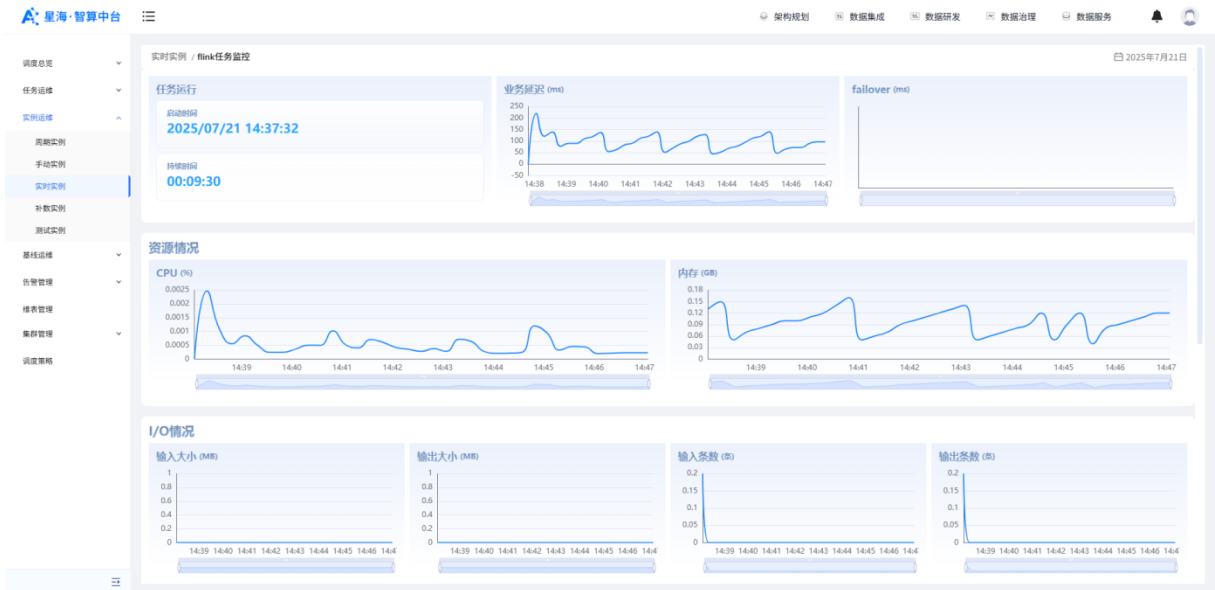
### 实时实例指标查看

实时实例运行起来之后，可以查看 flink 相关的运行指标，在列表页，点击【更多】按钮，点击【查看指标】即可跳转至指标页面。

注意：只有处于【运行中】状态的实时实例才可查看指标。



查看指标中可以查看任务运行情况、业务延迟情况、资源情况、I/O 情况、checkpoint 情况与返压情况等。



## 补数实例

对周期任务发起补数据操作后，将根据周期类型和业务日期产生不同的补数据实例。您可以根据业务需求对补数据实例进行运维管理，包括查看运行日志、重跑等操作。

注意：补数操作需要任务处于生效状态才能使用。

## 应用场景

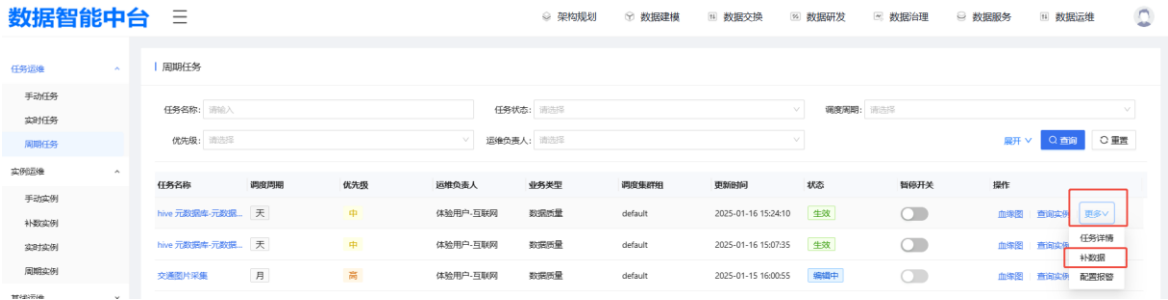
- 新开发的周期任务，只能从第二天开始调度。如果您需要马上查看历史分区数据，可以执行补数据操作。
- 上游依赖的任务重跑或补数据导致历史分区数据刷新。您可以通过补数据功能，回刷下游任务的历史分区数据。
- 业务的历史数据存在遗漏，需补充历史数据，相当于将写入数据至历史的对应时间分区。

业务历史数据有遗漏需要补充历史数据的示例说明：

- 步骤一：进入周期任务列表页，点击补数操作  
进入运维中心页面。

按照以下操作指引，进入周期任务列表页。

依次单击数据运维->任务运维->周期任务->操作栏更多->补数据



## ● 步骤二：填写补数日期

选中某生效任务，在其操作栏点击更多按钮再操作补数据功能，弹出弹窗



需要选中补数的业务日期，其中开始时间到结束时间都只能是历史时间，没法选中未来的时间节点，如当天为 10 号，我需要补 9 号的数据，则在任务生效的情况下，开始时间和结束时间都选择 9 号，即可生成 9 号的数实例。

补数任务一次性只能提交 50 个业务日期。

## ● 步骤三：运行补数实例

补数据实例生成后将自动触发运行。

### 补数实例入口

在数据中台首页，单击顶部菜单栏的数据运维。默认进入数据运维页面

按照以下操作指引，进入补数实例运维页面。

依次单击数据运维->实例运维->补数实例。



### 补数实例页面介绍

补数实例页面以列表方式，为您展示系统已生成的补数实例。补数实例页面主要由搜索及筛选区、补数实例列表组成。您可以在补数实例页面进行相关的运维操作。

#### 补数任务实例

任务名称	业务日期	运行状态	负责人	业务类型	开始运行时间	结束运行时间	运行时长(s)	消息	操作
<input type="checkbox"/> hive_logical	20240831	成功	体验用户-互联网	数据开发	2024-09-12 15:18:16	2024-09-12 15:20:19	123	任务运行成功	<a href="#">查询日志</a> <a href="#">查看任务</a> <a href="#">更多</a>
<input type="checkbox"/> hive_logical	20240901	成功	体验用户-互联网	数据开发	2024-09-11 20:17:01	2024-09-11 20:33:51	1010	任务运行成功	<a href="#">查询日志</a> <a href="#">查看任务</a> <a href="#">更多</a>
<input type="checkbox"/> hive_logical	20240904	成功	体验用户-互联网	数据开发	2024-09-11 20:17:01	2024-09-11 20:33:50	1009	任务运行成功	<a href="#">查询日志</a> <a href="#">查看任务</a> <a href="#">更多</a>
<input type="checkbox"/> hive_logical	20240905	成功	体验用户-互联网	数据开发	2024-09-11 20:17:01	2024-09-11 20:33:45	1004	任务运行成功	<a href="#">查询日志</a> <a href="#">查看任务</a> <a href="#">更多</a>
<input type="checkbox"/> hive_logical	20240907	成功	体验用户-互联网	数据开发	2024-09-11 20:17:01	2024-09-11 20:33:43	1002	任务运行成功	<a href="#">查询日志</a> <a href="#">查看任务</a> <a href="#">更多</a>
<input type="checkbox"/> hive_logical	20240903	成功	体验用户-互联网	数据开发	2024-09-11 20:17:01	2024-09-11 20:33:43	1002	任务运行成功	<a href="#">查询日志</a> <a href="#">查看任务</a> <a href="#">更多</a>
<input type="checkbox"/> hive_logical	20240902	成功	体验用户-互联网	数据开发	2024-09-11 20:17:01	2024-09-11 20:33:41	1000	任务运行成功	<a href="#">查询日志</a> <a href="#">查看任务</a> <a href="#">更多</a>
<input type="checkbox"/> hive_logical	20240906	成功	体验用户-互联网	数据开发	2024-09-11 20:17:01	2024-09-11 20:33:39	998	任务运行成功	<a href="#">查询日志</a> <a href="#">查看任务</a> <a href="#">更多</a>
<input type="checkbox"/> hive_logical	20240908	成功	体验用户-互联网	数据开发	2024-09-11 20:17:01	2024-09-11 20:33:38	997	任务运行成功	<a href="#">查询日志</a> <a href="#">查看任务</a> <a href="#">更多</a>

编号区域	描述
① 筛选区	<p>快捷筛选支持通过输入任务名称来筛选周期任务经过补数操作产生的补数实例，适用于知晓任务名称情况下来进行快捷筛选出当前任务进行补数操作后产生的所有补数实例信息。</p> <p>同时支持运行状态、负责人、业务类型、开始时间、结束时间、业务日期这几种维度下选择来进行列表内容的快捷筛选。</p> <ul style="list-style-type: none"> <li>● 运行状态：支持新建、启动中、运行中、停止中、停止、失败。</li> <li>● 负责人：可根据实时实例的运维负责人进行筛选。</li> <li>● 业务类型：支持数据交换、数据治理、数据开发这三种业务模块的筛选。</li> <li>● 业务日期：可根据补数实例的业务日期进行筛选。具体到某一天或某一业务时间区间（精确到秒）。</li> <li>● 开始时间：可根据补数实例的开始时间进行筛选。具体到某一天或某一业务时间区间（精确到秒）。</li> <li>● 结束时间：可根据补数实例结束时间进行筛选，具体到某一天或某一业务时间区间（精确到秒）。</li> </ul> <p>如果需要重新筛选，则可单击重置，即可快速清空所有筛选条件，恢复系统默认。</p>
② 补数实例列表	<p>补数实例页面以列表形式为您展示系统已生成的补数实例。列表操作栏为您展示补数实例支持的运维管理操作。</p>

# 告警管理

## 概述

任务监控功能支持您通过配置监控规则实现对离线任务的运行状态进行监控。在离线任务触发所配置的监控规则时，系统将根据所配置的接收人发送报警，以协助您保障任务的正常运行和数据的产出。

离线任务中支持对周期任务、手动任务直接进行监控告警配置。

## 告警管理入口

在数据中台首页，单击顶部菜单栏的数据运维。默认进入数据运维页面

按照以下操作指引，进入监控告警任务页面。

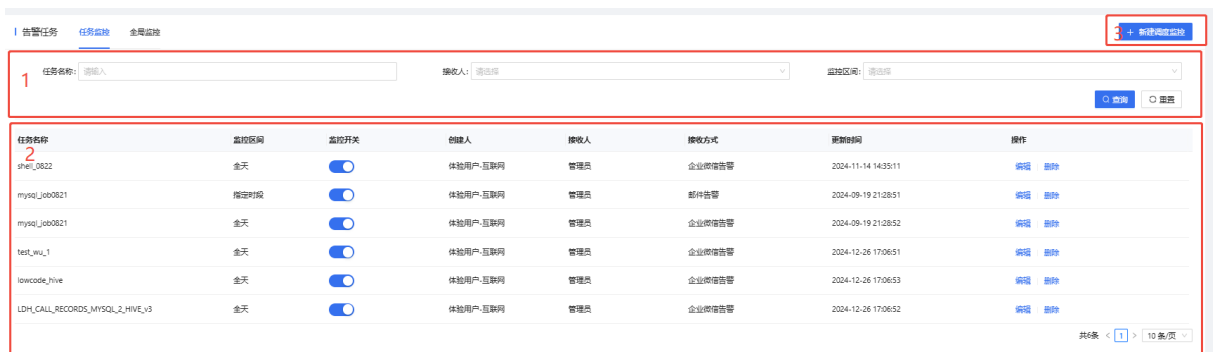
依次单击数据运维->告警管理->监控告警任务->teb 选中任务监控



## 监控告警任务页面介绍

在离线监控配置页面，您可以查看已生成的监控报警规则。

任务监控页面展示已配置的任务监控的监控对象、监控区间、创建人、接收人、接收方式、最近更新时间、监控开关的状态信息及支持的编辑和删除操作。



编号区域	描述
① 筛选区	快捷筛选支持通过输入任务名称来筛选此任务配置的监控告警有那些，适用于知晓任务名称情况下来进行快捷筛选出当前任务所属监控。

	<p>同时支持接收人、监控区间这两种维度下选择来进行列表内容的快捷筛选。</p> <ul style="list-style-type: none"> <li>● 接收人：即每条监控规则中配置的告警接收对象。</li> <li>● 监控区间：监控生效的时间区间，支持全天或指定时段这两种。</li> </ul> <p>如果需要重新筛选，则可单击重置，即可快速清空所有筛选条件，恢复系统默认。</p>
② 监控任务列表	在监控规则列表区域，您可编辑、删除监控规则和开启或关闭监控开关。
③ 新建监控区	支持任务的快捷新建操作。

## 配置监控告警任务-报警规则

离线任务监控管理支持批为单个任务批量配置多个监控报警规则，提高了配置监控报警规则的效率。

同时，我们支持为监控报警规则配置告警接收人，并且支持接收人可以指定不同的接收方式，以确保任务相关人员都能收到报警信息。

例如：

- 为责任人配置短信或邮件告警，这样责任人能了解所负责的任务的运行概况，但是不用第一时间进行处理。
- 为项目管理员配置企微告警，这样项目管理员能通过群里机器人的消息了解所管理的项目全局任务的整体运行情况，便于后续统计分析。

## 规则创建

1. 在任务监控页面。按照下图指引，进入新建离线任务监控页面。



2. 在新建任务监控页面中，配置报警原因步骤以及配置监控参数具体如下：

参数	描述
基本信息	选择需要配置监控规则的任务
新建报警原因	若需同时配置多个报警原因，您可单击进行添加。
报警信息-报警原因	报警原因支持配置出错、超时两种 <ul style="list-style-type: none"> <li>● 出错：任务运行过程中出错后即触发报警</li> <li>● 超时：设置超时原因须设置运行超过时间。从实例运行状态变成运行中开始计算，经过指定时长后任务未运行结束则触发报警。</li> </ul>
报警信息-报警频率	告警信息发送的频率。支持配置频率范围为多少分钟/次，默认值为 30 分钟一次。
报警信息-告警区间	设置监控生效的时间区间。支持全天或指定时段。指定时段时须配置起始时点和结束时间点。 <small>* 告警区间：<input type="radio"/> 全天 <input checked="" type="radio"/> 指定时段 开始时间 <input type="text"/> → 结束时间 <input type="text"/></small>
报警信息-报接收人	报警的接收人配置，只允许配置一个用户。可以是任务的责任人，也可以是项目的负责人。
报警信息-接收方式	支持邮件和企业微信两种。
报警信息-发送次数	报警触发时，报警信息发送的次数。支持填写整数，默认值为 3。

## 全局监控管理

### 概述

全局监控功能支持您通过控制内置的监控规则实现对全局相关类型的任务进行报警监控。在全局任务监控被触发时，系统将根据任务所配置负责人来进行报警发送，以协助您保障任务的正常运行和数据的产出。

全局任务中支持对周期任务暂停过长、以及实时任务的出错和超时直接进行监报告警配置。

## 全局监控入口

在数据中台首页，单击顶部菜单栏的数据运维。默认进入数据运维页面

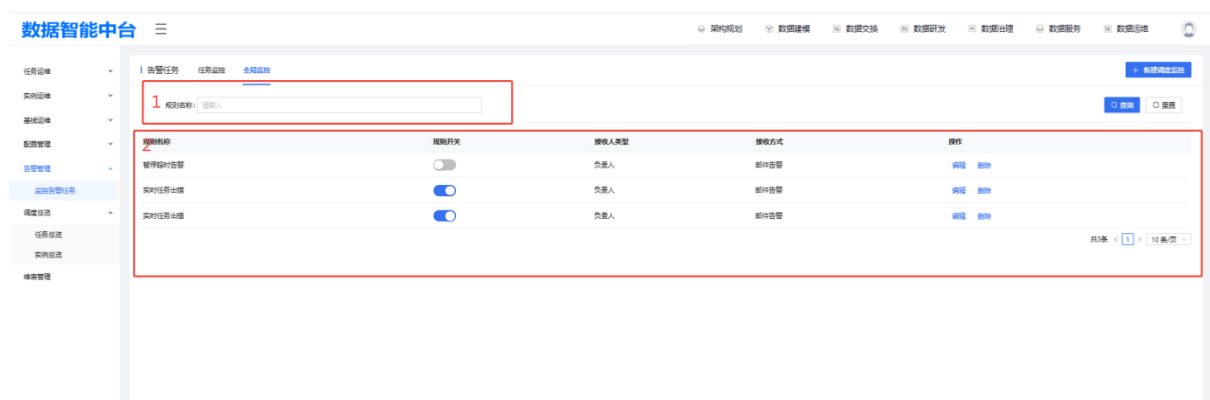
按照以下操作指引，进入监控告警任务页面。

依次单击数据运维->告警管理->监控告警任务->teb 选中全局监控



## 全局监控页面介绍

在全局监控配置页面，您可以查看内置好的全局监控报警任务，并控制是否打开这些监控任务。任务监控页面展示已配置的规则名称、规则开关、接收人类型、接收方式、以及支持的编辑操作。



编号区域	描述
① 筛选区	快捷筛选支持通过输入规则名称来筛选全局监控规则有那些，适用于知晓规则名称情况下来进行快捷筛选出所属监控。 如果需要重新筛选，则可单击重置，即可快速清空所有筛选条件，恢复系统默认。
② 监控任务列表	在监控规则列表区域，您可编辑监控规则、开启或关闭监控开关。

## 以周期任务暂停规则为例

### 前提条件

如果需要监控能成功生效，除了打开任务监控开关以外，还需要满足以下条件中的任意一个：

- 监控的任务类型已经在生产环境执行。

### 规则编辑

1. 在全局任务监控页面中，编辑报警规则的监控参数具体如下：

告警任务 - 修改全局监控配置

**基本信息**

规则名称: 暂停超时报警 告警接收人类型: 负责人

告警接收方式: 邮件告警 告警类型: 暂停超时报警

告警内容: 任务已暂停超过72小时,生成实际耗时超过72小时,请及时恢复,如在超长时间恢复会使资源循环,数据混乱。  
配置告警内容时,其中变量请用%{}代替

**告警配置**

超时时间: 72 小时 告警频率: 30 分钟

发送次数: 3 次

参数	描述
规则名称	不可修改,就是内置的全局监控规则中文名。
告警接收人类型	全局告警是针对所有实时任务和周期暂停任务的,所以
告警接收方式	支持邮件和企业微信两种。
告警类型	支持暂停超时告警,实时超时告警、实时出错告警三种。
告警内容	不可修改,默认按照内置模板内容进行变量生成。
超时时间	周期任务暂停设置的运行超过时间。从任务变成暂停状态开始计算,经过指定时长后任务还处于暂停状态则触发报警。
告警频率	告警信息发送的频率。支持配置频率范围为多少分钟/次,默认值为30分钟一次。
发送次数	报警触发时,报警信息发送的次数。支持填写整数,默认值为3。

## 数据治理

### 元数据

**元数据是描述数据的数据**,它描述了系统拥有什么数据、数据代表什么、数据的来源、数据如何在系统中迁移、数据之间的关系、如何使用数据等关键信息。只有知道了拥有哪些数据,才能对数据进行管理。元数据为数据管理提供基础支撑,为数据采集、数据质量、数据标准、数据资产、主数据等提供依据。

元数据模块负责从各业务系统中**抽取、加工、集中存储和管理元数据**,以支持数据治理,并加强组织内部数据的组织、检索和分析能力。

元数据主要涵盖以下几个功能模块：元数据总览、元模型管理、采集器管理、采集任务管理、元数据管理、多模态元数据采集。通过配置/使用元数据采集器、创建采集任务、配置任务调度等流程，从数据源采集元数据，提供全量采集或增量采集功能；对采集到的元数据统一汇总管理，为元数据授权以支撑元数据的使用、血缘分析等，同时提供元数据总览功能、元模型管理功能，支持多模态元数据采集。

元数据的主要采集流程为：

步骤	描述
步骤一（可选）：创建元模型 （元模型管理）	为元数据采集活动定义一个标准化的框架，定义要采集的元数据的结构和属性，明确属性的数据类型、值域、格式
步骤二（可选）：创建元数据采集器 （采集器管理）	配置元数据的来源、元模型以及采集逻辑
步骤三：创建采集任务（采集任务管理）	配置采集任务和采集调度，系统根据调度信息按照预定的任务设置和时间规则进行元数据采集
步骤四：查看并管理已采集的元数据，进行授权、血缘分析等 （元数据管理）	浏览、查询系统采集到的元数据，对元数据进行授权，查看和配置血缘关系

其中，如果使用内置元模型和内置元数据采集器，则不需要进行前两个步骤，直接从第三步开始操作即可。

数据、元数据、元模型的定义区分：

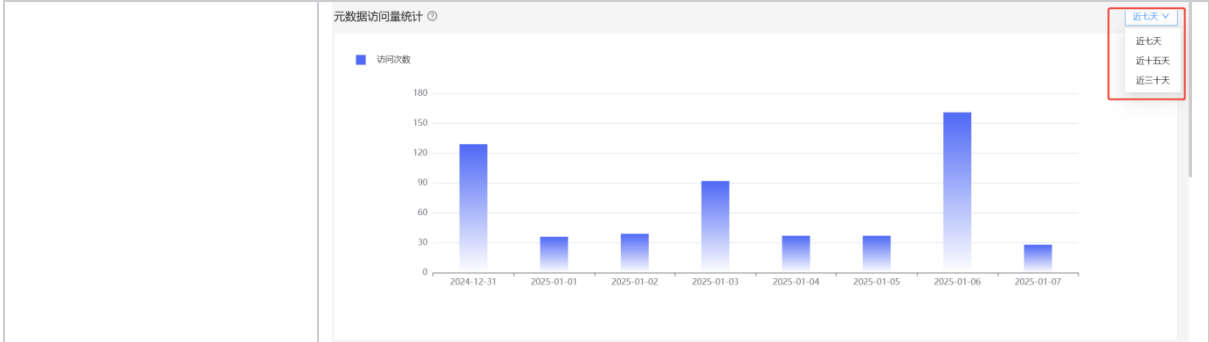
名词	解释
数据	数据是一种表示方法，是理解事实的符号，是以数字形式存储的信息
元数据	元数据是描述数据的数据，它描述了系统拥有什么数据、数据代表什么、数据的来源、数据如何在系统中迁移、数据之间的关系、如何使用数据等关键信息
元模型	元模型是描述元数据的数据。元模型为元数据采集活动定义了一个标准化的框架，定义了要采集的元数据的结构和属性，明确了属性的数据类型、值域、格式；元模型为采集器提供明确的采集目标和范围

## 元数据总览

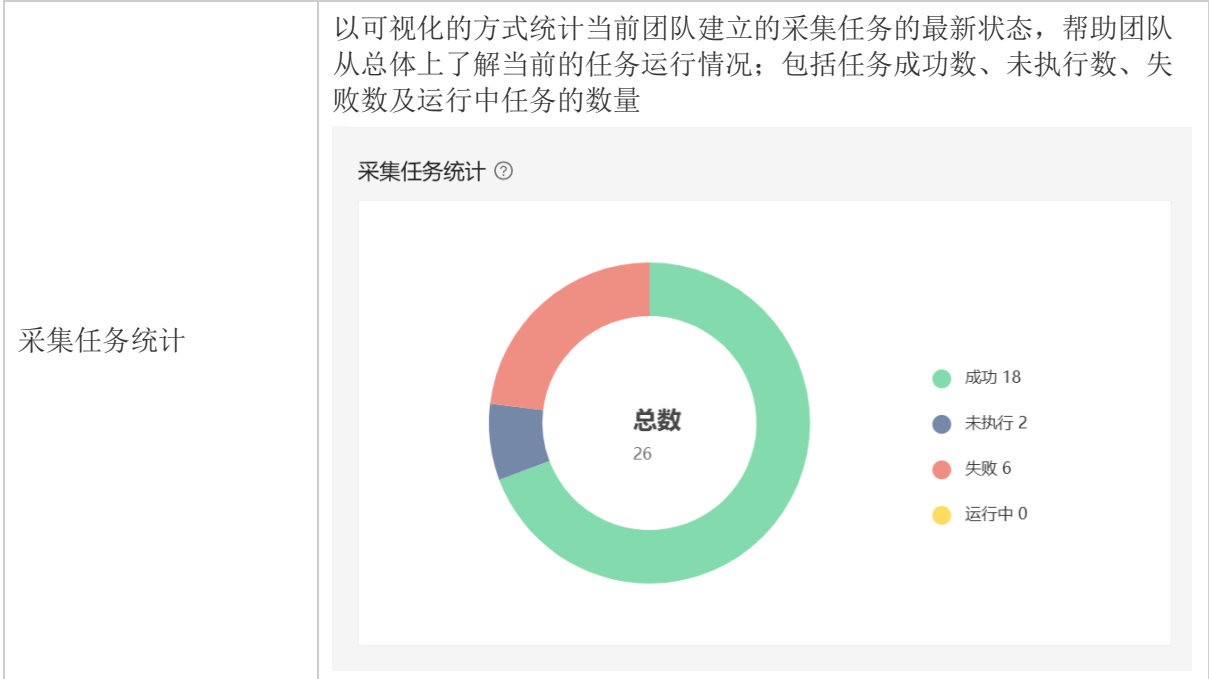
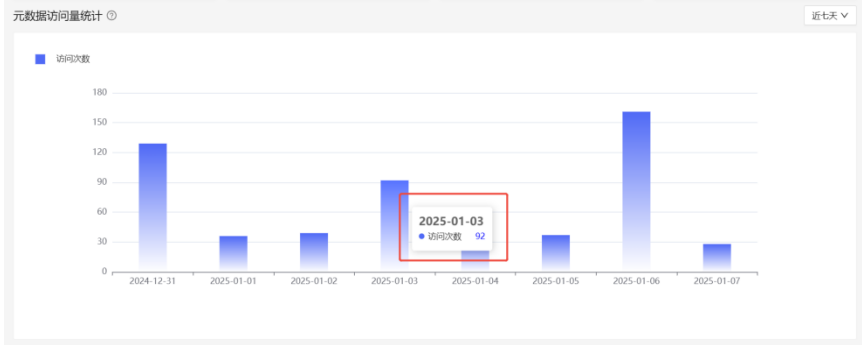
通过元数据总览界面呈现元数据全貌：集成元数据统计分析功能、元数据访问量统计、采集任务统计、元数据统计、热门元数据排行及最近采集的元数据总览，通过直观的可视化图表，可以更加直观的了解系统中的元数据总体情况。



模块	描述
元数据统计分析	<p>分为四个模块，分别展示产生元数据流程中每个关键环节的元数据统计情况，包括：已经采集的数据源数量，从数据源的维度展示元数据采集任务的完成情况；已构建的元模型数量，从建立元模型的维度展示元数据采集功能的全面性；已经完成的采集任务数量，展示系统的元数据采集能力；结构化元数据的数量，从最主要的元数据的维度展示系统在表维度上的采集能力</p>
元数据访问量统计	<p>以可视化图表的形式展示用户元数据访问次数的统计数量，从访问率的维度展示元数据的使用情况</p> <p>在右下角下拉菜单可以选择查看近七天、近十五天或近三十天的数据统计结果</p>



每日元数据访问量通过柱状图展示，鼠标悬停至柱状图上展示详细数值



元数据统计

在该模块汇总各个元模型采集到的元数据数量，展示内容包括元模型中文名、元模型英文名及元数据数量

	<p><b>元数据统计</b></p> <table border="1"> <thead> <tr> <th>元模型中文名</th> <th>元模型英文名</th> <th>元数据数量</th> </tr> </thead> <tbody> <tr> <td>标准Hive字段</td> <td>HiveField</td> <td>1701</td> </tr> <tr> <td>标准Hive表</td> <td>HiveTable</td> <td>249</td> </tr> <tr> <td>标准数据命名空间</td> <td>DataSourceSchema</td> <td>12</td> </tr> <tr> <td>标准数据源</td> <td>DataSource</td> <td>24</td> </tr> <tr> <td>标准Mysql表</td> <td>MySQLTable</td> <td>257</td> </tr> </tbody> </table> <p style="text-align: right;">上一页 <span style="border: 1px solid blue; padding: 0 2px;">1</span> 2 下一页</p>	元模型中文名	元模型英文名	元数据数量	标准Hive字段	HiveField	1701	标准Hive表	HiveTable	249	标准数据命名空间	DataSourceSchema	12	标准数据源	DataSource	24	标准Mysql表	MySQLTable	257						
元模型中文名	元模型英文名	元数据数量																							
标准Hive字段	HiveField	1701																							
标准Hive表	HiveTable	249																							
标准数据命名空间	DataSourceSchema	12																							
标准数据源	DataSource	24																							
标准Mysql表	MySQLTable	257																							
<p>热门元数据 Top5</p>	<p>统计最近一周内访问次数前 5 的元数据，按照次数由多到少的顺序排行，拉取并直观地展示用户最关心、利用率最高的元数据信息。元数据信息展示包括元数据名称、数据源类型、数据源名称及访问次数</p> <p><b>热门元数据 Top5</b> <span style="float: right;">🕒 最近一周</span></p> <table border="1"> <thead> <tr> <th>访问元数据</th> <th>数据源类型</th> <th>数据源名称</th> <th>访问次数</th> </tr> </thead> <tbody> <tr> <td>ods_call_records</td> <td>MySQL</td> <td>default_mysql</td> <td>101</td> </tr> <tr> <td>default_mysql</td> <td>MySQL</td> <td>default_mysql</td> <td>87</td> </tr> <tr> <td>default_hive</td> <td>Hive</td> <td>default_hive</td> <td>55</td> </tr> <tr> <td>dim_customers</td> <td>Hive</td> <td>default_hive</td> <td>35</td> </tr> <tr> <td>pro</td> <td>MySQL</td> <td>default_mysql</td> <td>35</td> </tr> </tbody> </table>	访问元数据	数据源类型	数据源名称	访问次数	ods_call_records	MySQL	default_mysql	101	default_mysql	MySQL	default_mysql	87	default_hive	Hive	default_hive	55	dim_customers	Hive	default_hive	35	pro	MySQL	default_mysql	35
访问元数据	数据源类型	数据源名称	访问次数																						
ods_call_records	MySQL	default_mysql	101																						
default_mysql	MySQL	default_mysql	87																						
default_hive	Hive	default_hive	55																						
dim_customers	Hive	default_hive	35																						
pro	MySQL	default_mysql	35																						
<p>最近采集元数据</p>	<p>展示最近执行采集任务采集到的元数据信息，根据采集时间距当前时间由短到长的顺序排序，展示内容包括元模型名称、元数据名称、数据源名称及更新时间</p> <p><b>最近采集元数据</b></p> <table border="1"> <thead> <tr> <th>元模型名称</th> <th>元数据名称</th> <th>数据源名称</th> <th>更新时间</th> </tr> </thead> <tbody> <tr> <td>HiveField</td> <td>end_point</td> <td>default_hive</td> <td>2025-01-09 16:12:29</td> </tr> <tr> <td>HiveField</td> <td>accident_type_id</td> <td>default_hive</td> <td>2025-01-09 16:12:29</td> </tr> <tr> <td>HiveField</td> <td>month_statistics_time</td> <td>default_hive</td> <td>2025-01-09 16:12:29</td> </tr> <tr> <td>HiveField</td> <td>event_description</td> <td>default_hive</td> <td>2025-01-09 16:12:29</td> </tr> <tr> <td>HiveField</td> <td>create_time</td> <td>default_hive</td> <td>2025-01-09 16:12:29</td> </tr> </tbody> </table> <p style="text-align: right;">上一页 <span style="border: 1px solid blue; padding: 0 2px;">1</span> 2 3 4 5 ... 711 下一页 <span style="border: 1px solid gray; padding: 0 2px;">5条/页</span> ▾</p>	元模型名称	元数据名称	数据源名称	更新时间	HiveField	end_point	default_hive	2025-01-09 16:12:29	HiveField	accident_type_id	default_hive	2025-01-09 16:12:29	HiveField	month_statistics_time	default_hive	2025-01-09 16:12:29	HiveField	event_description	default_hive	2025-01-09 16:12:29	HiveField	create_time	default_hive	2025-01-09 16:12:29
元模型名称	元数据名称	数据源名称	更新时间																						
HiveField	end_point	default_hive	2025-01-09 16:12:29																						
HiveField	accident_type_id	default_hive	2025-01-09 16:12:29																						
HiveField	month_statistics_time	default_hive	2025-01-09 16:12:29																						
HiveField	event_description	default_hive	2025-01-09 16:12:29																						
HiveField	create_time	default_hive	2025-01-09 16:12:29																						

## 元模型管理

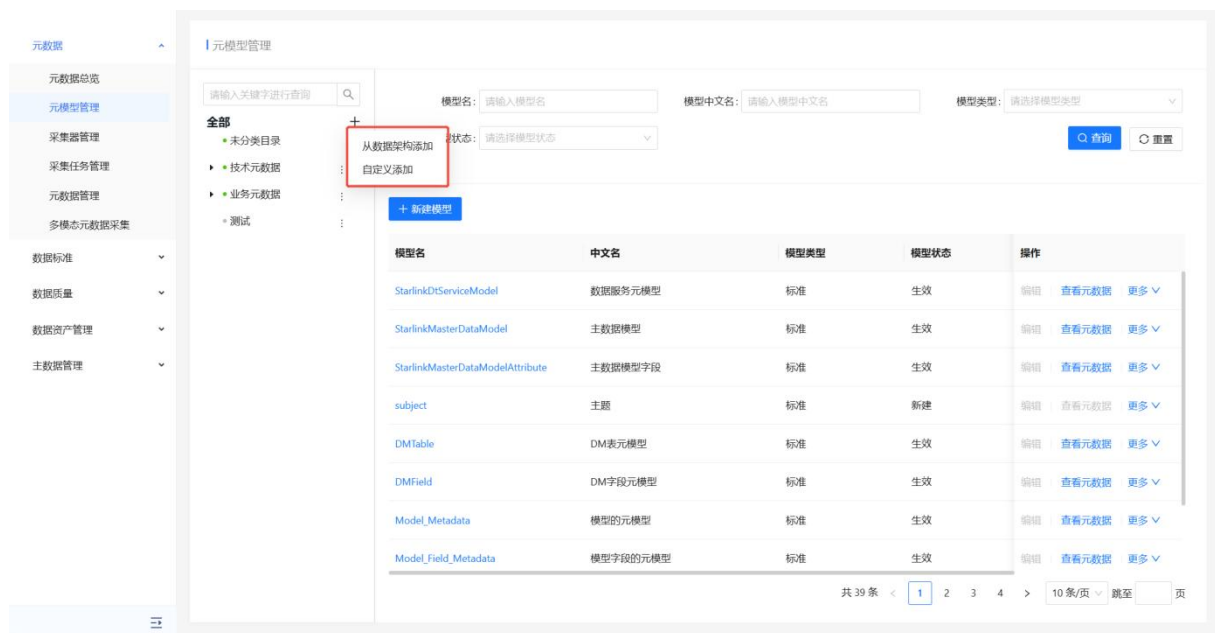
元模型目录按照元模型分类管理元模型，左侧目录区域默认显示全部目录，展示目录结构。

为应对元数据的分类分级需求，灵活响应业务或需求的变化，元模型管理模块提供灵活的目录管理能力。可以查看每个分类目录下的元模型，支持在目录中检索元模型，支持在现有目录结构中新建或修改以管理元模型目录。

系统内置了技术元数据和业务元数据目录，在技术元数据和业务元数据目录中包含了常见的元模型（模型类型为“标准”的均为内置元模型）。

### ● 新建目录

点击目录区域的加号【+】，可以选择从数据架构添加目录或自定义添加目录。



新建目录方式	方法
从数据架构添加	<p>选择【从数据架构添加】，需要选择目录密级，目录密级分为：绝密、机密、秘密、内部及公开五个级别；</p> <p>选择架构支持【级联选择】和【单独选择】两种方式：</p> <p>1) 级联选择。可以实现批量选择该架构层级及下级层次的所有架构</p>

	<div data-bbox="539 181 1114 808"> <p><b>从数据架构添加目录</b></p> <p>密级: 内部</p> <p><input checked="" type="radio"/> 级联选择 <input type="radio"/> 单独选择</p> <ul style="list-style-type: none"> <li>团队架构 <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> 主题 <ul style="list-style-type: none"> <li>业务主题域 <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> 产品域</li> <li><input checked="" type="checkbox"/> 公共域</li> </ul> </li> <li>应用主题域 <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> 渠道主题</li> <li><input checked="" type="checkbox"/> 产品/用户主题</li> </ul> </li> <li>来源系统 <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> 基础用户平台</li> <li><input checked="" type="checkbox"/> 运营平台</li> </ul> </li> </ul> </li> <li>层次 <ul style="list-style-type: none"> <li><input type="checkbox"/> 接口层(STG)</li> <li><input type="checkbox"/> 业务应用层(DWA)</li> <li><input type="checkbox"/> 公共维度层(DIM)</li> <li><input type="checkbox"/> 数据源层 (ODS)</li> <li><input type="checkbox"/> 明细层 (DWD)</li> <li><input type="checkbox"/> 汇总层 (DWS)</li> <li><input type="checkbox"/> 应用层 (ADS)</li> </ul> </li> <li>数据库 <ul style="list-style-type: none"> <li><input type="checkbox"/> MYSQL</li> <li><input type="checkbox"/> HIVE</li> <li><input type="checkbox"/> DORIS</li> </ul> </li> </ul> <p>取消 确定</p> </li></ul></div> <p>2) 单独选择。可以实现单独选择某层级架构</p> <div data-bbox="539 898 1114 1525"> <p><b>从数据架构添加目录</b></p> <p>密级: 内部</p> <p><input type="radio"/> 级联选择 <input checked="" type="radio"/> 单独选择</p> <ul style="list-style-type: none"> <li>团队架构 <ul style="list-style-type: none"> <li><input type="checkbox"/> 主题 <ul style="list-style-type: none"> <li>业务主题域 <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> 产品域</li> <li><input type="checkbox"/> 公共域</li> </ul> </li> <li>应用主题域 <ul style="list-style-type: none"> <li><input type="checkbox"/> 渠道主题</li> <li><input type="checkbox"/> 产品/用户主题</li> </ul> </li> <li>来源系统 <ul style="list-style-type: none"> <li><input type="checkbox"/> 基础用户平台</li> <li><input type="checkbox"/> 运营平台</li> </ul> </li> </ul> </li> <li>层次 <ul style="list-style-type: none"> <li><input type="checkbox"/> 接口层(STG)</li> <li><input type="checkbox"/> 业务应用层(DWA)</li> <li><input type="checkbox"/> 公共维度层(DIM)</li> <li><input type="checkbox"/> 数据源层 (ODS)</li> <li><input type="checkbox"/> 明细层 (DWD)</li> <li><input type="checkbox"/> 汇总层 (DWS)</li> <li><input type="checkbox"/> 应用层 (ADS)</li> </ul> </li> <li>数据库 <ul style="list-style-type: none"> <li><input type="checkbox"/> MYSQL</li> <li><input type="checkbox"/> HIVE</li> <li><input type="checkbox"/> DORIS</li> </ul> </li> </ul> <p>取消 确定</p> </li></ul></div>
自定义添加	<p>选择【自定义添加】，填写基础信息和自定义属性，自定义属性为目录的附加信息。基础信息部分需要填写目录名称（必填）、密级（必填）及描述；</p> <p>自定义属性为选填部分，可以通过点击【新增自定义属性】按钮实现自定义属性的新增和删除，支持新增多条自定义属性</p>

注册编辑 ✕

**基础信息**

• 目录名称:

• 密码:

描述:

0 / 200

**自定义属性**

[+ 新增自定义属性](#)

自定义属性:  ✕

自定义属性:  ✕

取消
确定

## ● 目录操作

针对现有目录，点击目录名右侧的按钮，可以实现对目录的管理和操作。

操作	描述
添加同级目录和添加子级目录	<p>为选中目录添加同级或子级目录，支持【从数据架构添加】和【自定义添加】</p> 
移动、编辑、发布、删除、下线	<p>对选中目录进行操作，通过点击相应选项，完成对目录的移动、编辑、发布、删除、下线</p> 

	
授权	<p>为团队或个人授予所选目录的管理和编辑权限。点击【授权】选项，在右侧抽屉展示目录的基本信息以及权限管理内容。在权限管理栏内，可以选择对【团队】权限进行管理和对【个人】权限进行管理，通过勾选的方式选择是否授予【管理】和【编辑】的权限。点击【确定】，完成目录的授权</p>

### 授权管理

基本信息

目录名称	密级	负责人
测试目录	公开	体验用户-互联网

权限管理

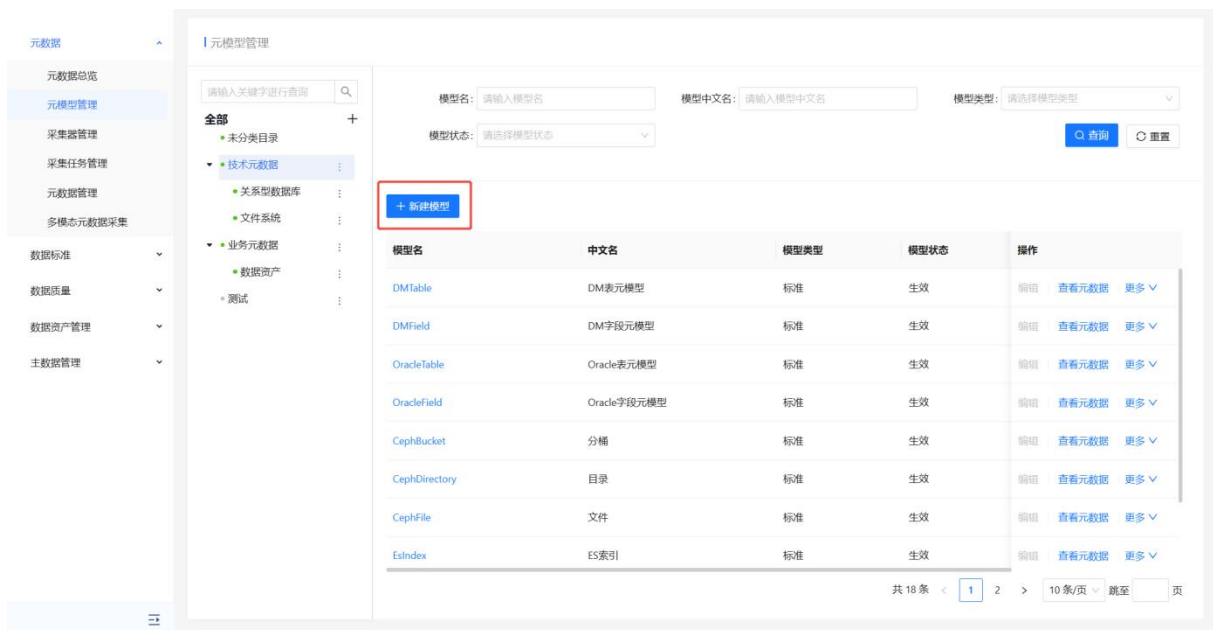
团队 个人

团队名称	管理	编辑
管算中台	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
AI中台	<input type="checkbox"/>	<input type="checkbox"/>
青海电信BOC	<input type="checkbox"/>	<input type="checkbox"/>
TeleBI分析团队	<input type="checkbox"/>	<input type="checkbox"/>
知识图谱	<input type="checkbox"/>	<input type="checkbox"/>

## 新建模型

为响应不同的元数据采集、数据管理及元数据管理等需求，支持新建自定义元模型。

点击【+新建模型】按钮，进入新建页面，填写模型信息，按照需求构建元模型。

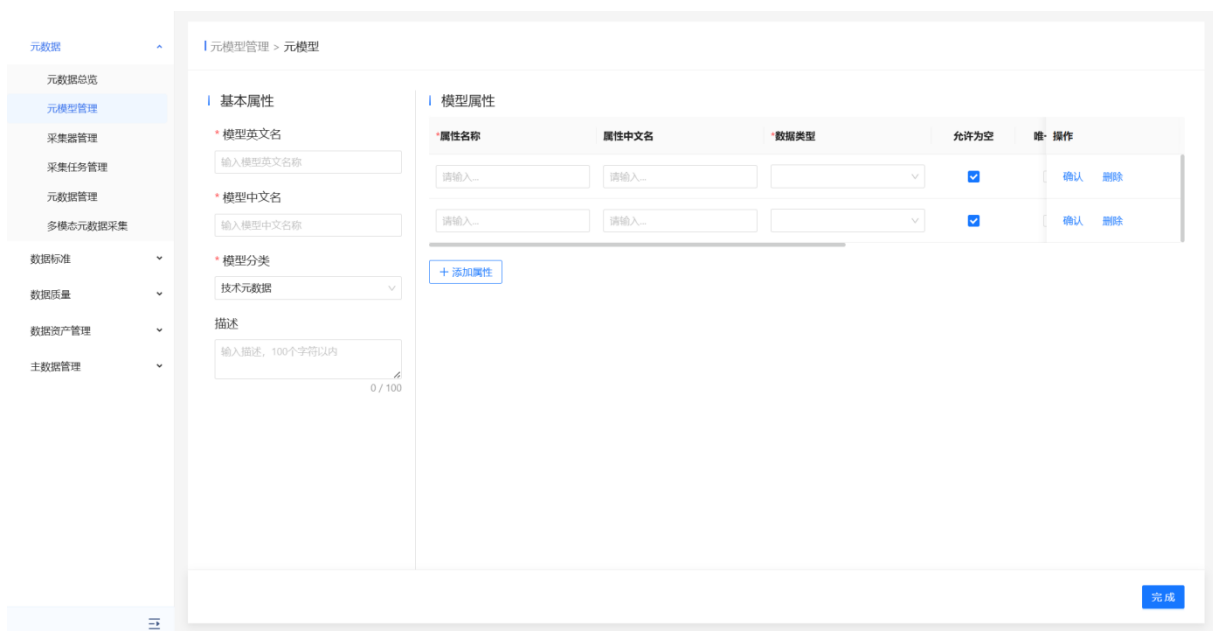


## ● 基本属性

填写模型英文名（必填）、模型中文名（必填）、模型分类（必填）及描述。

## ● 模型属性

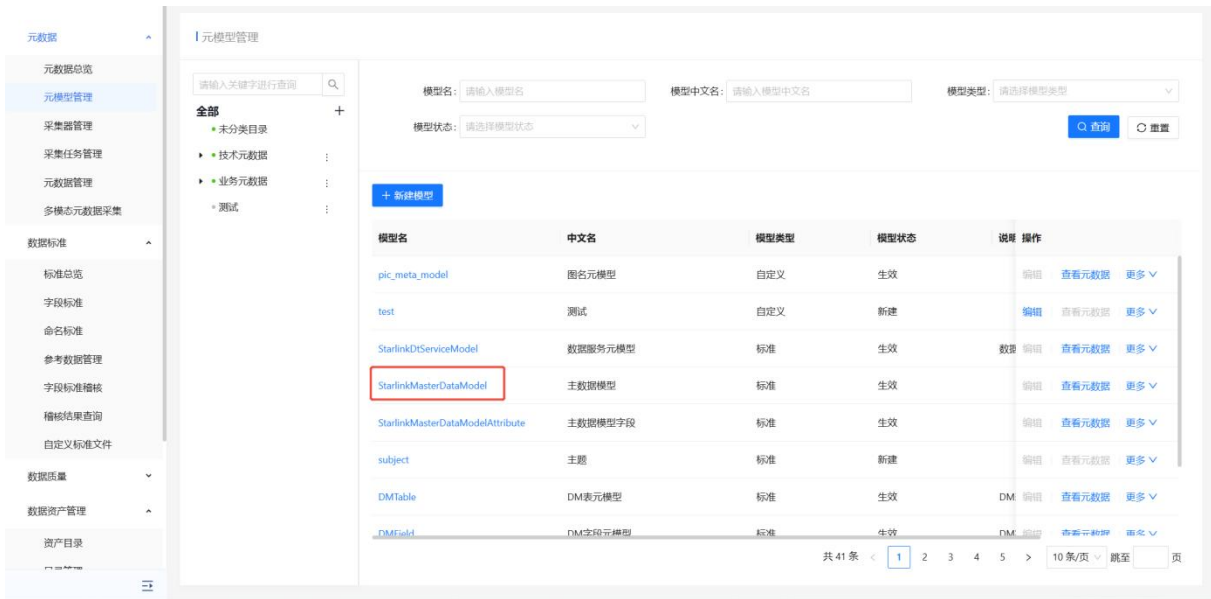
模型属性信息包括属性名称（必填）、属性中文名、数据类型（必填）、允许为空、唯一键、索引、默认值、说明及值域配置；可以对模型属性进行确认和删除；可以添加多条属性。



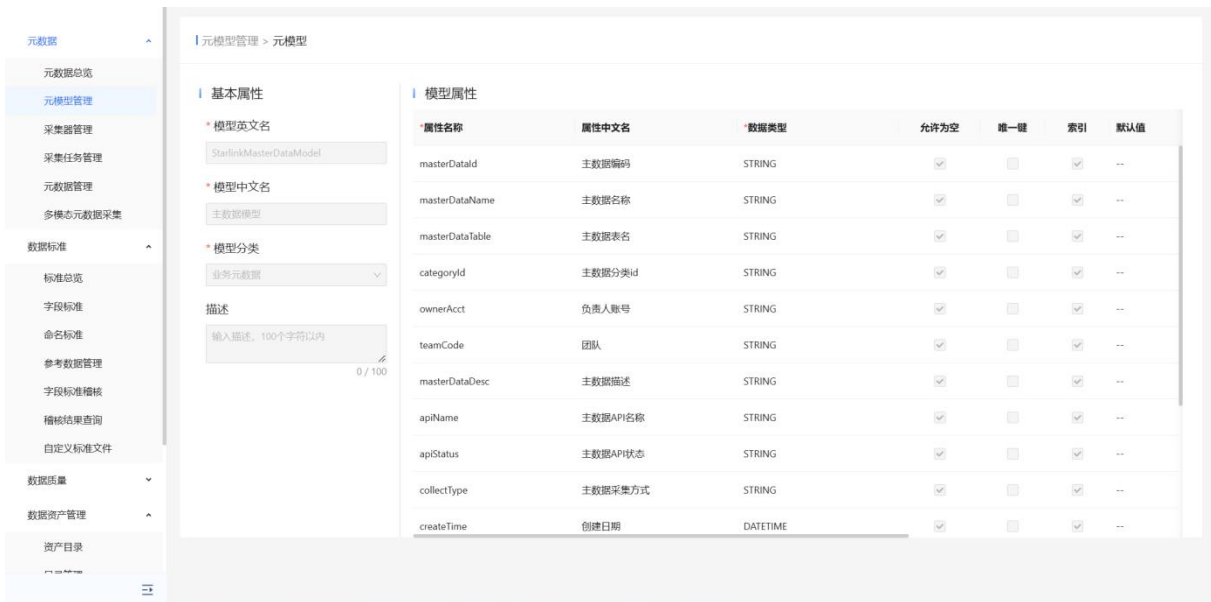
## 查看元模型

在元数据管理界面中，可以看到以列表形式展示出的元模型信息，包括模型名、中文名、模型类型及说明。

可以通过点击元模型的模型名，查看该元模型的具体信息



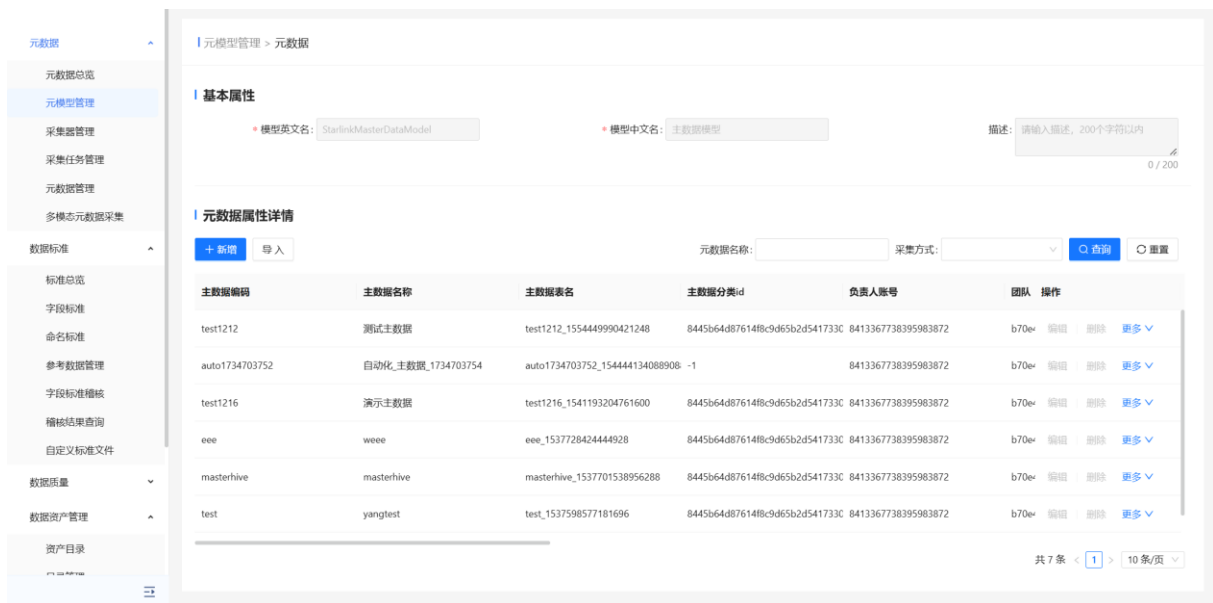
元模型的基本信息包括基本属性（模型英文名、模型中文名、模型分类、描述）和模型属性



## 查看元模型的元数据

支持查看使用该元模型采集到的全部元数据信息。点击操作下的【查看元数据】按钮，在新页面查看系统采集到的元数据。此部分展示的元数据与【元数据管理】模块下展示的元数据区别在于，此部分展示的是通过按照元模型字段采集到的元数据，而元数据管理模块是以树结构展示的 JDBC 类型元数据，且不包括非结构化元数据。

系统不支持对以采集方式获取的元数据进行编辑和删除，但支持对以手动新增和导入方式建立的元数据的编辑和删除。



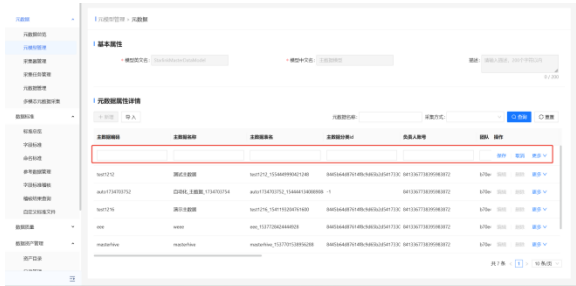
支持查看采集到的元数据的版本。点击【更多】->【版本管理】，在右侧弹窗中查看元数据的版本记录。



在查看元数据界面，支持手动新增元数据、导入元数据、查询元数据；手动添加和导入的元数据，不会被自动采集的元数据覆盖

操作	描述
新增	除了以采集的方式获取元数据来描述系统内的数据，还支持以自定义元数据的形式来丰富对系统内数据的描述。点击【+新增】按钮，手动填写元数

**据信息**



保存后，系统支持对手动新增元数据的编辑和删除

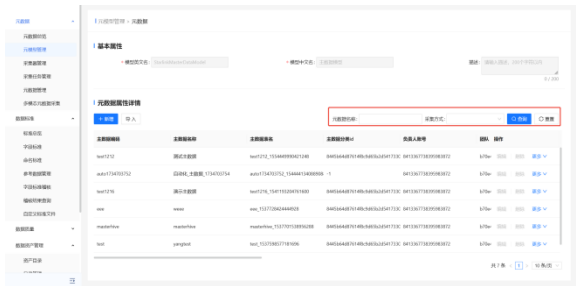
**导入**

支持以上传文件的形式导入元数据。点击【导入】按钮，根据弹窗提示下载模板，将填写好的模板文件上传至系统，完成元数据的导入



**查询**

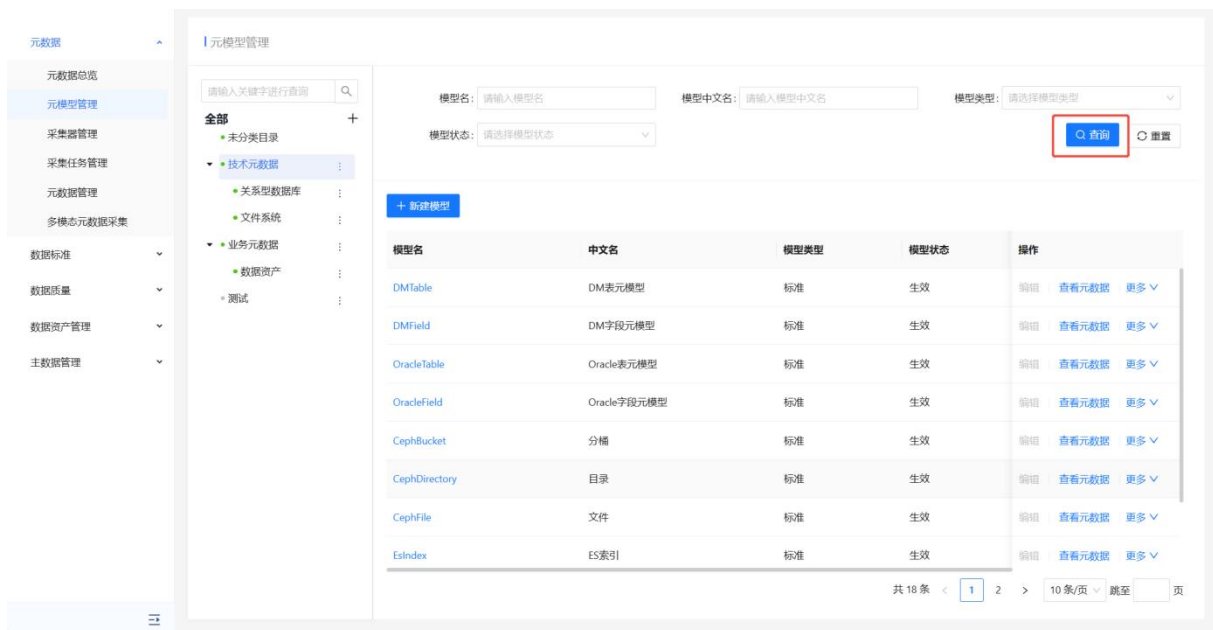
支持以填写元数据名称和选择采集方式的方式查询元数据



## 查询模型

面对海量的元模型，需要以简单直接的方式找到所需元模型。支持通过输入模型名、输入模型中文名、选择模型类型及选择模型状态的方式查询模型。

支持通过输入模型名、输入模型中文名、选择模型类型及选择模型状态的方式查询模型。点击【查询】按钮进行查询。



## 模型操作

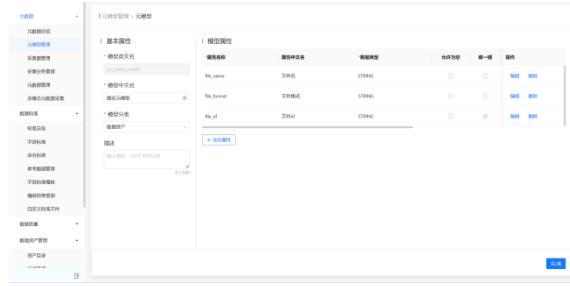
为应对查看元模型具体信息、对元模型进行维护等需求，需要对元模型进行操作。

支持对现有模型进行操作，操作行为包括编辑、查看元数据及更多（下线、删除、复制、变更及版本管理）。

操作	描述
编辑	<p>对现有元模型进行编辑，该功能仅对自定义模型生效。点击【编辑】按钮，进入元模型编辑界面，编辑内容同新建元模型</p> 
删除	<p>删除所选模型，该功能仅对自定义模型生效。点击【更多】-&gt;【删除】，在弹出窗口中确认是否删除模型，点击【确定】按钮完成删除</p> 

变更

变更所选元模型，该功能仅对自定义模型生效。点击【更多】->【变更】按钮，在新页面中修改元模型信息



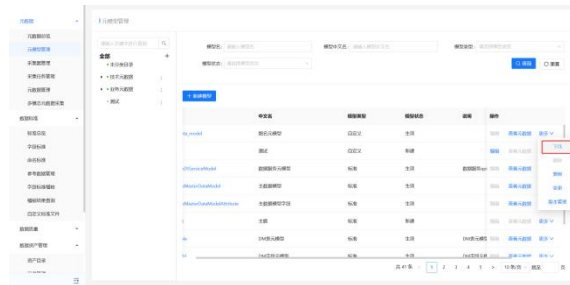
上线

将自定义的元模型上线，该功能仅对自定义模型生效。点击【更多】->【上线】按钮，完成元模型的上线



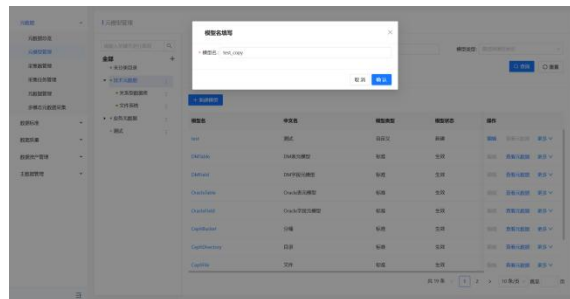
下线

将自定义的元模型下线，该功能仅对自定义模型生效。点击【更多】->【下线】按钮，完成元模型的下线



复制

复制所选模型。点击【复制】按钮，在弹窗中填写复制模型的模型名，点击【确定】完成模型复制



版本管理

查看所选模型的历史版本。点击【更多】->【版本管理】按钮，在右侧抽屉中查看模型的历史版本记录，版本记录包括历史版本及版本对比

版本记录			
版本	变更类型	变更人	变更时间
V3	新增	体验用户-互联网	2025-01-07 15:04:58
V2	修改	体验用户-互联网	2025-01-07 15:02:40
V1	新增	体验用户-互联网	2025-01-07 15:02:09

共 3 条 < 1 > 10 条/页

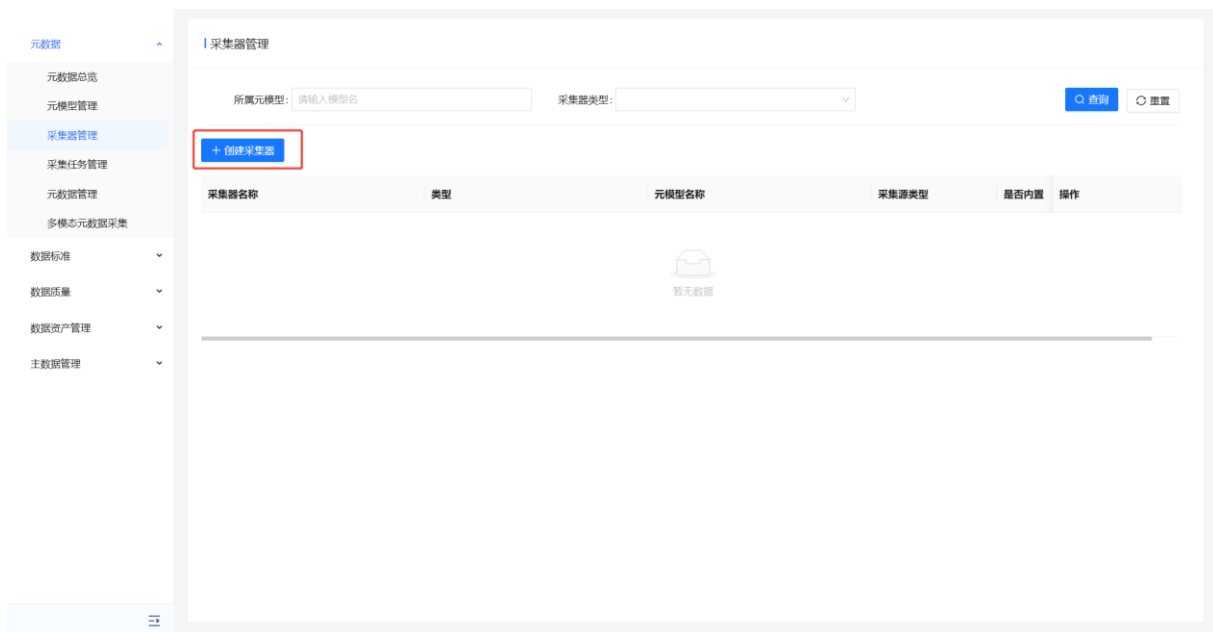
## 采集器管理

元数据采集器描述了描述**从哪里采集数据、怎样采数据、采什么数据**，通过配置采集任务，能够使采集器按一定时间规则从各种数据源中采集元数据。

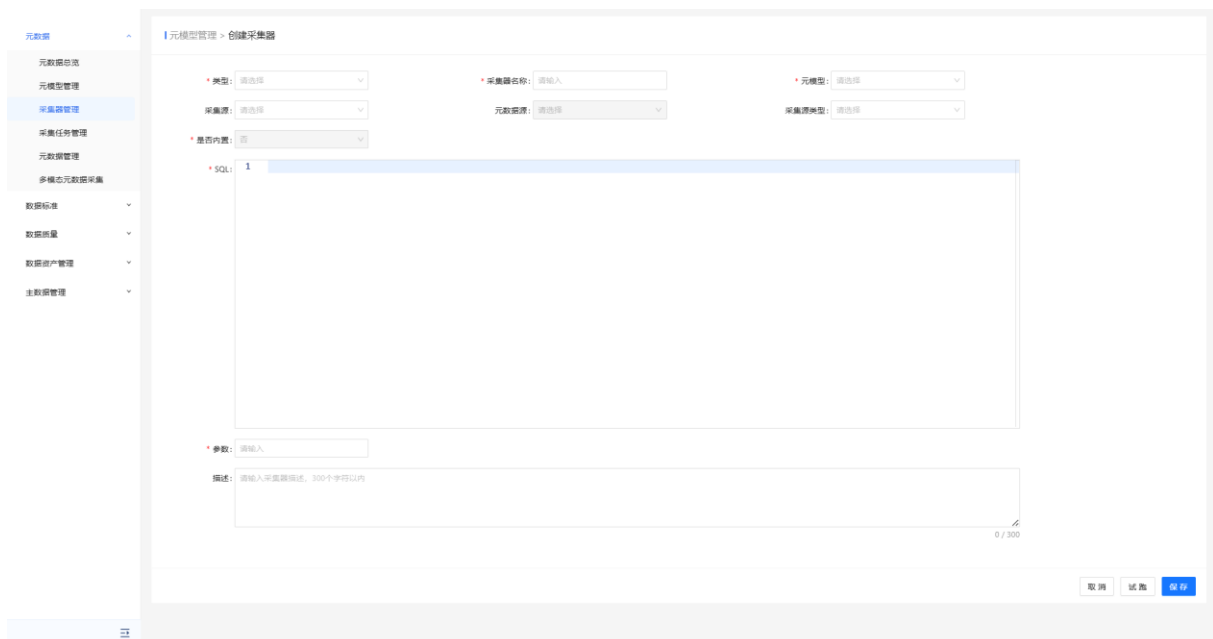
在采集器管理模块配置采集器的基本信息，包括：采集器类型、元模型、采集源、数据源、SQL 等信息，在配置采集器时，直接将采集器与元模型做绑定，后续配置采集任务时无需重新绑定元模型；

采集器分为系统**内置采集器**和**自定义采集器**。系统内置了常用数据源的元数据采集器，按照绑定的元模型批量采集数据源的表和字段，内置采集器不可修改；同时，为应对不同的业务需求，支持自定义创建元数据采集器，灵活提取数据源下的表、字段信息。

点击【+创建采集器】按钮创建新的采集器。



在创建采集器窗口中配置采集器相关信息，在此处**绑定元模型**，通过 SQL 代码描述采集器的功能，如采集哪些库表、哪些字段等。支持试跑查看采集器配置是否合理，点击【保存】按钮完成采集器的创建。



用户可以选择内置采集器，无需手动添加采集器。系统内置的采集器包括：

系统版本/数据库	Flink	Local
MySQL	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Doris	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Oracle	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Hive	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
DM	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
PgSQL	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
GaussDB	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
SqlServer	<input checked="" type="checkbox"/>	<input type="checkbox"/>
DB2	<input checked="" type="checkbox"/>	<input type="checkbox"/>
KingBase	<input checked="" type="checkbox"/>	<input type="checkbox"/>
OceanBase	<input checked="" type="checkbox"/>	<input type="checkbox"/>
ClickHouse	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Vertica	<input checked="" type="checkbox"/>	<input type="checkbox"/>
HanaDB	<input checked="" type="checkbox"/>	<input type="checkbox"/>
HBase	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Elasticsearch	<input type="checkbox"/>	<input checked="" type="checkbox"/>
MongoDB	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Ceph	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Cassandra	<input checked="" type="checkbox"/>	<input type="checkbox"/>

## 元数据采集

配置好采集器之后，需要在采集任务管理模块配置**采集任务**和**采集调度**，完成配置后系统即可根据调度信息按照预定的任务设置和时间规则进行元数据采集。

采集任务通过采集适配器连接到指定的数据源，执行采集任务。支持用户创建内置采集任务和自定义采集任务。内置采集任务通过选择数据源和环境，自动关联到数据源对应的内置元模型采集器，无需手动选择元模型采集器；自定义采集任务为特殊采集需求提供支撑，需要手动选择元模型采集器，采集与需求对应的元数据。

配置完采集任务后，需要配置调度，以按照预定的时间规则进行元数据采集。可以进行定时配置，实现元数据定时采集；也可以选择立即执行一次采集任务。目前支持 `mysql`、`hive`、`doris`、`DM` 等类型元数据采集。

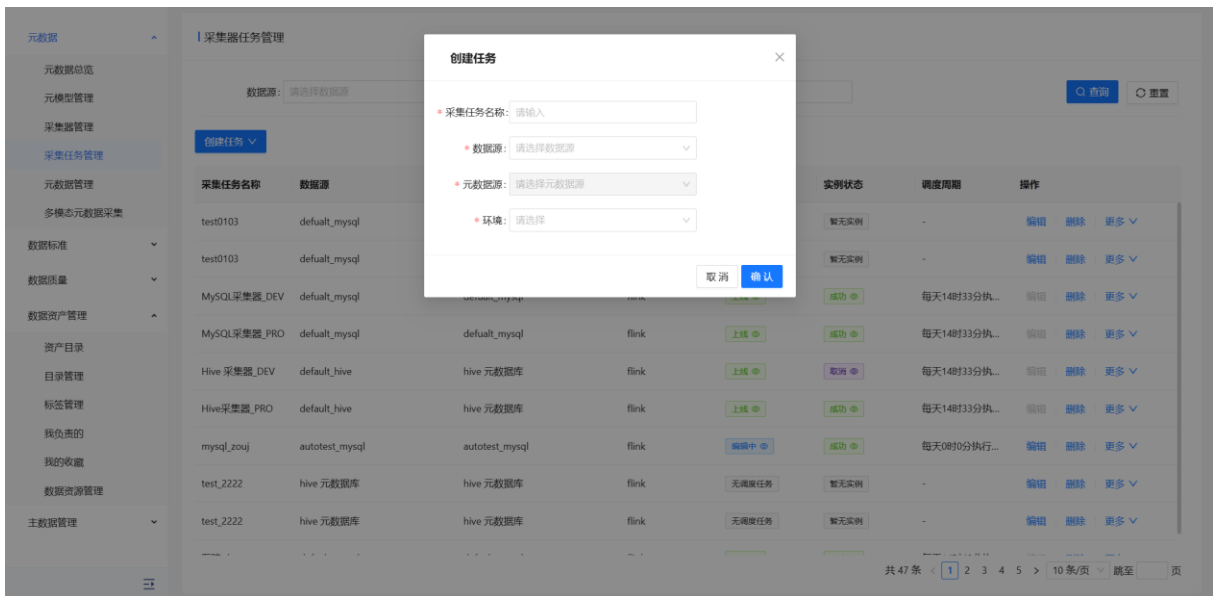
## 创建采集任务

配置采集任务信息，连接到相应数据源。

点击【+创建任务】按钮创建新的采集任务，可以选择创建内置采集器任务和自定义采集器任务。

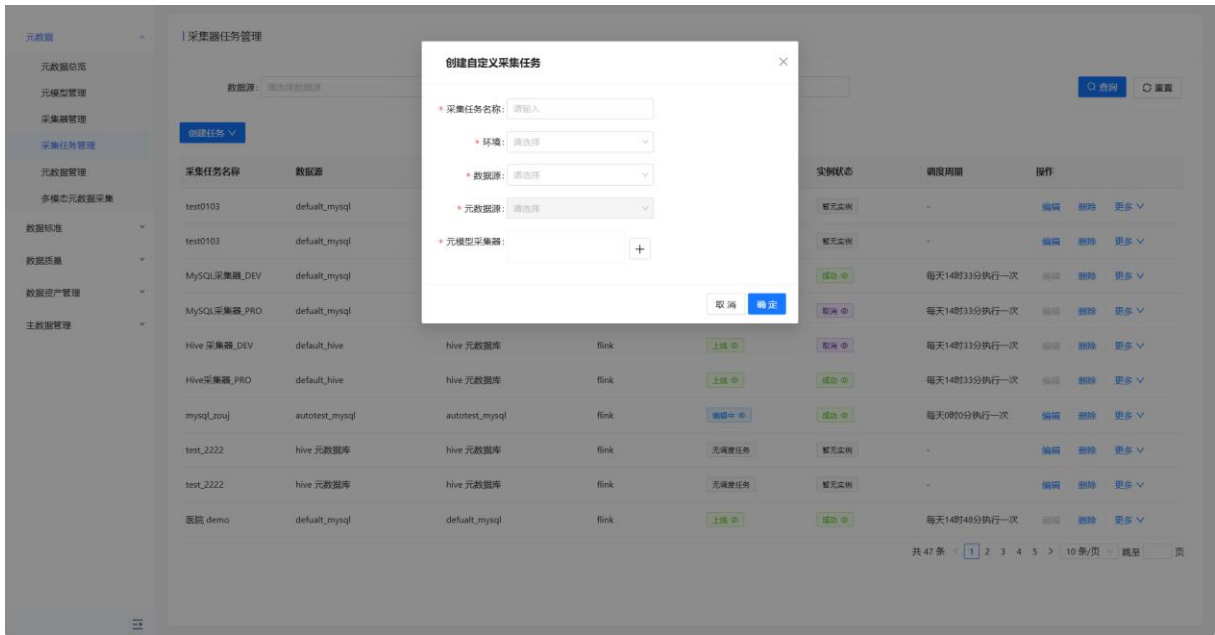
### ● 内置采集器任务

一般情况下，采集任务均使用内置采集器。用户仅需选择想要采集的数据源和环境（如果采集 hive 数据源，需要配置元数据源），数据源已与内置采集器匹配好，采集器同时采集所选数据源的表和字段。



### ● 自定义采集器任务

为应对特殊情况，如元数据中不包含的字段、查询表占用的空间、数据库版本迭代导致采集任务失败等特殊采集需求，可以通过自定义采集器任务解决。在弹窗中配置采集任务相关信息（如果采集 hive 数据源，需要配置元数据源），创建自定义采集任务时需要手动选择元模型采集器。



## 创建调度配置

配置采集任务的调度信息，系统将根据调度信息执行元数据的采集任务。

点击【更多】->【创建调度配置】按钮，在右侧抽屉中配置调度信息，包括基本信息、调度信息、任务信息及运行信息，点击【保存】按钮，完成调度配置，任务将根据新配置的信息进行采集

## 调度配置



### 基本信息

任务名称: default\_mysql-元数据采集任务

\* 优先级: 中

### 调度信息

\* 执行周期: 日

执行时间: 每天 00:00

crontab表达式: 0 0 0 \* \* ?

### 任务信息

\* 任务类型: FLINK

yarn队列:

### 运行信息

\* 调度集群组: default

业务时间偏移: -1

重试次数: 0

重试时间间隔(s): 30

\* 任务超时时长(s): 7200

取消

保存

## 采集任务操作

采集器管理界面展示采集任务名称、数据源、元数据源、采集器类型、调度状态、实例状态、调度周期及操作。操作动作支持编辑、删除及更多（创建调度配置、立即执行一次、调度上线及调度下线）。

操作	描述
编辑	编辑所选采集任务的配置信息。点击【编辑】按钮，在弹窗中编辑采集任务信息。点击【确认】完成任务编辑

	
删除	<p>删除所选采集任务。点击【删除】按钮，确认是否删除采集任务。</p> 
立即执行一次	<p>立即执行当前的采集任务。点击【更多】-&gt;【立即执行一次】按钮，执行采集任务，根据弹窗提示确认任务执行成功。</p> 
调度上线和调度下线	<p>上线或下线选中的调度任务。点击【更多】-&gt;【调度上线】按钮或【调度下线】按钮，对任务执行上线和下线操作。</p>

## 查询采集器任务

支持以输入数据源、选择任务名称的方式进行采集器任务查询。

采集任务名称	数据源	元数据源	采集器类型	调度状态	实例状态	调度周期	操作
test0103	default_mysql	default_mysql	flink	无调度任务	暂无实例	-	编辑 删除 更多
test0103	default_mysql	default_mysql	flink	无调度任务	暂无实例	-	编辑 删除 更多
MySQL采集器_DEV	default_mysql	default_mysql	flink	上线	成功	每天14时33分执行一次	编辑 删除 更多
MySQL采集器_PRO	default_mysql	default_mysql	flink	上线	取消	每天14时33分执行一次	编辑 删除 更多
Hive采集器_DEV	default_hive	hive 元数据库	flink	上线	取消	每天14时33分执行一次	编辑 删除 更多
Hive采集器_PRO	default_hive	hive 元数据库	flink	上线	成功	每天14时33分执行一次	编辑 删除 更多
mysql_zouj	autotest_mysql	autotest_mysql	flink	编辑	成功	每天08时0分执行一次	编辑 删除 更多
test_2222	hive 元数据库	hive 元数据库	flink	无调度任务	暂无实例	-	编辑 删除 更多
test_2222	hive 元数据库	hive 元数据库	flink	无调度任务	暂无实例	-	编辑 删除 更多
医院 demo	default_mysql	default_mysql	flink	上线	成功	每天14时48分执行一次	编辑 删除 更多

## 数据标准

数据标准模块通过建立数据标准体系，规范约束字段名称、数据类型、值域等属性信息，为目标数据制定字段标准的编码规则，来保障后续建模与应用过程中数据处理的一致性。

数据标准模块提供命名标准、参考标准管理功能，为字段标准提供参考依据；提供字段标准稽核及查询功能，把控系统数据的质量；提供自定义标准文件功能，为数据标准体系的建设提供参考依据。

数据标准的整体使用流程为：

步骤	描述
步骤一：上传标准文件（可选）	前期搜集字段标准相关标准文件，为字段标准提供参考依据；上传至平台“自定义标准文件”模块，方便查阅、共享
步骤二：新建命名标准（可选）	新建命名标准，为字段标准提供词根参考依据
步骤三：新建参考数据（可选）	新建参考数据，为字段标准提供规范的数据值域
步骤四：新建字段标准	新建字段标准，根据命名标准自动关联推荐选择规范的字段名称；如字段需要设定值域，可以选择引用参考数据
步骤五：数据建模引用字段标准（可选）	在数据建模时，可以引用已发布的数据标准
步骤六：新建字段标准稽核任务	新建标准稽核任务，配置监控对象和监控规则信息
步骤七：配置稽核任务调度	根据需求配置任务调度信息，按照配置信息定时执行稽核

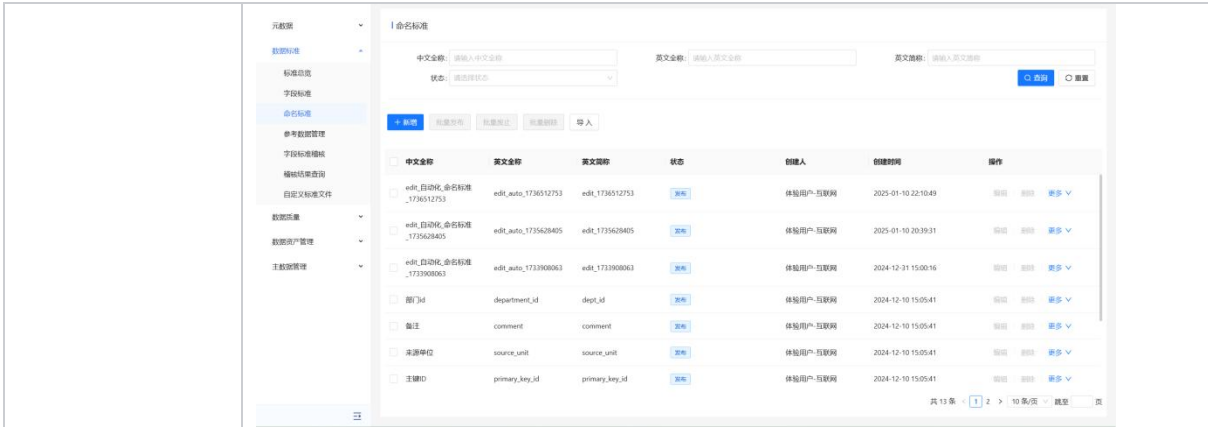
	任务
步骤八：查看稽核结果	查看字段稽核结果，如稽核未通过，查看具体有问题的字段标准及问题原因

## 标准总览

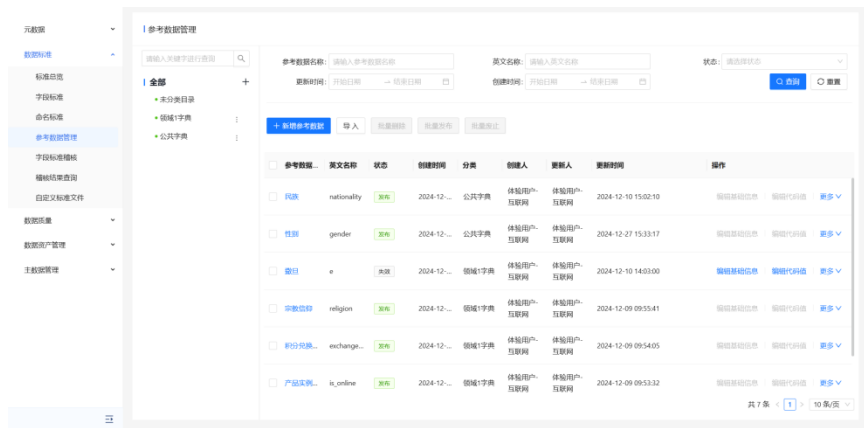
标准总览界面呈现标准全貌，集成了标准总览、标准引用情况、最热标准 top5、标准稽核情况的信息，以直观的可视化图表清晰地展示标准的具体情况，为使用者提供更加清晰的决策依据。



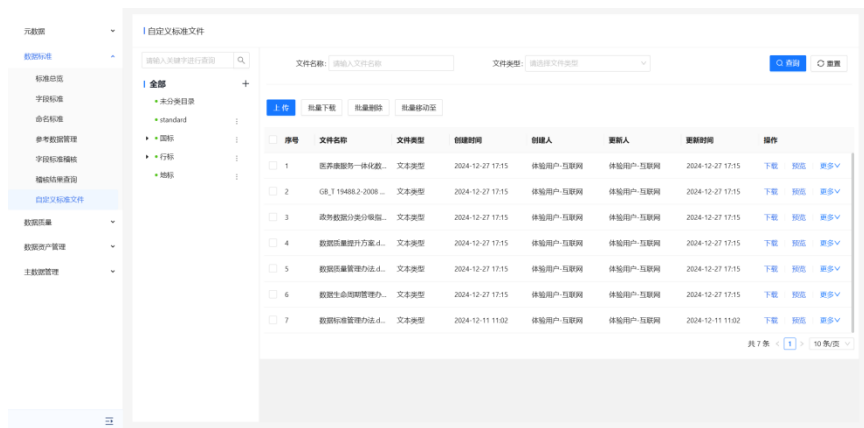
模块	描述
标准总览	<p>分为四个模块，统计并展示系统内现有已发布标准的数量，统计内容包括字段标准、命名标准、参考数据、标准文件；可以点击每个模块跳转到相应界面查看具体信息</p> <p>字段标准：</p> <p>命名标准：</p>



参考数据:



标准文件:



标准引用情况

两个部分分别展示了字段标准引用率及字段标准贯标率。字段标准引用率表示被引用的字段标准的数量占系统中所有已发布字段标准数量的比例；字段标准贯标率表示系统已发布的字段标准中被贯彻落实的标准比例

	<p><b>标准引用情况</b> <sup>Ⓞ</sup></p>  <p>字段标准引用率 <sup>Ⓞ</sup>      字段标准贯标率 <sup>Ⓞ</sup></p>										
<p>最热标准 top5</p>	<p>按照热度降序展示已发布标准中最热门五个标准</p> <p><b>最热标准top5</b> <sup>Ⓞ</sup></p>  <table border="1"> <tr><td>测试_日期(test_date)</td><td>9</td></tr> <tr><td>是否分支(is_branch)</td><td>6</td></tr> <tr><td>城市(city)</td><td>6</td></tr> <tr><td>类型_状态(type_status)</td><td>3</td></tr> <tr><td>性别(gender)</td><td>3</td></tr> </table>	测试_日期(test_date)	9	是否分支(is_branch)	6	城市(city)	6	类型_状态(type_status)	3	性别(gender)	3
测试_日期(test_date)	9										
是否分支(is_branch)	6										
城市(city)	6										
类型_状态(type_status)	3										
性别(gender)	3										
<p>标准稽核情况</p>	<p>该模块展示已配置监控任务数量、符合标准的监控数、不符合标准的监控数、符合标准的监控占比、不符合标准的监控占比</p> <p><b>标准稽核情况</b> <sup>Ⓞ</sup></p>  <table border="1"> <tr><td>已配置监控任务数量</td><td>2</td></tr> <tr><td>符合标准的监控数</td><td>1</td></tr> <tr><td>不符合标准的监控数</td><td>1</td></tr> <tr><td>符合标准的监控占比</td><td>50%</td></tr> <tr><td>不符合标准的监控占比</td><td>50%</td></tr> </table>	已配置监控任务数量	2	符合标准的监控数	1	不符合标准的监控数	1	符合标准的监控占比	50%	不符合标准的监控占比	50%
已配置监控任务数量	2										
符合标准的监控数	1										
不符合标准的监控数	1										
符合标准的监控占比	50%										
不符合标准的监控占比	50%										

## 字段标准

字段标准为字段提供标准化的约束法则，约束内容包括标准基础信息（标准中文名、标准英文名、所属目录）、属性信息（数据类型、值域）及自定义属性；在维护和新建字段标准时，均可使用命名标准提供的词根自动推荐字段标准的名称；规定字段标准的值域时，可以引用参考数据，为字段标准提供具体的值域范围。

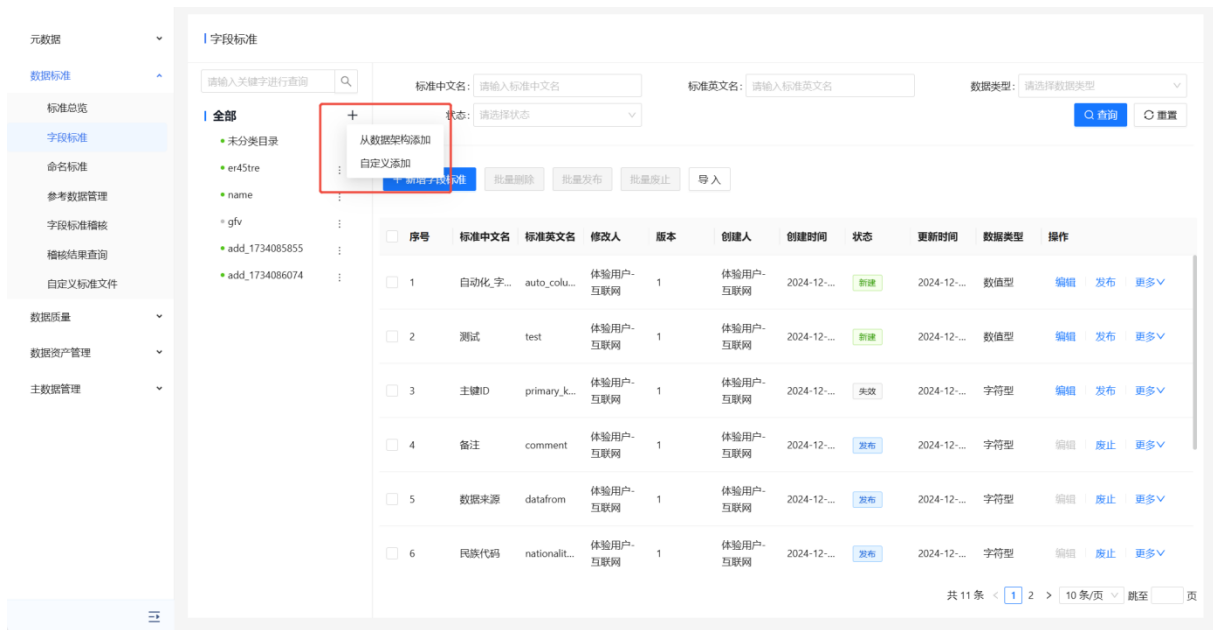
字段标准模块支持维护字段标准目录、新增字段标准、导入字段标准、查询字段标准、操作字段标准。

### 字段标准目录

左侧目录区域默认显示全部目录，可以检索目录或管理目录。

#### ● 新建目录

支持标准目录新建，点击目录区域的加号【+】，可以选择从数据架构添加目录或自定义添加目录。

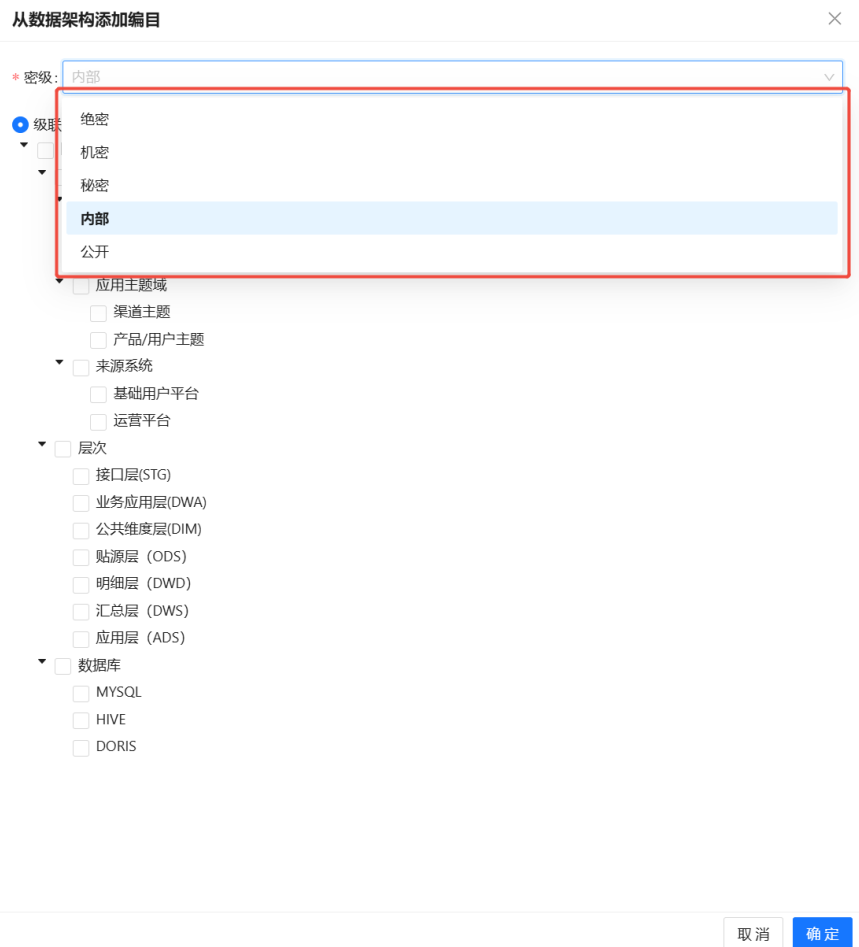


## 新建目录方式

## 方法

选择【从数据架构添加】，需要选择目录密级，目录密级分为：绝密、机密、秘密、内部及公开五个级别；

## 从数据架构添加



选择架构支持【级联选择】和【单独选择】两种方式。

【级联选择】可以实现批量选择该架构层级及下级层次的所有架构。

从数据架构添加编目 ×

\* 密级: 内部 ▾

级联选择  单独选择

- 团队架构
  - 主题
    - 业务主题域
      - 产品域
      - 公共域
    - 应用主题域
      - 渠道主题
      - 产品/用户主题
    - 来源系统
      - 基础用户平台
      - 运营平台
  - 层次
    - 接口层(STG)
    - 业务应用层(DWA)
    - 公共维度层(DIM)
    - 贴源层 (ODS)
    - 明细层 (DWD)
    - 汇总层 (DWS)
    - 应用层 (ADS)
  - 数据库
    - MYSQL
    - HIVE
    - DORIS

取消 确定

【单独选择】可以实现单独选择某层级架构。

	<div data-bbox="502 181 1367 1131"> <p><b>从数据架构添加编目</b> <span style="float: right;">×</span></p> <p>* 密级: 内部 <span style="float: right;">▼</span></p> <p><input type="radio"/> 级联选择 <input checked="" type="radio"/> 单独选择</p> <ul style="list-style-type: none"> <li>▼ <input type="checkbox"/> 团队架构 <ul style="list-style-type: none"> <li>▼ <input type="checkbox"/> 主题 <ul style="list-style-type: none"> <li>▼ <input type="checkbox"/> 业务主题域 <ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> 产品域</li> <li><input type="checkbox"/> 公共域</li> </ul> </li> <li>▼ <input type="checkbox"/> 应用主题域 <ul style="list-style-type: none"> <li><input type="checkbox"/> 渠道主题</li> <li><input type="checkbox"/> 产品/用户主题</li> </ul> </li> <li>▼ <input type="checkbox"/> 来源系统 <ul style="list-style-type: none"> <li><input type="checkbox"/> 基础用户平台</li> <li><input type="checkbox"/> 运营平台</li> </ul> </li> <li>▼ <input type="checkbox"/> 层次 <ul style="list-style-type: none"> <li><input type="checkbox"/> 接口层(STG)</li> <li><input type="checkbox"/> 业务应用层(DWA)</li> <li><input type="checkbox"/> 公共维度层(DIM)</li> <li><input type="checkbox"/> 贴源层 (ODS)</li> <li><input type="checkbox"/> 明细层 (DWD)</li> <li><input type="checkbox"/> 汇总层 (DWS)</li> <li><input type="checkbox"/> 应用层 (ADS)</li> </ul> </li> <li>▼ <input type="checkbox"/> 数据库 <ul style="list-style-type: none"> <li><input type="checkbox"/> MYSQL</li> <li><input type="checkbox"/> HIVE</li> <li><input type="checkbox"/> DORIS</li> </ul> </li> </ul> </li> </ul> </li> </ul> <p style="text-align: right;"><input type="button" value="取消"/> <input type="button" value="确定"/></p> </div> <p>点击【确定】完成目录的创建。</p>
自定义添加	<p>选择【自定义添加】，填写基础信息和自定义属性。</p> <p>基础信息部分需要填写目录名称（必填）、密级（必填）及描述；</p>

### 注册编目 ✕

**基础信息**

\* 目录名称:

\* 密级:

描述:  0 / 200

**自定义属性**

+ 新增自定义属性

自定义属性为选填部分，可以通过点击【新增自定义属性】按钮实现自定义属性的新增和删除，支持新增多条自定义属性。

注册编目 ×

**基础信息**

\* 目录名称:

\* 密级:

描述:  0 / 200

**自定义属性**

+ 新增自定义属性

自定义属性  🗑

自定义属性  🗑

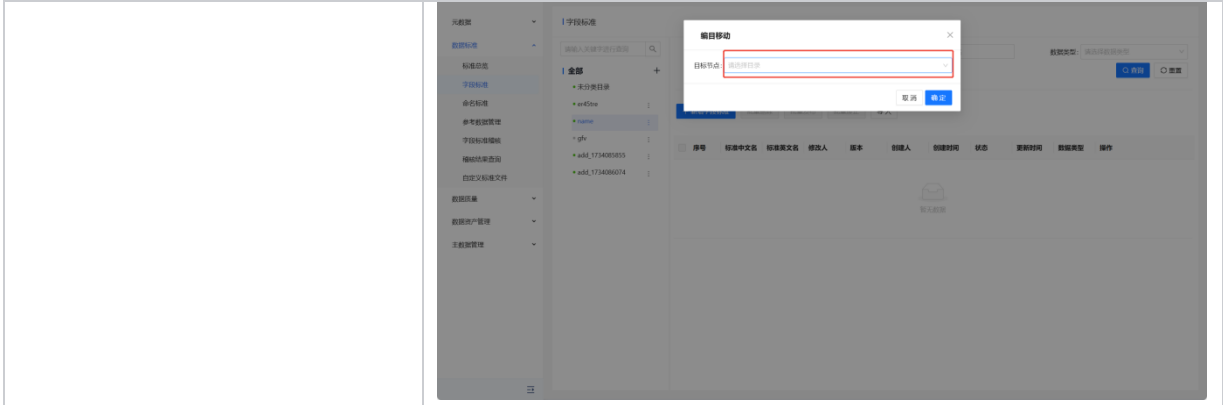
取消
确定

点击【确定】完成目录的创建。

● 目录操作

针对现有目录，点击目录名右侧的按钮，可以实现对目录的管理和操作。

操作	描述
添加同级目录和添加子级目录	<p>为选中目录添加同级或子级目录，支持【从数据架构添加】和【自定义添加】</p> 
移动	<p>移动所选目录至目标节点。点击【移动】选项，在弹窗的下拉菜单中选择想要将目录移动至的目标节点，点击【确定】，完成对目录的移动</p>



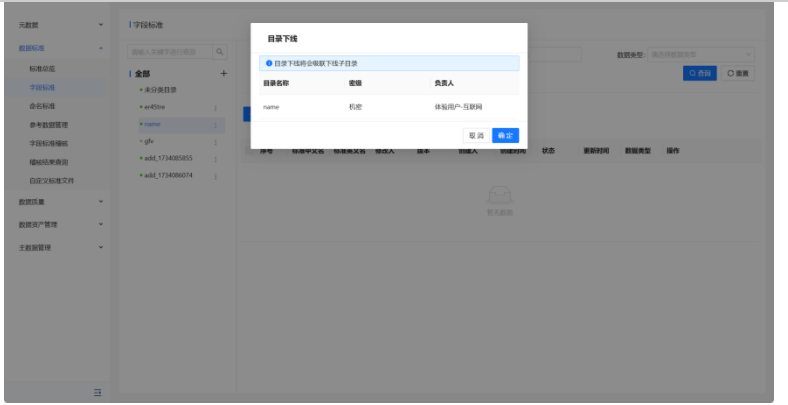
编辑

对所选目录进行编辑。点击【编辑】选项，在右侧弹窗中编辑目录信息，更改目录基础信息和自定义属性。点击【确定】，完成对目录的编辑



下线

将所选目录下线处理。点击【下线】选项，弹窗中展示已选择的目录信息，点击【确定】，完成目录的下线



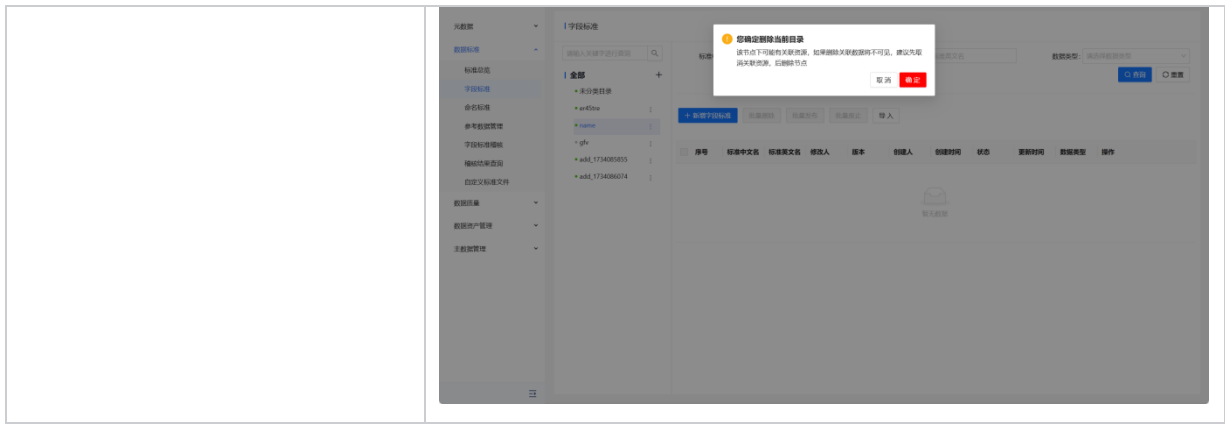
授权

为团队或个人授予所选目录的管理和编辑权限。点击【授权】选项，在右侧抽屉展示目录的基本信息以及权限管理内容。在权限管理栏内，可以选择对【团队】权限进行管理和对【个人】权限进行管理，通过勾选的方式选择是否授予【管理】和【编辑】的权限。点击【确定】，完成目录的授权



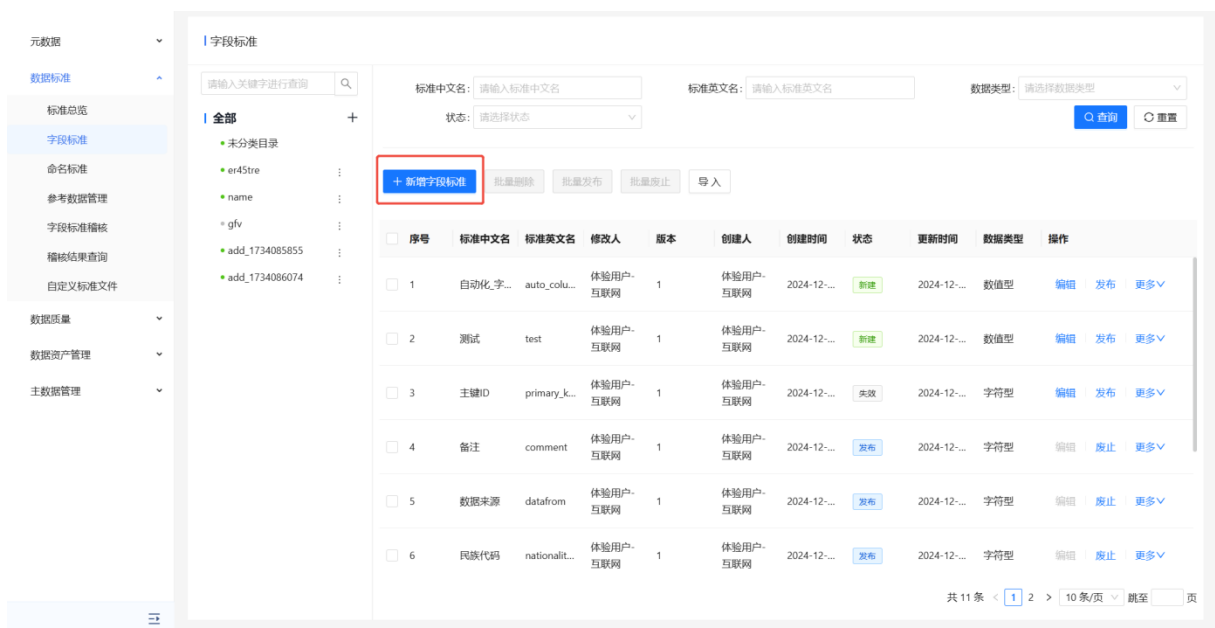
删除

删除所选目录。点击【删除】选项，根据弹窗提示选择是否确定删除当前目录，点击【确定】，完成目录的删除



## 新增字段标准

点击【+新增字段标准】按钮，进入新增页面，填写字段标准信息。



### 1、标准基础信息

填写标准基础信息，包括标准中文名（必填）、标准英文名（必填）、所属目录（必填）及描述。

元数据

数据标准

标准总览

字段标准

命名标准

参考数据管理

字段标准稽核

稽核结果查询

自定义标准文件

数据质量

数据资产管理

主数据管理

字段标准 > 新增字段标准

**标准基础信息**

\* 标准中文名:

\* 标准英文名:

\* 所属目录:

描述: 描述信息200个字符以内

0 / 200

**属性信息**

\* 数据类型:

\* 值域:  自定义  区间范围  无

**自定义属性**

字段名称	字段值	操作
+ 添加		

取消 确认

其中，关联字段所属目录有两种方式，一种通过在标准基础信息的所属目录下拉菜单中选择所属目录

元数据

数据标准

标准总览

字段标准

命名标准

参考数据管理

字段标准稽核

稽核结果查询

自定义标准文件

数据质量

数据资产管理

主数据管理

字段标准 > 新增字段标准

**标准基础信息**

\* 标准中文名:

\* 标准英文名:

\* 所属目录:

描述:

**属性信息**

\* 数据类型:

\* 值域:  自定义  区间范围  无

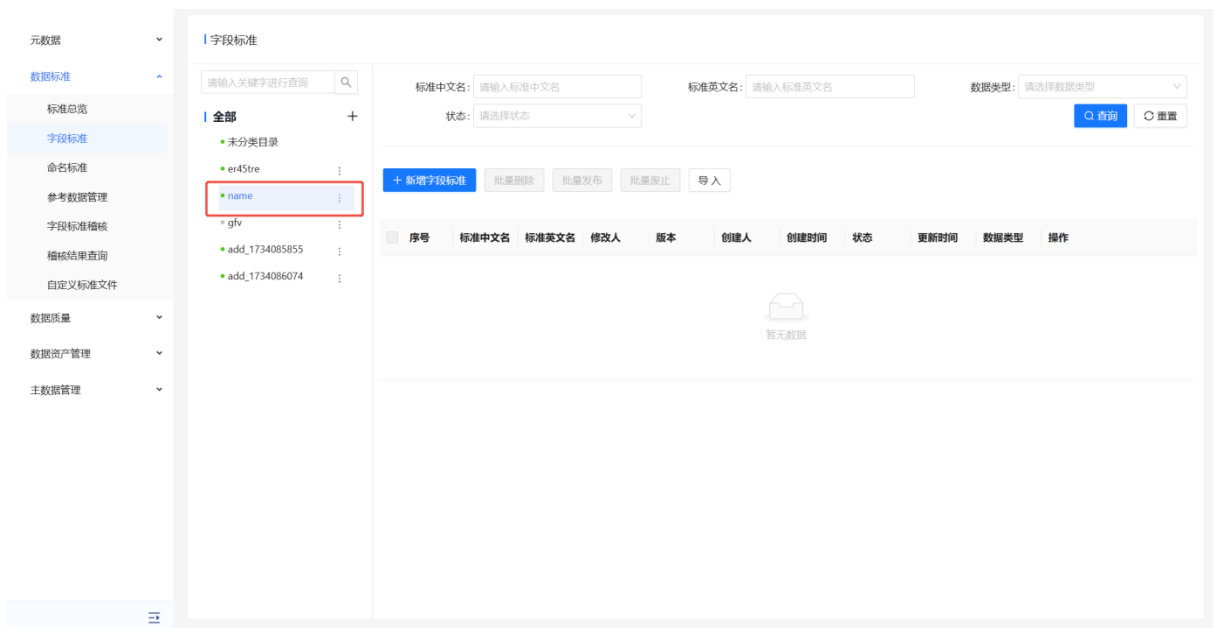
**自定义属性**

字段名称	字段值	操作
+ 添加		

er45tre  
name  
add\_1734085855  
add\_1734086074  
未分类目录

取消 确认

另一种在字段标准界面首先选中目标目录，再点击【+新增字段标准】按钮，则进入新增页面后，在所属目录位置上可以自动关联已选目录。



## 2、属性信息

属性信息部分包括数据类型（必填）和值域（必填）。

元数据

数据标准

- 标准总览
- 字段标准
- 命名标准
- 参考数据管理
- 字段标准稽核
- 稽核结果查询
- 自定义标准文件

数据质量

数据资产管理

主数据管理

字段标准 > 新增字段标准

**标准基础信息**

\* 标准中文名:

\* 标准英文名:

\* 所属目录: name

描述: 描述信息200个字符以内

0 / 200

**属性信息**

\* 数据类型:

\* 值域:  自定义  区间范围  无

**自定义属性**

字段名称	字段值	操作
+		

取消 确认

数据类型可以通过下拉菜单选择，可选择的数据类型包括数值型、字符型、日期型、日期时间型及布尔型。

元数据

数据标准

- 标准总览
- 字段标准
- 命名标准
- 参考数据管理
- 字段标准稽核
- 稽核结果查询
- 自定义标准文件

数据质量

数据资产管理

主数据管理

字段标准 > 新增字段标准

**标准基础信息**

\* 标准中文名:

\* 标准英文名:

\* 所属目录: name

描述: 描述信息200个字符以内

0 / 200

**属性信息**

\* 数据类型:

\* 值域:  自定义  区间范围  无

**自定义属性**

数值型

字符型

日期型

日期时间型

布尔型

取消 确认

### ● 数值型

若数据类型为数值型，则需要填写该字段的长度信息，精度信息为选填，值域信息包括自定义、引用参考数据、区间范围及无四种。

其中，自定义值域的方式为手动填写

元数据

数据标准

标准总览

字段标准

命名标准

参考数据管理

字段标准稽核

稽核结果查询

自定义标准文件

数据质量

数据资产管理

主数据管理

字段标准 > 新增字段标准

\* 标准英文名:

\* 所属目录:

描述: 描述信息200个字符以内

0 / 200

属性信息

\* 数据类型: 数值型

\* 长度:

精度:

\* 值域:  自定义  引用参考数据  区间范围  无

请输入自定义数值用逗号分隔

自定义属性

字段名称	字段值	操作
------	-----	----

取消 确认

引用参考数据的方式需点击下拉菜单，在右侧弹窗中选择想要引用的参考数据。支持选择参考数据的分类，支持参考数据的查询两种方式检索参考数据。点击【确认】完成参考数据的引用。

### 选择引用的参考数据



选择分类:

名称:

数据字典名称	英文简称	状态	创建人	创建时间	更新人	页
<input type="radio"/> 民族	nationality	<input type="button" value="发布"/>	体验用户-互联网	2024-12-10 15:02:10	体验用户-互联网	2 1
<input type="radio"/> 性别	gender	<input type="button" value="发布"/>	体验用户-互联网	2024-12-10 14:56:52	体验用户-互联网	2 1
<input type="radio"/> 撒旦	e	<input type="button" value="发布"/>	体验用户-互联网	2024-12-10 14:03:00	体验用户-互联网	2 1
<input type="radio"/> 宗教信仰	religion	<input type="button" value="发布"/>	体验用户-互联网	2024-12-09 09:55:41	体验用户-互联网	2 0
<input type="radio"/> 积分兑换渠道类型	exchange_channel_type	<input type="button" value="发布"/>	体验用户-互联网	2024-12-09 09:54:05	体验用户-互联网	2 0
<input type="radio"/> 产品实例状态	is_online	<input type="button" value="发布"/>	体验用户-互联网	2024-12-09 09:53:32	体验用户-互联网	2 0
<input type="radio"/> 学号	stu_nu	<input type="button" value="发布"/>	体验用户-互联网	2024-12-06 18:18:13	体验用户-互联网	2 1

<  >

区间范围的方式为手动填写值域区间

元数据

数据标准

- 标准总览
- 字段标准
- 命名标准
- 参考数据管理
- 字段标准稽核
- 稽核结果查询
- 自定义标准文件

数据质量

数据资产管理

主数据管理

字段标准 > 新增字段标准

**标准基础信息**

- \* 标准中文名:
- \* 标准英文名:
- \* 所属目录:
- 描述:  描述信息200个字符以内

**属性信息**

- \* 数据类型:
- \* 长度:
- 精度:
- \* 值域:  自定义  引用参考数据  区间范围  无
- ~

**自定义属性**

取消 确认

若字段标准无值域，则不显示具体值域信息。

### ● 字符型

若数据类型为数值型，则需要填写该字段的长度信息，值域信息包括自定义、引用参考数据及无三种。

元数据

数据标准

- 标准总览
- 字段标准
- 命名标准
- 参考数据管理
- 字段标准稽核
- 稽核结果查询
- 自定义标准文件

数据质量

数据资产管理

主数据管理

字段标准 > 新增字段标准

**标准基础信息**

- \* 标准中文名:
- \* 标准英文名:
- \* 所属目录:
- 描述:  描述信息200个字符以内

**属性信息**

- \* 数据类型:
- \* 长度:
- \* 值域:  自定义  引用参考数据  无
- 

**自定义属性**

字段名称	字段值	操作
------	-----	----

取消 确认

### ● 日期型、日期时间型

若数据类型为日期型和日期时间型，则需要选择该字段的格式，格式包括 YYYY-MM-DD、YYYYMMDD、YYYY/MM/DD、YYYYMM、YYYY-MM、YYYY/MM、YYYY、MM、DD 几种，值域信息包括自定义、区间范围及无三种。

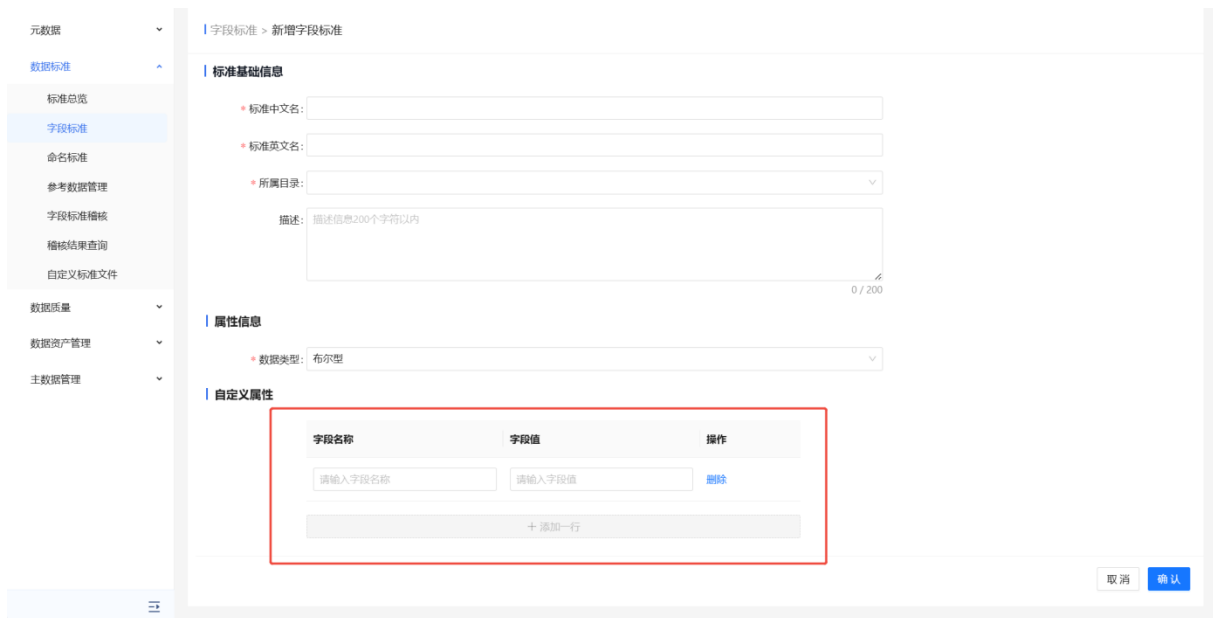


## ● 布尔型

若数据类型为布尔型，则仅需要选择数据类型。

## 3、自定义属性

支持自定义属性，点击【+添加一行】按钮添加字段名称和字段值。支持删除操作。



## 4、完成新增

点击【确认】按钮完成字段标准的新增。

### 6.3.2.3 导入字段标准

支持导入字段标准，点击【导入】按钮，右侧弹出抽屉页面。

元数据

数据标准

标准总览

字段标准

命名标准

参考数据管理

字段标准稽核

稽核结果查询

自定义标准文件

数据质量

数据资产管理

主数据管理

### 字段标准

请输入关键字进行查询

标准中文名: 请输入标准中文名

标准英文名: 请输入标准英文名

数据类型: 请选择数据类型

状态: 请选择状态

新增字段标准 批量删除 批量发布 批量废止 导入

序号	标准中文名	标准英文名	修改人	版本	创建人	创建时间	状态	更新时间	数据类型	操作
1	自动化字...	auto_colu...	体验用户-互联网	1	体验用户-互联网	2024-12-...	新建	2024-12-...	数值型	编辑 发布 更多
2	测试	test	体验用户-互联网	1	体验用户-互联网	2024-12-...	新建	2024-12-...	数值型	编辑 发布 更多
3	主键ID	primary_k...	体验用户-互联网	1	体验用户-互联网	2024-12-...	失效	2024-12-...	字符型	编辑 发布 更多
4	备注	comment	体验用户-互联网	1	体验用户-互联网	2024-12-...	发布	2024-12-...	字符型	编辑 废止 更多
5	数据来源	datafrom	体验用户-互联网	1	体验用户-互联网	2024-12-...	发布	2024-12-...	字符型	编辑 废止 更多
6	民族代码	nationalit...	体验用户-互联网	1	体验用户-互联网	2024-12-...	发布	2024-12-...	字符型	编辑 废止 更多

共 11 条 < 1 2 > 10 条/页 跳至 页

编辑导入字段标准，点击下拉菜单选择导入目录，与新增字段标准相同，同样支持两种目录选择方式，在此不再赘述。

## 导入字段标准



\* 目录选择:

\* 数据导入:

↓  
点击或拖动文件上传

还没有模版? 点击 [下载模版download](#)

注意事项:

- (1)仅支持上传“.xls”“.xlsx”格式文件，文件大小不超过10MB
- (2)请按照模板表头中列出的必填限制填写
- (3)导入仅支持新增
- (4)在导入数据时，如果存在重复的字典名称，则无法将其导入

确定

取消

数据导入支持下载模板，点击【[下载模板 download](#)】按钮下载参考模板，按照模板字段要求填充，点击或拖动文件至上传框进行文件上传，导入需校验导入内容是否符合格式，平台解析完成后实现字段标准导入。

**导入字段标准**

\* 目录选择:

\* 数据导入:

还没有模版? 点击 [下载模版download](#)

**下载**

字段标准导入模板 字段标准.xlsx  
[打开文件](#)

[查看更多](#)

点击或拖动文件上传

**注意事项:**

- (1)仅支持上传“.xls”“.xlsx”格式文件，文件大小不超过10MB
- (2)请按照模板表头中列出的必填限制填写
- (3)导入仅支持新增
- (4)在导入数据时，如果存在重复的字典名称，则无法将其导入

**确定**

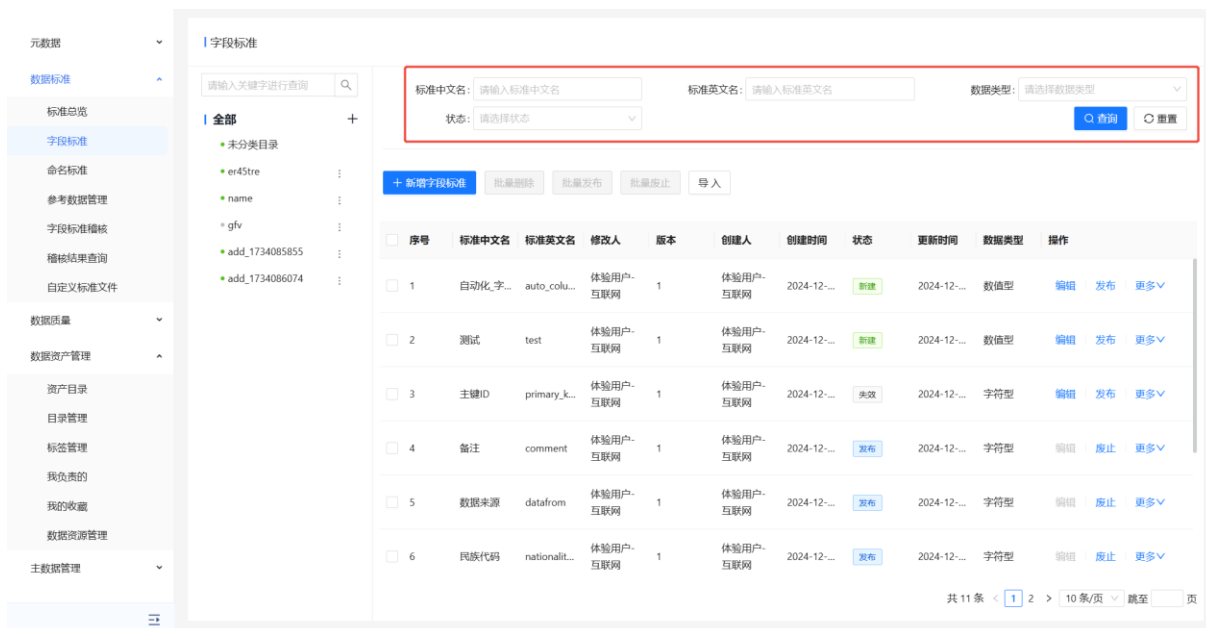
取消

点击【确定】按钮完成字段标准的导入。

### 查询字段标准

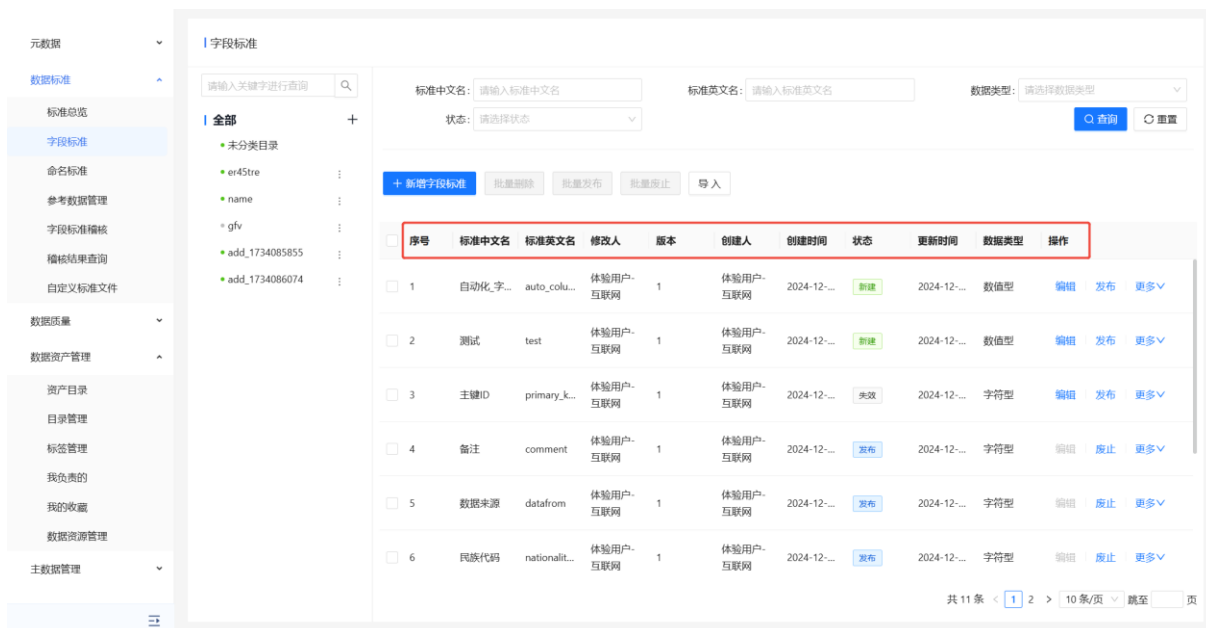
支持通过输入标准中文名、输入标准英文名、选择数据类型及选择状态的方式查询字段标准。

点击【查询】按钮进行查询。



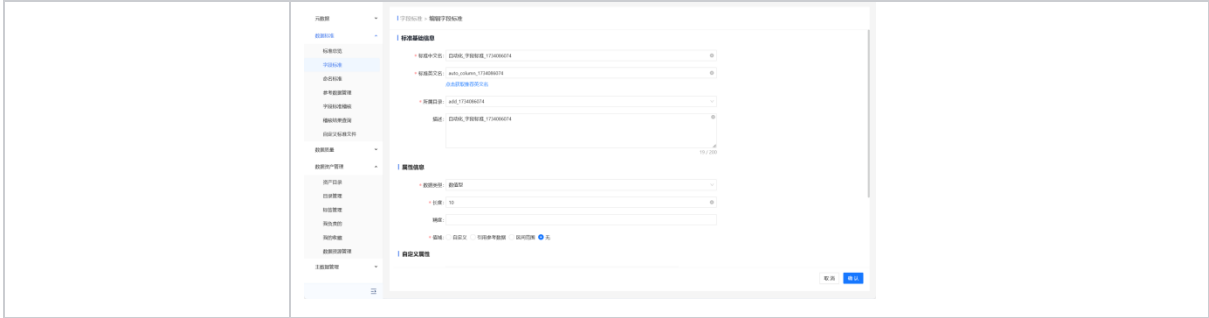
## 字段标准操作

页面默认展示全部字段标准，包括序号、标准中文名、标准英文名、修改人、版本、创建人、创建时间、状态、更新时间、数据类型及操作。其中字段标准状态分为新建、发布及失效三种，字段标准废止后，将失效，失效状态下可以进行编辑再重新发布。



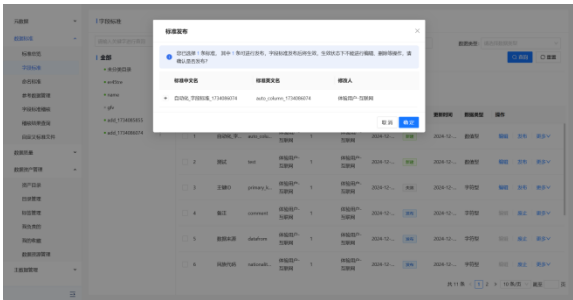
支持对字段标准进行操作，操作行为包括编辑、发布及更多（版本管理、查看详情和删除）。

操作	描述
编辑	编辑所选字段标准。点击【编辑】按钮，进入字段标准编辑界面，编辑内容同新增字段标准。点击【确认】完成字段编辑



将所选字段标准发布上线，即时生效，支持对该字段标准的引用。点击【发布】按钮，在弹窗界面确认选择发布的字段标准。点击【确认】完成字段发布

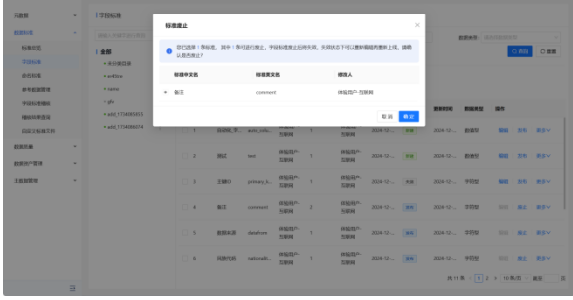
发布



注意，字段标准经发布后，无法进行编辑。如需编辑已发布的字段标准，需先废止该字段标准，再进行编辑，编辑后重新发布

废止所选字段标准，不再支持对该标准的引用。点击【废止】按钮，在弹窗中确认是否废止标准，点击【确定】完成标准废止

废止



版本管理

查看所选标准的版本变更历史。点击【更多】->【版本管理】按钮，在右侧抽屉中查看字段标准的版本记录，版本记录包括历史版本及版本对

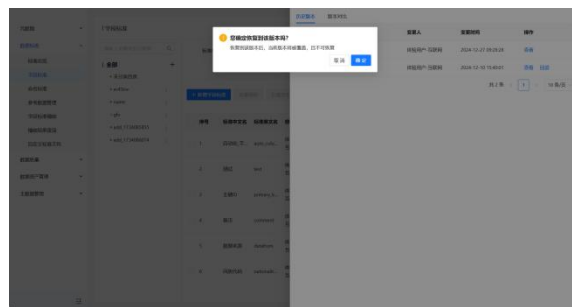
比。

版本	变更类型	变更人	变更时间	操作
V2	发布	体验用户-互联网	2024-12-27 09:28:28	<a href="#">查看</a>
V1	发布	体验用户-互联网	2024-12-10 15:40:01	<a href="#">查看</a> <a href="#">回退</a>

历史版本记录了该字段标准所有的变更信息，点击【查看】，进入该版本详细信息界面

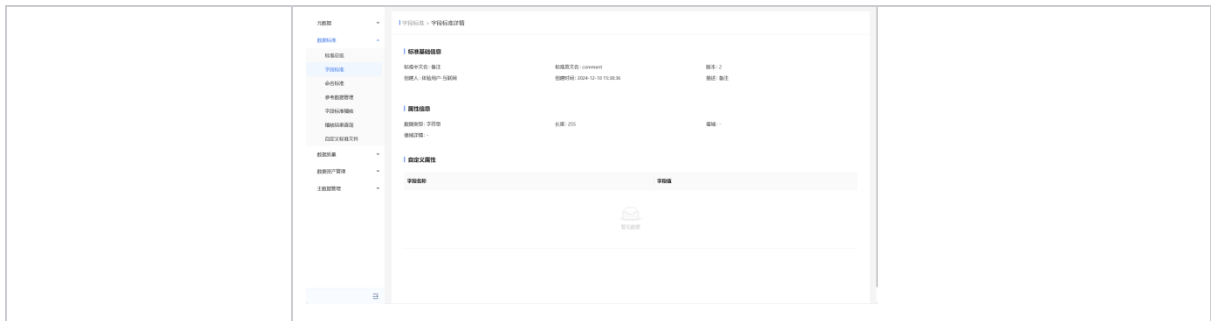


点击【回退】，弹出确认对话框，支持回滚到过去其他版本。点击【确认】完成回滚，回滚后，当前版本将被覆盖，且不可恢复。



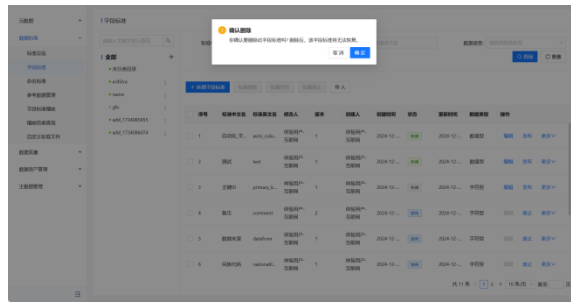
查看详情

查看所选标准的具体信息。点击【更多】->【查看详情】，进入字段标准详情展示页，展示内容包括标准基础信息、属性信息及自定义属性



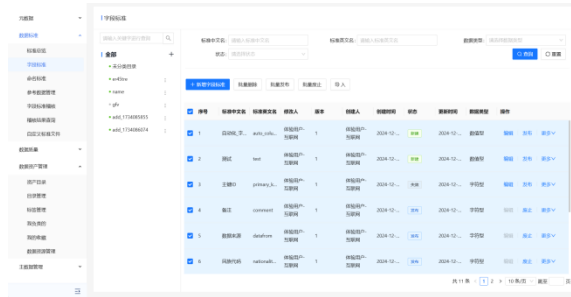
删除

删除所选字段标准。点击【更多】->【删除】，在弹出窗口中确认是否删除字段，点击【确定】按钮完成删除



批量操作

批量处理多个字段标准。通过多选或全选字段标准，实现批量操作，批量操作包括批量删除、批量发布及批量废止



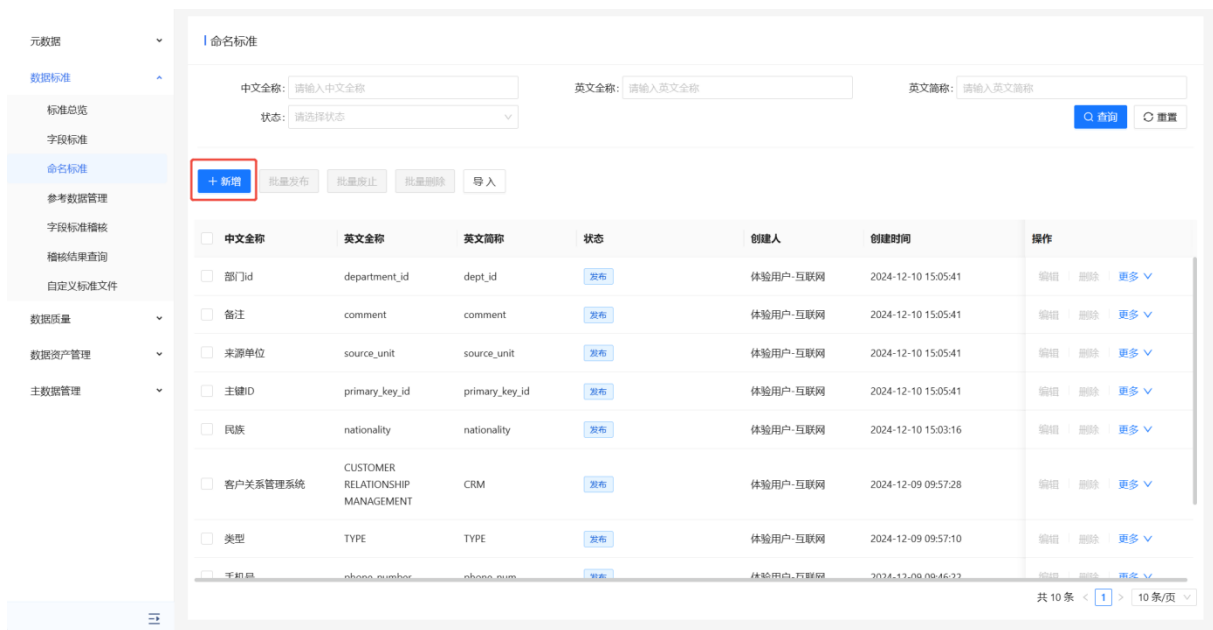
## 命名标准

命名标准为数据元素（如字段、指标及维度等）的命名提供基础词汇和结构化的命名方法；命名标准支撑参考数据的命名，以关联词根的方式为字段标准的命名提供参考。

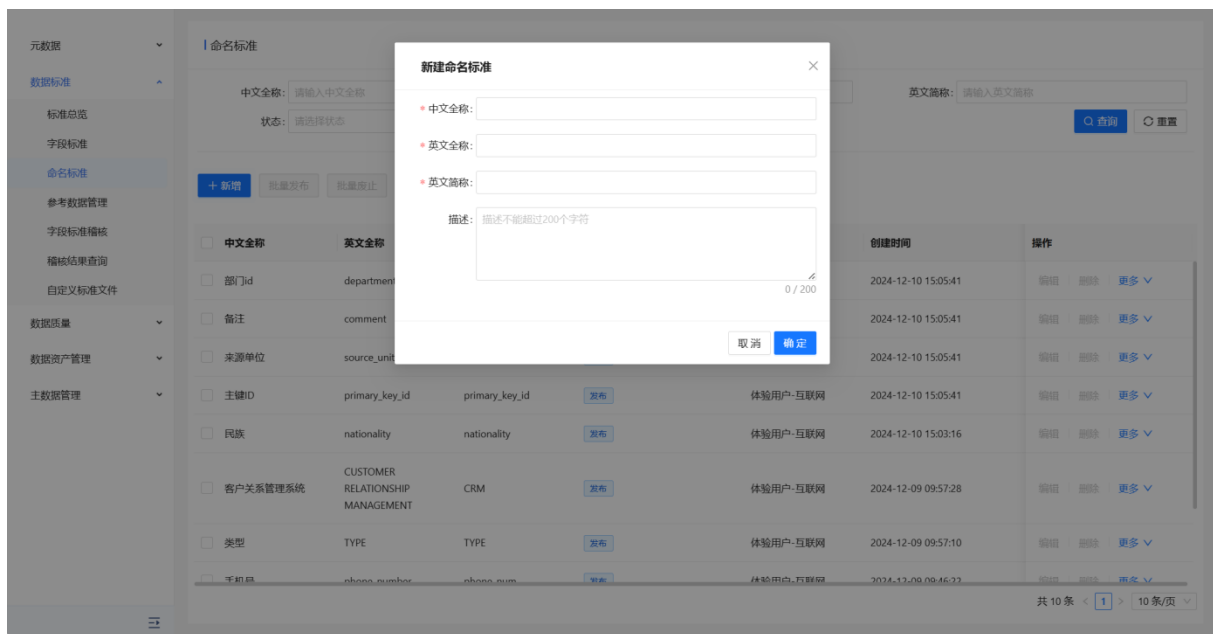
命名标准模块统一管理命名标准，支持命名标准的发布、废止、修改、新增、导入及删除等操作。

## 新增命名标准

点击【新增】按钮

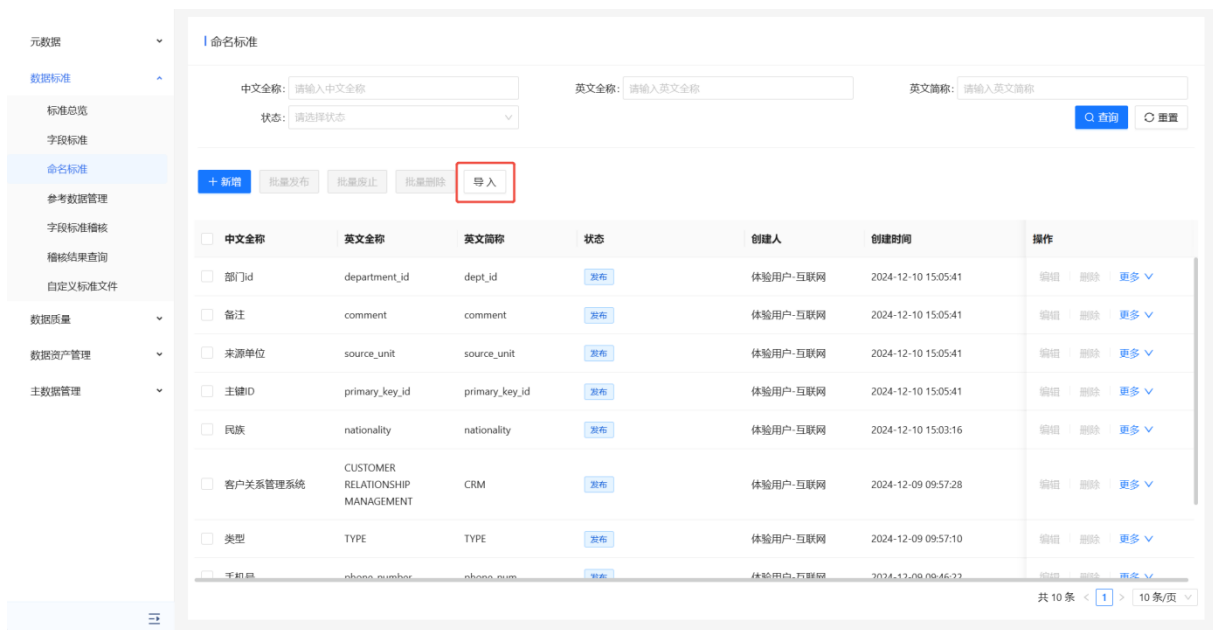


在弹窗中输入新增命名标准的中文全称（必填）、英文全称（必填）、英文简称（必填）及描述。点击【确定】按钮，完成命名标准的新增。

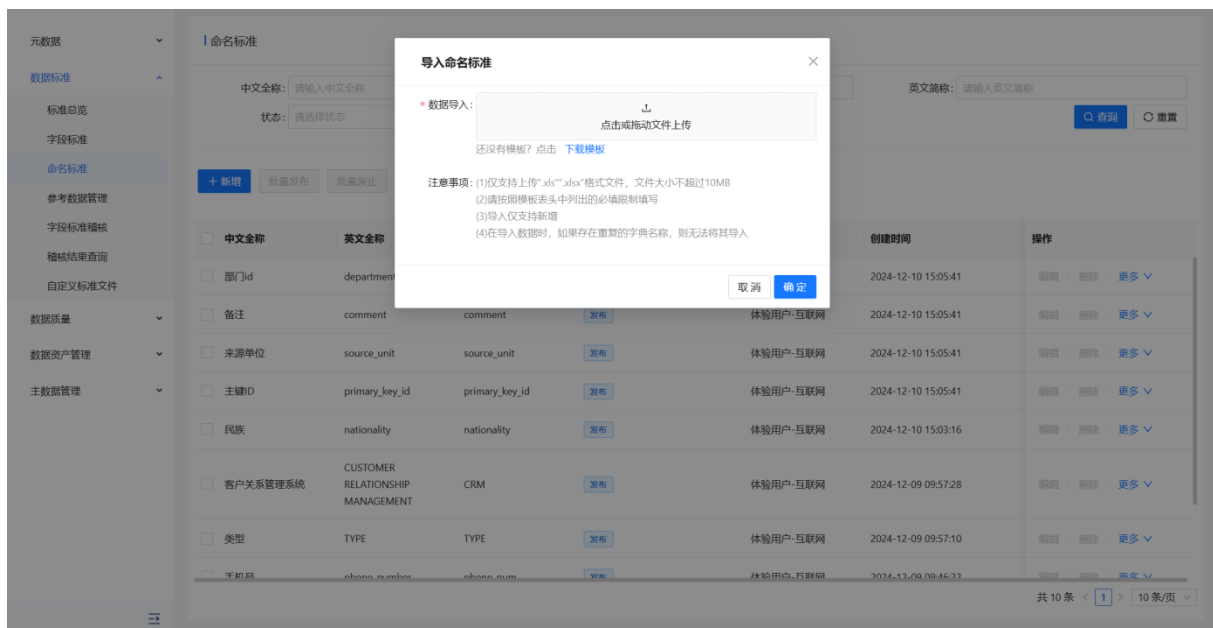


## 导入命名标准

支持导入字段标准，点击【导入】按钮，在弹窗中进行导入操作。



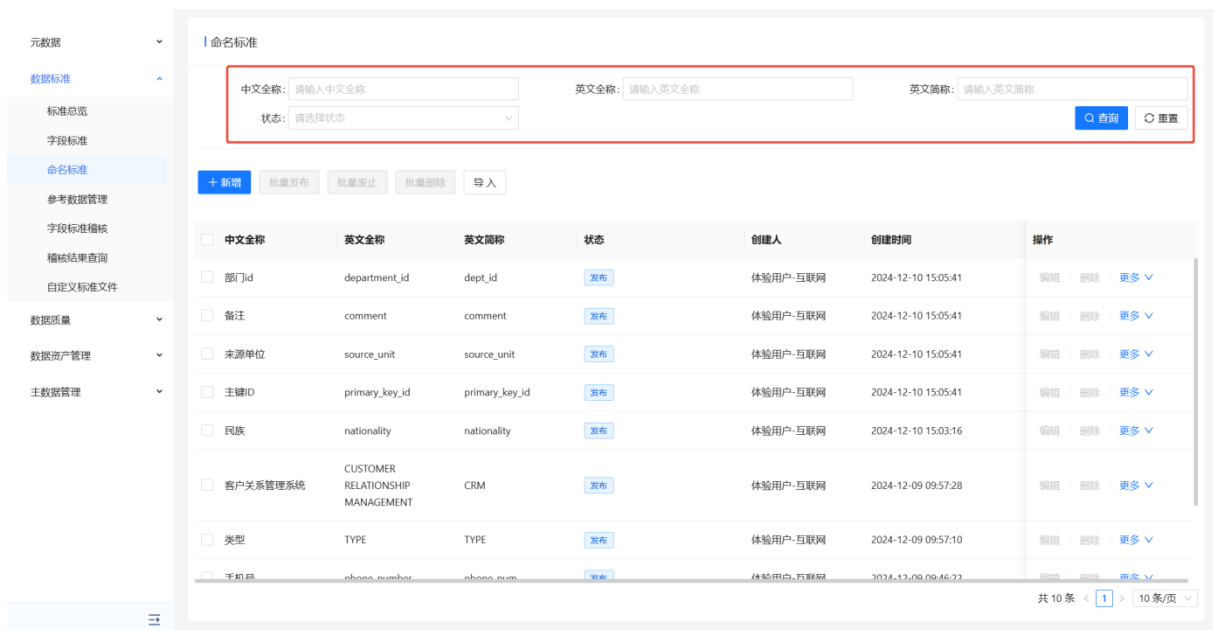
数据导入支持下载模板，点击【下载模板】按钮下载参考模板，按照模板字段要求填充，点击或拖动文件至上传框进行文件上传，导入需校验导入内容是否符合格式，平台解析完成后实现命名标准导入。



点击【确定】按钮完成命名标准导入。

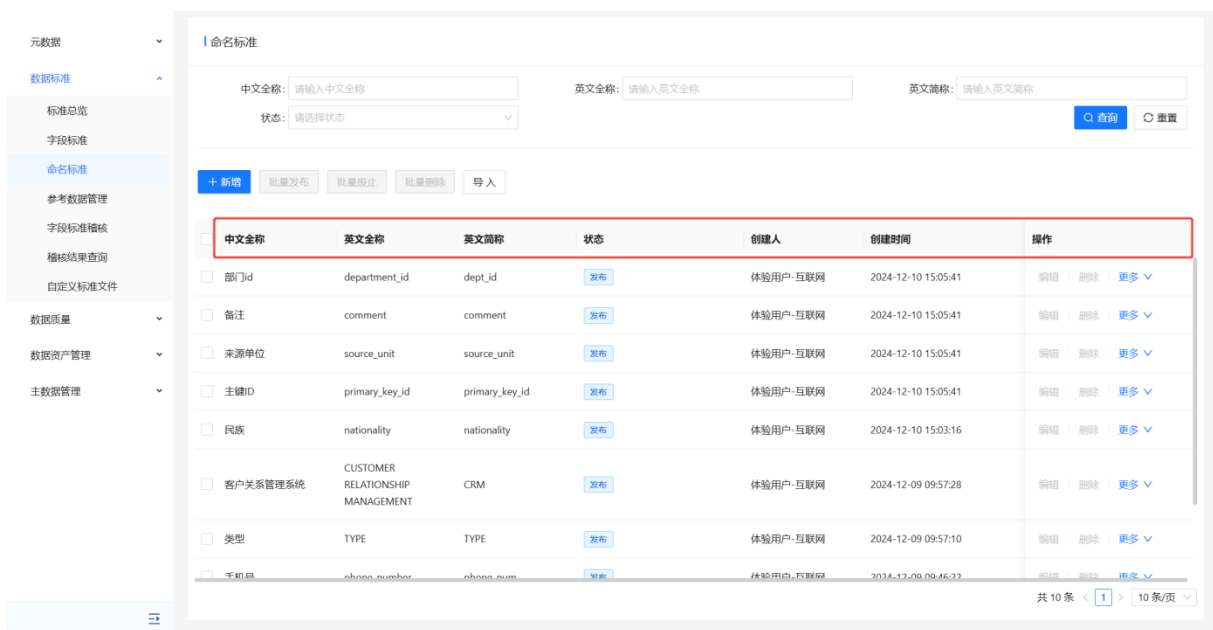
## 查询命名标准

支持通过输入中文全称、输入英文全称、输入英文简称及选择状态的方式查询命名标准。点击【查询】按钮进行查询。



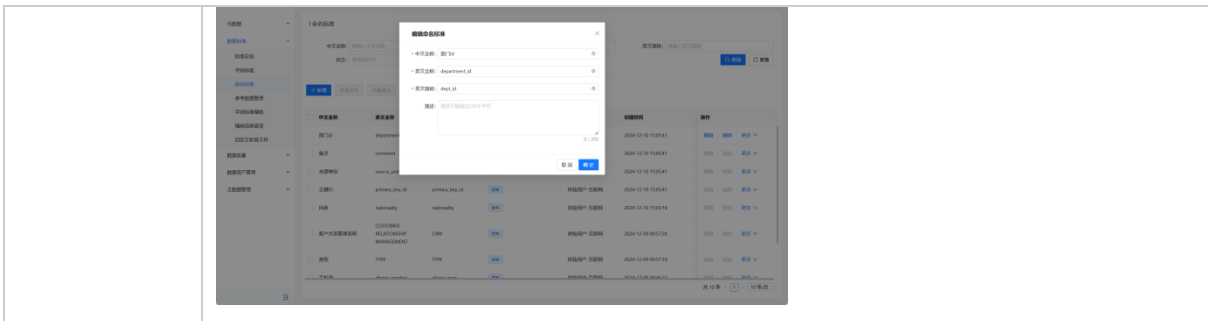
## 命名标准操作

页面默认展示全部字段标准，包括序中文全称、英文全称、英文简称、状态、创建人、创建时间及操作。



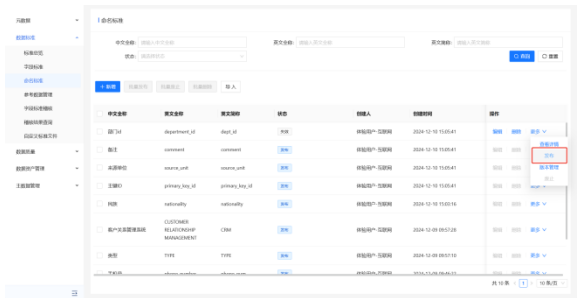
其中字段标准状态分为新建、发布及失效三种，字段标准废止后，将失效，失效状态下可以进行编辑再重新发布。

操作	描述
编辑	编辑所选命名标准。点击【编辑】按钮，进入命名标准编辑弹窗，编辑内容同新增命名标准。点击【确认】完成命名标准编辑

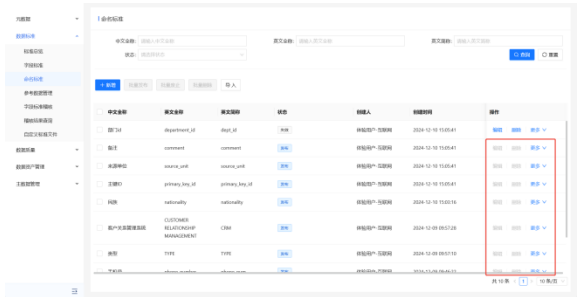


将所选命名标准发布上线，即时生效，支持对该字段标准的引用。点击【更多】->【发布】按钮，在弹窗界面确认选择发布的命名标准。点击【确认】完成命名标准发布

发布

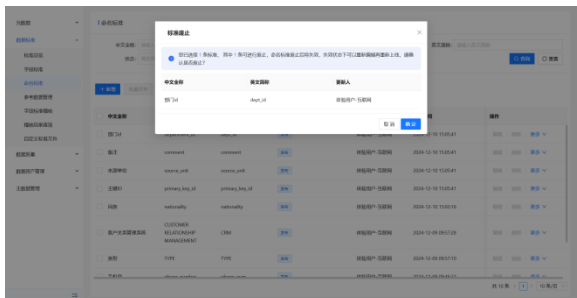


注意，命名标准经发布后，无法进行编辑。如需编辑已发布的命名标准，需先废止该命名标准，再进行编辑，编辑后重新发布



废止

废止所选命名标准，不再支持对该标准的引用。点击【更多】->【废止】按钮，在弹窗中确认是否废止标准，点击【确定】完成标准废止



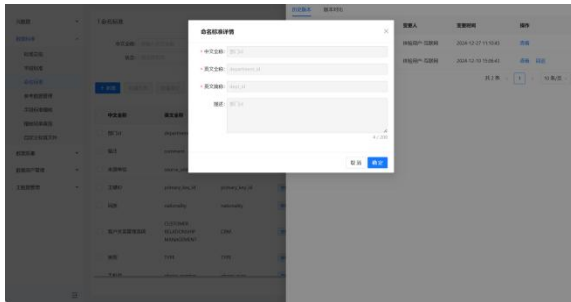
版本管理

查看所选标准的版本变更历史。点击【更多】->【版本管理】按钮，在右侧抽屉中查看命名标准的版本记录，版本记录包括历史版本和版本对比

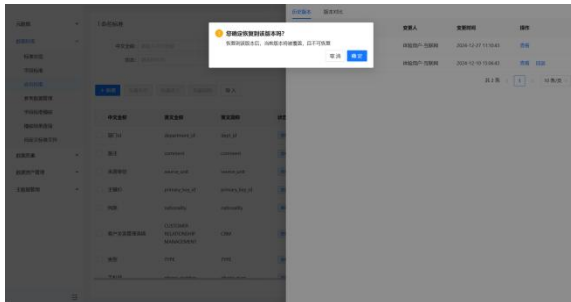
版本	变更类型	变更人	变更时间	操作
V2	发布	体验用户-互联网	2024-12-27 11:10:43	<a href="#">查看</a>
V1	发布	体验用户-互联网	2024-12-10 15:06:43	<a href="#">查看</a> <a href="#">回退</a>

共 2 条 < 1 > 10 条/页

历史版本记录了该命名标准所有的变更信息，点击【查看】，在弹窗中查看该版本详细信息

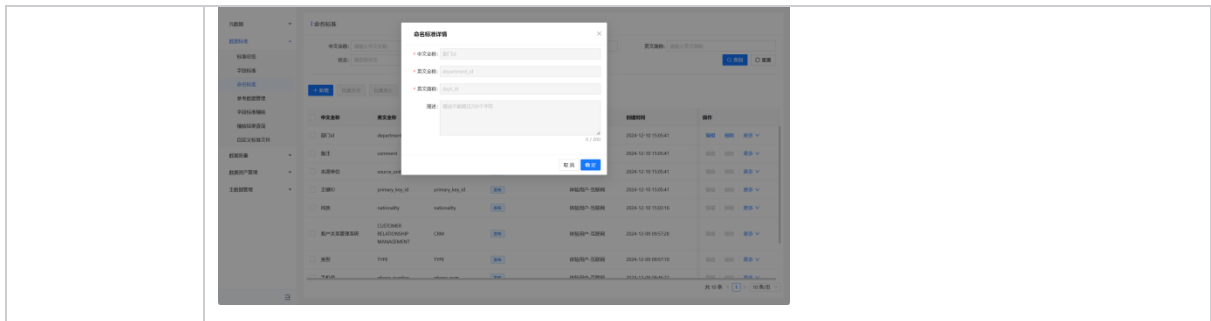


点击【回退】，弹出确认对话框，支持回滚到过去其他版本。点击【确定】完成回滚，回滚后，当前版本将被覆盖，且不可恢复



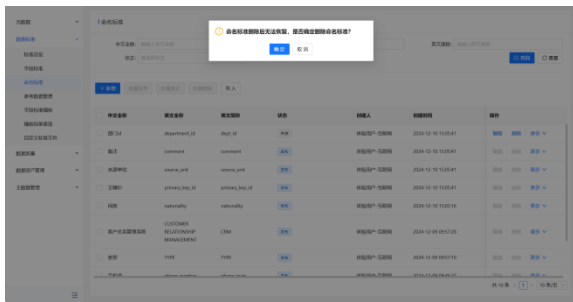
查看详情

查看所选命名标准的具体信息。点击【更多】->【查看详情】，进入命名标准详情展示弹窗，展示内容包括命名标准中文全称、英文全称、英文简称及描述



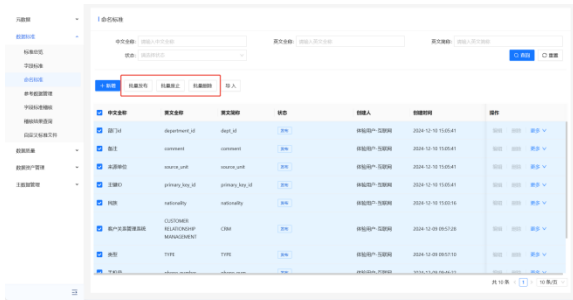
删除所选命名标准。点击【更多】->【删除】，在弹出窗口中确认是否删除标准，点击【确定】按钮完成删除。

删除



批量处理多个命名标准。支持对多个命名标准进行批量处理。通过多选或全选命名标准，实现批量操作，批量操作包括批量发布、批量废止及批量删除

批量操作



## 参考数据管理

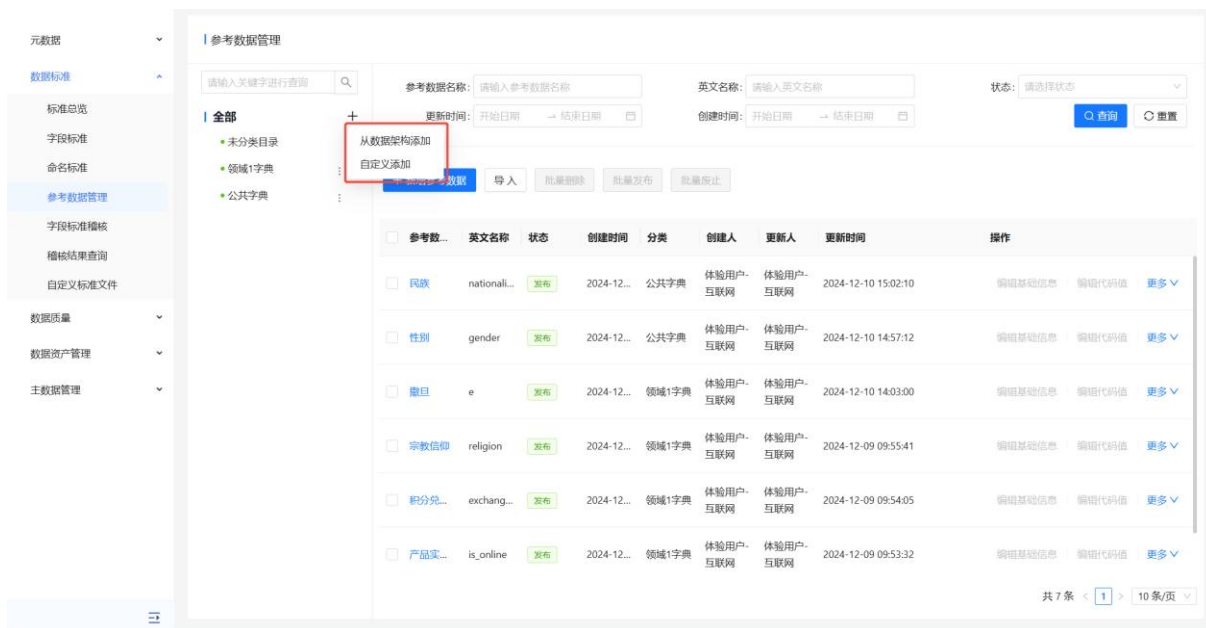
参考数据是数据的字典，是将组织内数据与组织外信息关联起来的代码表，定义了数据的值域；参考数据被用于字段标准的支撑，字段标准通过引用参考数据来确定具体的值域范围。

参考数据管理模块统一维护和管理参考数据，支持参考数据的分类、新增、导入、编辑、删除、发布、废止等操作。

## 参考数据目录

### ● 新建目录

支持标准目录新建，点击目录区域的加号【+】，可以选择从数据架构添加目录或自定义添加目录。

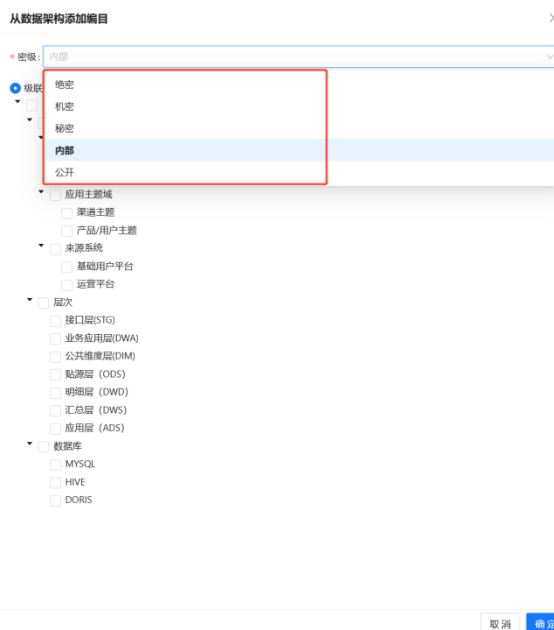


### 新建目录方式

### 方法

选择【从数据架构添加】，需要选择目录密级，目录密级分为：绝密、机密、秘密、内部及公开五个级别；

### 从数据架构添加



选择架构支持【级联选择】与【单独选择】两种方式。

【级联选择】可以实现批量选择该架构层级及下级层次的所有架构。

	 <p><b>【单独选择】</b> 可以实现单独选择某层级架构。</p>  <p>点击<b>【确定】</b> 完成目录的创建。</p>
自定义添加	<p>选择<b>【自定义添加】</b>，填写基础信息与自定义属性。 基础信息部分需要填写目录名称（必填）、密级（必填）及描述；</p>

## 注册编目



### 基础信息

\* 目录名称:

\* 密级:

描述:

0 / 200

### 自定义属性

+ 新增自定义属性

取消

确定

自定义属性为选填部分，可以通过点击【新增自定义属性】按钮实现自定义属性的新增和删除，支持新增多条自定义属性。

注册编目 ×

### 基础信息

\* 目录名称:

\* 密级:

描述:  0 / 200

### 自定义属性

[+ 新增自定义属性](#)

自定义属性
🗑

自定义属性
🗑

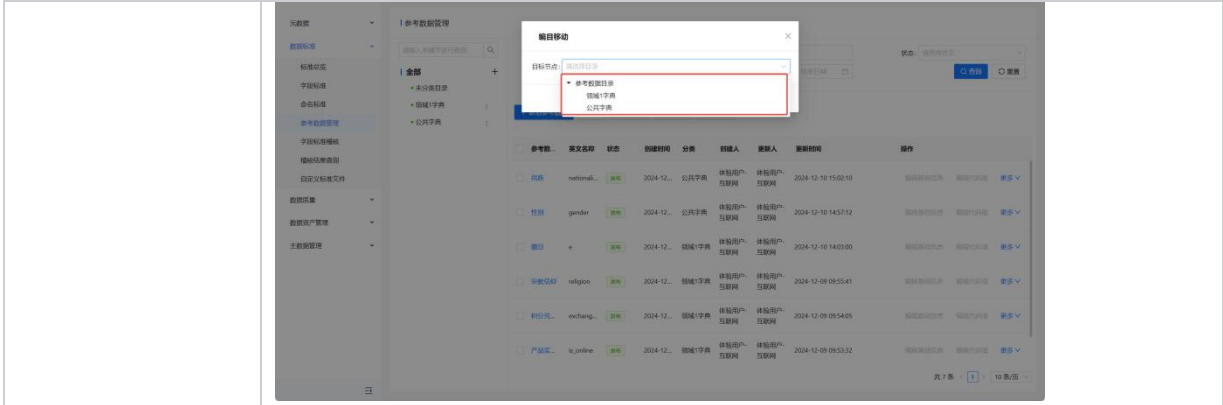
取消
确定

点击【确定】完成目录的创建。

## ● 目录操作

针对现有目录，点击目录名右侧的按钮，可以实现对目录的管理和操作。

操作	描述
添加同级目录和添加子级目录	<p>为选中目录添加同级或子级目录，支持【从数据架构添加】和【自定义添加】</p> 
移动	<p>移动所选目录至目标节点。点击【移动】选项，在弹窗的下拉菜单中选择想要将目录移动至的目标节点，点击【确定】，完成对目录的移动</p>



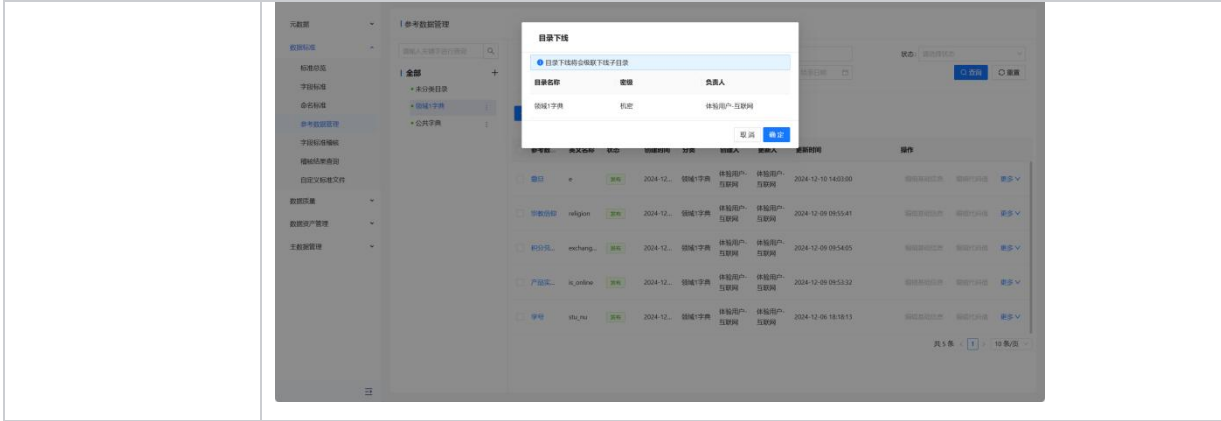
编辑

对所选目录进行编辑。点击【编辑】选项，在右侧弹窗中编辑目录信息，更改目录基础信息和自定义属性。点击【确定】，完成对目录的编辑



下线

将所选目录下线处理。点击【下线】选项，弹窗中展示已选择的目录信息，点击【确定】，完成目录的下线



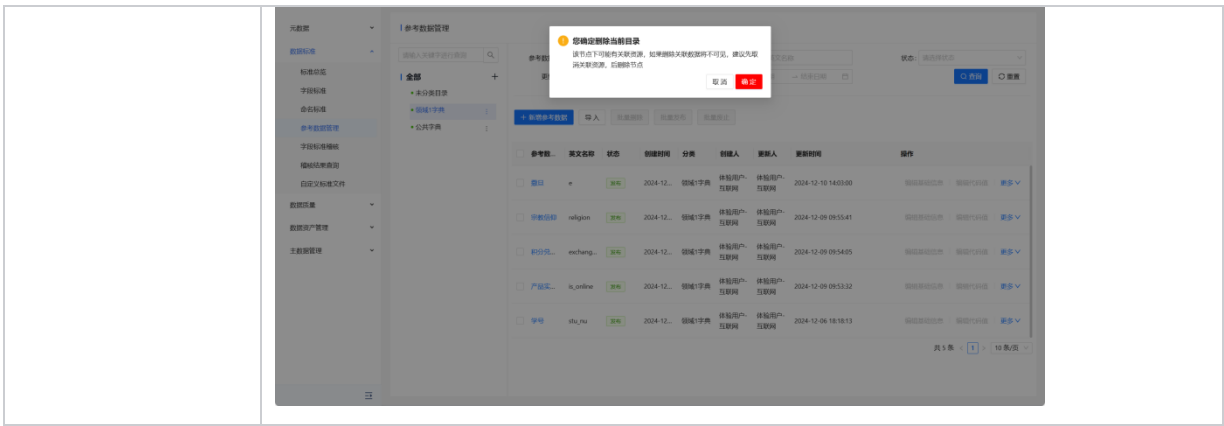
为团队或个人授予所选目录的管理和编辑权限。点击【授权】选项，在右侧抽屉展示目录的基本信息以及权限管理内容。在权限管理栏内，可以选择对【团队】权限进行管理和对【个人】权限进行管理，通过勾选的方式选择是否授予【管理】和【编辑】的权限。点击【确定】，完成目录的授权

授权



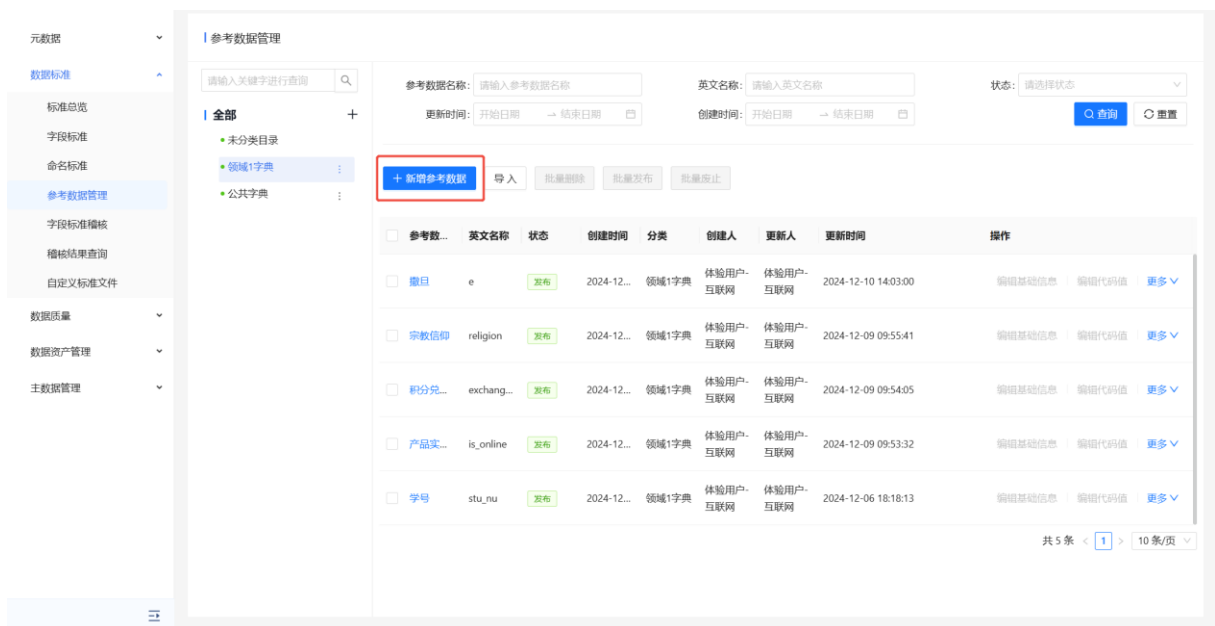
删除所选目录。点击【删除】选项，根据弹窗提示选择是否确定删除当前目录，点击【确定】，完成目录的删除

删除



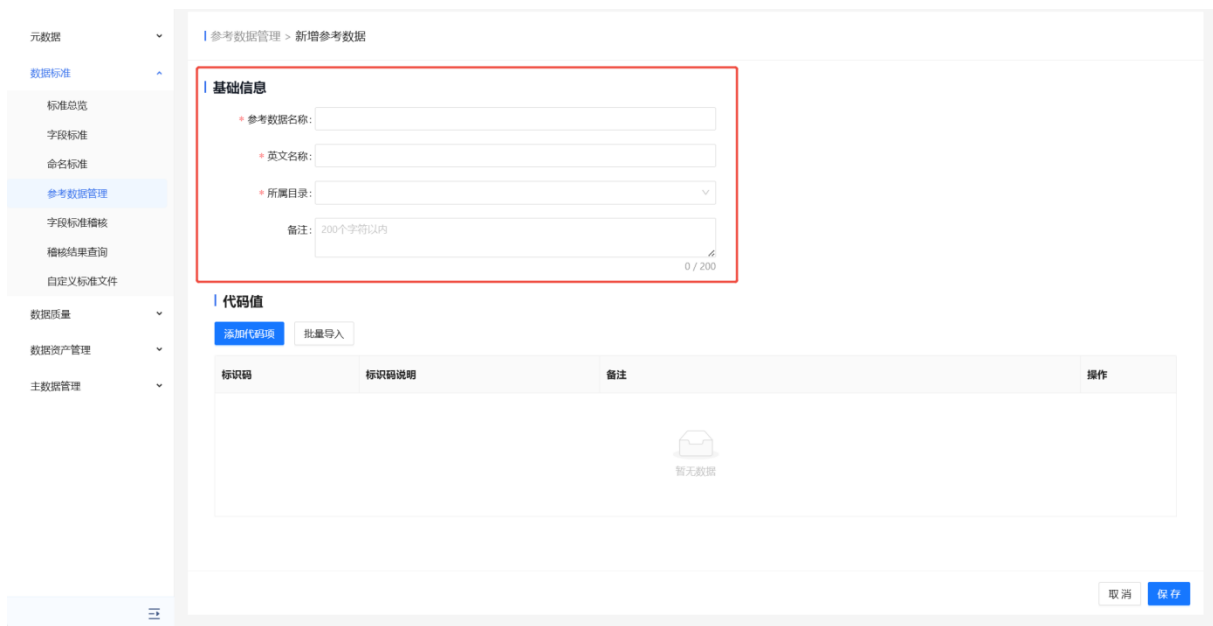
## 新增参考数据

点击【+新增参考数据】按钮，进入新增页面，填写参考数据信息。



### 1、基础信息

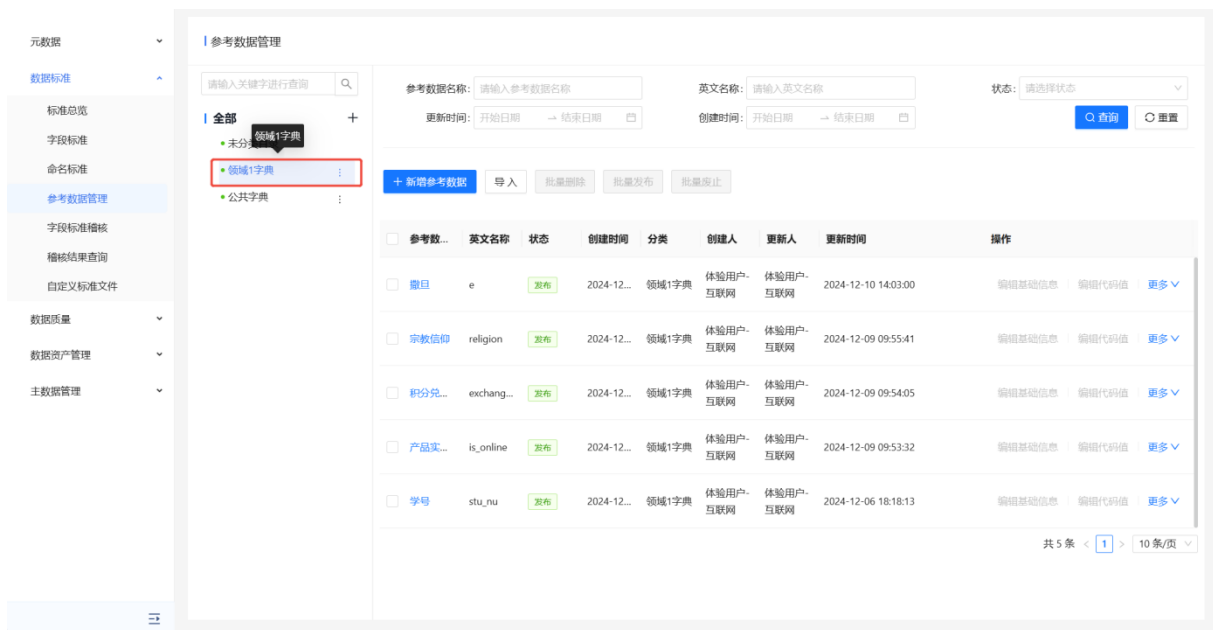
填写参考数据基础信息，包括参考数据名称（必填）、英文名称（必填）、所属目录（必填）及备注。



其中，关联字段所属目录有两种方式，一种通过在标准基础信息的所属目录下拉菜单中选择所属目录



另一种在字段标准界面首先选中目标目录，再点击【+新增字段标准】按钮，则进入新增页面后，在所属目录位置上可以自动关联已选目录。

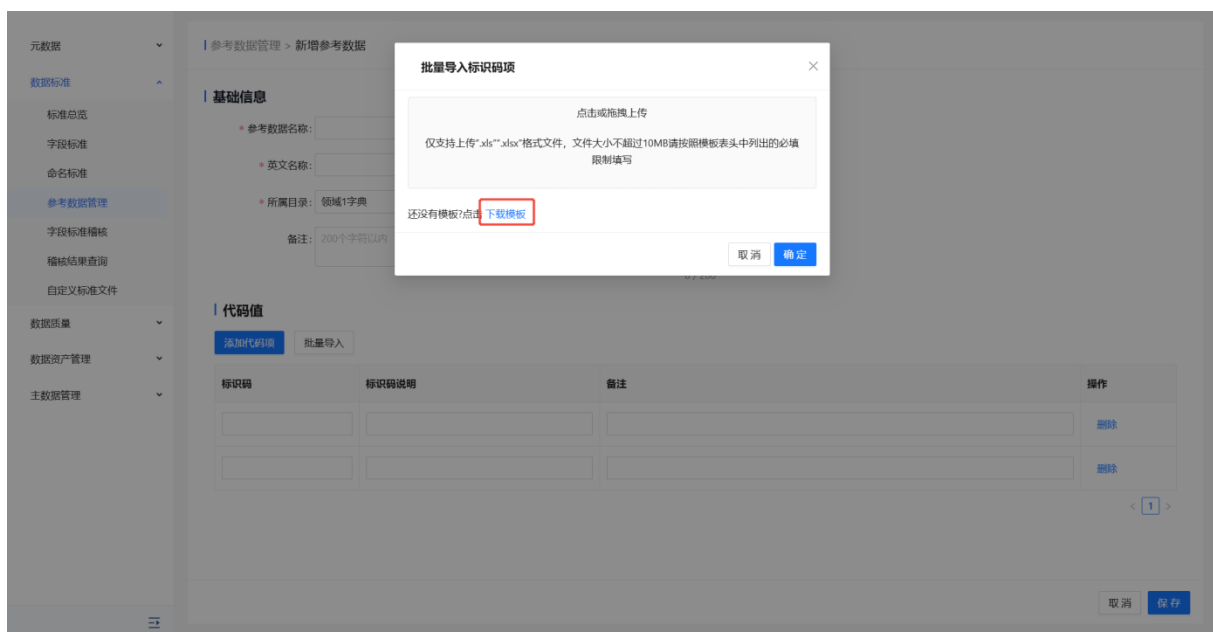


## 2、代码值

支持添加代码项和批量导入两种方式录入参考数据代码值。代码值信息包括标识码、标识码说明、备注及操作。点击【添加代码项】按钮，手动输入代码值信息。



点击【批量导入】按钮，在弹窗中上传文件。数据导入支持下载模板，点击【下载模板】按钮下载参考模板，按照模板字段要求填充，点击或拖动文件至上传框进行文件上传。



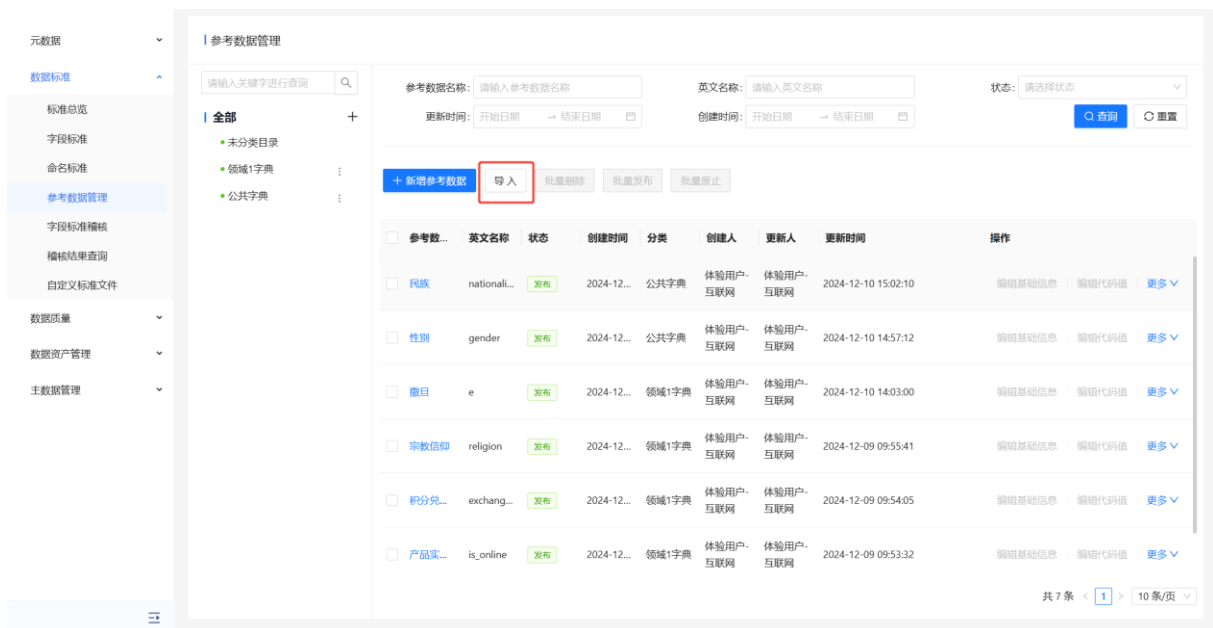
点击【确定】按钮完成批量导入内容的上传。

### 3、完成新增

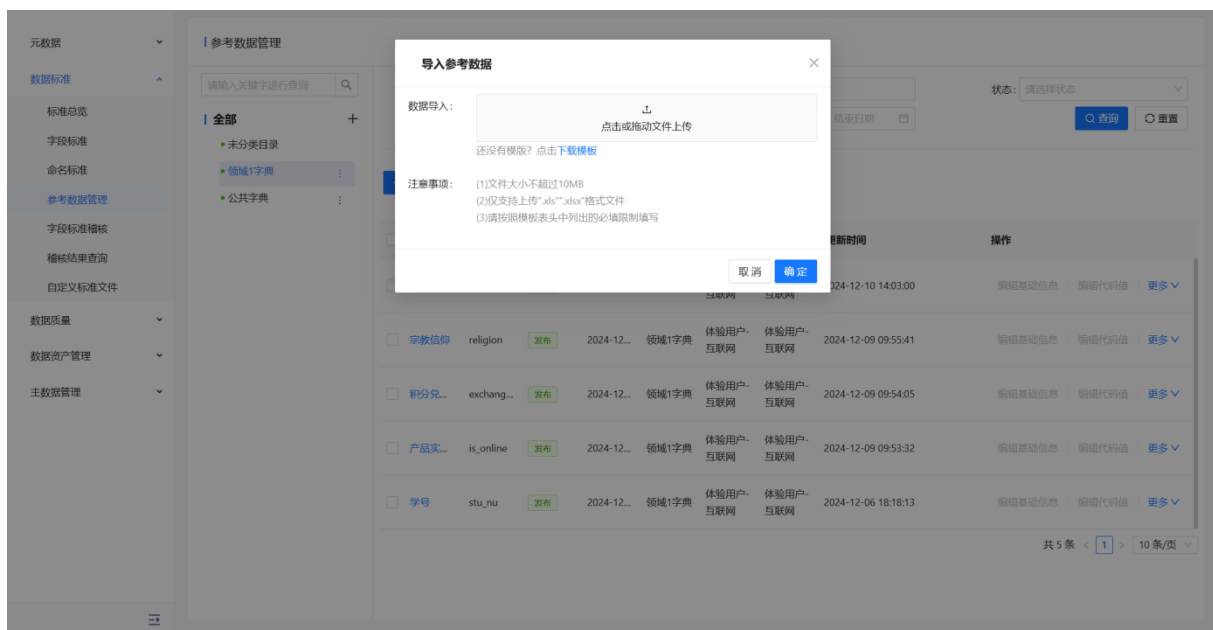
点击【保存】按钮完成参考数据的新增。

## 导入参考数据

支持导入字段标准，在左侧目录中选择导入节点，点击【导入】按钮。



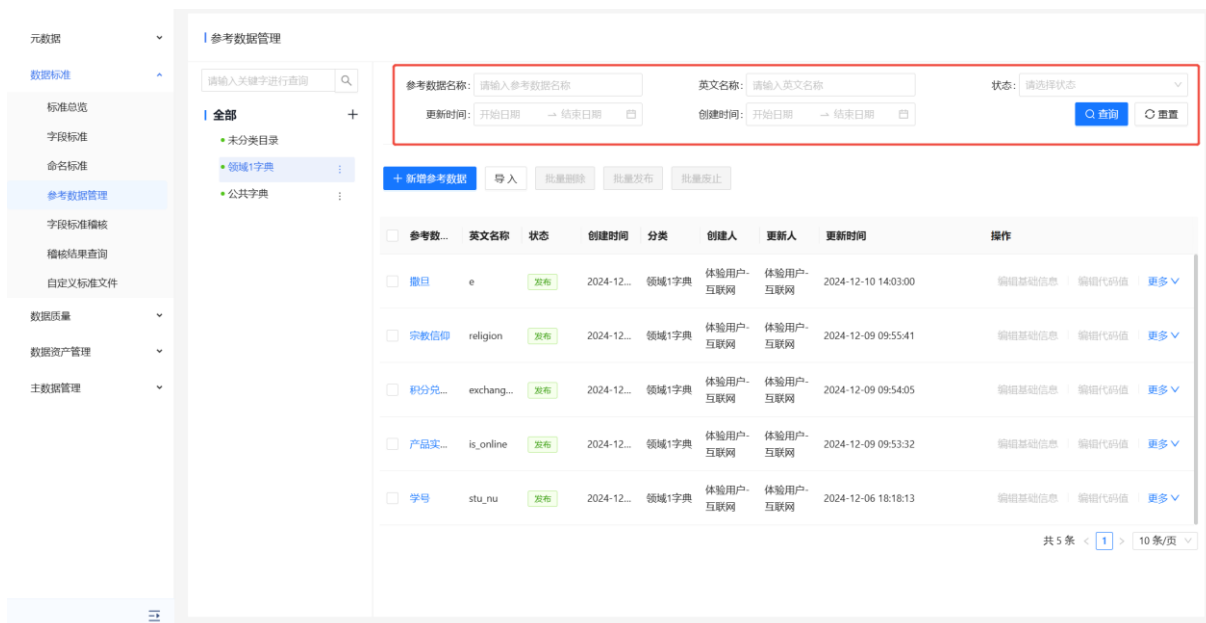
数据导入支持下载模板，点击【下载模板】按钮下载参考模板，按照模板字段要求填充，点击或拖动文件至上传框进行文件上传。



点击【确定】按钮完成字段标准的导入。

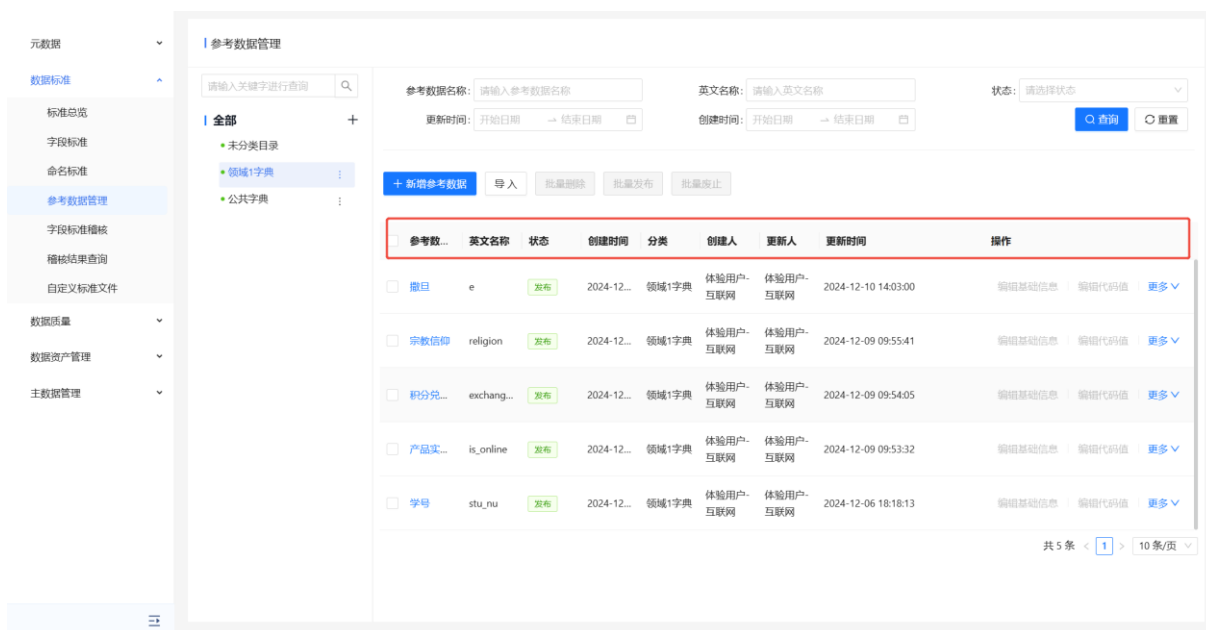
## 查询参考数据

支持通过输入参考数据名称、输入英文名称、选择状态、选择更新时间范围及选择创建时间范围的方式查询参考数据。点击【查询】按钮进行查询。

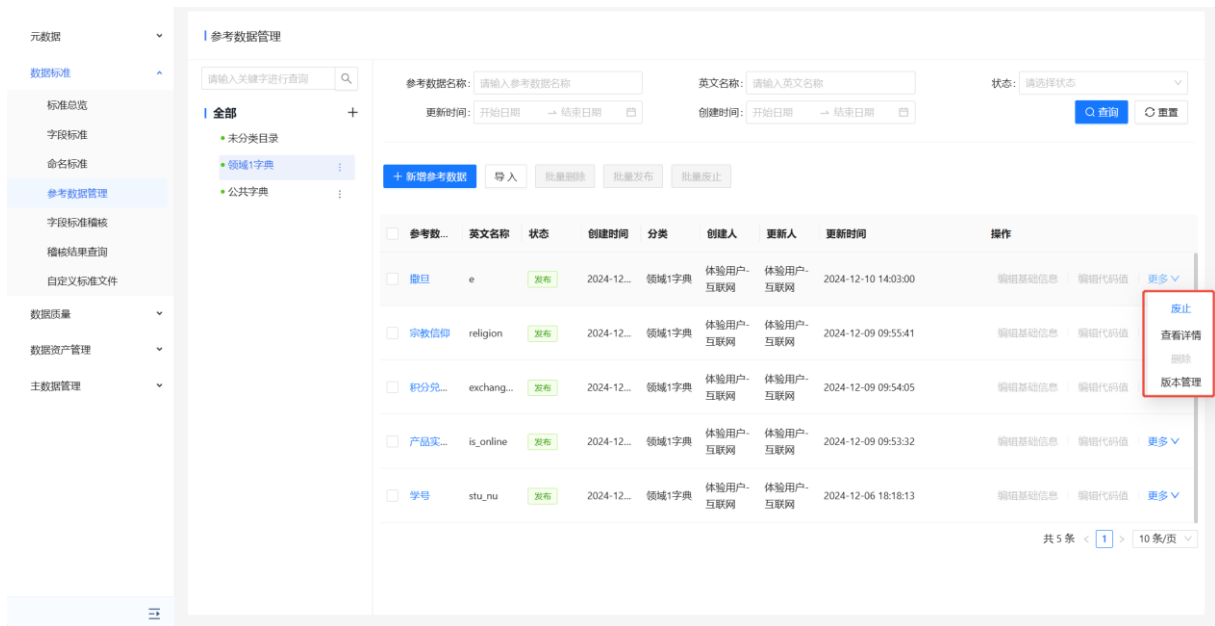


## 参考数据操作

页面默认展示全部参考数据，包括参考数据名称、英文名称、状态、创建时间、分类、创建人、更新人、更新时间及操作。



支持对参考数据进行操作，操作行为包括编辑基础信息、编辑代码值及更多（废止、查看详情、删除及版本管理）。



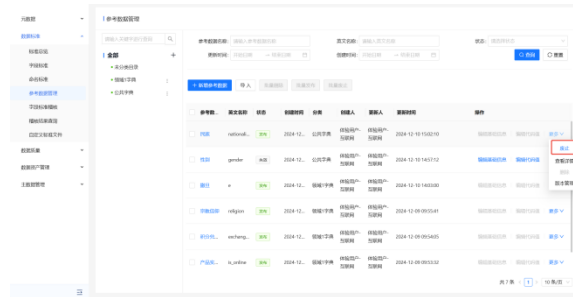
操作	描述
<p>编辑基础信息</p>	<p>编辑所选参考数据基础信息。点击【编辑基础信息】按钮，进入参考数据编辑界面，编辑参考数据的基础信息。点击【保存】完成字段编辑</p> 
<p>编辑代码值</p>	<p>编辑所选参考数据代码值。点击【编辑代码值】按钮，进入参考数据编辑界面，编辑参考数据的代码值。与新增参考数据的操作相同，支持添加代码项和批量导入，可以对代码项进行编辑和删除操作。点击【保存】完成参考数据的编辑</p> 

发布

将所选参考数据发布上线，即时生效，支持对该参考数据的引用。点击【发布】按钮，在弹窗界面确认选择发布的参考数据。点击【确认】完成发布

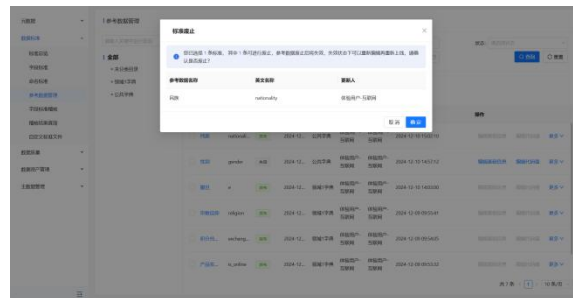


注意，参考数据经发布后，无法进行编辑。如需编辑已发布的参考数据，需先废止该参考数据，再进行编辑，编辑后重新发布。



废止

废止所选参考数据，不再支持对该参考数据的引用。点击【废止】按钮，在弹窗中确认是否废止，点击【确定】完成参考数据废止



版本管理

查看所选参考了数据的版本变更历史。点击【更多】->【版本管理】按钮，在右侧抽屉中查看参考数据的版本记录，版本记录包括历史版本和版本对比

版本记录

历史版本 版本对比

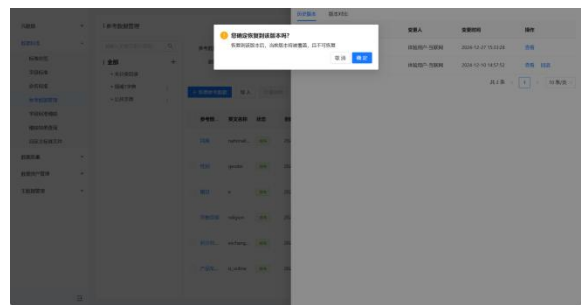
版本	变更类型	变更人	变更时间	操作
V2	发布	体验用户-互联网	2024-12-27 15:33:28	查看
V1	发布	体验用户-互联网	2024-12-10 14:57:52	查看 回退

共 2 条 < 1 > 10 条/页

历史版本记录了该参考数据所有的变更信息，点击【查看】，进入该版本详细信息界面

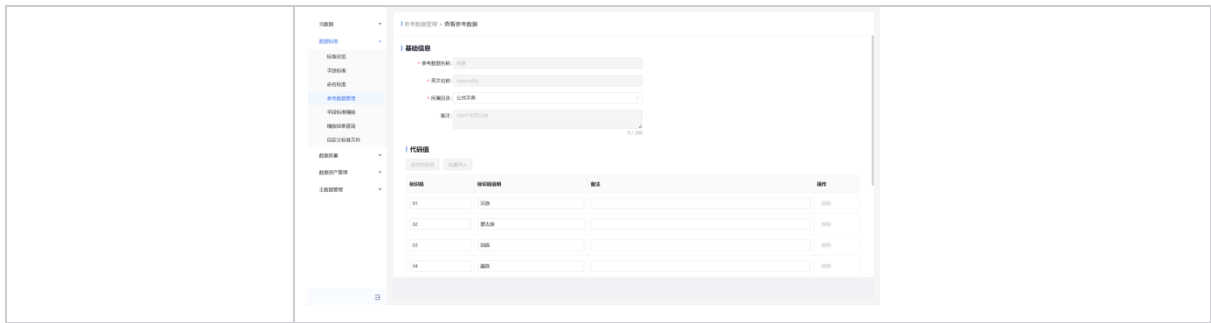


点击【回退】，弹出确认对话框，支持回滚到过去其他版本。点击【确认】完成回滚，回滚后，当前版本将被覆盖，且不可恢复。



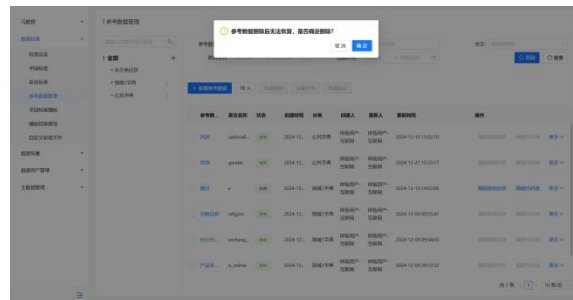
查看详情

查看所选参考数据的具体信息。点击【更多】->【查看详情】，进入参考数据详情展示页，展示内容包括基础信息和代码值。



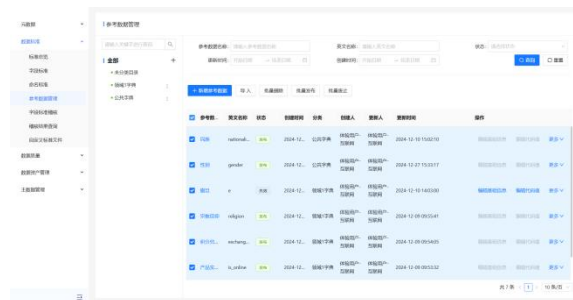
删除所选参考数据。点击【更多】->【删除】，在弹出窗口中确认是否删除参考数据，点击【确定】按钮完成删除。

删除



批量处理多个参考数据。支持对多个参考数据进行批量处理。通过多选或全选参考数据，实现批量操作，批量操作包括批量删除、批量发布及批量废止。

批量操作



## 字段标准稽核

字段标准稽核为字段标准的质量提供保障，通过建立监控任务，按照一定的时间周期执行字段标准的稽核任务。

字段标准稽核的流程为：

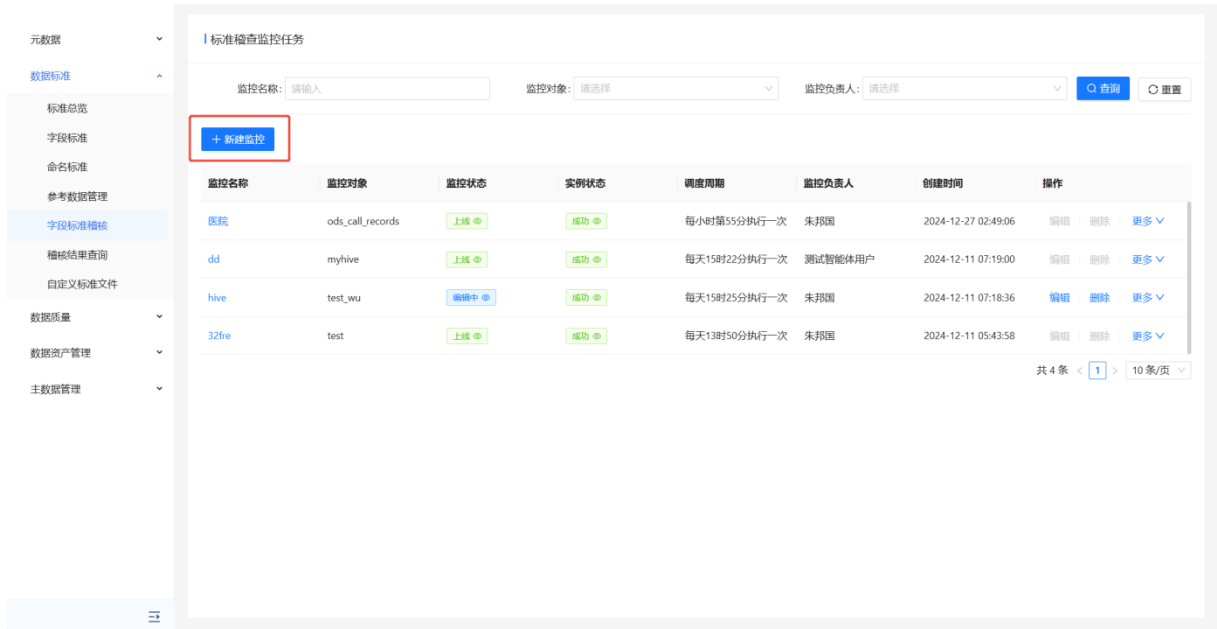
步骤	描述
步骤一：新建监控任务	设置监控对象，选择需要监控的数据源、库、表、字段；设置监控规则，为字段配置用于稽核的标准
步骤二：创建调度	配置字段稽核监控任务的信息，包括执行周期、任务类型、运行信息等
步骤三：调度上线/调度下线	将配置好调度信息的监控任务上线/下线，该任务则会根据调度信息按周期执行/停止执行标准稽核。稽核结果需

要在稽核结果查询模块查看

## 新增稽核任务

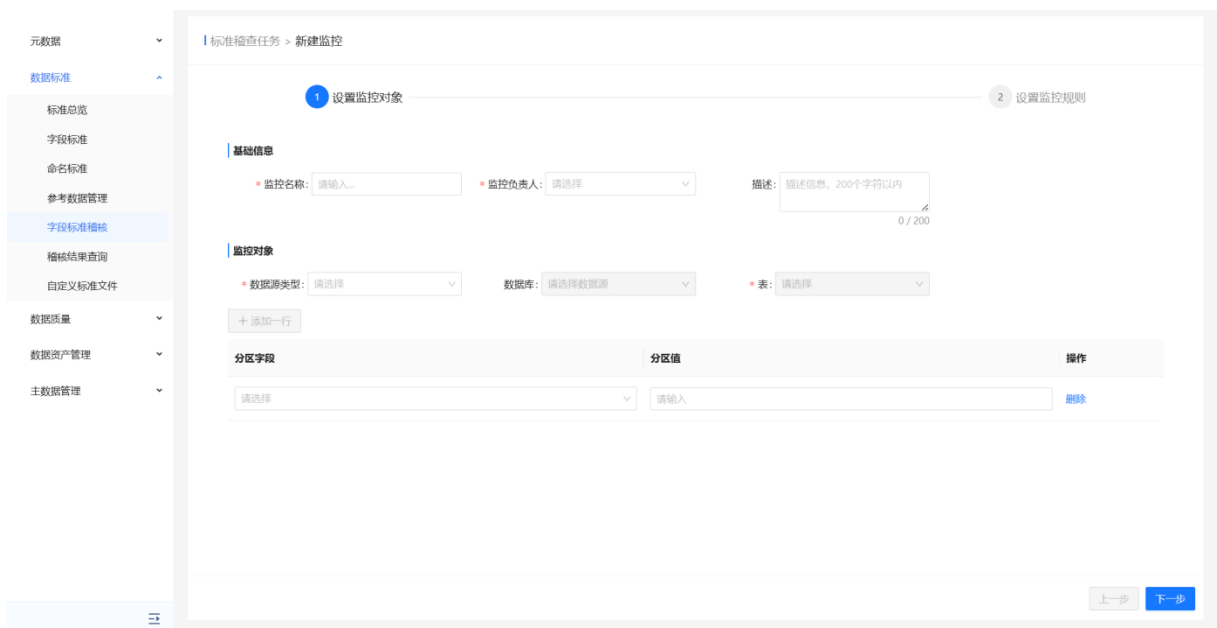
### 1、新建任务

点击【+新建监控】按钮，进入新建监控业务界面



### 2、设置监控对象

配置基础信息和监控对象。基础信息包括监控名称（必填）、监控负责人（必选）及描述；监控对象包括数据源类型（必选）、数据库、表（必选）、分区字段及分区值，支持对分区字段和分区值进行删除操作。



点击【下一步】按钮，设置监控规则。

### 3、设置监控规则

监控规则界面展示该数据表内编号、字段编码、字段名称、字段类型及标准内容。

标准稽查任务 > 新建监控

设置监控对象 2 设置监控规则

批量配置

编号	字段编码	字段名称	字段类型	标准内容
1	name	-	string	请选择
2	id	-	string	请选择
3	age	-	string	请选择

共 3 条 < 1 > 10 条/页

上一步 完成

点击标准内容下拉框，显示该字段中所有的标准中文名、标准英文名、数据类型、修改人、版本、创建人、创建时间及状态，可以根据分类或标准中文名进行检索。

## 选择标准内容



选择分类: 请选择

请输入标准中文名



	标准中文名	标准英文名	数据类型	修改人	版本	创建人	创建时间	状态
<input type="radio"/>	备注	comment	字符型	体验用户-互联网	2	体验用户-互联网	2024-12-10 15:38:36	<a href="#">发布</a>
<input type="radio"/>	数据来源	datafrom	字符型	体验用户-互联网	1	体验用户-互联网	2024-12-10 15:37:05	<a href="#">发布</a>
<input type="radio"/>	民族代码	nationality_code	字符型	体验用户-互联网	1	体验用户-互联网	2024-12-10 15:36:12	<a href="#">发布</a>
<input type="radio"/>	产品实例状态_测试	is_online_test	字符型	体验用户-互联网	2	体验用户-互联网	2024-12-09 09:51:14	<a href="#">发布</a>
<input type="radio"/>	手机号_测试	phone_num_test	字符型	体验用户-互联网	2	体验用户-互联网	2024-12-09 09:50:33	<a href="#">发布</a>
<input type="radio"/>	当天5G流量_测试	traffic_5g_today_test	数值型	体验用户-互联网	2	体验用户-互联网	2024-12-09 09:47:37	<a href="#">发布</a>
<input type="radio"/>	编号_测试	id_test	数值型	体验用户-互联网	2	体验用户-互联网	2024-12-06 18:16:59	<a href="#">发布</a>

共 7 条

1

关闭

确认

选择需要稽核的标准内容后，点击【确认】，表示需要稽核该标准。

### 选择标准内容



选择分类: 请选择

请输入标准中文名



	标准中文名	标准英文名	数据类型	修改人	版本	创建人	创建时间	状态
<input type="radio"/>	备注	comment	字符型	体验用户-互联网	2	体验用户-互联网	2024-12-10 15:38:36	<a href="#">发布</a>
<input type="radio"/>	数据来源	datafrom	字符型	体验用户-互联网	1	体验用户-互联网	2024-12-10 15:37:05	<a href="#">发布</a>
<input type="radio"/>	民族代码	nationality_code	字符型	体验用户-互联网	1	体验用户-互联网	2024-12-10 15:36:12	<a href="#">发布</a>
<input type="radio"/>	产品实例状态_测试	is_online_test	字符型	体验用户-互联网	2	体验用户-互联网	2024-12-09 09:51:14	<a href="#">发布</a>
<input type="radio"/>	手机号_测试	phone_num_test	字符型	体验用户-互联网	2	体验用户-互联网	2024-12-09 09:50:33	<a href="#">发布</a>
<input type="radio"/>	当天5G流量_测试	traffic_5g_today_test	数值型	体验用户-互联网	2	体验用户-互联网	2024-12-09 09:47:37	<a href="#">发布</a>
<input checked="" type="radio"/>	编号_测试	id_test	数值型	体验用户-互联网	2	体验用户-互联网	2024-12-06 18:16:59	<a href="#">发布</a>

共 7 条 < 1 >

关闭

确认

启用完成后点击【完成】按钮，则该任务创建完成，后续在监控任务界面进行调度。

元数据

数据标准

- 标准总览
- 字段标准
- 命名标准
- 参考数据管理
- 字段标准稽核
- 稽核结果查询
- 自定义标准文件

数据质量

数据资产管理

主数据管理

标准检查任务 > 新建监控

1 设置监控对象

2 设置监控规则

批量配置

编号	字段编码	字段名称	字段类型	标准内容
1	name	-	string	编号_测试
2	id	-	string	请选择
3	age	-	string	请选择

共 3 条 < 1 > 10 条/页

上一步 完成

## 创建调度

点击【更多】->【创建调度】，在右侧抽屉中进行调度配置。配置信息包括基本信息、调度信息、任务信息及运行信息。

## | 基本信息

任务名称: 1-稽核任务

\* 优先级: 中

## | 调度信息

\* 执行周期: 日

执行时间: 每天 00:00

crontab表达式: 0 0 0 \* \* ?

## | 任务信息

\* 任务类型: FLINK

yarn队列:

## | 运行信息

\* 调度集群组: default

业务时间偏移: -1

重试次数: 0

重试时间间隔(s): 30

\* 任务超时时长(s): 7200

取消

保存

在基本信息中配置任务名称和任务优先级，任务优先级分为高、中、低及最低四种。

## | 基本信息

任务名称: 1-稽核任务

## | 调度信息

\* 优先级: 中

高

**中**

低

最低

\* 执行周期:

执行时间:

crontab表达式: 0 0 0 \* \* ?

## | 任务信息

\* 任务类型: FLINK

yarn队列:

## | 运行信息

\* 调度集群组: default

业务时间偏移: -1

重试次数: 0

重试时间间隔(s): 30

\* 任务超时时长(s): 7200

取消

保存

在调度信息中配置执行周期和执行时间，执行周期分为分钟、小时、日及月四种。

## 调度配置



### 基本信息

任务名称: 1-稽核任务

\* 优先级: 中

### 调度信息

\* 执行周期: 日

执行时间: 分钟  
小时

crontab表达式: 日

月

### 任务信息

\* 任务类型: FLINK

yarn队列:

### 运行信息

\* 调度集群组: default

业务时间偏移: -1

重试次数: 0

重试时间间隔(s): 30

\* 任务超时时长(s): 7200

取消

保存

在任务信息中配置 yarn 队列

基本信息

任务名称: 1-稽核任务

\* 优先级: 中

调度信息

\* 执行周期: 日

执行时间: 每天 00:00

crontab表达式: 0 0 0 \* \* ?

任务信息

\* 任务类型: FLINK

yarn队列:

运行信息

\* 调度集群组: default

业务时间偏移: -1

重试次数: 0

重试时间间隔(s): 30

\* 任务超时时长(s): 7200

取消

保存

在运行信息中配置调度集群组、业务时间偏移、重试次数、重试时间间隔及任务超时时长

## | 基本信息

任务名称: 1-稽核任务

\* 优先级: 中

## | 调度信息

\* 执行周期: 日

执行时间: 每天 00:00

crontab表达式: 0 0 0 \* \* ?

## | 任务信息

\* 任务类型: FLINK

yarn队列:

## | 运行信息

\* 调度集群组: default

业务时间偏移: -1

重试次数: 0

重试时间间隔(s): 30

\* 任务超时时长(s): 7200

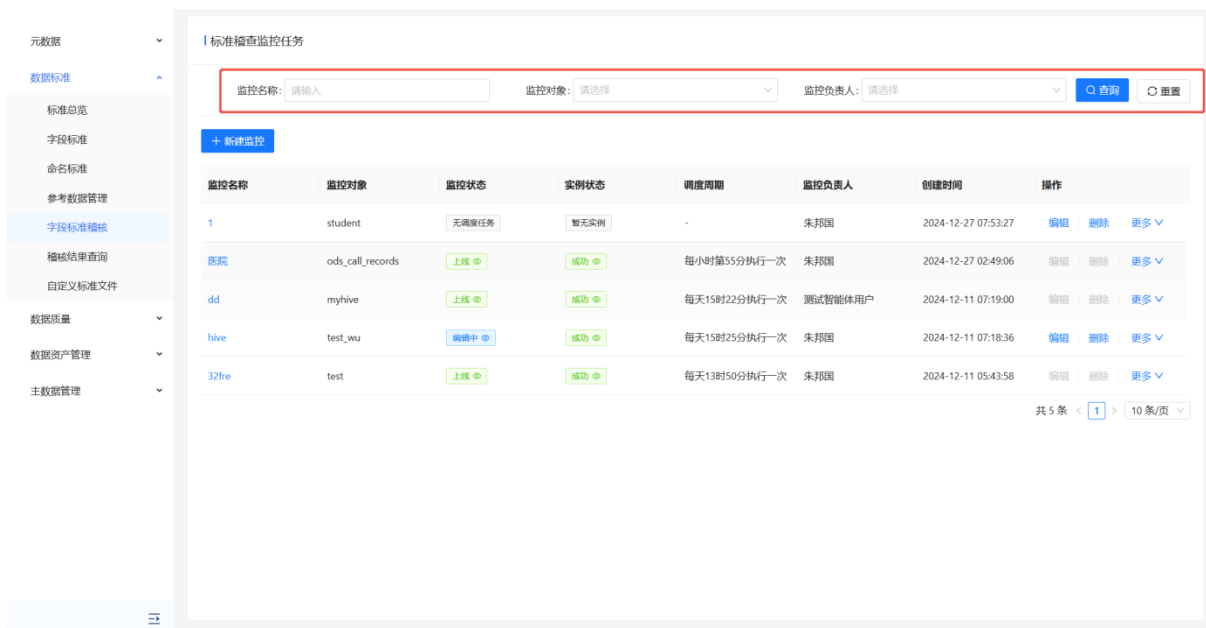
取消

保存

点击【保存】完成调度配置。

## 查询稽核任务

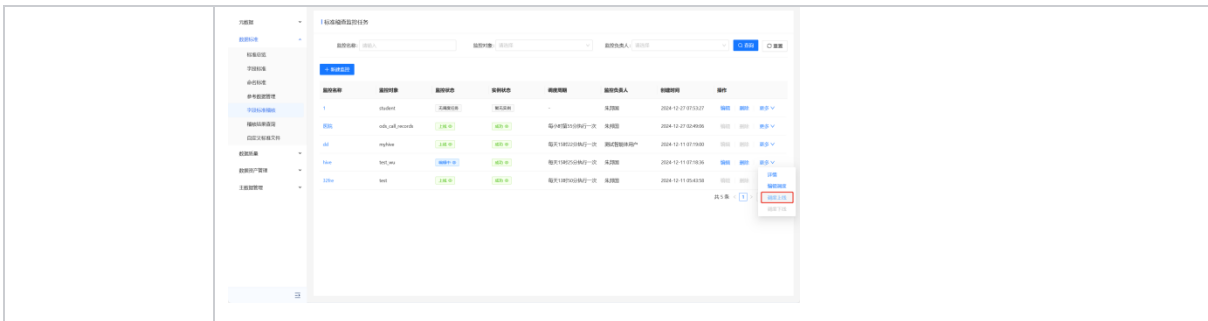
支持输入监控名称、监控对象及监控负责人进行稽核任务的查询。



## 稽核任务操作

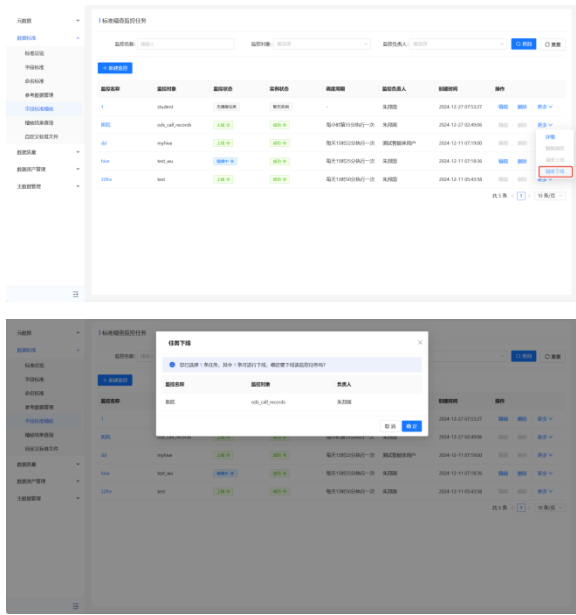
支持编辑、删除及更多（详情、创建调度、调度上线及调度下线）操作。对稽核任务进行调度，对于未调度上线的任务进行编辑。

操作	描述
编辑	<p>编辑所选稽核任务。点击【编辑】按钮，进入标准稽查任务的编辑界面，编辑稽查任务的基础信息。点击【下一步】-&gt;【保存】完成任务编辑。</p> 
调度上线	<p>将所选稽核任务上线，即时生效，支持对该字段标准的引用。点击【更多】-&gt;【调度上线】按钮，在弹窗界面确认选择上线的任务。点击【确认】完成字调度上线。</p> 



废止所选稽核任务，不再支持对该标准的引用。点击【更多】->【调度下线】按钮，在弹窗界面确认选择下线的任务。点击【确认】完成字调度下线。

调度下线



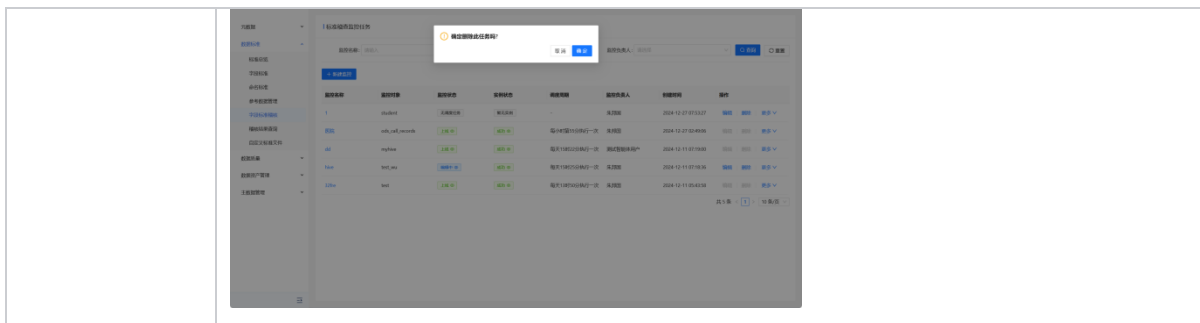
查看所选稽核任务的具体信息。点击【更多】->【详情】，进入稽核任务详情展示页，展示内容包括基础信息和监控规则。

查看详情



删除

删除所选稽核任务。点击【删除】按钮，在弹出窗口中确认是否删除任务，点击【确定】按钮完成删除。

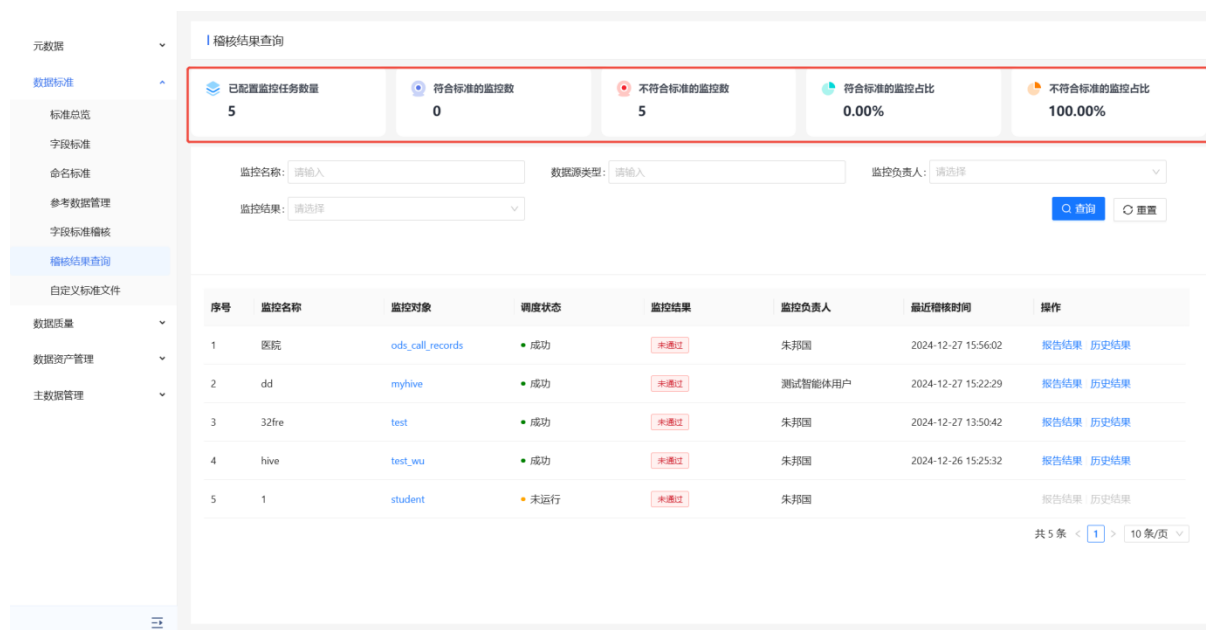


## 稽核结果查询

稽核结果查询模块提供对字段标准稽核任务结果的查看功能，可以查看每次的报告结果和历史稽核任务的结果；支持对稽核任务的查询检索；界面展示目前已配置的监控任务情况和任务执行情况，包括已配置监控任务数量、符合标准的监控数、不符合标准的监控数、符合标准的监控占比、不符合标准的监控占比。

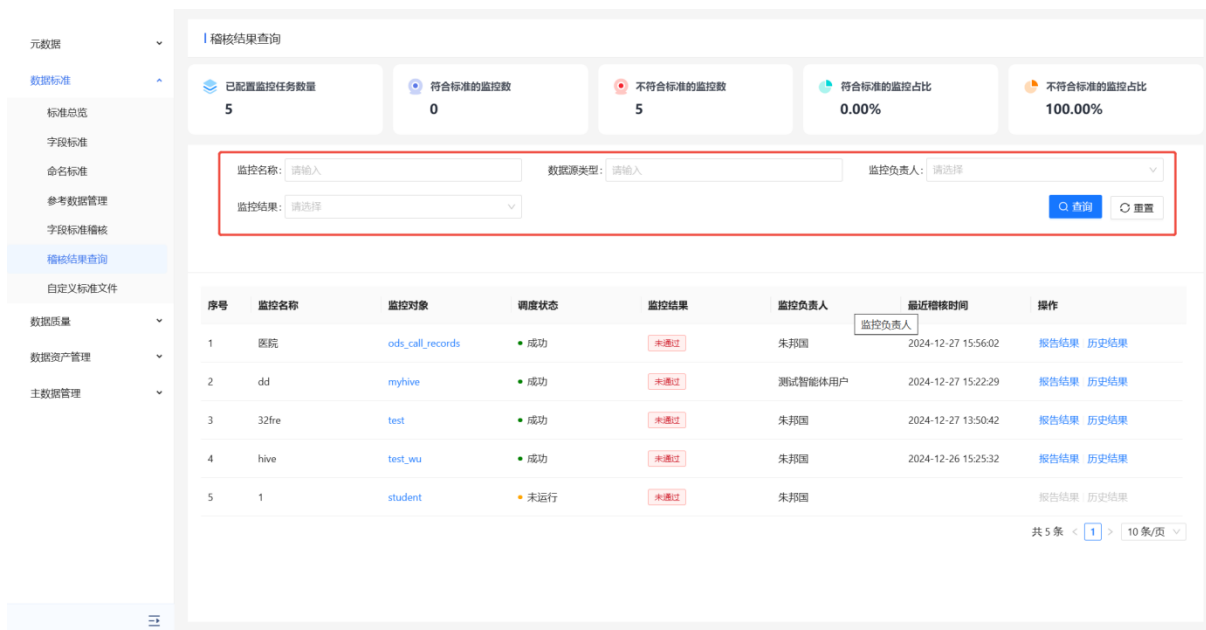
## 稽核结果任务展示面板

集合结果查询页面展示稽核任务展示面板，包括已配置监控任务数量、符合标准的监控数、不符合标准的监控数、符合标准的监控占比及不符合标准的监控占比。



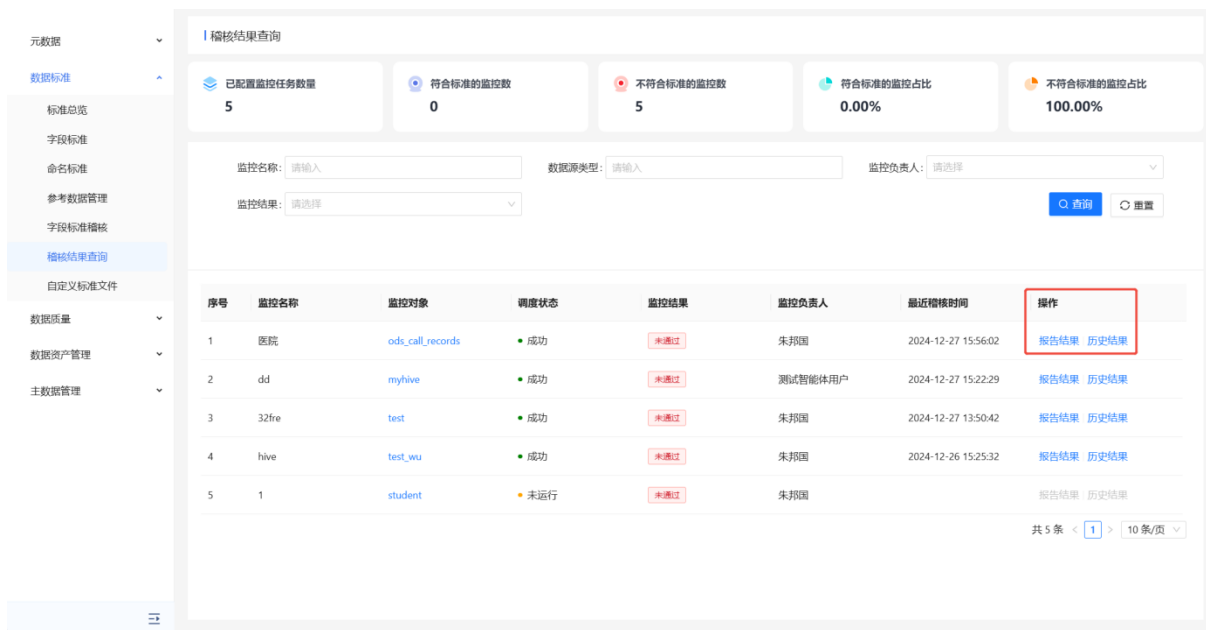
## 查询稽核结果

支持通过监控名称、数据源类型、监控负责人及监控结果进行查询。



## 稽核结果操作

支持查看报告结果和历史结果。



操作	描述
查看报告结果	查看字段标准稽核的结果。点击【报告结果】按钮，进入报告结果详情页面

校验规则	校验规则数量	校验失败数	合格率	通过率	合格率
1	50	1	0	100.00%	0.00%

校验项名称	校验项名称	校验项名称	操作
校验项名称	校验项名称	校验项名称	详情

点击【详情】按钮，查看元数据结果及数据结果

**详情** ×

**元数据**

校验属性	详情	监控结果
数据类型	varchar	通过
长度	255	未通过

**数据**

校验属性	详情	监控结果	操作
数据类型校验	未通过数据 0/25	通过	查看详情
字段长度校验	未通过数据 24/25	未通过	查看详情

查看字段标准稽核的历史结果。点击【历史结果】按钮，进入历史结果详情页面

查看历史结果

序号	稽核名称	稽核内容	稽核状态	稽核结果	稽核负责人	稽核稽核时间	操作
1	稽核	校验项名称	未通过	未通过	吴国超	2024-12-27 15:54:02	查看详情
2	稽核	校验项名称	未通过	未通过	吴国超	2024-12-27 14:54:02	查看详情
3	稽核	校验项名称	未通过	未通过	吴国超	2024-12-27 13:54:02	查看详情
4	稽核	校验项名称	未通过	未通过	吴国超	2024-12-27 12:54:02	查看详情
5	稽核	校验项名称	未通过	未通过	吴国超	2024-12-27 11:54:02	查看详情
6	稽核	校验项名称	未通过	未通过	吴国超	2024-12-27 10:54:02	查看详情

点击【查看结果】按钮，跳转至本条结果的报告结果详情界面，查看历史报告结果详情



## 数据质量

数据质量模块能够按照内置或自定义规则，对具体的字段内容或各类聚合计算指标（例如波动性等）进行质量稽核，并形成相应质量稽核报告，数据质量提供规则模板、任务监控、报告查询、数据质量态势感知等内容。

数据质量的整体使用流程为：

步骤	描述
步骤一：配置质量稽核规则	用户可查看中台内置的系统模板规则，或在已有规则模板的基础上自行定制个性化稽核规则，也可按照系统规范语法，自行编写规则 SQL 语句并保存为模板。
步骤二：创建质量监控任务	针对结构化数据表或多模态数据创建质量监控任务，从系统内置或自定义模板中选择相应稽核规则，配置稽核任务调度信息。
步骤三：查看质量稽核结果	可按照账期查看质量监控任务的稽核结果，并可查看质量异常数据明细。
步骤四：查看整体质量稽核态势	支持用户选择并查看不同时间粒度的中台整体质量稽核任务各类统计数据

## 数据质量总览

数据质量总览界面呈现数据质量全貌，集成了数据质量统计总览、问题数据占比统计、问题趋势、稽核任务数趋势、稽核任务情况明细等信息，以直观的可视化图表清晰地展示数据质量稽核的具体情况。



模块	描述
数据质量总览	<p>此部分主要展示了一定时间范围内的生效稽核规则数量、校验数据总量、问题数据总量、合格数据总量、问题数据占比、合规数据占比等统计值，用户选择不同时间段，统计数值也会相应联动变化</p>
问题数据占比统计	<p>分别以规则类型（行规则、列规则、表规则）以及校验类型（完整性、准确性、一致性、及时性、有效性、唯一性）两个维度整体统计一定时间范围的问题类型占比</p>
问题数历史趋势	<p>以折线图的形式展现一定时间范围的稽核不通过的数据量走势，为用户提供质量趋势判断依据</p> <p><b>问题数历史趋势</b></p>



## 规则模板

规则模板为用户提供数据质量稽核中需要使用到稽核规则，对于结构化数据，当前共有 38 类系统内置稽核规则，按照规则类型分为行规则、列规则、表规则，按照校验类型分为完整性、准确性、一致性、及时性、有效性及唯一性；对于非结构化数据，当前共有 33 类系统内置稽核规则，按照数据类型分为文档类、视频类、音频类、图片类稽核规则。

当前支持用户自行创建结构化数据的自定义规则模板，可通过内置的基础规则或自定义 SQL 语句的方式按需创建自定义规则模板，自定义规则模板支持编辑与删除操作。

规则模板

系统模板 | 自定义模板

规则类型: 全部 | 数据源类型: 数据库 | 模板名称: 规则模板

校验类型: 完整性 | 准确性 | 一致性 | 及时性 | 有效性 | 唯一性

规则名称	校验类型	规则类型	数据源类型	描述	最后修改人	最后修改时间
数据完整性及时性校验	及时性	表规则	MYSQL/HIVE/DORIS	数据完整性及时性校验	系统	2024-12-06 18:29:01
准确性校验-完整性	准确性	列规则	MYSQL/HIVE/DORIS	准确性校验-完整性	系统	2024-12-06 18:29:01
准确性校验-准确性	准确性	列规则	MYSQL/HIVE/DORIS	准确性校验-准确性	系统	2024-12-06 18:29:01
准确性校验-完整性	准确性	列规则	MYSQL/HIVE/DORIS	准确性校验-完整性	系统	2024-12-06 18:29:01
准确性校验-及时性	准确性	列规则	MYSQL/HIVE/DORIS	准确性校验-及时性	系统	2024-12-06 18:29:01
准确性校验-及时性	准确性	列规则	MYSQL/HIVE/DORIS	准确性校验-及时性	系统	2024-12-06 18:29:01
字段关联一致性校验	一致性	表规则	MYSQL/HIVE/DORIS	字段关联一致性校验	系统	2024-12-06 18:29:01
字符串包含	有效性	行规则	MYSQL/HIVE/DORIS	字符串包含	系统	2024-12-06 18:29:01
字段唯一性校验	唯一性	表规则	MYSQL/HIVE/DORIS	字段唯一性校验	系统	2024-12-06 18:29:01
全量唯一性校验	唯一性	表规则	MYSQL/HIVE/DORIS	全量唯一性校验	系统	2024-12-06 18:29:01

共 38 条 | 1 | 2 | 3 | 4 | 10 条/页 | 首页 | 尾页

## 系统内置规则

进入界面后默认展示系统内置的各类稽核规则，可通过上方的规则类型、数据源类型、模板名称、校验类型等筛选符合条件的内置稽核规则，内置稽核规则分为结构化数据的数据库表规则，

以及多模态数据规则。当前共有 38 类结构化表类数据稽核规则，以及 33 类多模态数据稽核规则

模板名称	校验类型	规则类型	数据源类型	描述	最后修改人	最后修改时间
数据库连接及时性校验	及时性	表规则	MYSQL,HIVE,DORIS	数据库连接及时性校验	系统	2024-12-06 18:29:01
波动性校验-老基比	准确性	列规则	MYSQL,HIVE,DORIS	波动性校验-老基比	系统	2024-12-06 18:29:01
波动性校验-间隔周环比	准确性	列规则	MYSQL,HIVE,DORIS	波动性校验-间隔周环比	系统	2024-12-06 18:29:01
波动性校验-新老基比	准确性	列规则	MYSQL,HIVE,DORIS	波动性校验-新老基比	系统	2024-12-06 18:29:01
波动性校验-环比	准确性	列规则	MYSQL,HIVE,DORIS	波动性校验-环比	系统	2024-12-06 18:29:01
波动性校验-同比	准确性	列规则	MYSQL,HIVE,DORIS	波动性校验-同比	系统	2024-12-06 18:29:01
字段关联一致性校验	一致性	表规则	MYSQL,HIVE,DORIS	字段关联一致性校验	系统	2024-12-06 18:29:01
字段非空占	准确性	行规则	MYSQL,HIVE,DORIS	字段非空占	系统	2024-12-06 18:29:01
字段唯一性校验	唯一性	表规则	MYSQL,HIVE,DORIS	字段唯一性校验	系统	2024-12-06 18:29:01
全表唯一性校验	唯一性	表规则	MYSQL,HIVE,DORIS	全表唯一性校验	系统	2024-12-06 18:29:01

点击蓝色模板名称，可以弹出规则详情查看界面，系统内置规则仅支持查看，不支持前端的编辑、删除等操作

**规则模版详情**

模版名称: 波动性校验-定基比

校验类型: 准确性

规则类型: 列级

数据源类型: MYSQL,HIVE,DORIS

描述: 波动性校验-定基比

负责人: admin

## 用户自定义规则

点击【自定义模板】标签进入自定义规则模板界面，可查看对应的模板名称、校验类型、规则类型、数据源类型、描述、最后修改人、最后修改时间等信息，可通过上方的数据源类型、模板名称等筛选符合条件的内置稽核规则

模板名称	校验类型	规则类型	数据源类型	描述	最后修改人	最后修改时间	操作
固定电话	准确性	行规则	MYSQL,DORIS,HIVE	固定电话校验规则	体验用户-互娱网	2025-01-09 10:26:45	编辑 删除
自定义SQL规则	准确性	自定义SQL规则	MySql,Doris,Hive	-	体验用户-互娱网	2025-01-09 10:22:33	编辑 删除
身份证校验	完整性	行规则	MYSQL	33	体验用户-互娱网	2024-12-30 13:57:52	编辑 删除

操作	描述
新建	点击【新建】按钮，右侧弹出配置项界面，需要配置模板名称、规则类型、基础规则、数据源类型、校验类型等必填信息，模板描述为选填。当选择的基础规则不同时，需要配置的具体规则内容会有不同，基础规则可选项包括行规则、列规则、表规则、表间规则、自定义 SQL

数据 新建自定义规则模版 ×

\* 模版名称:

\* 规则类型: 行规则 ▼

\* 基础规则: 正则校验 ▼

\* 正则表达式:

\* 数据源类型: MYSQL × HIVE × DORIS ×

\* 校验类型: 完整性 ▼

模版描述:  0 / 200

取消 确定

当选择的基础规则为自定义 SQL 时，支持用户根据平台提供的 SQL 语法规范自行编写规则实现的 SQL 语句，系统会识别出语句中的自定义参数，并需要对参数配置参数名、参数类型、默认值、参数描述等信息，支持多个参数的配置

×

新建自定义规则模版

\* 模版名称:

\* 规则类型: 自定义SQL规则

\* 数据源类型: MYSQL × HIVE × DORIS ×

\* 校验类型: 完整性

```

1 <!--
2 # 请参考如下所示自定义SQL 填写您自己的SQL。注意参数应该使用'$'符号标记,遵循freemarker语
3 # 系统支持输入参数:
4 #         sourceTable: 绑定表名
5 #         sourceTableColumns: 绑定表字段
6 #         filterConditions: 过滤分区
7 #         ruleInstanceCode: 当前规则实例编码
8 #         testRun: 是否试跑
9 # 必要输出视图:
10 #         quality_rule_error_details_out: 固定输出视图名
11 #         ruleInstanceCode: 当前规则实例编码 (从系统参数获取)
12 #         total: 校验数据条数 (可能是聚合值)
13 #         failed: 校验失败的数据条数 (可能是聚合值)
14 #         error_msg: 校验失败信息
15 #         input_data: 校验结果的输入数据 (可能是聚合值)
16 #         source_data: 校验结果的数据 (可能是聚合值)
17 -->
18 CREATE TEMPORARY VIEW temp_view AS
19 SELECT
20 <#if sourceTableColumns?has_content>
21     <!-- 输出元表内容, 逗号分隔 --><!-- 分隔符, 如果同时有列和函数则额外输出“,” -->
```

模版参数列表

参数编码: name

\* 参数名:  \* 参数类型: ▼ 默认值:

参数描述:  0 / 200

参数编码: age

\* 参数名:  \* 参数类型: ▼ 默认值:

参数描述:  0 / 200

模版描述:  0 / 200

取消
确定

编辑

点击【编辑】按钮, 可对已经创建的自定义模板进行二次编辑

	<div data-bbox="486 212 1364 1691"><h3>编辑自定义规则模板</h3><p>* 模版名称: 固定电话</p><p>* 规则类型: 行规则</p><p>* 基础规则: 正则校验</p><p>* 正则表达式: (0\d{2,3})-\d{7,8}</p><p>* 数据源类型: MYSQL × DORIS × HIVE ×</p><p>* 校验类型: 准确性</p><p>模版描述: 固定电话校验规则</p><p>8 / 200</p><p>取消 确定</p></div>
删除	点击【删除】按钮，会弹出二次确认框，此时点击确定即可删除自定义规则模板



## 任务监控

质量稽核任务为数据的质量提供保障，通过建立监控任务，按照一定的时间周期执行数据质量的稽核任务。

创建质量稽核任务的流程为：

步骤	描述
步骤一：新建监控任务	设置监控对象，选择需要监控的数据源、库、表；设置监控规则，为字段配置用于稽核的质量规则
步骤二：创建调度	配置质量稽核监控任务的信息，包括执行周期、任务类型、运行信息等
步骤三：调度上线/调度下线	将配置好调度信息的监控任务上线/下线，该任务则会根据调度信息按周期执行/停止执行质量稽核。稽核结果需要在报告查询界面查看

质量任务

监控名称:  数据源类型:  监控负责人:

监控名称	监控对象	已启用/规则总数	任务状态	实例状态	调度周期	监控负责人	创建时间	修改时间	操作
test_1127	industry_starlink_transport...	1/1	<span>编辑中</span>	<span>成功</span>	每小时第0分执行一次	管理员	2025-11-27 22:50:58	2025-11-28 01:23:34	<a href="#">编辑</a> <a href="#">删除</a> <a href="#">更多</a>
aaacsssss	businessdata	1/1	<span>编辑中</span>	<span>失败</span>	每年11月6日14时2分执行...	管理员	2025-11-05 18:09:25	2025-11-06 13:58:45	<a href="#">编辑</a> <a href="#">删除</a> <a href="#">更多</a>
ABC	industry_starlink_transport...	1/1	<span>编辑中</span>	<span>失败</span>	每天0时0分执行一次	冯洪涛	2025-10-24 19:55:00	2025-11-14 15:20:09	<a href="#">编辑</a> <a href="#">删除</a> <a href="#">更多</a>
阿达曼的	dtcommon_starlink	1/1	<span>无调度任务</span>	<span>暂无实例</span>	-	冯洪涛	2025-10-21 17:10:48	2025-10-21 17:10:50	<a href="#">编辑</a> <a href="#">删除</a> <a href="#">更多</a>
密稽核测试	industry_starlink_transport...	1/1	<span>上线</span>	<span>失败</span>	每天0时0分执行一次	冯洪涛	2025-10-16 18:20:09	2025-10-16 22:27:38	<a href="#">编辑</a> <a href="#">删除</a> <a href="#">更多</a>
事件名称稽核0925	businessdata	1/1	<span>上线</span>	<span>失败</span>	每天0时0分执行一次	冯洪涛	2025-09-25 23:16:17	2025-09-25 23:16:38	<a href="#">编辑</a> <a href="#">删除</a> <a href="#">更多</a>
事件名称稽核080080	pro	1/1	<span>编辑中</span>	<span>成功</span>	每天0时0分执行一次	冯洪涛	2025-08-08 19:24:27	2025-08-08 19:24:29	<a href="#">编辑</a> <a href="#">删除</a> <a href="#">更多</a>
事件名称稽核	industry_starlink_transport...	1/1	<span>编辑中</span>	<span>失败</span>	每天22时44分执行一次	冯洪涛	2025-08-07 22:41:59	2025-08-07 22:42:01	<a href="#">编辑</a> <a href="#">删除</a> <a href="#">更多</a>
数据质量稽核	industry_starlink_transport...	1/1	<span>编辑中</span>	<span>失败</span>	每天0时0分执行一次	管理员	2025-06-12 16:14:13	2025-06-12 16:14:16	<a href="#">编辑</a> <a href="#">删除</a> <a href="#">更多</a>
值域规则	honda	1/1	<span>无调度任务</span>	<span>暂无实例</span>	-	王飞	2025-05-29 16:17:25	2025-05-29 16:17:29	<a href="#">编辑</a> <a href="#">删除</a> <a href="#">更多</a>

共 25 条 < 1 2 3 > 10 条/页 跳至  页

用户在此界面可以创建数据质量稽核任务，当前支持创建结构化数据对象、主数据、元数据、

# 数据标准等质量稽核任务

## 新增稽核任务

### 1、新建任务

点击【+新建监控】按钮，进入新建监控业务界面

监控名称	监控对象	已启用/规则总数	任务状态	实例状态	调度周期	监控负责人	创建时间	修改时间	操作
test_1127	industry_starlink_transport...	1/1	编辑中	成功	每小时第0分执行一次	管理员	2025-11-27 22:50:58	2025-11-28 01:23:34	编辑 删除 更多
aaassssss	businessdata	1/1	编辑中	失败	每年11月6日14时2分执行...	管理员	2025-11-05 18:09:25	2025-11-06 13:58:45	编辑 删除 更多
ABC	industry_starlink_transport...	1/1	编辑中	失败	每天0时0分执行一次	冯洪涛	2025-10-24 19:55:00	2025-11-14 15:20:09	编辑 删除 更多
阿达曼的	dtcommon-starlink	1/1	无调度任务	暂无实例	-	冯洪涛	2025-10-21 17:10:48	2025-10-21 17:10:50	编辑 删除 更多
图稽核测试	industry_starlink_transport...	1/1	上线	失败	每天0时0分执行一次	冯洪涛	2025-10-16 18:20:09	2025-10-16 22:27:38	编辑 删除 更多
事件名称稽核0925	businessdata	1/1	上线	失败	每天0时0分执行一次	冯洪涛	2025-09-25 23:16:17	2025-09-25 23:16:38	编辑 删除 更多
事件名称稽核080080	pro	1/1	编辑中	成功	每天0时0分执行一次	冯洪涛	2025-08-08 19:24:27	2025-08-08 19:24:29	编辑 删除 更多
事件名称稽核	industry_starlink_transport...	1/1	编辑中	失败	每天22时44分执行一次	冯洪涛	2025-08-07 22:41:59	2025-08-07 22:42:01	编辑 删除 更多
数据质量稽核	industry_starlink_transport...	1/1	编辑中	失败	每天0时0分执行一次	管理员	2025-06-12 16:14:13	2025-06-12 16:14:16	编辑 删除 更多
值域规则	honda	1/1	无调度任务	暂无实例	-	王飞	2025-05-29 16:17:25	2025-05-29 16:17:29	编辑 删除 更多

### 2、设置监控对象

配置基础信息和监控对象。基础信息包括监控名称（必填）、监控负责人（必选）及描述；当数据类型为数据库表时，监控对象配置项包括、数据源类型（必选）、数据库、表（必选）、过滤筛选条件

基础信息

- 监控名称:
- 监控负责人:
- 监控任务描述:

设置监控对象

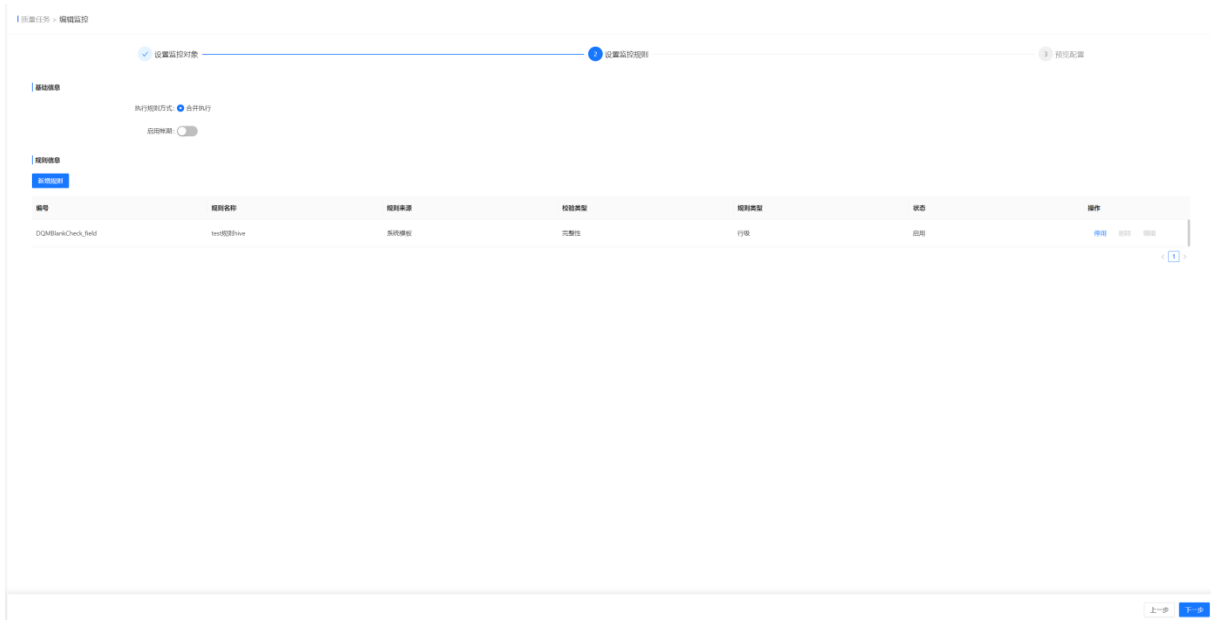
- 数据源类型:
- 数据库:
- 表:
- 过滤筛选条件:

后续操作示例将展示数据库表类型的配置步骤

点击【下一步】按钮，设置监控规则。

### 3、设置监控规则

监控规则界面展示规则的规则名称、规则来源、校验类型、规则类型、状态等。



点击【新建规则】，弹出规则绑定界面，依次选择模板类型、校验类型、规则类型等基础配置后，可选择分属于以上维度的具体规则模板，用户可自定义规则名称，以及具体的规则内容配置，可选输入规则描述，点击确定后即可保存，新建规则默认为启用状态

模版类型:  内置模版  自定义模版

\* 校验类型:  完整性  准确性  有效性  及时性  唯一性  一致性

\* 规则类型:  行规则  自定义SQL规则  列规则  表规则  表间规则

\* 规则名称:

\* 规则模版:

\* 检验字段:

字段名	字段类型	字段描述
<input checked="" type="radio"/> age	int	age
<input type="radio"/> name	string	name
<input type="radio"/> id	int	id

< 1 >

\* 参考数据:

参考数据中文名	参考数据英文名	状态
+ 民族	nationality	2
+ 性别	gender	2
+ 撒旦	e	3
+ 宗教信仰	religion	2
+ 积分兑换渠道类型	exchange_channel_type	2
+ 产品实例状态	is_online	2
+ 学号	stu_nu	2

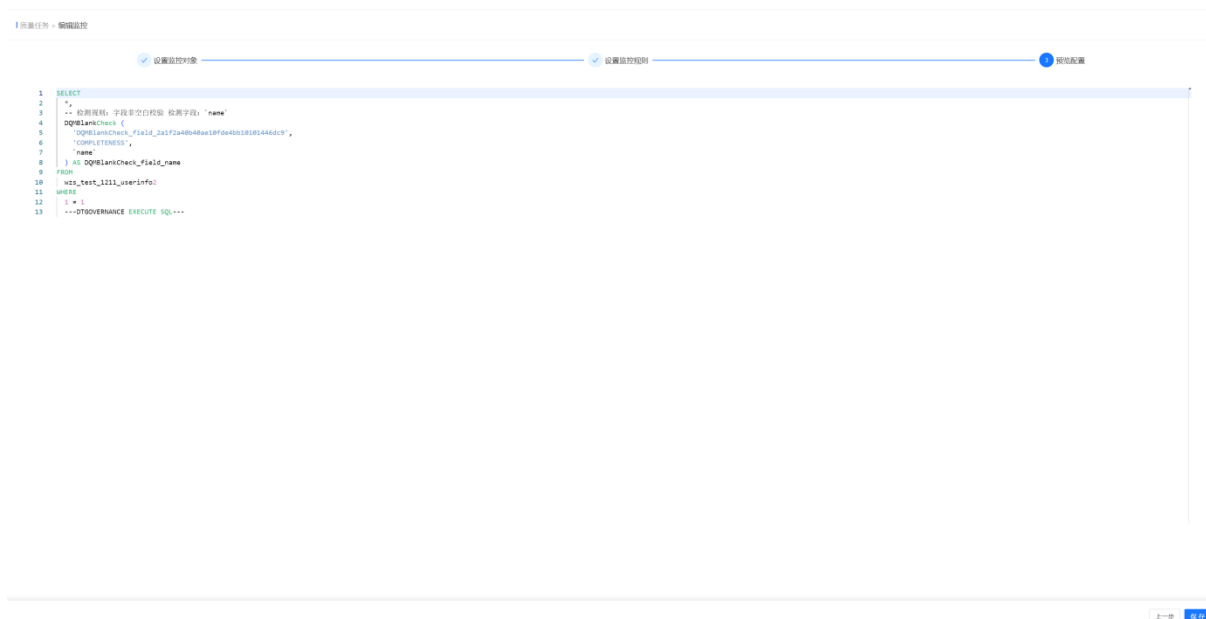
< 1 >

请输入参考数据

描述:

0 / 200

完成规则配置后点击下一步，进入规则配置预览界面，在此界面可以查看已配置的稽核规则的具体 SQL 语句，点击保存即可完成质量稽核任务的配置



```
1 SELECT
2 *
3 -- 校验规则：字段非空校验 校验字段: 'name'
4 DQ@lanhcheck (
5   'DQ@lanhcheck_field_2a1f2a48048e18f0e40b018181446c3',
6   'COMPLETENESS',
7   'name'
8 ) AS DQ@lanhcheck_field_name
9 FROM
10 wds_test_1211_userinfo
11 WHERE
12 1 = 1
13 ---DTGOVERNANCE EXECUTE SQL---
```

任务创建完成后，后续在监控任务界面进行调度配置。

## 创建调度

点击【更多】->【创建调度】，在右侧抽屉中进行调度配置。配置信息包括基本信息、调度信息、任务信息及运行信息

## | 基本信息

任务名称: 1-稽核任务

\* 优先级: 中

## | 调度信息

\* 执行周期: 日

执行时间: 每天 00:00

crontab表达式: 0 0 0 \* \* ?

## | 任务信息

\* 任务类型: FLINK

yarn队列:

## | 运行信息

\* 调度集群组: default

业务时间偏移: -1

重试次数: 0

重试时间间隔(s): 30

\* 任务超时时长(s): 7200

取消

保存

在基本信息中配置任务名称和任务优先级，任务优先级分为高、中、低及最低四种

基本信息

任务名称: 1-稽核任务

调度信息

\* 优先级: 中

高

**中**

低

最低

\* 执行周期:

执行时间:

crontab表达式: 0 0 0 \*\* ?

任务信息

\* 任务类型: FLINK

yarn队列:

运行信息

\* 调度集群组: default

业务时间偏移: -1

重试次数: 0

重试时间间隔(s): 30

\* 任务超时时长(s): 7200

取消

保存

在调度信息中配置执行周期和执行时间，执行周期分为分钟、小时、日及月四种

## 调度配置



### 基本信息

任务名称: 1-稽核任务

\* 优先级: 中

### 调度信息

\* 执行周期: 日

执行时间: 分钟  
小时

crontab表达式: 日

月

### 任务信息

\* 任务类型: FLINK

yarn队列:

### 运行信息

\* 调度集群组: default

业务时间偏移: -1

重试次数: 0

重试时间间隔(s): 30

\* 任务超时时长(s): 7200

取消

保存

在任务信息中配置 yarn 队列

基本信息

任务名称: 1-稽核任务

\* 优先级: 中

调度信息

\* 执行周期: 日

执行时间: 每天 00:00

crontab表达式: 0 0 0 \* \* ?

任务信息

\* 任务类型: FLINK

yarn队列:

运行信息

\* 调度集群组: default

业务时间偏移: -1

重试次数: 0

重试时间间隔(s): 30

\* 任务超时时长(s): 7200

取消

保存

在运行信息中配置调度集群组、业务时间偏移、重试次数、重试时间间隔及任务超时时长

## | 基本信息

任务名称: 1-稽核任务

\* 优先级: 中

## | 调度信息

\* 执行周期: 日

执行时间: 每天 00:00

crontab表达式: 0 0 0 \* \* ?

## | 任务信息

\* 任务类型: FLINK

yarn队列:

## | 运行信息

\* 调度集群组: default

业务时间偏移: -1

重试次数: 0

重试时间间隔(s): 30

\* 任务超时时长(s): 7200

取消

保存

点击【保存】完成调度配置

## 查询质量稽核任务

支持输入监控名称、监控对象及监控负责人进行稽核任务的查询

序号	监控名称	监控对象	已报警/预期总数	任务状态	实例状态	调度周期	监控负责人	创建时间	修改时间	操作
1	test监控A	java	1/1	编辑中	就绪	每天10时55分执行一次	体验用户-互联网	2025-01-10 16:53:47	2025-01-10 16:53:47	编辑 删除 更多
2	test监控B	java	1/1	编辑中	就绪	每天10时55分执行一次	体验用户-互联网	2025-01-10 16:48:17	2025-01-10 16:48:17	编辑 删除 更多
3	test监控	java	1/1	上线中	就绪	每天10时45分执行一次	体验用户-互联网	2025-01-10 16:42:30	2025-01-10 16:42:30	编辑 删除 更多
4	test	java	1/1	上线中	就绪	每天10时37分执行一次	体验用户-全部产品	2025-01-10 10:34:22	2025-01-10 10:34:22	编辑 删除 更多
5	账单数据质量稽核	java	5/5	上线中	就绪	每天15时21分执行一次	体验用户-互联网	2024-12-26 14:43:25	2025-01-06 17:30:29	编辑 删除 更多
6	自定义任务编辑7140714024407227	java	1/1	编辑中	就绪	每天2时0分执行一次	测试组	2024-12-18 16:48:02	2025-01-07 09:53:17	编辑 删除 更多
7	自定义任务编辑7140548881419751	java	1/1	编辑中	就绪	每天2时0分执行一次	测试组	2024-12-18 14:56:27	2024-12-18 14:56:28	编辑 删除 更多
8	自定义任务编辑7113804362493241	java	1/1	编辑中	繁忙实例	每天2时0分执行一次	测试组	2024-12-11 16:01:03	2024-12-11 16:01:04	编辑 删除 更多
9	自定义任务编辑71138028515071365	java	1/1	无报警任务	繁忙实例	-	测试组	2024-12-11 15:40:52	2024-12-11 15:40:53	编辑 删除 更多
10	自定义任务编辑71138022121219126	java	1/1	编辑中	就绪	每天2时0分执行一次	测试组	2024-12-11 15:30:21	2024-12-11 15:30:22	编辑 删除 更多

## 稽核任务操作

支持编辑、删除及更多（详情、创建调度、调度上线及调度下线）操作。对稽核任务进行调度，对于未调度上线的任务进行编辑

操作	描述
编辑	<p>编辑所选稽核任务。点击【编辑】按钮，进入标准稽查任务的编辑界面，编辑稽查任务的基础信息、规则配置、依次点击【下一步】-&gt;【保存】完成任务编辑</p> 
调度上线	<p>将所选稽核任务上线，即时生效。点击【更多】-&gt;【调度上线】按钮，出现上线成功的提示，即完成了调度上线操作</p> 
调度下线	<p>下线该稽核任务的调度配置。点击【更多】-&gt;【调度下线】按钮，在弹窗界面二次确认是否要下线调度。点击【确认】完成调度下线</p>

确定下线此任务吗?

取消

确定

查看所选稽核任务的具体信息。直接点击蓝色的稽核任务名称，进入稽核任务详情展示页，展示内容包括基础信息、稽核规则以及执行配置

监控任务 > test监控doris

基础信息 规则 执行配置

### 监控对象

数据源类型: DORIS

数据源: M0001bd0d6e47bd0f2b82d164ceca7e2dda71

库: pro

表: DTEXCHANGE\_DORIS01

### 基础信息

监控名称: test监控doris

监控对象描述:

监控负责人: 体验用户-互联网

创建时间: 2025-01-10 16:53:47

修改时间: 2025-01-10 16:53:47

监控任务 > 任务的稽核规则

基础信息 规则 执行配置

编号	规则名称	参数	规则来源	校验类型	规则类型
DQMNormalCheck_66d3687a...	字段非空校验	["show", true, "parameterCode", "checkFieldName", "parameterDesc", "校验字段", ...]	系统稽核	完整性	行级
DQMEnumCheck_66d371587...	枚举校验	["show", true, "parameterCode", "checkFieldName", "parameterDesc", "校验字段", ...]	系统稽核	准确性	行级
DQMLengthCheck_66d41660...	字段长度校验	["show", true, "parameterCode", "checkFieldName", "parameterDesc", "校验字段", ...]	系统稽核	有效性	行级
MAX_653099464a3856a1183...	最大值	["show", true, "parameterCode", "checkFieldName", "parameterDesc", "校验字段", ...]	系统稽核	准确性	列级
MIN_3736f5d81afca02ab03...	最小值	["show", true, "parameterCode", "checkFieldName", "parameterDesc", "校验字段", ...]	系统稽核	准确性	列级

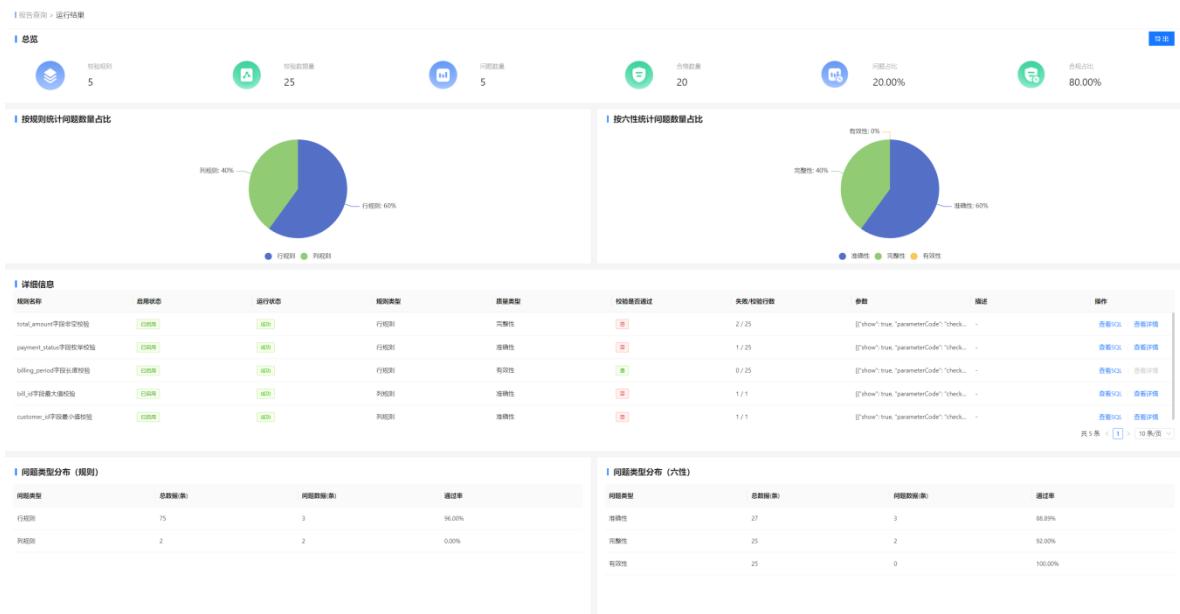
查看详情

	<p>监控任务 &gt; 账单数据质量稽核</p> <p>基础信息 规则 <b>执行配置</b></p> <p>规则执行方式: 合并执行</p> <p>合并后SQL预览:</p> <pre> 1 SELECT 2 * 3 -- 检测规则: 字段非空校验 检测字段: `total_amount` 4 DQWNotNullCheck ( 5   'DQWNotNullCheck_field_9b8f1e11076371959b1ebee961f3f4f6', 6   'COMPLETENESS', 7   `total_amount` 8 ) AS DQWNotNullCheck_field_total_amount, 9 -- 检测规则: 枚举校验 检测字段: `payment_status` 10 DQWEnumCheck ( 11   'DQWEnumCheck_field_01511567fe87e84fa0d153b32e2aeb2f', 12   'ACCURACY', 13   `payment_status`, 14   'PAID,UNPAID' 15 ) AS DQWEnumCheck_field_payment_status, 16 -- 检测规则: 字段长度校验 检测字段: `billing_period` 17 DQWLengthCheck ( 18   'DQWLengthCheck_field_416d00180a869231cfff6dc548cad2da', 19   'VALIDITY', 20   `billing_period`, </pre>
删除	<p>删除所选稽核任务。点击【删除】按钮，在弹出窗口中确认是否删除任务，点击【确定】按钮完成删除</p> 
测试执行	<p>当新建的稽核任务配置完成后需要手动测试时，可以点击测试执行，提示成功，此时将会忽略调度配置，直接执行一次稽核任务</p> 

## 报告查询

报告查询界面会展示全量稽核任务的运行与稽核情况，用户可以查看每个质量监控任务最新账期实例的执行结果以及稽核结果，也可以在此查看历史账期的执行情况。

质量报告分为总览、问题数量占比统计、异常质量详细信息、问题类型分布统计等部分，用户可以直观的看到本次质量稽核的数据统计，并对具体的质量异常数据进行分析改善。质量报告支持导出。



此界面可以查看结构化数据、主数据的质量稽核报告，点击【报告结果】按钮，可以查看最新一次账期的稽核结果

**报告概览**

监控开始时间: 2025-01-17 14:46:21 | 监控名称: 账目核对 | 监控结果: 全部

数据源类型: 账目核对 | 监控负责人: 账目核对

**数据质量** | 元数据质量 | 数据分布质量 | 多维数据质量

序号	规则名称	监控对象	质量状态	监控结果	监控负责人	最近稽核时间	操作
1	sql监控divs	DBEXCHANGE_DORIS1	通过	通过	体验用户-互联网	2025-01-17 14:46:21	查看详情 历史结果
2	sql监控hive	wsq_tmt_1211_userinfo2	通过	通过	体验用户-互联网	2025-01-16 23:00:10	查看详情 历史结果
3	sql监控	wsq_tmt_1211_userinfo2	通过	通过	体验用户-互联网	2025-01-16 23:00:10	查看详情 历史结果
4	sql	wsq_tmt_1211_userinfo2	通过	通过	体验用户-全部产品	2025-01-17 10:37:14	查看详情 历史结果
5	账单数据质量稽核	ods_billing_hive	通过	通过	体验用户-互联网	2025-01-16 23:00:10	查看详情 历史结果
6	由源数据质量稽核1134511602407227	dtxchange_mysql_source_20240919	通过	通过	账目核对	2025-01-10 22:31:54	查看详情 历史结果
7	由源数据质量稽核11345548681619781	dtxchange_mysql_source_20240919	通过	通过	账目核对	2024-12-18 14:57:20	查看详情 历史结果
8	由源数据质量稽核11339040524078141	dtxchange_mysql_source_20240919	通过	暂无数据	账目核对	-	查看详情 历史结果
9	由源数据质量稽核11339038515073365	dtxchange_mysql_source_20240919	无数据	暂无数据	账目核对	-	查看详情 历史结果
10	由源数据质量稽核1133902212119128	dtxchange_mysql_source_20240919	通过	暂无数据	账目核对	2024-12-11 15:31:26	查看详情 历史结果

共 12 条 < > 列表页 筛选 语言

## 查看最新质量报告

报告模块	描述
数据质量总览	<p>此部分主要展示了当前账期的生效稽核规则数量、校验数据总量、问题数据总量、合格数据总量、问题数据占比、合规数据占比等统计值</p>
问题数据占比统计	<p>分别以规则类型（行规则、列规则、表规则）以及校验类型（完整性、准确性、一致性、及时性、有效性、唯一性）两个维度统计当前账期的质量问题类型占比</p>

<p>详细信息</p>	<p>这里会展示稽核任务中全量稽核规则的执行情况，包括规则名称、启用状态、运行状态、规则类型、质量类型、校验是否通过、失败/校验行数（即未通过校验的数据行数/参与稽核的总数据行数）、参数、描述等信息</p> <p>对于未通过的稽核规则，可以点击查看详情展示具体的未通过数据详情，包括输入数据、错误信息、完整信息</p> <p>点击查看 SQL，可以展示该稽核规则的具体执行语句</p> <pre> SQL详情 1 SELECT 2 *, 3 -- 检测规则: 字段非空校验 检测字段: `total_amount` 4 DQ\$NonnullCheck ( 5   'DQ\$NonnullCheck_field_9b8f1a11076371059b1e9a961f3f4f6', 6   'COMPLETENESS', 7   `total_amount` 8 ) AS DQ\$NonnullCheck_field_total_amount 9 FROM 10 ods_billing_hive 11 WHERE 12 1 = 1 13 ---DTGOVERNANCE EXECUTE SQL---</pre>
<p>问题类型分布统计</p>	<p>以折线图的形式展现一定时间范围的质量稽核任务的数量走势，为用户提供质量趋势判断依据</p>
<p>稽核任务情况统计</p>	<p>分别以规则类型（行规则、列规则、表规则）以及校验类型（完整性、准确性、一致性、及时性、有效性、唯一性）两个维度统计当前账期的质量问题具体情况，包括问题类型、总数据条数、问题数据条数、通过率等信息</p>

### 查看历史质量报告

点击【历史结果】按钮，即可进入当前稽核任务的历史账期稽核报告列表，能够看到每个账期的稽核任务的执行状态、监控结果等情况，点击查看结果即可进入相应稽核账期的质量报告详情界面，界面内容参考 6.4.1.4.1 内容

报告查询

监控开始时间: 请选择日期 监控名称: 请输入 监控结果: 全部

数据源类型: 请选择 监控负责人: 请选择

数据源类型: 请选择 监控负责人: 请选择

序号	监控名称	监控对象	健康状态	监控结果	监控负责人	最近稽核时间	操作
1	test监控	DTXCHANGE_D08501	健康	成功	林超用户-互联网	2025-01-17 14:46:21	查看详情 历史记录
2	test监控	ws_test_1211_userinfo2	健康	成功	林超用户-互联网	2025-01-16 23:00:10	查看详情 历史记录
3	test监控	ws_test_1211_userinfo2	健康	成功	林超用户-互联网	2025-01-16 23:00:10	查看详情 历史记录
4	test	ws_test_1211_userinfo2	健康	成功	林超用户-全部产品	2025-01-17 10:37:14	查看详情 历史记录
5	账单数据质量稽核	odc_billng_bill	健康	成功	林超用户-互联网	2025-01-16 23:00:10	查看详情 历史记录
6	自由职业任务稽核(11345116824807227)	dtxchange_mysql_source_20240819	健康	成功	朱莉霞	2025-01-19 22:31:54	查看详情 历史记录
7	自由职业任务稽核(11345048881619781)	dtxchange_mysql_source_20240819	健康	成功	朱莉霞	2024-12-18 14:57:20	查看详情 历史记录
8	自由职业任务稽核(11339040024878141)	dtxchange_mysql_source_20240819	健康	失败	朱莉霞	-	查看详情 历史记录
9	自由职业任务稽核(11339028515873365)	dtxchange_mysql_source_20240819	无健康数据	暂无数据	朱莉霞	-	查看详情 历史记录
10	自由职业任务稽核(113390227219128)	dtxchange_mysql_source_20240819	健康	成功	朱莉霞	2024-12-11 15:31:26	查看详情 历史记录

共 12 条 < 1 > 10 条/页 < 筛选 > 页

报告查询 > 历史稽核

序号	监控名称	是否任务	监控对象	执行状态	稽核结果	稽核负责人	执行稽核时间	操作
1	账单数据质量稽核	否	odc_billng_bill	成功	失败	林超用户-互联网	2025-01-16 15:21:13	查看详情
2	账单数据质量稽核	否	odc_billng_bill	成功	失败	林超用户-互联网	2025-01-15 15:21:11	查看详情
3	账单数据质量稽核	否	odc_billng_bill	成功	失败	林超用户-互联网	2025-01-14 15:21:11	查看详情
4	账单数据质量稽核	否	odc_billng_bill	成功	失败	林超用户-互联网	2025-01-13 15:21:10	查看详情
5	账单数据质量稽核	否	odc_billng_bill	成功	失败	林超用户-互联网	2025-01-10 15:21:11	查看详情
6	账单数据质量稽核	否	odc_billng_bill	成功	失败	林超用户-互联网	2025-01-09 15:21:13	查看详情
7	账单数据质量稽核	否	odc_billng_bill	成功	失败	林超用户-互联网	2025-01-08 15:21:13	查看详情
8	账单数据质量稽核	否	odc_billng_bill	成功	失败	林超用户-互联网	2025-01-07 15:21:11	查看详情
9	账单数据质量稽核	否	odc_billng_bill	成功	失败	林超用户-互联网	2025-01-06 15:21:11	查看详情
10	账单数据质量稽核	否	odc_billng_bill	成功	失败	林超用户-互联网	2025-01-05 15:21:13	查看详情

共 23 条 < 1 > 2 > 10 条/页 < 筛选 > 页

## 查询质量稽核报告

可通过监控开始时间、监控名称、监控结果、数据源类型、监控负责人等条件筛选相应的稽核任务及其报告

报告查询

监控开始时间: 请选择日期 监控名称: 请输入 监控结果: 全部

数据源类型: 请选择 监控负责人: 请选择

数据源类型: 请选择 监控负责人: 请选择

序号	监控名称	监控对象	健康状态	监控结果	监控负责人	最近稽核时间	操作
1	test监控	DTXCHANGE_D08501	健康	成功	林超用户-互联网	2025-01-17 14:46:21	查看详情 历史记录
2	test监控	ws_test_1211_userinfo2	健康	成功	林超用户-互联网	2025-01-16 23:00:10	查看详情 历史记录
3	test监控	ws_test_1211_userinfo2	健康	成功	林超用户-互联网	2025-01-16 23:00:10	查看详情 历史记录
4	test	ws_test_1211_userinfo2	健康	成功	林超用户-全部产品	2025-01-17 10:37:14	查看详情 历史记录
5	账单数据质量稽核	odc_billng_bill	健康	成功	林超用户-互联网	2025-01-16 23:00:10	查看详情 历史记录
6	自由职业任务稽核(11345116824807227)	dtxchange_mysql_source_20240819	健康	成功	朱莉霞	2025-01-19 22:31:54	查看详情 历史记录
7	自由职业任务稽核(11345048881619781)	dtxchange_mysql_source_20240819	健康	成功	朱莉霞	2024-12-18 14:57:20	查看详情 历史记录
8	自由职业任务稽核(11339040024878141)	dtxchange_mysql_source_20240819	健康	失败	朱莉霞	-	查看详情 历史记录
9	自由职业任务稽核(11339028515873365)	dtxchange_mysql_source_20240819	无健康数据	暂无数据	朱莉霞	-	查看详情 历史记录
10	自由职业任务稽核(113390227219128)	dtxchange_mysql_source_20240819	健康	成功	朱莉霞	2024-12-11 15:31:26	查看详情 历史记录

共 12 条 < 1 > 2 > 10 条/页 < 筛选 > 页

## 数据资源管理

数据资源管理是在元数据基础上提供的企业数据目录管理模块，涵盖全局数据检索、元数据详情查看、个人数据管理等功能。可以帮助用户统一查询浏览，更好地理解和使用数据。

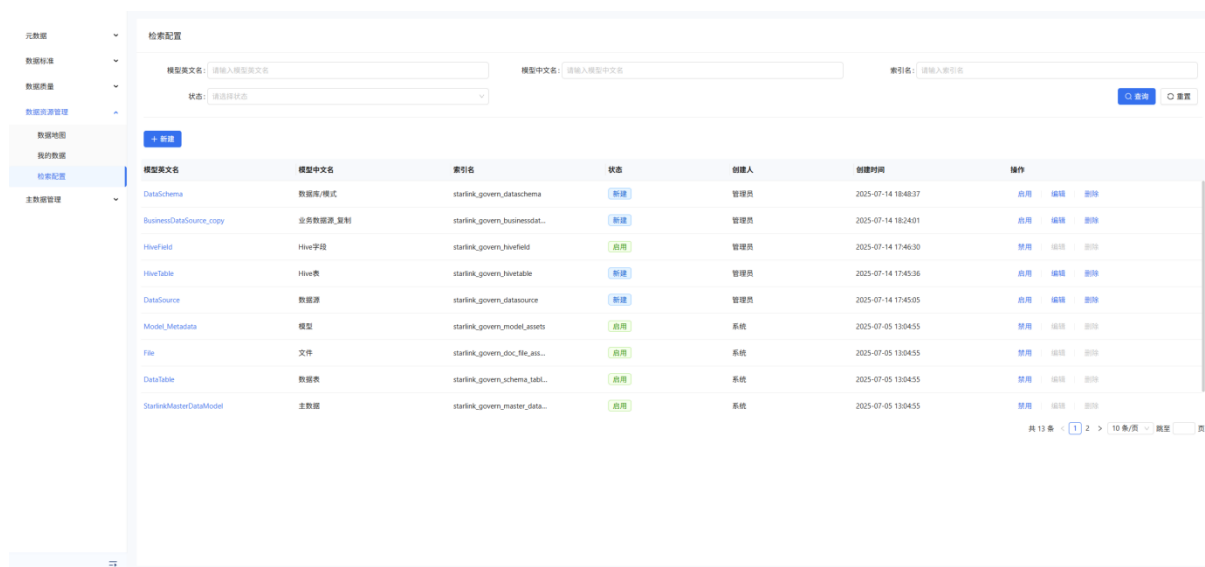
### 数据资产管理使用流程:

步骤	描述
步骤一：元数据采集/多模态元数据采集	元数据采集是数据资产使用的前提，需在 <b>6.2 元数据管</b>

集	理模块配置好元数据采集任务。
步骤二（可选）：检索配置	
步骤三：数据查询检索	通过元数据采集来的数据信息会接入到数据地图，用户可以在数据地图中快速查找支持的各类数据，并可进行查看、收藏、创建 API 等操作。
步骤四：个人数据管理	用户可以在个人数据中心查看个人负责和收藏的数据资源，以及所负责资源的访问情况。

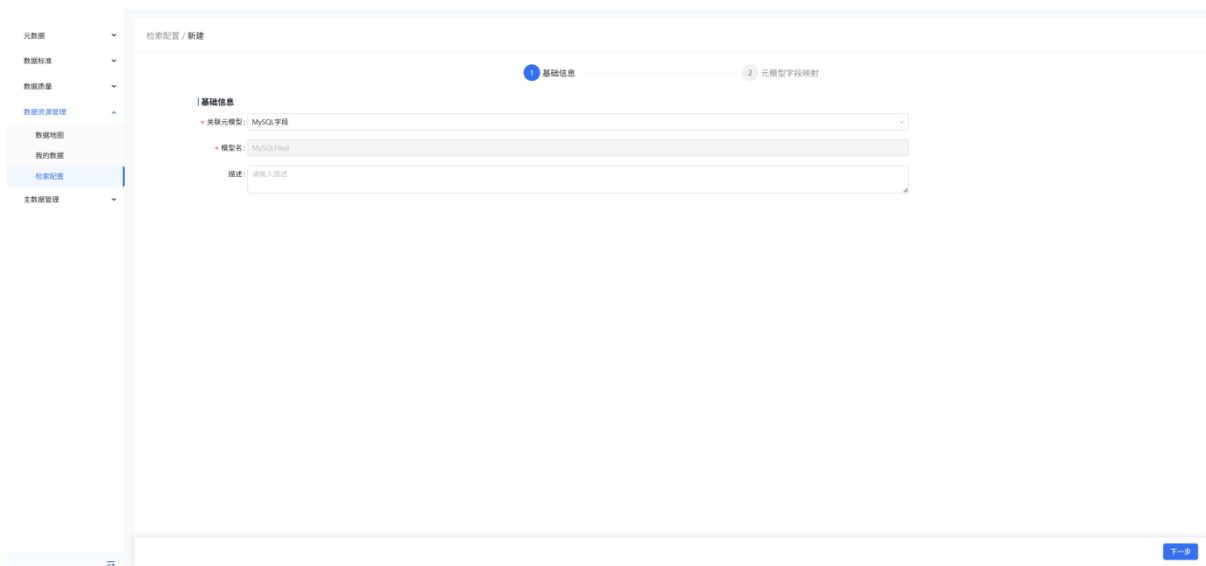
### 1.1.1.1. 检索配置

检索配置用于控制在数据地图展示和搜索的资源内容，与元模型紧密关联。状态为“启用”的检索模型即可在数据地图展示和检索，目前系统内置的模型、主数据以及图片、音频等多模态数据都可以在此处进行控制。

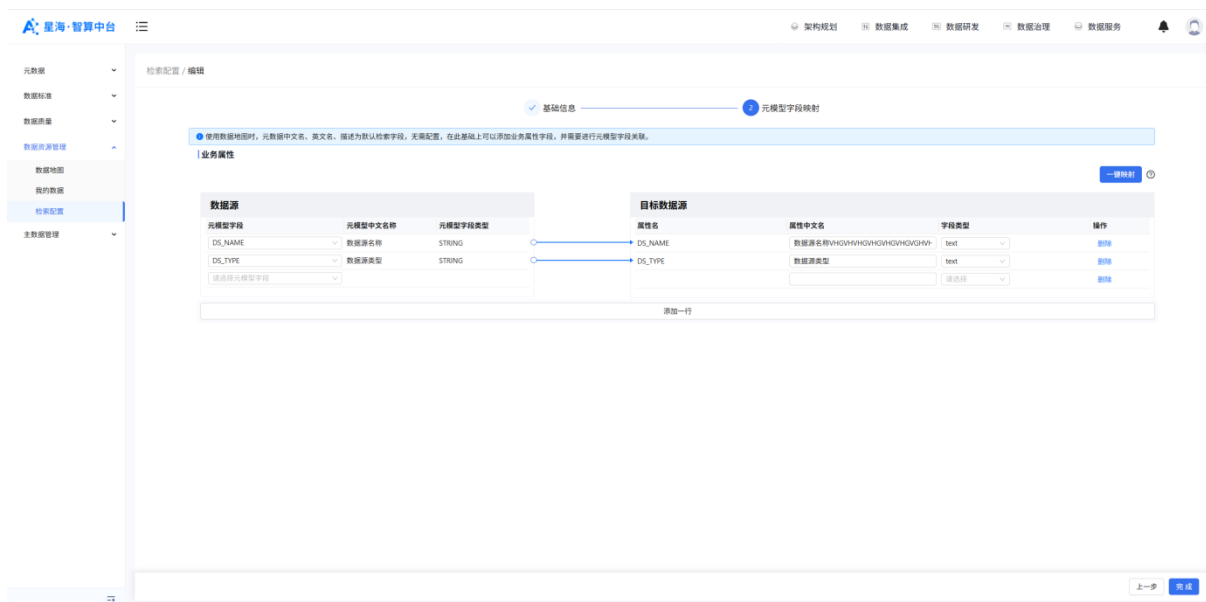


如果用户想新建一种资源类型并在数据地图展示，可以新建检索项并启用。

点击页面【新建】按钮，选择关联的元模型，点击【下一步】，可以配置元数据的其他检索项。



在配置时需注意，元数据中文名、英文名、描述为默认检索字段，无需配置，在此基础上可以添加业务属性字段，并需要进行元模型字段关联。



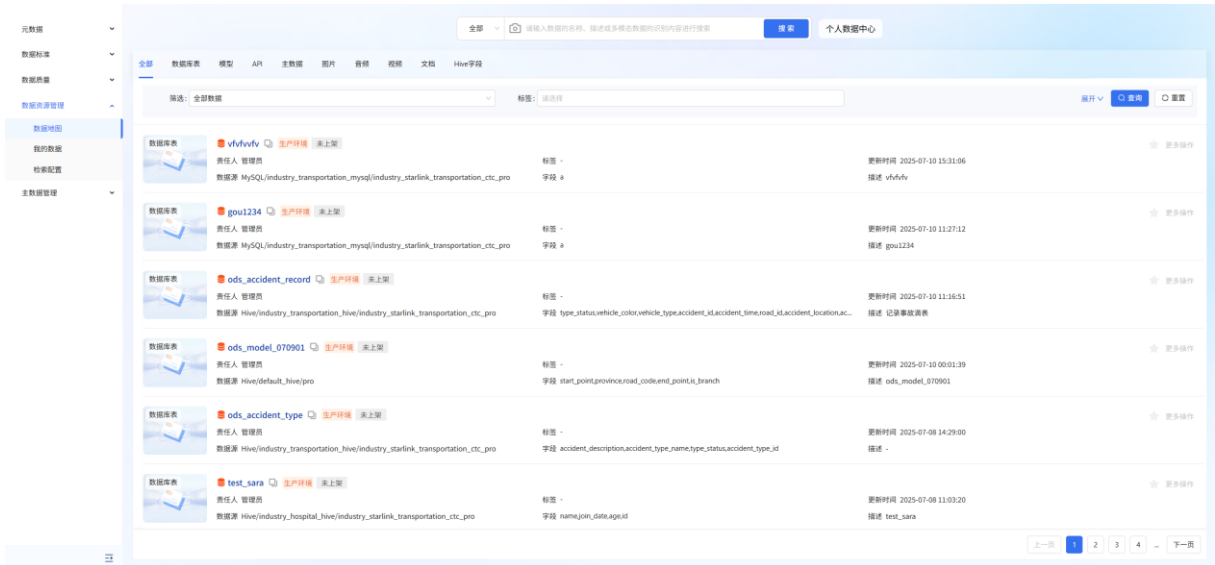
配置好检索字段后，点击【完成】完成创建，可在列表中点击【启用】，该类型的资源即可在数据地图中展示。

### 1.1.1.2. 数据地图

经过元数据采集后的数据即可接入数据地图，可以在数据地图进行检索和详请查看，当前系统支持接入以下类型的数据资源：

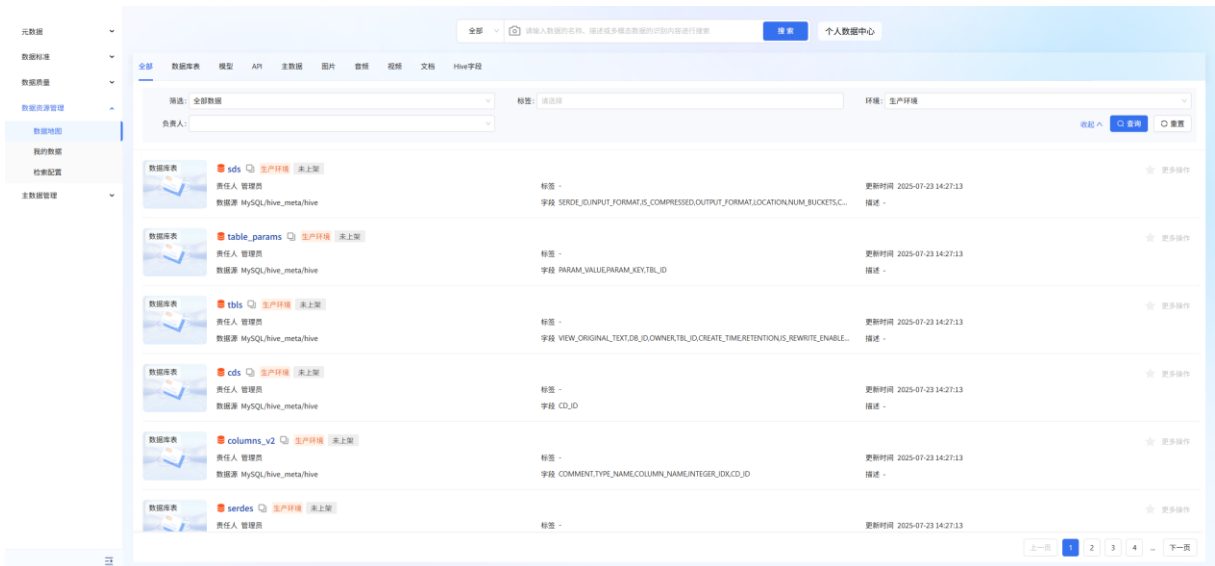
- 数据库表
- 模型（发布状态）
- API（发布状态）

- 主数据
- 视频
- 文档

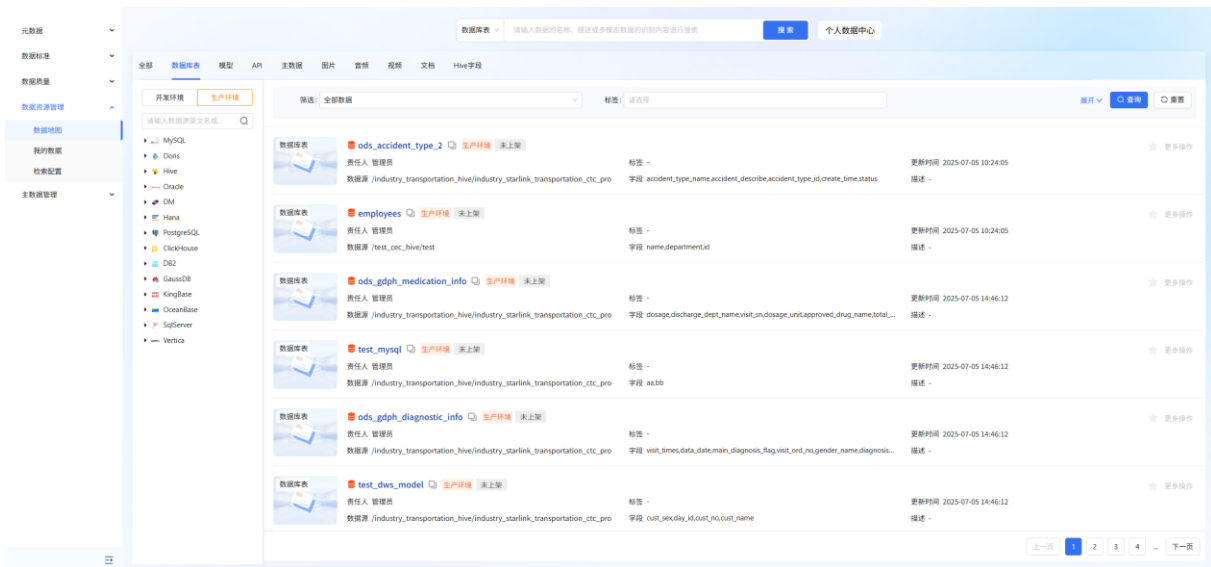


数据地图提供统一的检索功能，页面上侧为统一检索入口，输入资产名称等关键字，点击搜索按钮可以进行检索。

点击列表区域的展开按钮，可以触发高级检索，支持使用环境、团队、负责人、资产标签等字段对资源进行高级搜索。



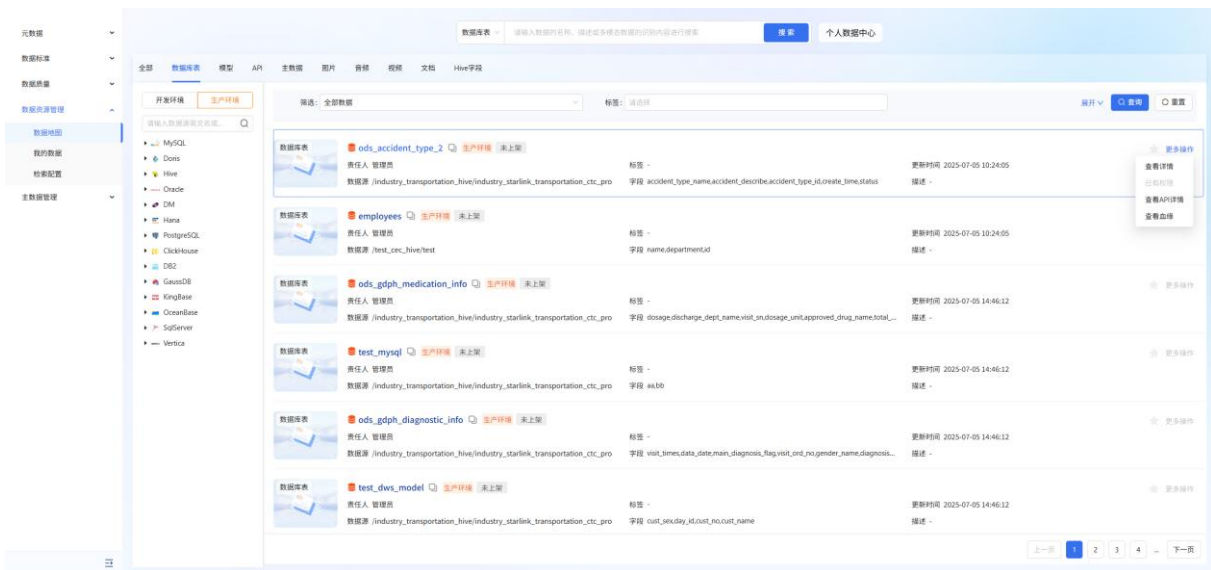
数据地图支持在目录树中选择任意数据节点，或切换不同的数据类型，可以查看数据资源实体情况。



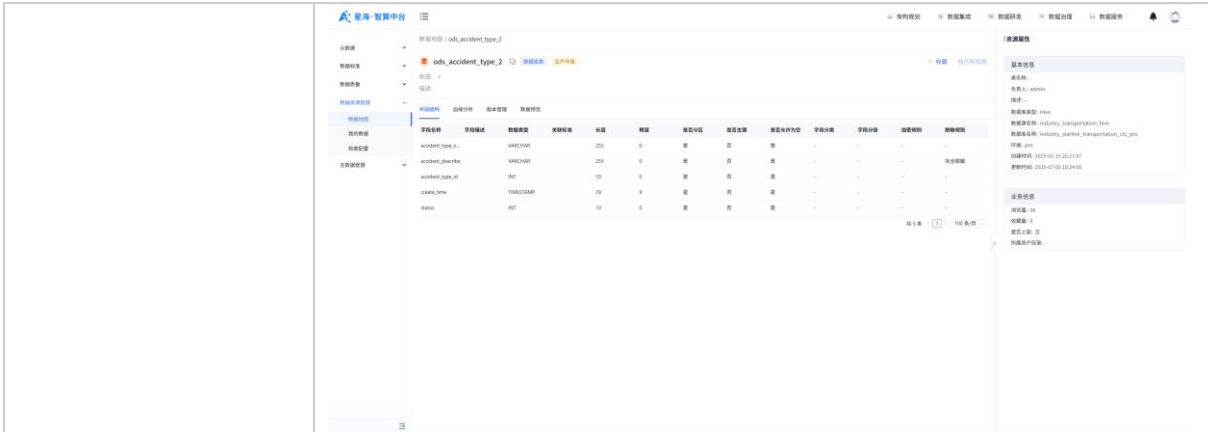
在查找到的资产列表中，单击目标对象名称，即可进入详情页面查看资产元数据信息。

### ● 资源操作

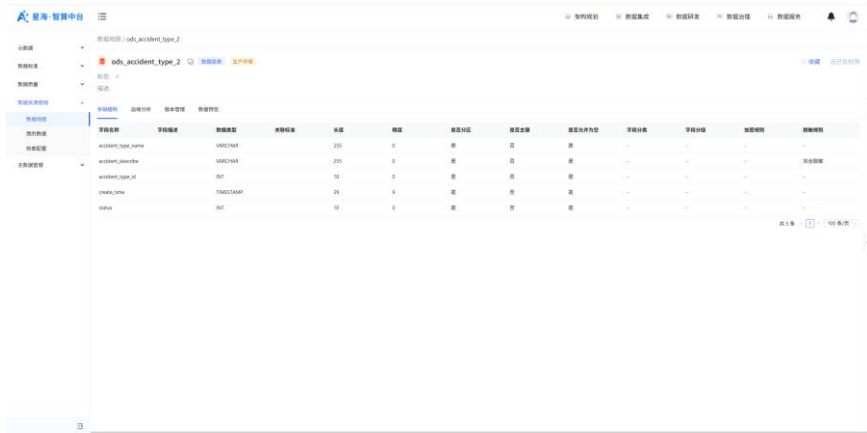
点击目标资源的【更多操作】按钮，可以进行资源详情查看、权限申请、创建 API 等操作。



操作	描述
查看详情	点击【更多操作】->【查看详情】，查看资源详情信息，包括资源基本信息（所属目录、描述）、资产属性（基本信息、基本属性）等



点击箭头，可以收起右侧资源属性菜单栏

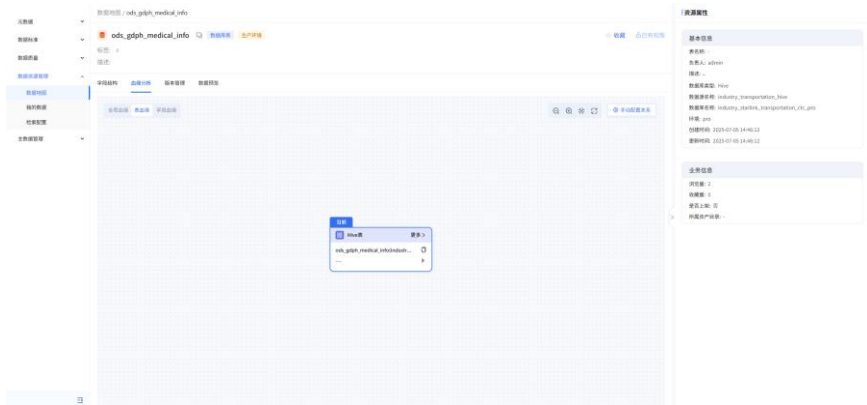


权限申请（仅限数据库表）

点击【更多操作】->【权限申请】，在右侧弹窗中为团队或个人申请数据库表的权限，选择具体的数据库表、权限类型以及申请时常，填写申请理由，即可进行权限申请。

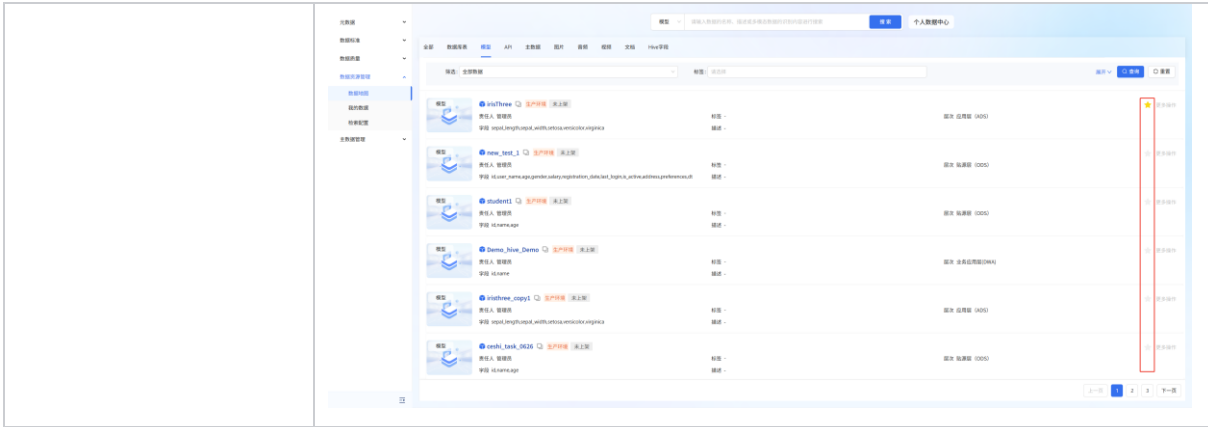
查看血缘（仅限数据库表）

点击【更多操作】->【查看血缘】，可以快速定位至数据库表的血缘分析 tab 页



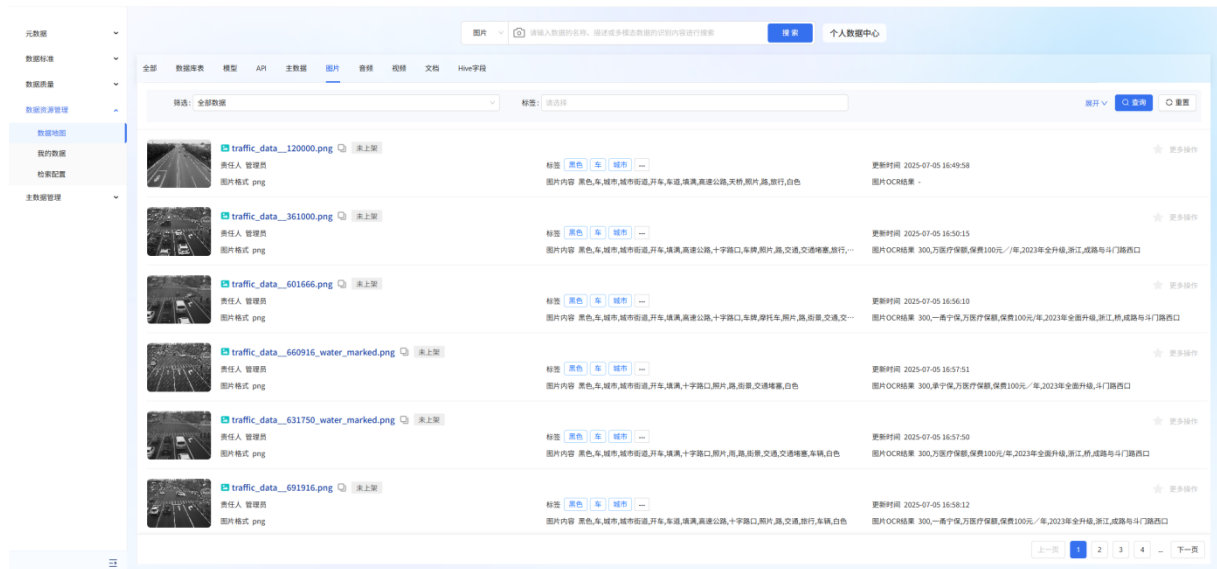
收藏

点击目标资源右侧的星星，收藏目标资源



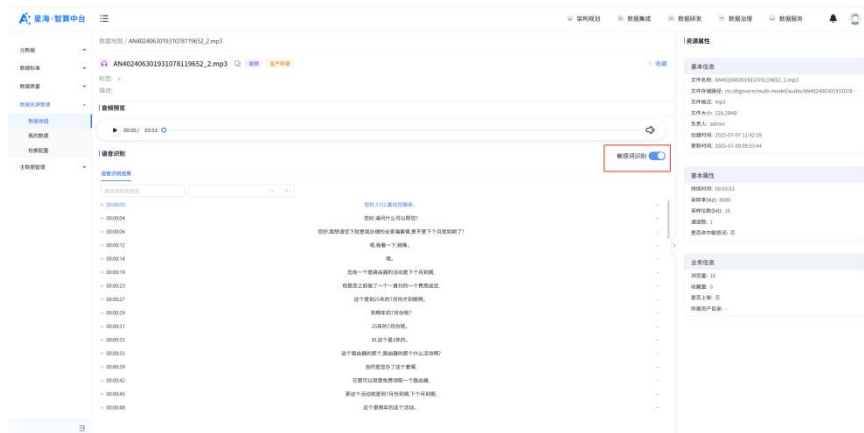
## 多模态检索

对于非结构化数据（图片、视频、音频、文档），系统支持内容检索，通过输入关键词，可以实现文搜图/视频，同时系统支持图搜图/视频，并支持多模态数据的详情查看。

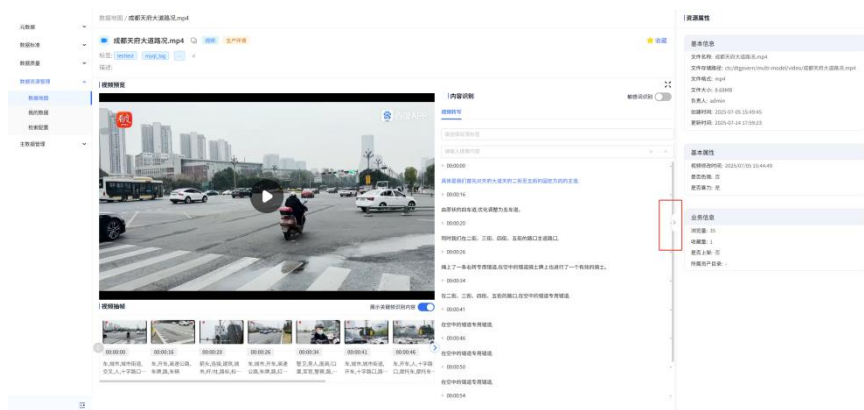


操作	描述
查看详情	<p>点击【更多操作】-&gt;【查看详情】，查看资源详情信息，包括资源基本信息（所属目录、密级、是否发布为资产、描述）、音频预览、语音识别结果、资产属性等信息</p> 

在音频的语音识别结果部分，支持对标签段落的选择以及敏感词识别

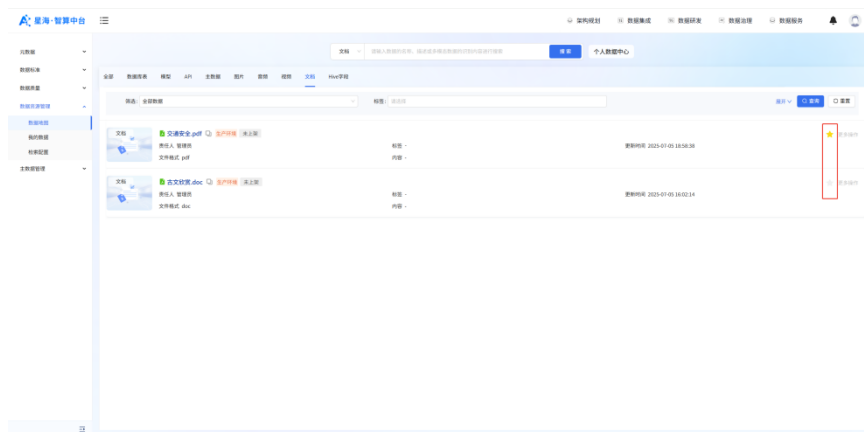


点击箭头，可以收起右侧资产属性菜单栏



点击目标资源右侧的星星，收藏目标资源

收藏

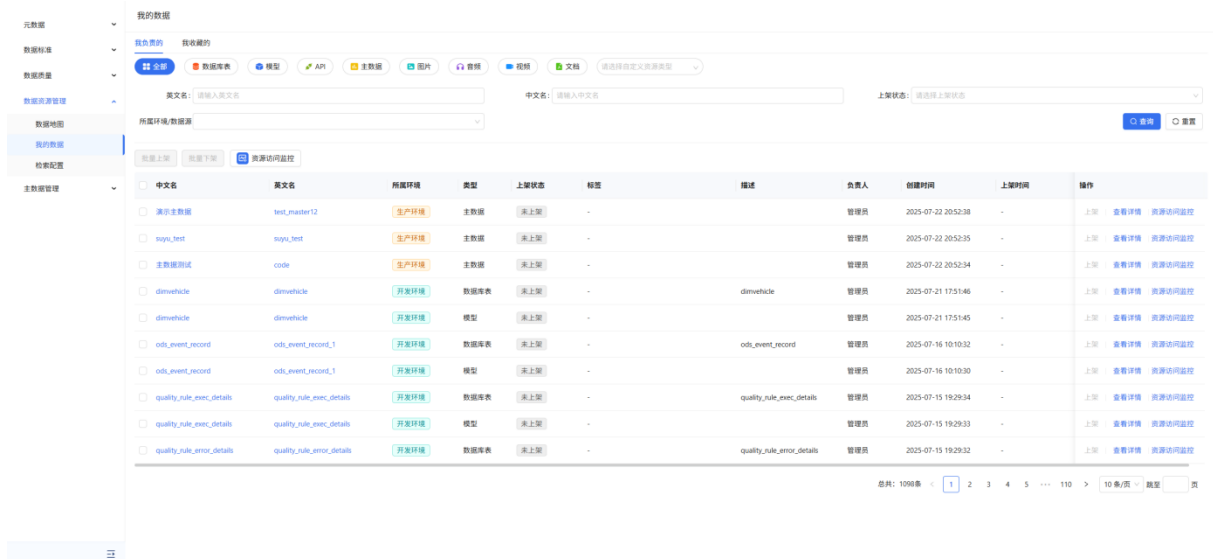


### 1.1.1.3. 我的数据

支持用户在我的数据相关页面查看数据信息：包括我负责的数据以及我收藏的数据，同时支持查看负责的数据的访问情况。

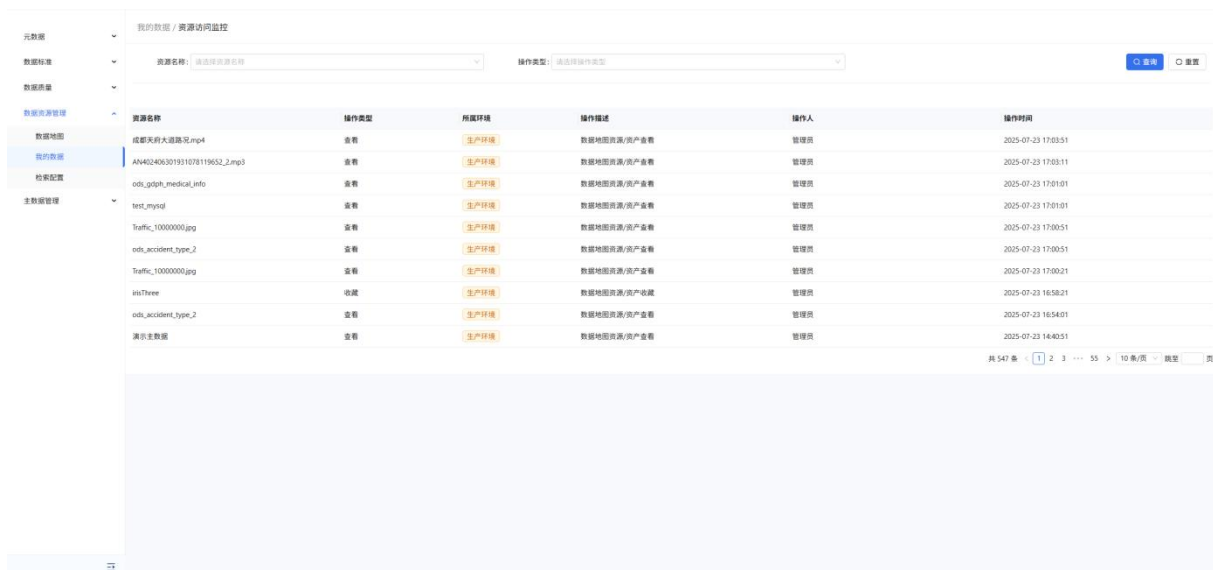
### 1.1.1.3.1. 我负责的

点击我负责的，展示当前用户负责的资源，支持按照资源类型查看资源列表，支持详情查看，上架到资产门户（本期暂不支持），同时支持查看负责的资产的访问情况。



### 资源访问监控

点击【资源访问监控】按钮，进入资源访问监控界面，界面按照最新访问时间逆序展示了所有资源被访问的情况，展示内容包括资源名称、操作类型、操作人、操作时间等。



### 资源操作

操作	描述
批量上架/下架（暂不支持）	点击资源中文名左侧的选择框，批量选择需要上架/下架的资源，点击【批量上架】/【批量下架】按钮，完成资源上架/下架至资产门户。

上架/下架（暂不支持）	在目标资源的操作菜单点击【上架】/【下架】按钮，完成资源上架/下架至资产门户。
查看详情	点击资源中文名，或者在所选标签的操作栏点击【查看详情】，查看资源详情，展示内容与数据地图保持一致。
资源访问监控	点击某个资源对应操作项中的【资源访问监控】，可以查看单条资源的访问情况。

### 1.1.1.3.2. 我收藏的

点击我收藏的，展示当前账户下收藏的资产，同个人负责的资产，可以进行搜索和筛选，也可以取消收藏。

中文名	英文名	所属环境	类型	上架状态	标签	描述	负责人	创建时间	上架时间	操作
基础API0705	test0705	生产环境	API	未上架	-		管理员	2025-07-09 16:00:57	-	查看详情 取消收藏
ods_event_record	ods_event_record	生产环境	模型	未上架	-		管理员	2025-07-07 14:42:49	-	查看详情 取消收藏
2301011008348745.mp3	2301011008348745.mp3	生产环境	音频	未上架	-		管理员	2025-07-07 11:41:25	-	查看详情 取消收藏
master_070701	master_070701	生产环境	主数据	未上架	myql_tag		管理员	2025-07-07 11:24:12	-	查看详情 取消收藏
图片元数据表	meta_data_image	生产环境	数据表	未上架	-		管理员	2025-07-05 17:43:23	-	查看详情 取消收藏
traffic_data_90000.png	traffic_data_90000.png	生产环境	图片	未上架	旅行 车辆		管理员	2025-07-05 17:21:59	-	查看详情 取消收藏
交通安全.pdf		生产环境	文档	未上架	-		管理员	2025-07-05 16:45:46	-	查看详情 取消收藏
Traffic_100000000.jpg	Traffic_100000000.jpg	生产环境	图片	未上架	旅行 车辆		管理员	2025-07-05 16:39:00	-	查看详情 取消收藏
kuosue_myql_test	kuosue_myql_test	生产环境	数据表	未上架	-		管理员	2025-07-05 16:29:50	-	查看详情 取消收藏
asset_model_070501	asset_model_070501	生产环境	模型	未上架	-		管理员	2025-07-05 16:28:21	-	查看详情 取消收藏

## 资源操作

操作	描述
查看详情	点击资源中文名，或者在所选标签的操作栏点击【查看详情】，查看资源详情，展示内容与数据地图保持一致。
取消收藏	点击目标资源右侧的星星，取消收藏目标资源。

## 服务开发

### 创建 API

数据服务模块提供了一种快速且用户友好的方式，将数据表转换为数据 API，支持通过可视化配置的向导模式、高级用户的自定义 SQL 脚本模式以及拖拉拽交互的编排模式三种途径来创建 API。

使用数据服务创建 API 前，需先将您的数据库通过资源管理进行纳管，以此作为数据服务 API 的数据来源。开发数据服务 API 时，数据服务将访问该数据源获取数据表的

Schema 信息，帮助您进一步设置请求和返回参数。调用数据服务 API 时，数据服务将访问该数据源来执行查询请求。

### 向导模式：快速生成数据 API

对于不需要编码能力的用户，数据服务提供了直观的向导模式，允许通过简单的点击和选择操作，快速将各类数据源下的数据表转换成数据 API。这个过程完整步骤化，用户只需按照提示逐步完成配置，即可轻松创建一个功能完备的 API。非常适合那些希望迅速搭建 API 而不具备编程背景的用户。

### 脚本模式：自定义 SQL 查询

为了满足高阶用户的个性化需求，数据服务还提供了脚本模式，让用户能够编写自定义的 SQL 查询语句来构建 API。这种模式支持复杂的查询逻辑，包括多表关联、复杂条件筛选以及聚合函数等，为用户提供更高的灵活性和控制力。

### 3) 编排模式：拖拉拽式创建 API

编排模式通过拖拉拽式的图形界面，让用户能够直观地创建和管理 API，无需编写代码。这种方式极大地简化了 API 的开发过程，使得即使是非技术人员也能轻松构建复杂的 API 逻辑。

## 测试 API

对已经创建的 API 进行测试运行，检查 API 的可用性。

## 使用 API

完成创建的 API 可以通过上线操作上线至 API 网关。上线成功后，其他服务才可使用该 API。API 上线成功后，还可以上架功能上架到数据集集市和资产，实现企业内部的数据共享数据，最终形成数据闭环。

## 服务管理

### API 集市

集成多种 API 资源，用户可以在其中发现、测试、集成所需的 API 到自己的应用上。该功能包含查询、新增和编辑 API 集市。

提供统一的 API 集市入口，支持按分类浏览与搜索 API，集中展示 API 的基本信息、责任人、更新时间及使用状态

提供标准化的 API 详情页，展示接口的基础信息、调用方式、请求参数、返回参数及错误码说明，并支持接口文档下载，降低调用与对接成本。

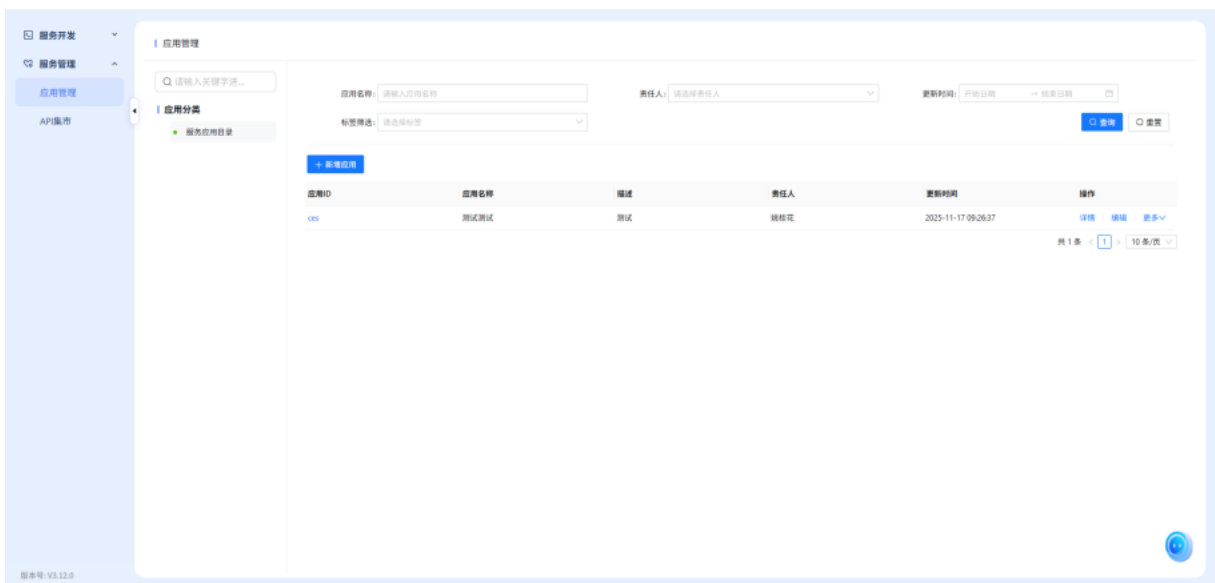


## 应用管理

应用管理涉及对使用 API 或数据服务的应用程序的管理和控制。这可能包括应用注册、授权、配额设置、访问控制等方面。

支持业务应用的统一注册与维护，集中管理应用名称、应用标识、负责人、标签及分类信息，形成 API 与数据服务的标准消费主体。

支持将 API 按应用进行授权管理，记录授权状态、申请人、授权时间及有效期，实现 API 使用的可控、可追溯。



## 平台管理

### 我的权限

我的权限以“当前登录者”为第一视角，汇总展示其在平台内已获得的全部数据权限。

### 操作步骤

菜单选择平台管理-->【我的权限】-->点击【数据权限】

## 字段说明

字段	取值示例	业务含义
数据类型	MySQL / Hive / Kafka	数据源技术类型
数据源名称	trade_center_mysql	平台内唯一标识
数据表名	订单主表	业务语义名称
数据表英文名	dwd_trade_order_df	物理表名
权限类型	查看 / 编辑 / 管理	平台定义的三级权限
来源	个人 / 团队	授权路径
是否被删除	是 / 否	数据源或表在物理库是否已删除
环境	DEV / PRO	逻辑隔离标识
有效期截止	2026-06-30 23:59:59	权限失效时间

**权限信息展示**：以列表形式清晰呈现当前用户所拥有的权限详情，涵盖数据类型、数据源名称、数据表中文名与英文名等关键信息，使用户快速了解自身可操作的数据范畴。例如，用户能直观看自己在某个特定数据库中，对于“订单表”（英文名 Orders）具有查看和编辑权限，而对“财务表”仅能查看。

**权限类型区分**：明确标注权限类型为查看、编辑或管理，用户依据相应权限执行操作，如查看权限仅允许读取数据，编辑权限可修改数据内容，管理权限则能对数据及相关设置进行全方位管控，包括权限分配等。

**来源标识**：标明权限来源于个人授权还是团队共享，方便用户追溯权限获取途径，理解权限范围 and 限制。若权限来自团队，当团队配置变动时，用户可及时知晓自身权限的潜在变化。

**状态与环境标注**：标注数据源或表是否被删除，避免用户对已失效数据进行无效操作。同时区

分环境为开发库（DEV）或生产库（PRO），确保用户在不同环境下的操作符合规定，防止误操作导致生产数据异常。例如，仅允许在开发库进行测试性质的数据修改。

**有效期管理：**展示权限有效期截止日期，临近到期时平台可提醒用户申请延期或重新授权，保障用户在业务周期内的正常使用，避免因权限中断影响工作连续性。

## 成员管理

成员管理面向“以人为维度”的授权场景，解决“给指定账号开通哪些表什么权限”的问题。页面布局与团队管理保持 90% 一致，降低学习成本。

### 操作步骤

菜单选择平台管理-->点击【成员管理】

### 列表字段

字段	说明	检索能力
用户名	登录账号	无
用户中文名	账户中文名称	模糊
手机号	联系方式	无
是否管理员	是：可授权他人；否：仅能查看自己 权限	无

**成员信息展示：**列表展示团队成员的用户名、用户中文名和手机号，方便管理员快速识别和联系成员。例如，通过中文名能精准定位成员，手机号则用于紧急沟通或权限变更通知。

**权限查看与编辑：**可查看成员的权限详情，包括所拥有的数据源、数据表及对应权限类型等。

管理员能根据成员岗位变动或业务需求调整权限，如将新入职数据分析师的权限从仅查看部分数据表扩展为可编辑多个关键数据表。

**管理员标识：**明确标注成员是否为管理员，保障平台管理职责明晰，避免权限滥用。