



大数据管理平台 DataWings

用户手册

天翼云科技有限公司





版权所有 © 天翼云科技有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

目录页

目录页.....	3
1 产品介绍.....	11
1.1 产品定义.....	11
1.2 产品优势.....	11
1.3 应用场景.....	12
1.4 核心能力.....	13
2 计费说明.....	15
2.1 产品规格.....	15
2.2 计费模式.....	15
2.3 公测申请.....	16
2.4 续订/退订.....	18
3 快速入门.....	19
3.1 客户控制台.....	20
3.2 创建集群.....	27
3.3 引擎配置.....	30
3.4 创建数据源.....	32
3.5 添加库表 HDFS 权限.....	37
4 用户指南.....	39
4.1 首页.....	40

4.2 数据源.....	41
4.2.1 数据源概述.....	41
4.2.2 数据源管理.....	42
4.3 数据建模.....	47
4.3.1 数据建模概述.....	47
4.3.2 建模首页.....	50
4.3.3 数据标准.....	51
4.3.3.1 命名词典.....	51
4.3.3.2 字段标准.....	56
4.3.3.3 标准值域.....	62
4.3.3.4 度量单位.....	65
4.3.3.5 规范管理.....	70
4.3.4 数仓规划.....	74
4.3.4.1 业务分类.....	74
4.3.4.2 数仓分层.....	80
4.3.4.3 规划配置.....	88
4.3.4.4 数仓要素.....	90
4.3.5 数据指标.....	97
4.3.5.1 原子指标.....	97
4.3.5.2 业务限定.....	105
4.3.5.3 统计周期.....	108

4.3.5.4 维度.....	111
4.3.5.5 派生指标.....	116
4.3.5.6 复合指标.....	123
4.3.6 维度建模.....	130
4.3.6.1 正向建模.....	130
4.3.6.1.1 可视化建模.....	130
4.3.6.1.1.1 可视化建模概述.....	130
4.3.6.1.1.2 新建接入表.....	131
4.3.6.1.1.3 新建维度表.....	137
4.3.6.1.1.4 新建事实表.....	145
4.3.6.1.1.5 新建汇总表.....	151
4.3.6.1.1.6 新建应用表.....	159
4.3.6.1.1.7 新建其他表.....	165
4.3.6.1.2 查看表详情.....	172
4.3.6.1.3 管理表模型.....	173
4.3.6.2 逆向建模.....	177
4.4 数据集成.....	183
4.4.1 数据集成概述.....	183
4.4.2 集成集群管理.....	186
4.4.2.1 资源组.....	186
4.4.2.2 集群注册.....	192

4.4.3 同步任务	198
4.4.3.1 任务创建	198
4.4.3.2 任务管理	205
4.5 数据处理	210
4.5.1 数据开发	210
4.5.1.1 项目管理	210
4.5.1.2 workflow 开发	213
4.5.1.3 workflow 模板	220
4.5.1.4 数据开发节点介绍	223
4.5.1.4.1 SparkSql 节点	223
4.5.1.4.2 数据质量节点	225
4.5.1.4.3 通用处理节点	226
4.5.1.4.4 任务依赖节点	230
4.5.2 任务调度	232
4.5.2.1 调度大屏	232
4.5.2.2 任务运维	235
4.5.3 即席查询	241
4.5.3.1 用户空间	241
4.5.3.2 数据库	243
4.5.4 配置中心	245
4.6 数据治理	249

4.6.1 数据调研	249
4.6.2 数据质量	255
4.5.2.1 数据质量概述	255
4.5.2.2 数据质量大屏	259
4.5.2.3 规则管理	261
4.5.2.3.1 数据质量规则列表	261
4.5.2.3.2 创建数据质量规则	262
4.5.2.3.3 试运行质量规则	268
4.5.2.4 模板管理	269
4.5.2.5 实例管理	272
4.5.2.6 抑制管理	275
4.5.2.7 质检报告	276
4.5.2.8 数据质量分析	277
4.6.3 数据治理 360	280
4.6.3.1 数据治理 360 概述	280
4.6.3.2 开启使用	281
4.6.3.3 健康诊断	282
4.6.3.4 治理项问题	284
4.6.3.5 治理项配置	286
4.6.3.6 健康指南	288
4.6.3.7 白名单	289

4.7 数据地图	290
4.7.1 数据地图概述	290
4.7.2 数据发现	291
4.7.3 地图配置	295
4.7.4 资产首页	299
4.7.5 数据总览	300
4.7.6 数据检索	301
4.7.7 我的数据	304
4.8 数据服务	305
4.8.1 数据服务概述	305
4.8.2 服务开发	306
4.8.3 服务管理	321
4.8.4 服务监控	331
4.8.5 服务配置	334
4.9 运维中心	335
4.9.1 查看运维大屏	335
4.9.2 告警管理	337
4.9.2.1 通道及对象管理	337
4.9.2.2 告警实例	340
4.9.2.3 值班表	341
4.9.3 资源运维	344

4.9.4 监控规则管理	345
4.9.5 操作日志	348
4.10 系统管理	350
4.10.1 权限管理	350
4.10.1.1 角色管理	350
4.10.2 集群管理	354
4.10.2.1 集群注册	354
4.10.2.2 引擎管理	357
4.10.3 流程中心	362
4.10.3.1 我的申请	362
4.10.3.2 我的审批	365
4.10.3.3 我的导入	368
4.10.3.4 权限申请	370
4.10.3.5 消息中心	373
5 最佳实践	374
5.1 获取翼 MapReduce 数据湖集群信息	374
5.2 电商行业数据建模实践	380
6 常见问题	388
6.1 常见产品问题	388
6.2 购买类	390
6.3 计费类	391

1 产品介绍

1.1 产品定义

本章节主要介绍大数据管理平台 DataWings 的产品定位和能力矩阵。

产品定位

大数据管理平台 DataWings，是一站式、开箱即用的大数据开发与数据资产管理工具，包括数据建模、数据集成、数据处理、数据治理、数据地图、数据服务等能力，致力于解决企业在数字化转型时面临的数据采集、开发、管理、治理等方面的问题，帮助企业高效上云，提升数据资产价值，构建数据智能应用，实现数据驱动业务增长。

产品能力矩阵



1.2 产品优势

本章节主要介绍大数据管理平台 DataWings 的产品优势。

一、一站式全流程

提供集数据建模、数据集成、数据处理、数据治理、数据地图、数据服务于一体的大数据管理平台，无需切换多个系统使用，提高用户的工作效率。

二、开箱即用低门槛

支持全流程可视化开发，提供低代码、零代码的使用模式，多模块支持表单式操作，降低上手难度，提高开发效率。

三、全周期数据治理

强调过程治理及持续治理，贯穿数据全生命周期；有机结合事前规范定义，事中监控并及时阻断脏数据，事后健康分诊断，并提供综合治理方案。

四、高安全高保障

提供统一的安全认证、租户隔离，并支持详细的操作日志分析、实时多渠道的监报告警，使得数据安全、服务状态和任务运行能得到有效保障。

1.3 应用场景

本章节主要介绍大数据管理平台 DataWings 的应用场景。

一、构建数据仓库+数据分析

- 通过数据建模模块，进行数仓建设。
- 通过数据集成模块，进行数据同步。
- 通过数据处理模块，进行报表开发。
- 通过即席查询模块，进行数据探索分析。

二、日志分析

- 通过数据集成模块，进行数据接入与转换。
- 通过数据地图模块，进行数据资产统一管理。

- 通过即席查询模块，进行数据分析。

三、全链路数据治理

- 通过数据建模模块，制定数据标准、自顶向下规划数仓结构、梳理数据指标、构建不同分层的表模型发布到引擎。
- 通过数据集成模块，进行数据同步。
- 通过数据处理模块，进行数据开发。
- 通过数据质量模块，进行质量规则配置及质量校验。
- 通过数据服务模块，进行对外共享数据。
- 通过数据地图模块，展示数据资产。

四、数据共享交换

通过数据服务模块，提供全面的服务发布及共享能力，帮助企业统一管理面向内外部的 API 服务。通过标准、规范的服务开发过程，实现 API 开发、API 测试、API 发布、API 管理、API 授权等全流程，为上层业务提供灵活的数据查询，打破数据孤岛，实现数据高效共享。

五、数据资产管理

通过数据地图模块，对数据进行全面盘点，形成数据地图，为业务应用和数据获取夯实基础，帮助业务人员快速精确查找目标数据。

1.4 核心能力

本章节主要介绍大数据管理平台 DataWings 的核心能力。

一、标准化数据建模

支持制定数据标准、自顶向下规划数仓结构、梳理数据指标、构建不同分层的表模型发布到引擎。适用于建设数仓，方法论与工具相结合，赋能业务场景。

二、全域数据集成

提供全域数据集成工具。致力于满足数据采集、汇聚和分发的要求，为不同数据源构建传输通道。适用于数据同步、数据分发等应用场景。

三、一站式可视化数据开发

基于多种大数据引擎实现 workflow 脚本编排，满足复杂业务需要，支持低代码生成 ETL，用户完全不感知底层技术。适用于数仓的开发与构建，及业务应用搭建。

四、全周期数据治理

基于数据全生命周期进行治理，涵盖前向、后向等多维治理流程，实现数据的可查，可控，可管，可用。基于标准化治理体系建设，改善数据质量，保证数据的安全性，推进数据资产的整合、开发，充分发挥数据价值，为数据共享、开放、数据挖掘分析应用提供支撑。

数据治理包括数据调研、数据质量、治理 360 等模块。数据调研支持在数据接入前进行全方位的数据资产盘点，提供资产目录配置、模板配置、录入数据等功能。数据质量支持基于数据质量评估标准和管理规范，及时发现、监测定位、跟踪解决各类数据质量问题，形成数据质量问题的闭环处理，以保证数据质量的稳定可靠。数据治理 360 通过对平台中研发与存储维度的潜在问题进行诊断，基于健康分评估模型与元数据，自动从全局与个人视角进行量化评估，给出健康分与相应治理指南，帮助用户对平台内资源的使用和价值进行有效的管理。

五、统一数据资产管理

支持数据发现、数据总览、数据检索、元数据管理等功能，高效管理数据资产。适用于数据量庞大的企业，对元数据进行快速的查询，及基于数据资产，完成数据的检索、资产详情查看等

场景。

六、全面数据服务与共享

统一管理面向内外部 API 服务，为上层业务提供灵活的数据查询。适用于通过标准化、规范化的数据开发手段，实现 API 开发、API 测试、API 发布、API 管理、API 授权全流程功能场景。

2.1 产品规格

本章节主要介绍大数据管理平台 DataWings 的产品规格。

DataWings 各版本功能集及说明

版本	包含模块
公测版本	数据集成、数据建模、数据处理、数据地图、数据调研、数据质量、数据治理 360、数据服务、运维中心、数据源管理。

◇ 说明

公测期间，默认开通公测版本，所有模块功能均可使用。

2 计费说明

2.2 计费模式

本章节主要介绍大数据管理平台 DataWings 的计费说明。

计费项

购买大数据管理平台 DataWings 的费用包含两个部分：

- 大数据管理平台 DataWings 版本服务费用。

- 资源组费用（集成资源组、调度资源组、服务资源组）。

计费方式

大数据管理平台 DataWings 当前仅支持包周期的计费方式。

- 包年/包月：根据版本/资源组的购买时长，一次性支付费用。最短时长为 1 个月，实际可订购时长以页面显示为准。

说明

公测期间可免费申请试用大数据管理平台 DataWings。

2.3 公测申请

本章节主要介绍公测期间申请试用大数据管理平台 DataWings 的步骤。

申请试用步骤

- 1、登录并进入官网首页。
- 2、在官网首页，单击左上角“产品”，在大数据分类中，找到大数据管理平台 DataWings 产品，点击进入。

查看全部云产品 >

- 推荐
- 计算 >
- 存储
- 网络
- 云终端 >
- CDN与边缘
- 容器与中间件
- 数据库 >
- 大数据 >**
- 人工智能

输入产品名称

数据计算与分析

- 翼MapReduce
- 云搜索服务
- 数据湖探索
- 数据仓库服务

数据开发与治理

- 数据治理中心 DataArts Studio
- 大数据管理平台 DataWings **NEW**

3、在【大数据管理平台 DataWings】产品详情页界面，单击【申请开通】。

公测

大数据管理平台 DataWings

天翼云大数据管理平台 DataWings，是一站式、开箱即用的大数据开发与数据资产管理工具，包括数据建模、数据集成、数据处理、数据治理、数据地图、数据服务等能力，致力于解决企业在数字化转型时面临的数据采集、开发、管理、治理等方面的问题，帮助企业高效上云，实现数据驱动业务增长。

[申请开通](#) [管理控制台](#) [产品文档 >](#)



算力套餐
让算力触手可及

云上钜惠
领取5万元大礼包更实惠

云主机特惠
新老同享云主机3折起

- 动态 【产品优势】DataWings产品能力优势
- 动态 【产品规格】DataWings产品规格说明
- 动态 【快速入门】DataWings产品快速入门

4、填写并提交申请信息。



5、等待客户经理联系和确认开通。

2.4 续订/退订

本章节主要介绍大数据管理平台 DataWings 的续订/退订操作。

续订操作

到期说明

- 包周期产品在到期前的第 7 天、第 3 天和第 1 天，向您发送到期通知。
- 到期前未续费的话，到期后您将无法使用产品，资源组将会被冻结。
- 到期后，将为您继续保留 15 天，如果 15 天内您仍未完成续订操作，您将无法再使用产品功能，资源组将会自动销毁。

操作步骤

- 1、在天翼云官网，点击右上角的“管理中心”。



2、左侧导航栏点击“订单管理”-“续订管理”，输入筛选条件，找到需要续订的产品，点击“手动续订”，按提示完成续订操作即可。



退订说明

公测期间暂不支持自行退订，您可以联系客户经理进行退订。

注意

退订成功后，将立刻销毁用户数据和资源组，大数据底座任务需要您自行前往停止或关闭。

3 快速入门

3.1 客户控制台

客户控制台为您提供查看全局概览和管理资源组能力。

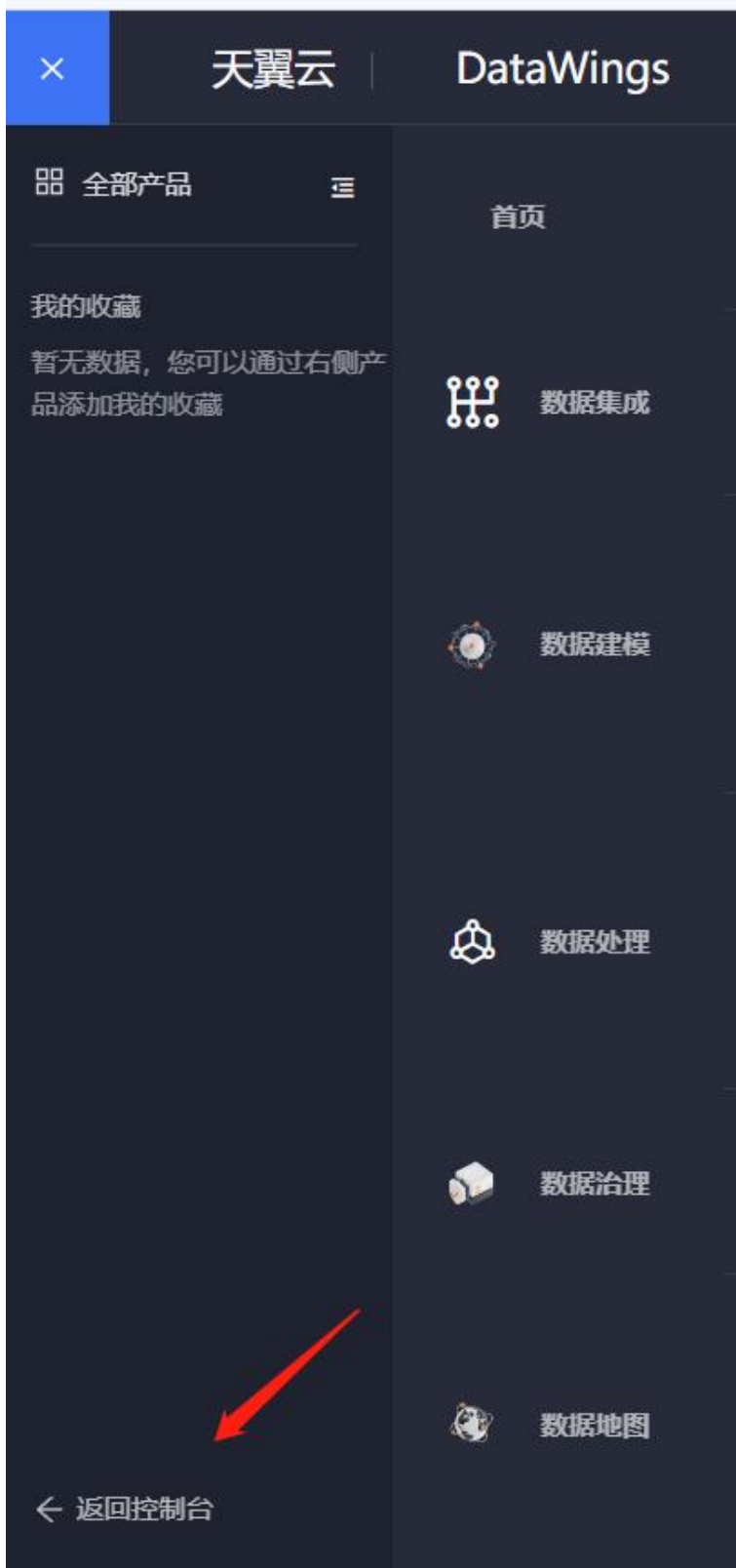
概览页

进入概览页

方式一：在产品详情页点击“管理控制台”，默认进入概览页。

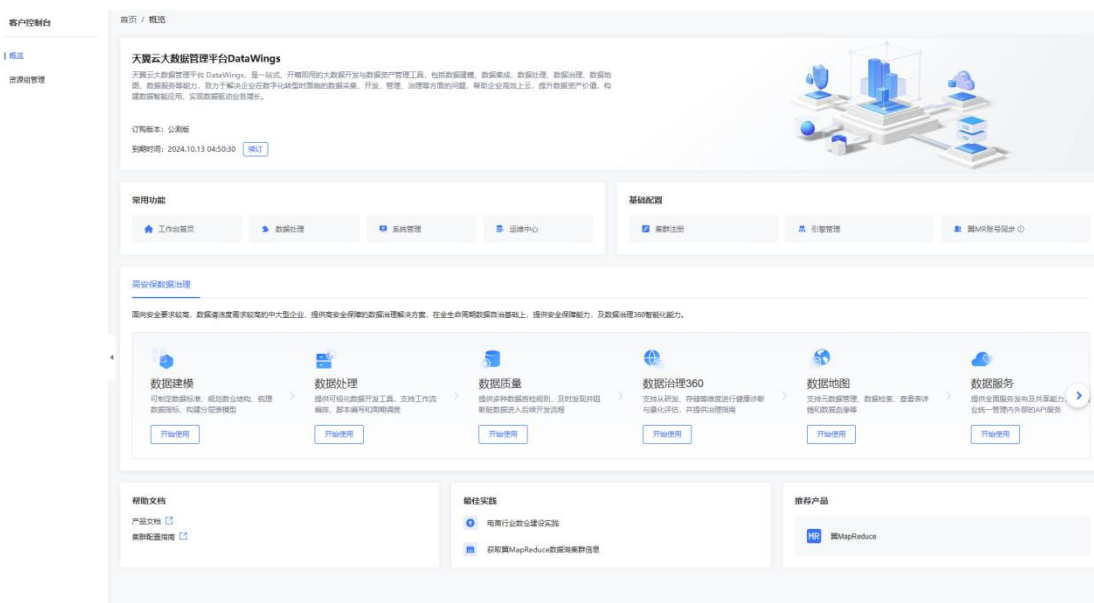


方式二：在产品页面内，单击左上方的图标，点击“返回控制台”，进入概览页。



功能说明

- 支持查看已购产品的版本和到期时间等信息。
- 支持跳转至常用功能页面：工作台首页，数据处理、系统管理、运维中心。
- 支持跳转至基础配置页面：集群注册、引擎管理、翼 MR 账号同步。
- 支持查看各模块介绍，点击“开始使用”可跳转至相应模块页面。
- 支持跳转查看帮助文档、最佳实践、推荐产品等。



资源组管理

点击“资源组管理”，您可以在资源组管理页面查看和操作集成资源组、调度资源组、服务资源组和数据源资源组信息。

资源组列表

通过切换资源组类型，您可以查看不同资源组类型的资源组列表信息，每个资源组下有 1 个或者多个资源。

资源组信息：包括资源组名称、状态、已购规格、数量、网络配置、创建时间等信息。

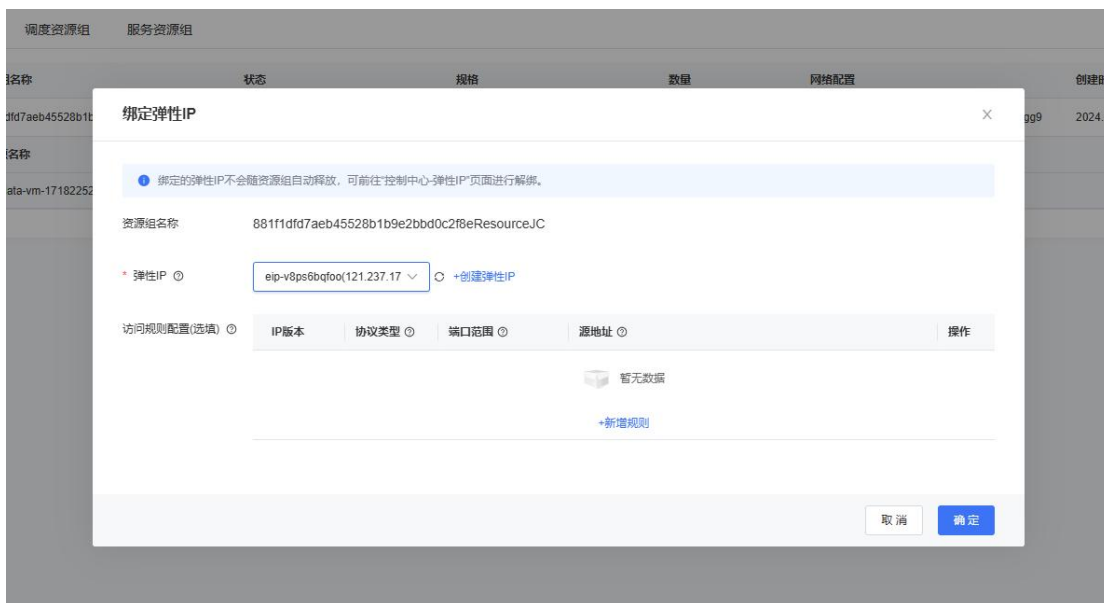
资源信息：包括资源名称、内网 IP、外网 IP、IPv6 地址等信息。

资源名称	状态	规格	数量	网络配置	创建时间	操作								
881f1dfd7aeb45528b1b9e2bbd0c2f8eResourceJC	运行中	2C16G	1个	VPC vpc-ve9d7gqh9 子网 subnet-6e1c0c3g9	2024.08.13 04:47:13	详情								
<table border="1"> <thead> <tr> <th>资源名称</th> <th>内网IP</th> <th>外网IP</th> <th>操作</th> </tr> </thead> <tbody> <tr> <td>sgdata-em-17182252-45-e872x</td> <td>192.168.8.22</td> <td>121.237.179.223</td> <td>绑定弹性IP 解除弹性IP 更多</td> </tr> </tbody> </table>							资源名称	内网IP	外网IP	操作	sgdata-em-17182252-45-e872x	192.168.8.22	121.237.179.223	绑定弹性IP 解除弹性IP 更多
资源名称	内网IP	外网IP	操作											
sgdata-em-17182252-45-e872x	192.168.8.22	121.237.179.223	绑定弹性IP 解除弹性IP 更多											

绑定弹性 IP

如果您对公网访问有要求，可以给资源绑定弹性 IP，绑定弹性 IP 前，您需要前往[弹性 IP](#) 进行购买。

绑定弹性 IP 时，您可以同时创建访问规则，也可以在绑定弹性 IP 后再配置访问规则。



IP版本	协议类型	端口范围	源地址	操作
IPv4	ANY	1-65535	116.22.252.236/32	删除
IPv4	ANY	请输入端口号	请输入IP	删除

+新增规则

◇ 注意

资源组到期或者退订时自动释放，弹性 IP 不会随资源组自动释放，需要前往[弹性 IP](#)页面进行解绑。

解绑弹性 IP

绑定弹性 IP 后，如果您不再需要使用弹性 IP，可以解绑弹性 IP。



绑定 IPv6 带宽

前提条件

- 当前资源实例已被分配 IPv6 地址。
- 当前 VPC 已创建 IPv6 网关。
- 已购买 IPv6 带宽。

◇ 说明

订购产品和资源组时 VPC 和子网已启用 IPv6，资源实例才会被分配 IPv6 地址。

若当前资源实例已被分配 IPv6 地址，您可以点击【绑定 IPv6 带宽】按钮绑定已购买的 IPv6 带宽，绑定成功后，您可以通过访问 IPv6 地址访问资源实例。

若当前资源实例无 IPv6 地址，【绑定 IPv6 带宽】按钮将置灰无法点击。

解绑 IPv6 带宽

若当前实例已绑定 IPv6 带宽，您不再期望使用 IPv6 能力，您可以点击【解绑 IPv6 带宽】按钮，解除 IPv6 带宽和当前资源实例的绑定关系。

配置访问规则

您可以配置 IPv4/IPv6 的端口访问规则。



配置访问规则

资源组名称 881f1dfd7aeb45528b1b9e2bbd0c2f8eResourceDD

访问规则配置(选项)	IP版本	协议类型	端口范围	源地址	操作
	IPv4	ANY	1-65535	113.68.91.231/32	删除
	IPv6	TCP	请输入端口号	请输入IP	删除

+新增规则

配置资源组 Host 映射

若资源组需要通过域名方式访问其他服务/资源，您可以在资源组管理页面管理资源实例的 Host 映射关系。

若您已注册翼 MapReduce 集群且创建了对应引擎，集群的 Host 映射关系已自动同步。

批量配置资源组 Host 映射

点击【批量配置 Host】按钮，您可以输入 Host 映射关系，点击确定后，映射关系将会对当前资

源组下的所有资源生效。



单独配置 Host 映射

选择某一资源实例，点击【更多】-【配置 Host】，您可以单独配置当前资源实例的 Host 映射关系。



查看资源实例 Host 映射

选择某一资源实例，点击【更多】-【查看 Host 详情】，可以查看当前资源实例的所有 Host 映射关系。



3.2 创建集群

本章节主要介绍大数据管理平台 DataWings 的集群注册操作说明。


背景信息

为了能够更好的使用大数据管理平台的能力，若您需要使用数据处理模块，需先购买翼 MR 大数据集群，并绑定大数据引擎。DataWings 支持自动同步和手动创建两种集群注册方式。

前提条件

已购买开通[翼 MapReduce 产品](#)。

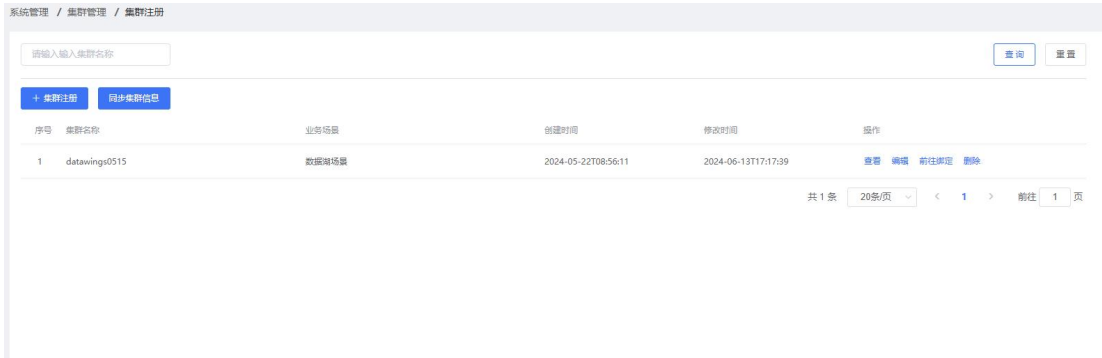
进入页面

单击左上方的图标，选择全部产品 > 系统管理 > 集群管理 > 集群注册。

同步集群信息

- 点击“同步集群信息”，DataWings 将自动同步当前主账号下已购买开通的翼 MapReduce 集群信息（当前仅支持数据湖集群的自动同步）。

- 当翼 MR 集群信息变更时，您可以再次点击同步按钮更新集群信息，您也可以点击编辑手动更新集群信息。
- 如您需要注册 Doris 集群信息，需要手动点击“编辑”，输入 Doris 集群信息。



操作说明

操作	说明
查看	查看当前集群信息。
编辑	编辑当前集群信息。
前往绑定	点击“前往绑定”打开引擎管理页面，您可以在引擎管理创建大数据引擎。
删除	删除当前集群，已绑定引擎的集群不支持删除。

手动注册集群

- 手动注册集群信息需要提前获取翼 MapReduce 集群信息，您可以前往已购买的翼 MapRuce 集群的 Manager 控制台获取相关信息。
- 集群信息获取可参考[获取翼 MapReduce 数据湖集群信息](#)。

基本信息

* 集群类型

* 集群场景

* 集群名称 0/30

Hadoop

* 配置文件

Yarn

* 版本选择

* 配置文件

Hive

* 版本选择

* 配置文件

Spark

* 版本选择

* 配置文件

Kerberos

是否开启Kerberos

Doris

是否开启Doris

资源组网络配置

hosts

① 请按照格式填写主机配置：'IP地址 主机名'，IP地址只能包含数字和 '.'，主机名只能包含字母、数字、 '-' 或 '_'，每行一个配置，用换行符分隔。

后续步骤

同步/手动注册集群完成后，您需要前往[引擎管理](#)创建引擎，为集群绑定独享调度资源组。

3.3 引擎配置

本章节主要介绍大数据管理平台 DataWings 的引擎配置操作。


前提条件

- 已购买独享调度资源组，并确保调度资源组与翼 MapReduce 集群间网络可连通（在同一 VPC 下）。

使用限制

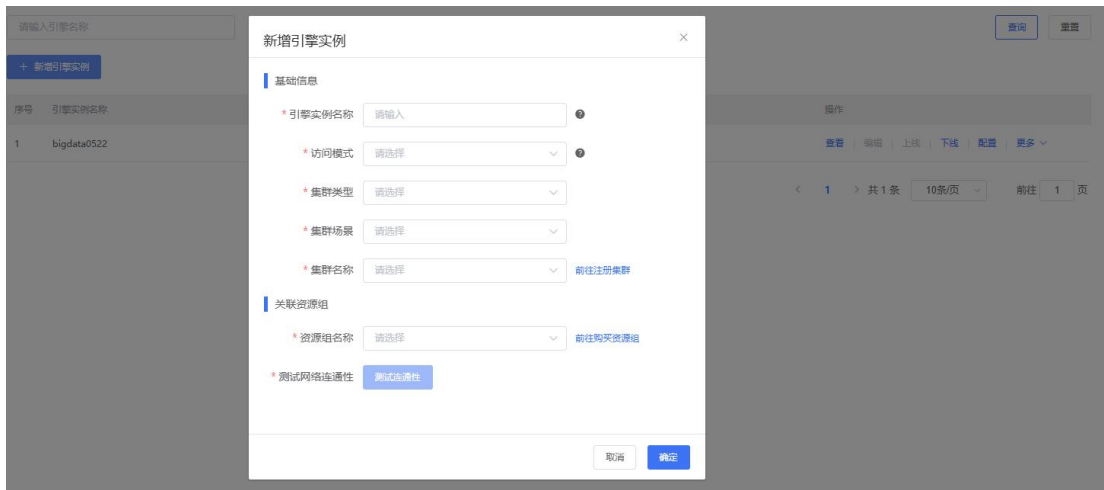
- 仅支持使用独享调度资源组在 DataWings 运行翼 MapReduce 任务。

页面进入

单击左上方的图标，选择全部产品 > 系统管理 > 集群管理 > 引擎管理。

功能说明

新增引擎实例



- 1、配置引擎基本信息。

参数	说明
引擎实例名称	当前引擎在 DataWings 的名称，用于标识当前引擎的配置信息。
访问模式	简单模式：DataWings 用户绑定同一个翼 MR 集群用户（LDAP 用户），所有引擎任务使用同一个集群用户身份提交并执行任务。
集群类型	当前仅支持翼 MapReduce 集群。
集群场景	支持数据湖场景和数据分析场景。
集群名称	已注册的集群名称，详情请见 集群注册 。

2、配置资源组信息。

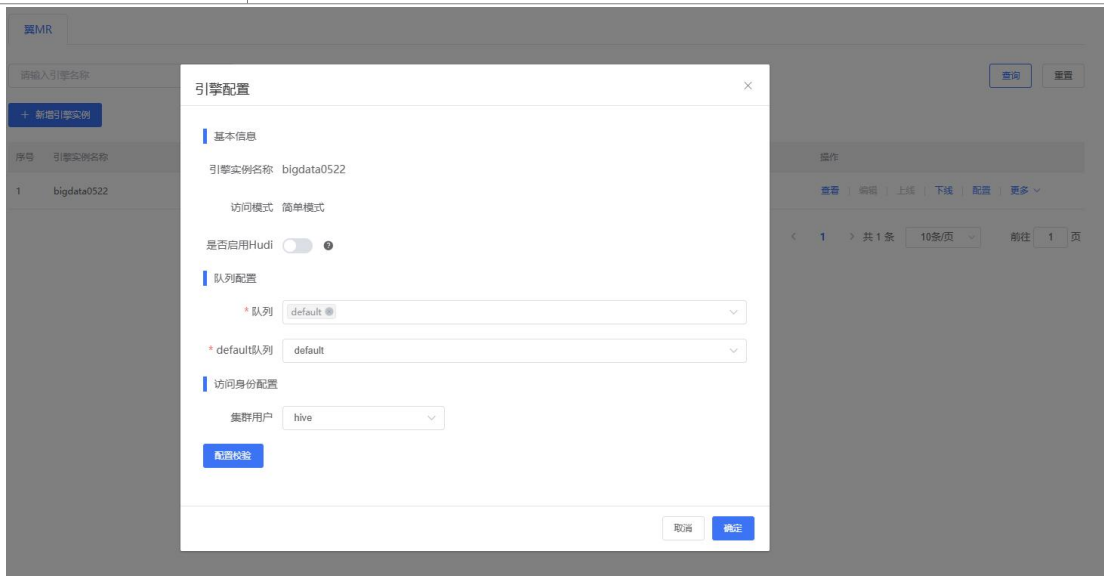
资源组名称：已购买的独享调度资源组。

测试连通性：您需要确保调度资源组与集群之间网络是互通的，才能成功创建引擎以及执行引擎任务。

3、配置引擎信息。

参数	说明
是否启用 Hudi	若您购买的翼 MapReduce 集群包含 Spark-Hudi 客户端，您可以选择启用 Hudi，启用后在即席查询模块，您可以创建 Hudi 类型的数据分析任务。

参数	说明
队列	当前引擎可以使用的集群队列列表。
default 队列	当前引擎默认使用的队列。
访问身份配置	集群用户 (LDAP 用户) 默认配置为 hive 用户, 您可以重新输入以修改集群用户。



3.4 创建数据源

本章节介绍大数据管理平台 DataWings 创建数据源操作说明。

内置数据源

内置数据源是在集群注册并创建引擎实例后自动创建的数据源, 当前内置数据源包括 Hive 和 Doris 两种数据源。当集群注册里的集群信息包含 Hive 和 Doris 的配置信息后, 将会自动创建对应的数据源, 无需手动再前往数据源管理创建, 以下是内置数据源详细信息。

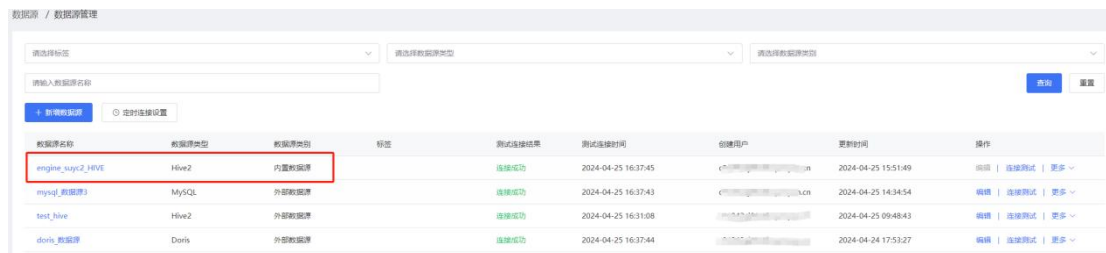
前置条件

已购买开通翼 MapReduce 产品。

操作步骤

数据源管理支持内置 Hive 和 Doris 两种数据源的纳管。

1. 先进行集群注册，具体可参考[创建集群](#)。
2. 进行引擎配置，具体可操控[引擎配置](#)。
3. 引擎创建完成后，将会自动创建对应的 Hive 和 Doris 数据源（Doris 数据源信息需要在集群注册里手动填写）。
4. 在“系统管理”-“数据源”-“数据源管理”界面会显示对应的内置数据源信息。



数据源名称	数据源类型	数据源类别	状态	测试连接结果	测试连接时间	创建用户	更新时间	操作
engine_sap2_HIVE	Hive2	内置数据源	成功	连接成功	2024-04-25 16:37:45	c@10.10.10.10	2024-04-25 15:51:49	编辑 连接测试 更多 >
mysql_数据源1	MySQL	外部数据源	成功	连接成功	2024-04-25 16:37:43	c@10.10.10.10	2024-04-25 14:34:54	编辑 连接测试 更多 >
test_hive	Hive2	外部数据源	成功	连接成功	2024-04-25 16:31:08	c@10.10.10.10	2024-04-25 09:48:43	编辑 连接测试 更多 >
doris_数据源	Doris	外部数据源	成功	连接成功	2024-04-25 16:37:44	c@10.10.10.10	2024-04-24 17:53:27	编辑 连接测试 更多 >

外置数据源

数据源管理模块支持新增多种关系型数据库、大数据存储、文件服务器和消息队列，其中，关系型数据库支持 MySQL、Oracle、SQL server、PostgreSQL、Greenplum、达梦等，大数据存储支持 HDFS、Hive、GaussDB、Doris、Hudi、HBase 等，文件服务器类型支持 FTP，NoSQL 型数据库支持 MongoDB，消息队列支持 Kafka。

前置条件

1. 需购买弹性 IP 与 NAT 网关。
2. 当 VPC 内的云主机需要主动访问 Internet，可以使用弹性 IP 绑定云主机实现。但是如果 VPC 内需要主动访问 Internet 的云主机过多时，为了节省弹性 IP 资源并且避免云主机 IP

直接暴露在公网上，可以使用 NAT 网关的 SNAT 功能，构建 VPC 主动访问公网出口。

3. 具体场景请参考[应用场景](#)的构建 VPC 云上网络出口场景。
4. 具体操作步骤请参考[使用 SNAT 访问公网](#)。

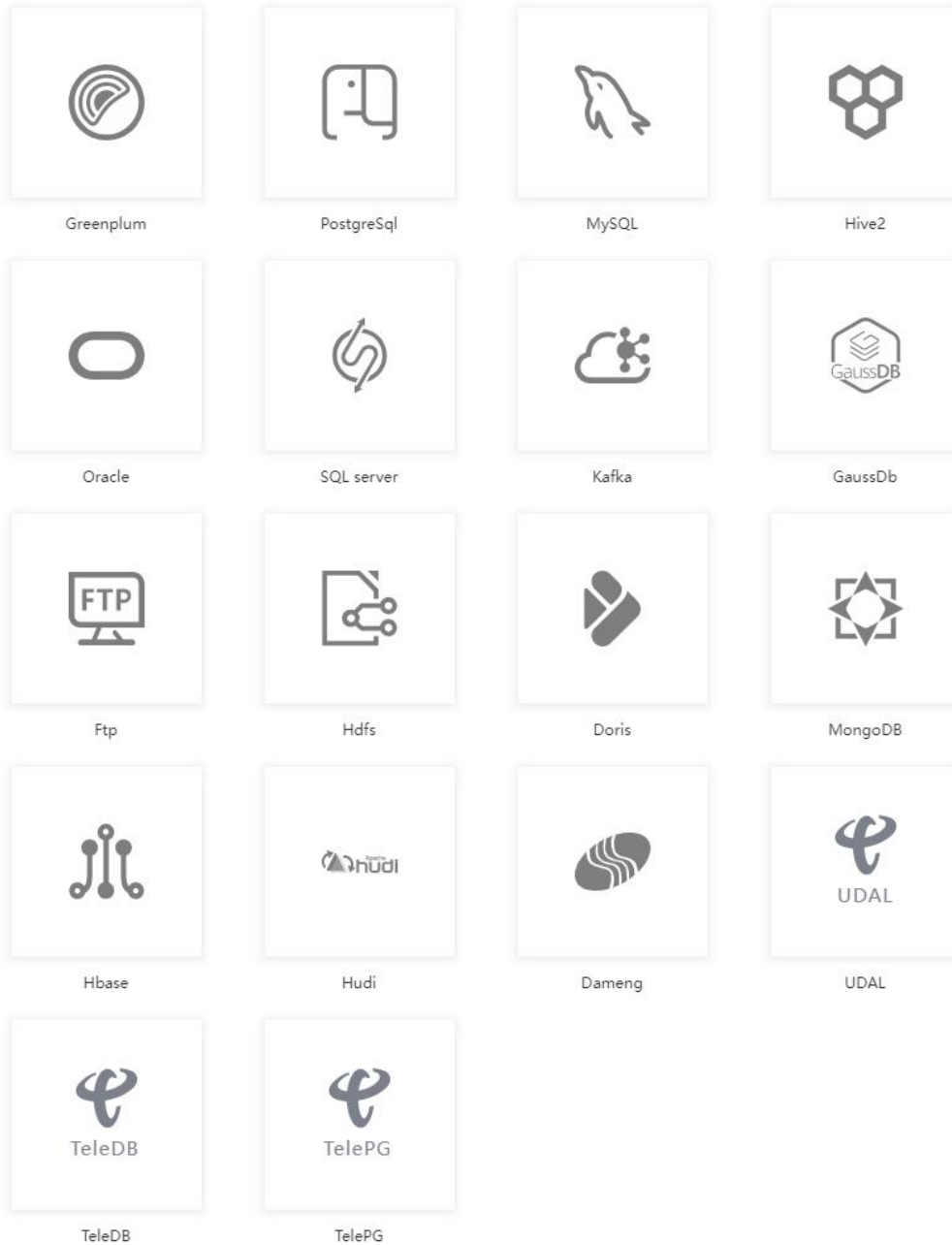
新增数据源

新增包括以下几个环节：

1. 选择新增的数据源类型。
2. 填写数据源连接信息。
3. 测试连接。
4. 如果测试通过，则数据源新增成功；反之，检查并重新编辑数据源连接信息再进行连通性测试后提交。

选择新增的数据源类型

根据需要选择创建的数据源类型。



取消

下一步

填写数据源连接信息

根据要求填写数据源连接信息（此处界面以mysql为例），填写数据源名称、标签、数据源描述，选择数据源驱动，填写主机名、端口、用户名和密码，若需指定连接的数据库时可填写具

体的数据库名称，若无需指定则可不填写（将获取数据源下所有的数据库）。

新增数据源 X

数据类型

* 数据源名称 0/128

标签

数据源描述 0/300

* 数据源驱动

* 主机名

* 端口号

* 用户名

* 密码

数据库名称

jdbc连接参数

URL预览

连接测试

新增数据源在保存之前必须进行一次连接测试，确保可连通，才可创建成功，否则无法创建。

编辑数据源

数据源类型	MySQL
* 数据源名称	MySQL_test_collection_5 23/128
标签	mysql ×
数据源描述	<div style="border: 1px solid #ccc; height: 30px;"></div> 0/300
* 数据源驱动	8.0.30
* 主机名	192.168.61.44
* 端口号	50001
* 用户名	datawings
* 密码 🗑
数据库名称	<div style="border: 1px solid #ccc; height: 20px;"></div>
jdbc连接参数	<div style="border: 1px solid #ccc; height: 20px;"></div>
URL预览	jdbc:mysql://192.168.61.44:50001/?

连接测试 取消 确定

提交

点击提交，即可完成数据源的创建。

3.5 添加库表 HDFS 权限

本节介绍如何在翼 MapReduce 数据湖集群中授予 LDAP 用户 HDFS 权限。

背景

在数据处理中执行计算引擎任务或者在数据建模常见表模型的业务中，DataWings 使用引擎实例

中配置的 LDAP 用户进行任务执行, 若提示该 LDAP 用户没有 HDFS 权限, 您可以在翼 MapReduce 产品界面进行相关设置。

前提条件

已购买翼 MapReduce 数据湖集群。

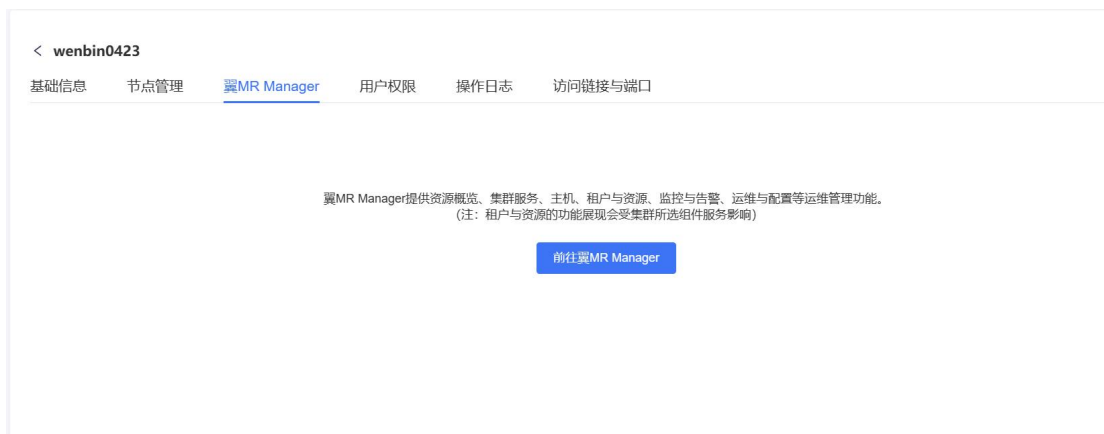
操作步骤

1. 点击[翼 MapReduce](#)进入翼 MapReduce 控制台。



集群名称	集群类型	付费类型	状态	创建时间	操作
wenbin0423	数据湖	包年包月	运行中	2024.04.22 20:52	续订 退订 删除
wenbin0422	数据湖	包年包月	已终止	2024.04.22 20:07	续订 退订 删除
wenbinemr	数据湖	包年包月	已终止	2024.04.16 18:27	续订 退订 删除
wenbin0416	数据湖	包年包月	已终止	2024.04.15 20:06	续订 退订 删除
wenbin0415	数据湖	包年包月	已终止	2024.04.15 19:33	续订 退订 删除
wenbin	数据湖	包年包月	已终止	2024.03.21 18:39	续订 退订 删除

2. 点击集群名称-翼 MR Manager, 进入运维管理平台。



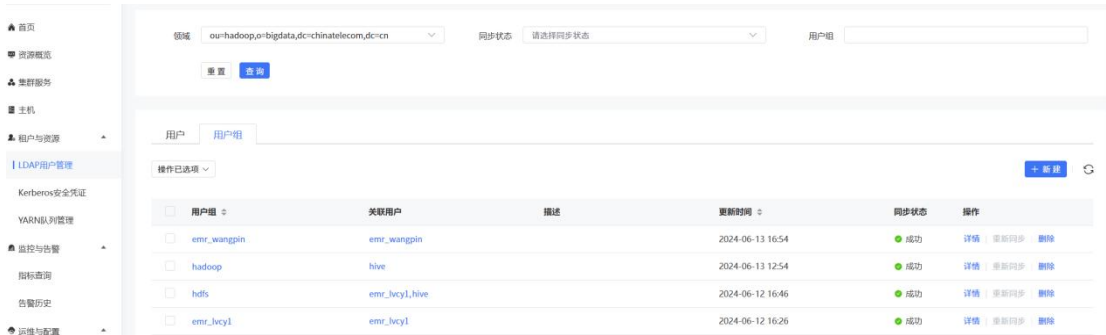
< wenbin0423

基础信息 节点管理 翼MR Manager 用户权限 操作日志 访问链接与端口

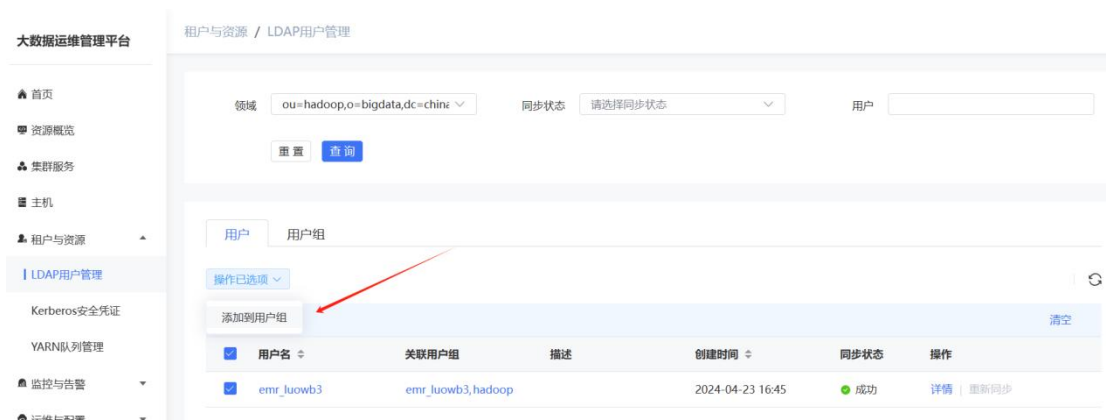
翼MR Manager提供资源概览、集群服务、主机、租户与资源、监控与告警、运维与配置等运维管理功能。
(注: 租户与资源的功能展现会受集群所选组件服务影响)

[前往翼MR Manager](#)

3. 进入 LDAP 用户管理页面, 点击“用户组”-“新建”按钮创建 hadoop 用户组。



4. 点击“用户”切换到用户页面，选择需要授权的 LDAP 用户，点击“操作已选项”-“添加到用户组”，将当前用户关联到 hadoop 用户组。



后续步骤

1. 在引擎配置时，指定 DataWings 执行任务的 LDAP 用户为当前已授权用户。
2. 数据建模创建物理表时，将会拥有 hadoop 用户组的 HDFS 权限。
3. 在数据处理、即席查询任务中，将会拥有 hadoop 用户组的 HDFS 权限。
4. 此介绍仅为基础的 HDFS 权限授权操作，如果您需要更细粒度的 HDFS 权限管控，可以在翼 MapReduce 中进行精细化配置。

4 用户指南

4.1 首页

DataWings 首页主要起到信息概览与操作引导作用，帮助用户了解个人信息、支持用户快速前往功能模块、最佳实例或帮助文档等页面进行数据作业或信息查询。

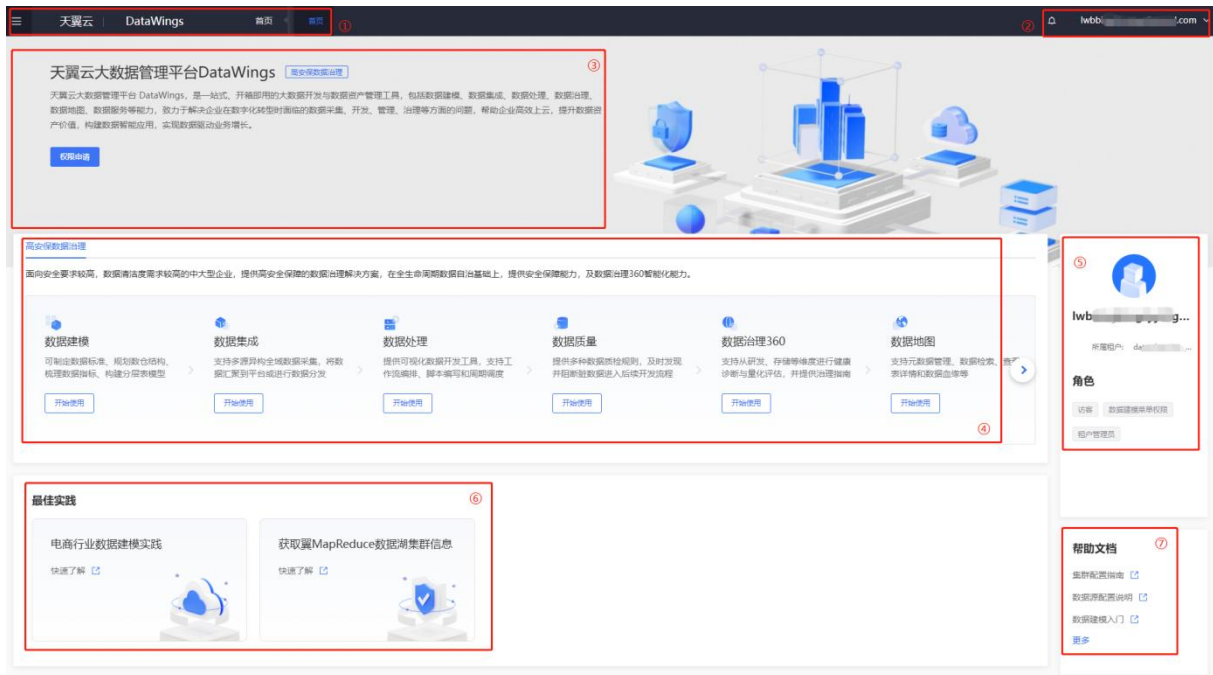
进入页面

方法 1：用户登录 DataWings 后，将默认进入首页。

方法 2：用户可在菜单栏中选择“首页”，进入该页面。

区域说明

区域编号	功能	说明
①	菜单栏	支持菜单查看与功能跳转。
②	站内信&登录信息	支持查看站内信、用户登出与密码修改。
③	产品信息介绍	DataWings 产品与版本介绍，并支持快速跳转权限申请页面。
④	功能指引	按版本展示，支持快速跳转至目标功能页面。
⑤	用户信息概览	展示用户基础信息以及被分配的角色权限。
⑥	最佳实践	支持跳转查看 DataWings 最佳实践介绍。
⑦	帮助文档	支持跳转查看产品帮助文档。



4.2 数据源

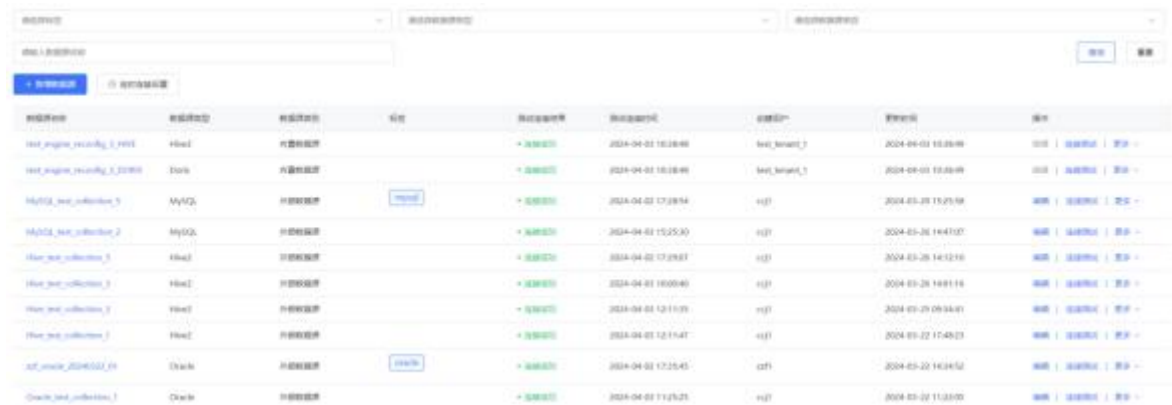
4.2.1 数据源概述

本章节主要介绍大数据管理平台 DataWings 的数据源模块。

数据源模块用于统一管理平台的数据源连接信息，在本模块添加的数据源可在数据集成、数据地图、数据处理、数据服务、数据安全等模块中使用。

功能入口

单击左上方的☰图标，选择全部产品 > 系统管理 > 数据源，进入数据源管理页面。



数据源名称	数据源类型	数据源状态	名称	操作按钮	创建时间	更新时间	操作
test_engine_testdb1_1993	Hive	连接成功		连接成功	2024-04-02 10:28:49	2024-04-02 10:28:49	编辑 连接测试 删除
test_engine_testdb1_22993	Oracle	连接成功		连接成功	2024-04-02 10:28:49	2024-04-02 10:28:49	编辑 连接测试 删除
MySQL_test_collection_5	MySQL	连接成功		连接成功	2024-04-02 17:28:54	2024-03-28 15:25:58	编辑 连接测试 删除
MySQL_test_collection_2	MySQL	连接成功		连接成功	2024-04-02 17:25:30	2024-03-26 16:47:07	编辑 连接测试 删除
Hive_test_collection_5	Hive	连接成功		连接成功	2024-04-02 17:28:57	2024-03-28 14:32:16	编辑 连接测试 删除
Hive_test_collection_3	Hive	连接成功		连接成功	2024-04-02 18:00:40	2024-03-28 14:31:16	编辑 连接测试 删除
Hive_test_collection_3	Hive	连接成功		连接成功	2024-04-02 12:11:35	2024-03-29 09:34:41	编辑 连接测试 删除
Hive_test_collection_3	Hive	连接成功		连接成功	2024-04-02 12:11:47	2024-03-22 17:45:23	编辑 连接测试 删除
test_engine_testdb1_21	Oracle	连接成功		连接成功	2024-04-02 17:25:45	2024-03-22 14:34:52	编辑 连接测试 删除
Oracle_test_collection_1	Oracle	连接成功		连接成功	2024-04-02 17:25:25	2024-03-22 11:33:05	编辑 连接测试 删除

4.2.2 数据源管理

数据源管理支持界面式管理数据源，包括新增、编辑、查询、删除等管理功能。数据源类型分为内置数据源和外部数据源，内置数据源需通过集群注册功能配置 Hive 组件信息，并成功创建对应引擎后，系统将自动创建内置数据源；外部数据源可通过数据源管理界面，直接新增外部数据源。

数据源管理列表

数据源名称	数据源类型	数据源类别	标签	测试连接结果	测试连接时间	创建用户	更新时间	操作
test_engine_reconfiq_3_HIVE	Hive2	内置数据源		• 连接成功	2024-04-03 10:38:48	test_tenant_1	2024-04-03 10:38:49	编辑 连接测试 更多 >
test_engine_reconfiq_3_DORIS	Doris	内置数据源		• 连接成功	2024-04-03 10:38:48	test_tenant_1	2024-04-03 10:38:49	编辑 连接测试 更多 >
MySQL_test_collection_5	MySQL	外部数据源	mysql	• 连接成功	2024-04-02 17:28:54	cgj1	2024-03-29 15:25:58	编辑 连接测试 更多 >
MySQL_test_collection_2	MySQL	外部数据源		• 连接成功	2024-04-03 15:25:30	cgj1	2024-03-26 14:47:07	编辑 连接测试 更多 >
Hive_test_collection_5	Hive2	外部数据源		• 连接成功	2024-04-02 17:29:07	cgj1	2024-03-26 14:12:10	编辑 连接测试 更多 >
Hive_test_collection_3	Hive2	外部数据源		• 连接成功	2024-04-03 10:00:46	cgj1	2024-03-26 14:01:16	编辑 连接测试 更多 >
Hive_test_collection_2	Hive2	外部数据源		• 连接成功	2024-04-03 12:11:35	cgj1	2024-03-25 09:34:41	编辑 连接测试 更多 >
Hive_test_collection_1	Hive2	外部数据源		• 连接成功	2024-04-03 12:11:47	cgj1	2024-03-22 17:48:23	编辑 连接测试 更多 >
zsf_oracle_20240322_01	Oracle	外部数据源	oracle	• 连接成功	2024-04-02 17:35:45	zsf1	2024-03-22 14:34:52	编辑 连接测试 更多 >
Oracle_test_collection_1	Oracle	外部数据源		• 连接成功	2024-04-03 11:25:25	cgj1	2024-03-22 11:22:05	编辑 连接测试 更多 >

新增数据源

数据源管理模块支持新增多种关系型数据库、大数据存储、文件服务器和消息队列，其中，关系型数据库支持 MySQL、Oracle、SQL server、PostgreSQL、Greenplum、达梦等，大数据存储支持 HDFS、Hive、GaussDB、Doris、Hudi、HBase 等，文件服务器类型支持 FTP，NoSQL 型数据库支持 MongoDB，消息队列支持 Kafka。

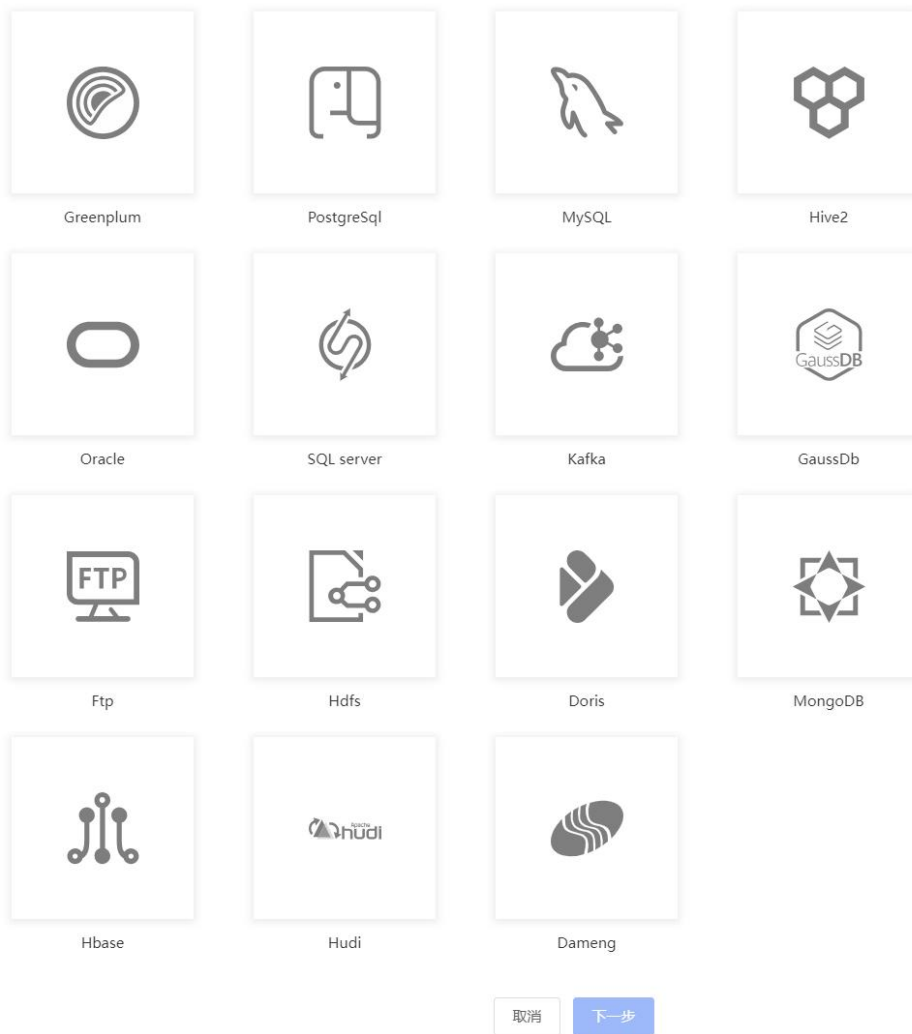
新增包括以下几个环节：

- 1.选择新增的数据源类型；
- 2.填写数据源连接信息；
- 3.连接测试；
- 4.如果测试通过，则数据源新增成功；反之，检查并重新编辑数据源连接信息再进行连通性测试。

选择新增的数据源

新增数据源

×



填写数据源连接信息

根据要求填写数据源连接信息（此处界面以 MySQL 为例）：

新增数据源

×

数据源类型	MySQL	▼
* 数据源名称	<input type="text"/>	0/128
标签	<input type="text" value="标签"/>	
数据源描述	<input type="text"/>	0/50
* 数据源驱动	请选择	▼
* 主机名	<input type="text"/>	
* 端口号	<input type="text"/>	
* 用户名	<input type="text"/>	
* 密码	<input type="password"/>	/?
数据库名称	<input type="text"/>	
jdbc连接参数	<input type="text"/>	
URL预览	jdbc:mysql:///?/?	
数据来源单位	<input type="text"/>	
所属应用系统	<input type="text"/>	
业务含义	<input type="text"/>	
安全性要求	<input type="text"/>	
提供方式	请选择	▼

连接测试

新增和编辑的数据源在保存之前必须进行一次连接测试，确保可联调，否则无法保存。

编辑数据源

数据源类型	MySQL
* 数据源名称	MySQL_test_collection_5 23/128
标签	mysql ×
数据源描述	<input type="text"/> 0/300
* 数据源驱动	8.0.30
* 主机名	192.168.61.44
* 端口号	50001
* 用户名	datawings
* 密码 <input type="text"/>
数据库名称	<input type="text"/>
jdbc连接参数	<input type="text"/>
URL预览	jdbc:mysql://192.168.61.44:50001/?

连接测试 取消 确定

搜索数据源

支持通过数据源类型、数据源名称、标签等对数据源进行组合搜索，支持关键词模糊搜索。用户输入搜索条件后，点击【查询】按钮，列表中将展示符合条件的数据源。点击【重置】按钮，可清空搜索条件。

编辑数据源

点击数据源对应操作栏的【编辑】按钮，或者直接点击数据源名称，将弹出数据源编辑弹窗，在弹窗中可编辑数据源的配置信息，编辑过的数据源需至少进行一次连接测试，点击【确定】按钮，即可保存最新填写的信息。

注：内置数据源通过引擎管理模块添加，数据源管理支持展示内置数据源信息，不支持编辑。

数据源名称	数据源类型	数据源类别	标签	测试连接结果	测试连接时间	创建用户	更新时间	操作
test_engine_reconfig_3_HIVE	Hive2	内置数据源		• 连接成功	2024-04-03 10:38:48	test_tenant_1	2024-04-03 10:38:49	编辑 连接测试 更多 >
test_engine_reconfig_3_DORIS	Doris	内置数据源		• 连接成功	2024-04-03 10:38:48	test_tenant_1	2024-04-03 10:38:49	编辑 连接测试 更多 >
MySQL_test_collection_5	MySQL	外部数据源	mysql	• 连接成功	2024-04-02 17:28:54	cyj1	2024-03-29 15:25:58	编辑 连接测试 更多 >
MySQL_test_collection_2	MySQL	外部数据源		• 连接成功	2024-04-03 15:25:30	cyj1	2024-03-26 14:47:07	编辑 连接测试 更多 >
Hive_test_collection_5	Hive2	外部数据源		• 连接成功	2024-04-02 17:29:07	cyj1	2024-03-26 14:12:10	编辑 连接测试 更多 >

删除数据源

点击数据源对应操作栏的【删除】按钮，将弹出二次提示弹窗，若点击【确定】按钮，将删除对应的数据源，若点击【取消】按钮，则返回列表页面。

注：数据源被删除后无法恢复，可能导致引用数据源的任务或 API 出错，请谨慎操作。内置数据源无法被删除，需要在集群管理-引擎管理处操作引擎下线后并删除引擎，数据源管理会自动删除对应的内置数据源。

复制数据源

点击数据源对应操作栏的【复制】按钮，将弹出“复制数据源”弹窗，可输入新的数据源名称，修改标签和描述等信息。填写完毕后点击【确定】按钮，将复制产生新的数据源，其配置信息与原数据源相同，若点击【取消】按钮，则返回列表页面。

复制数据源 ×

原数据源 MySQL_test_collection_5

* 数据源名称 0/128

标签

数据源描述 0/300

4.3 数据建模

4.3.1 数据建模概述

数据建模支持将无序、杂乱、繁琐、庞大且难以管理的数据，进行结构化有序的管理。使企业数据产生更多的价值，将数据价值最大化。


功能概述

- 数据建模支持制定并沉淀企业数据标准、数仓规划设计、数据指标定义、维度建模。
- 通过维度建模的正向建模可以将设计产出的接入表、维度表、明细表、汇总表和应用表等物化到相应的数据源。通过维度建模的逆向建模可以将其他建模工具建的存量物理表反向建模到维度建模中，进行统一管理。该功能无需再次执行建模操作，即可快速将已有模型反向建模至维度建模中，可节省大量的时间成本。

前提条件

在进行数据建模前，租户管理员需先选择规划模板。

1. 进入数仓规划。

点击左上方的，选择全部产品 > 数据建模 > 数仓规划，进入数仓规划。

2. 查看并选择模板。



- 点击“查看并选择模板”，可点击模板下拉框进行选择。

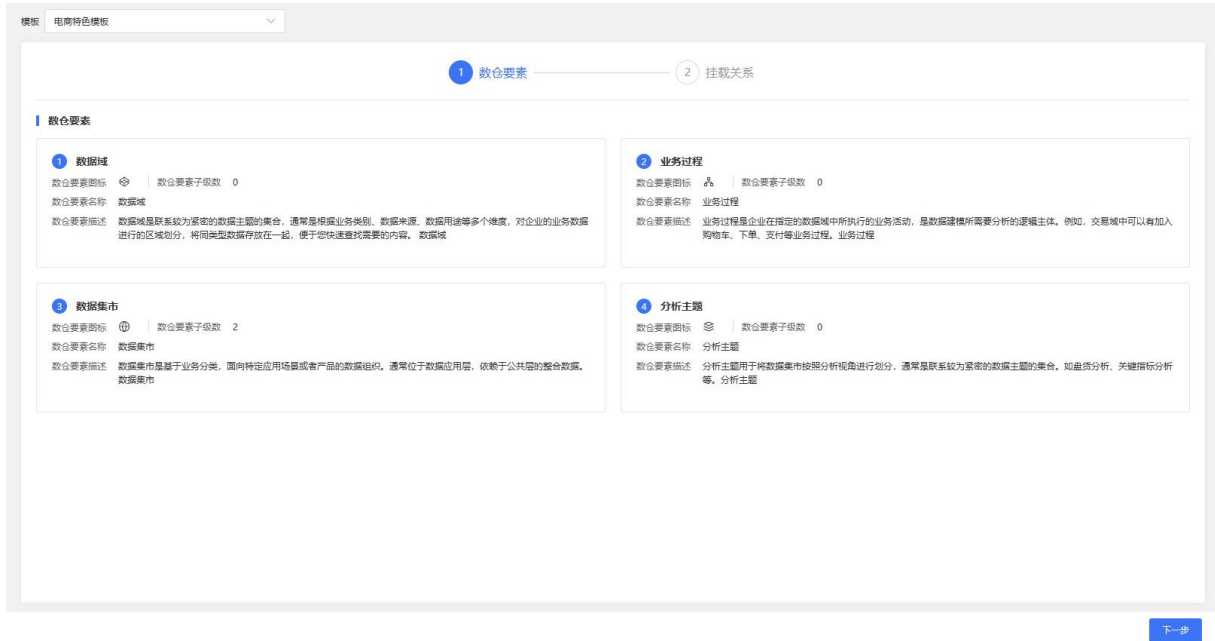


说明

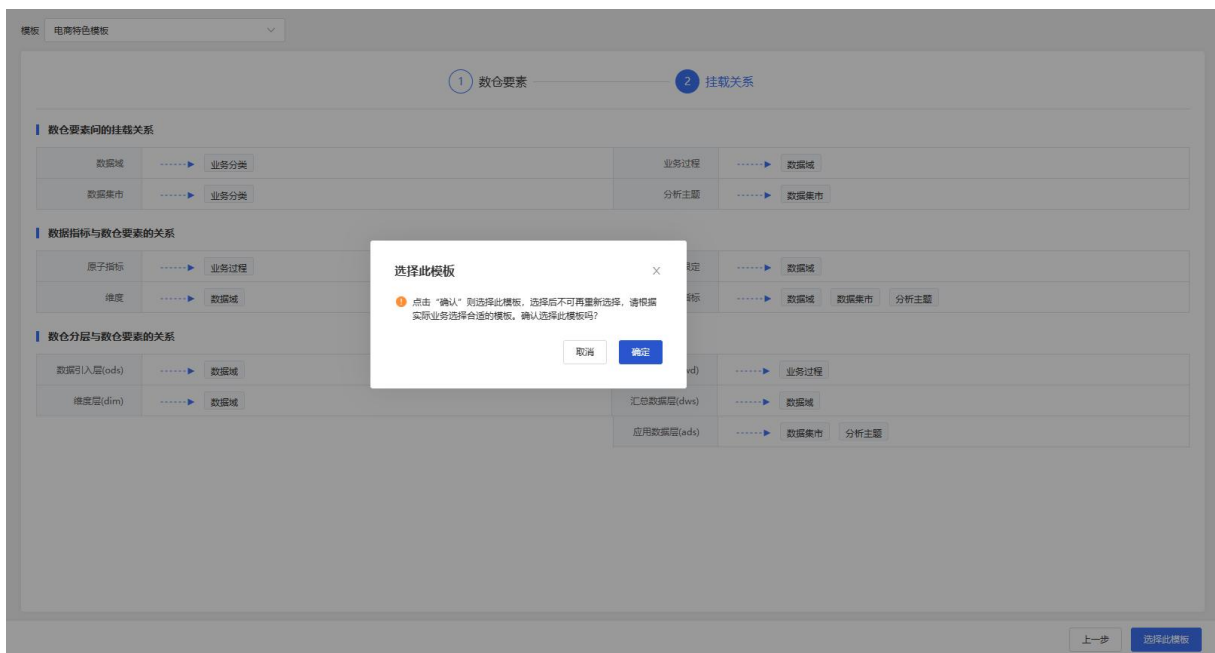
只有管理员可选择，一般用户只可以查看。

- 选择 1 个合适的模板，并点击下一步。

如下以租户管理员选择“电商特色模板”为例：




3. 点击“选择此模板”，并点击“确认”，则规划模板生效。



4.3.2 建模首页

建模首页，展示整个数据建模的功能架构图，点击各子模块可以跳转至对应的功能页面。同时展示原子指标、派生指标、各个分层的数据表模型数量和最新表模型发布情况，全局把控建模动态。

进入建模首页

单击左上方的图标，选择全部产品 > 数据建模 > 建模首页，默认进入建模首页。

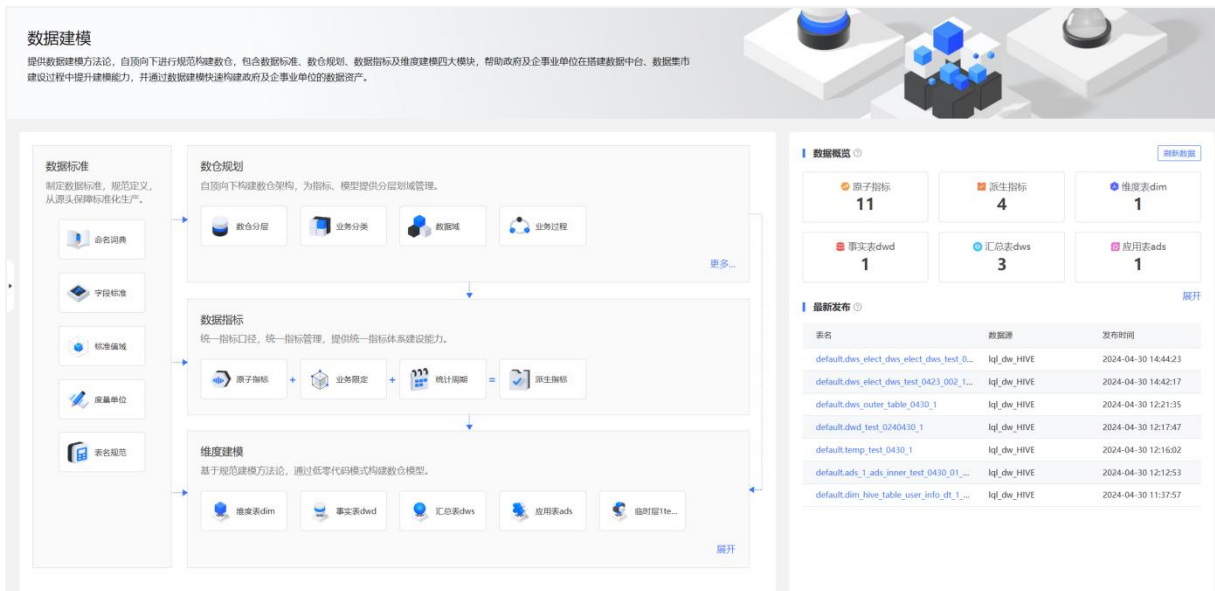
功能说明

在建模首页：分三个区域：上方横栏、左侧、右侧。

上方横栏：总体介绍数据建模包含的功能模块及用途。

左侧：功能架构图、功能说明及子模块跳转，点击各子模块可跳转到对应的功能页面。

右侧：查看当前租户内创建的原子指标、派生指标及各个分层的数据表模型数量、最新表模型发布情况，全局把控建模动态。



4.3.3 数据标准

4.3.3.1 命名词典


本章节主要介绍命名词典的用途及命名词典的操作步骤。

功能概述


命名词典为您提供管理各个数据要素，如业务名词、物理表、字段、原子指标的词根和词缀及其标准化翻译功能。在各个数据要素命名时，可以参考命名词典进行命名。命名词典模块支持新建、修改、删除命名词典，同时支持批量导入导出命名词典。

创建命名词典

1. 进入命名词典。

点击左上方的图标，选择全部产品 > 数据建模 > 数据标准，在数据标准页面左侧抽屉栏，单击命名词典，进入命名词典页面。

2. 创建一级分类。

- a. 在命名词典页面左侧的分类，鼠标悬浮至默认的父亲级分类“全部”右侧，点击“新建分类”。
- b. 在弹出的对话框中，输入分类名称，单击“确定”。一级分类创建成功后，可在左侧区域下方的目录树下方查看已创建的一级目录。

3. 创建命名词典，在“新建命名词典”对话框配置基本信息。

参数	描述
应用范围	<p>必选，选择该命名词典的应用范围。</p> <ul style="list-style-type: none"> 以应用范围选择“业务分类”为例：在建设业务分类时，如果没有先在命名词典建立记录，则该业务分类会打上“未在命名词典注册”标签。如果在命名词典上已注册，再去业务分类建该分类，则会打上“符合标准”标签。 如果应用范围选择“其它”时，则不对各类名词做约束，仅在命名词典记录。
英文缩写	必填，最大字符数为 128；此项确定后不可再修改。
英文名称	必填，最大字符数为 512；此项确定后不可再修改。
中文名称	必填，按照键入前文字规则输入，最大字符数为 512。
分类	必选，默认选择“命名词典”，归属于“全部”这个分类。您可从已创建的分类中选择。
备注	选填，最大字符数为 2048。

新增命名词典 ×

* 应用范围

* 英文缩写 ?


* 英文名称 ?


* 中文名称

* 分类

备注 0/2048

批量导入命名词典

如您有大量命名词典待创建, 可点击图标  , DataWings 为您提供了导入模板, 您可按照模板表格填写好待导入的命名词典后批量导入。

1. 参考上述步骤进入命名词典页面后, 单击  。
2. 下载模板。
3. 在本地填写模板。
4. 选择文件, 目前仅支持 excel 文件。
5. 点击“确定”。



查看导入结果:

- 点击弹窗的导入单号, 进入导入单详情。
- 通过全部产品 > 系统管理 > 流程中心 > 我的导入, 点击导入单号, 进入导入单详情。
- 点击“完成”, 刷新当前页面查看结果。

命名词典管理

- 批量导出, 点击  批量导出。



导出已选: 则按照用户勾选的记录导出;

导出已搜: 则按搜索结果导出;

导出全部：则导出全部记录。

- 批量删除，选择列表上需删除的命名词典，点击“批量删除”。

命名词典 ▲

命名词典为您提供管理各个数据要素，如业务名词、物理表、字段的词根和词素及其标准化翻译功能。在各个数据要素命名时，可以参考命名词典进行命名。您可以按 [命名词典](#)

新建命名词典		批量删除		请选择应用范围		请输入英文缩写/英文名称/中文名称搜索		查询	重置		
序号	英文缩写	英文名称	中文名称	分类	应用范围	创建人	创建时间	修改人	修改时间	备注	操作
<input checked="" type="checkbox"/>	1	dfdf	dfdf	策略	命名词典	存储策略	pocfr	2024-04-20 17:05:11	pocfr	2024-04-20 17:05:17	引用分析 修改 删除
<input checked="" type="checkbox"/>	2	xsbm	xsbm	学生报名	学生	数据域	wangy...	2023-11-07 13:45:49	fur33	2024-04-20 15:19:51	数据子域 引用分析 修改 删除
<input type="checkbox"/>	3	dwd	DataWareh...	明细数据层	系统内置分类	数仓分层	system	2023-04-14 08:38:53	system	2024-04-16 00:00:00	引用分析 修改 删除
<input type="checkbox"/>	4	dws	DataWareh...	汇总数据层	系统内置分类	数仓分层	system	2023-04-14 08:38:53	system	2024-04-16 00:00:00	引用分析 修改 删除
<input type="checkbox"/>	5	ads	Application...	应用数据层	系统内置分类	数仓分层	system	2023-04-14 08:38:54	system	2024-04-16 00:00:00	引用分析 修改 删除
<input type="checkbox"/>	6	dim	Dimension	维度层	系统内置分类	数仓分层	system	2023-04-14 08:38:54	system	2024-04-16 00:00:00	引用分析 修改 删除
<input type="checkbox"/>	7	ods	Operational...	数据引入层	系统内置分类	数仓分层	system	2023-04-14 08:38:53	system	2024-04-16 00:00:00	引用分析 修改 删除
<input type="checkbox"/>	8	cjfb	cjfb	成绩发布	命名词典	数据域	wangy...	2023-11-07 13:45:50	wangy...	2023-11-07 13:45:50	数据子域 引用分析 修改 删除
<input type="checkbox"/>	9	XLJY	XLJY	学历教育	命名词典	分析主题	wangy...	2023-11-07 13:45:49	wangy...	2023-11-07 13:45:49	专题域 引用分析 修改 删除
<input type="checkbox"/>	10	JYFZPJ	JYFZPJ	教育发展评价	命名词典	分析主题	wangy...	2023-11-07 13:45:49	wangy...	2023-11-07 13:45:49	专题域 引用分析 修改 删除

共46条 < 1 2 3 4 5 > 10 前往 1 页

- 引用分析，点击列表右侧的“引用分析”查看。

英文名称	中文名称	英文缩写	引用类型
test.dim_hive_in...	用户信息表数据di...	/	表模型

引用分析

- 表模型 (存储策略) 1个

关闭

后续步骤

业务分类、数仓分层、字段标准、原子指标、维度，可视化建模的命名可参照命名词典。

4.3.3.2 字段标准

本章节主要介绍字段标准的用途及操作步骤。

功能概述

字段标准是对含义相同但字段名称不同的数据进行统一规范管理的数据准则。可将多个表中含义相同但字段名不同的内容进行关联，并对该字段制定相关的取值范围、度量单位、标准值域等内容。

前提条件（建议）


建议先去“命名词典”注册词根词缀，再建字段标准，以达到规范命名的目的。

如果字段标准的组成元素不在命名词典注册，则会打上“词根词缀未注册”或“词根词缀部分注册”标签，提醒用户去注册。

如果组成元素已在命名词典注册，则会打上“词根词缀已注册”标签。

创建字段标准

1. 进入字段标准。

点击左上方的，选择全部产品 > 数据建模 > 数据标准；在数据标准页面左侧抽屉栏，单击字段标准，进入字段标准页面。

2. 创建字段标准。

单击“新建字段标准”，在“新建字段标准”对话框配置基本信息。

参数	描述
标准编码	自定义编码；必填，按照键入前文字规则输入，最大字符数为 128； 此项确定后不可修改。
英文缩写	必填，按照键入前文字规则输入，最大字符数为 128； 此项确定后不可修改。
英文名称	必填，按照键入前文字规则输入，最大字符数为 512； 此项确定后不可修改。
中文名称	必填，按照键入前文字规则输入，最大字符数为 512。
数据类型	必选，下拉框单选，可按照实际需要选择对应的数据类型，含已经定义好长度和精度的 DECIMAL 类型。
非空	选填，引用该标准的字段值是否可以为空。默认为空。
默认值	选填，引用该标准的字段值未填写时的默认值。
格式校验	填写正则表达式。
业务定义	选填，最大字符数为 2048。

方式 1：不在命名词典注册词根词缀，按照上述参数要求，直接新建字段标准。

字段标准列表 新建字段标准 x

建议先去命名词典 [注册词根词缀](#)，再建字段标准，规范命名！ 查看命名词典

* 标准编码

* 英文缩写

* 英文名称

* 中文名称

* 数据类型

非空 引用该标准的字段值是否可以为空

默认值 0/1024

格式校验 0/512 复制

业务定义 0/2048

取消 确定

点击“确定”后，在新的弹窗点击“不去注册，并确认”，则直接生效。

字段标准列表 新建字段标准 x

建议先去命名词典 [注册词根词缀](#)，再建字段标准，规范命名！ 查看命名词典

* 标准编码

* 英文缩写

* 英文名称

* 中文名称

* 数据类型

非空 引用该标准的字段值是否可以为空

默认值 0/1024

格式校验 0/512 复制

业务定义 0/2048

取消 确定

温馨提示

该字段标准的
英文名称 **user_type** 的分词 **user, type**
英文缩写 **user_type** 的分词 **user, type**
在词根词缀的英文缩写、英文名称中均不存在；建议先去命名词典注册词根词缀，再建字段标准，规范命名！

取消 去注册词根词缀 不去注册，并确认

方式 2：在命名词典注册词根词缀，再新建字段标准。

字段标准列表 新建字段标准 x

建议先去命名词典 [注册词根词缀](#)，再建字段标准，规范命名! [查看命名词典](#)

* 标准编码

* 英文缩写

* 英文名称

* 中文名称

* 数据类型

非空 引用该标准的字段值是否可以为空

默认值 0/1024

格式校验 0/512

业务定义 0/2048

点击“确定”后，在新的弹窗点击“去注册词根词缀”，则系统自动跳转到命名词典注册词根词缀的页面。

温馨提示 ×

该字段标准的
英文名称 **user_age** 的分词 **user, age**
英文缩写 **user_age** 的分词 **user, age**
在词根词缀的英文缩写、英文名称中均不存在；建议先去命名词典注册词根词缀，再建字段标准，规范命名！

系统自动跳转到命名词典注册词根词缀的页面，填写命名词典各参数；点击“确定”，则该词根词缀注册完成。

新增命名词典 ×

* 应用范围 × ▼

* 英文缩写 ?

* 英文名称 ?

* 中文名称 ×

* 分类

备注 0/2048

再返回新建字段标准弹窗，点击“确定”，则生成符合规范的字段标准。

批量导入字段标准

如您有大量字段标准待创建，则可选择“导入字段标准”，DataWings 为您提供了导入模板，您可按照模板表格填写好待导入的字段标准后批量导入。

1. 参考上述步骤进入字段标准页面后，单击“导入字段标准”。
2. 下载模板。
3. 在本地填写模板。
4. 选择文件，目前仅支持 excel 文件。
5. 点击“确定”。



查看导入结果：

- 点击弹窗的导入单号，进入导入单详情。
- 通过全部产品 > 系统管理 > 流程中心 > 我的导入，点击导入单号，进入导入单详情。
- 点击“完成”，刷新当前页面查看结果。

管理字段标准

- 修改：修改时，标准编码、英文缩写与英文名称不可修改。
- 删除：若有引用，删除所有引用关系后，才可以删除该字段标准；若无引用，进行二次确认，则可删除。
- 导出字段标准：有 3 种导出模式。可按需选择。



后续步骤

- 在“可视化建模”的字段设置中引用字段标准。
- 在“创建数据质量规则”中建字段标准的格式校验规则。

4.3.3.3 标准值域

本章节介绍标准值域的用途、标准值域的增删改操作及后续步骤。


功能概述

标准值域是字段标准的取值范围，在标准值域中可设置字段标准可选择的数据内容及范围。

标准值域与枚举维度相似，但数据治理的作用要强于枚举维度，标准值域可以与数据质量打通，某个字段被关联标准值域后，该字段的取值范围必须在标准值域内，如取值范围不在其内，需要马上治理。

创建标准值域

1. 进入标准值域。

点击左上方的，选择全部产品 > 数据建模 > 数据标准；在数据标准页面左侧抽屉栏，单击标准值域，进入标准值域页面。

2. 创建标准值域。

单击“新建标准值域”，在“新建标准值域”对话框配置基本信息。

参数	描述
基本信息-英文名称	必填，按照键入前文字规则输入，最大字符数为 512； 此项确定后不可再修改。
基本信息-中文名称	必填，按照键入前文字规则输入，最大字符数为 512。
基本信息-值域描述	非必填，对值域的描述，最大字符数为 2048。
编码信息	<ul style="list-style-type: none"> ● 编码取值：自定义编码；必填，最大字符数为 128。 ● 编码英文名：非必填，最大字符数为 512。 ● 编码中文名：非必填，最大字符数为 512。 ● 编码含义：非必填，最大字符数为 2048。

新增标准值域
×

基本信息

* 英文名称

* 中文名称

值域描述 0/2048

编码信息

* 编码取值	编码英文名	编码中文名	编码含义	操作
<input style="width: 95%;" type="text"/>	<input style="width: 95%;" type="text"/>	<input style="width: 95%;" type="text"/>	<input style="width: 95%;" type="text"/>	

+ 添加

取消
确定

标准值域新建和编辑时支持添加多个值域编码。点击“确定”后，系统自动刷新界面，可查看

列表。

标准值域 ▲

标准值域是字段标准的取值范围，在标准值域中可设置字段标准可选择的数据内容及范围。标准值域与枚举维度相似，但数据治理的作用要源于枚举维度。标准值域可以与数据质量打通，某个字段被设定为标准值域后，该字段的取值范围必须在标准值域内，如取值范围不在其内，需要马上治理，如性别字段设置标准值域取值为男和女；则数据仅能存储男或女，存在其他数据时，需要马上治理。您可以[新建标准值域](#)

[新建标准值域](#) [导入标准值域](#) [导出标准值域](#) [查询](#) [重置](#)

序号	英文名称	中文名称	描述	创建人	创建时间	修改人	修改时间	操作
1	gender	性别		zzfl	2024-03-27 19:49:24	zzfl	2024-03-27 19:49:24	修改 删除

序号	编码取值	编码英文名	编码中文名	编码含义
1	f	female	女性	
2	m	male	男性	

批量导入标准值域

如您有大量标准值域待创建，则可选择“导入标准值域”，DataWings 为您提供了导入模板，您可按照模板表格填写好待导入的标准值域后批量导入。

1. 参考上述步骤进入标准值域页面后，单击“导入标准值域”。
2. 下载模板。
3. 在本地填写模板。
4. 选择文件，目前仅支持 excel 文件。
5. 点击“确定”。

导入

导入模式 跳过已存在的同名对象

1. 请按照模板格式准备要导入的数据
[下载模板](#)

2. 选择需要导入的文件
[选取文件](#)

*仅支持导入excel文件

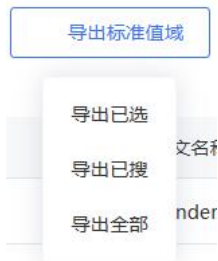
[取消](#) [确定](#)

查看导入结果：

- 点击弹窗的导入单号，进入导入单详情。
- 通过全部产品 > 系统管理 > 流程中心 > 我的导入，点击导入单号，进入导入单详情。
- 点击“完成”，刷新当前页面查看结果。

管理标准值域

- 修改：修改时，英文名称不可修改。
- 删除：进行二次确认，则可删除。
- 导出标准值域：有 3 种导出模式。可按需选择。



后续步骤

- 在可视化建模的字段设置中引用标准值域。
- 在创建数据质量规则中建值域的校验规则。

4.3.3.4 度量单位

度量单位即数据字段参数的数量单位。本章节主要介绍度量单位的创建方法。

功能概述

数据字段参数的数量单位(如元、个、米等)。

创建度量单位

1. 进入度量单位。

点击左上方的■，选择全部产品 > 数据建模 > 数据标准；在数据标准页面左侧抽屉栏，单击度量单位，进入度量单位页面。

2. 创建度量单位。

单击“新建度量单位”，在“新建度量单位”对话框配置基本信息。

参数	描述
英文缩写	必填，按照键入前文字规则输入，最大字符数为 128。 此项确定后不可再修改。
英文名称	必填，按照键入前文字规则输入，最大字符数为 512。 此项确定后不可再修改。
中文名称	必填，按照键入前文字规则输入，最大字符数为 512。
分类	必选，下拉框单选，度量单位的类型。 货币单位 对象量词 比例单位 排各单位 时间单位 长度单位 其他

新增度量单位 ×

* 英文缩写

* 英文名称

* 中文名称

* 分类

批量导入度量单位

如您有大量度量单位待创建, 则可选择“导入度量单位”, DataWings 为您提供了导入模板, 您可按照模板表格填写好待导入的度量单位后批量导入。

1. 参考上述步骤进入度量单位页面后, 单击“导入度量单位”。
2. 下载模板。
3. 在本地填写模板。
4. 选择文件, 目前仅支持 excel 文件。
5. 点击“确定”。



查看导入结果：

- 点击弹窗的导入单号，进入导入单详情。
- 通过全部产品 > 系统管理 > 流程中心 > 我的导入，点击导入单号，进入导入单详情。
- 点击“完成”，刷新当前页面查看结果。

管理度量单位

- 修改：修改时，英文缩写、英文名称不可修改。
- 删除：若已经被模型或原子指标引用，需要删除所有引用关系后，才可以删除该度量单位。
若无引用，进行二次确认，则可删除。
- 导出度量单位：有 3 种导出模式。可按需选择。



后续步骤

- 在可视化建模的字段设置中引用度量单位。

- 在原子指标配置中引用度量单位。

4.3.3.5 规范管理

规范管理用于规范表模型及指标的命名，制定表模型及指标的命名规范。并提供表模型“待治理清单”。


功能概述

规范管理用于规范表模型及指标的命名，制定表模型及指标的命名规范，在模型和指标创建时，系统会根据规范推荐表名和自动生成指标名，使用方能通过表名和指标名快速了解对应的业务信息。不符合表名规范的表，则在“待治理清单”中展示，形成治理闭环。

表名规范

创建表名规范规则

1. 进入表名规范。

点击左上方的，选择全部产品 > 数据建模 > 数据标准；在数据标准页面左侧抽屉栏，单击规范管理，默认进入“表名规范”页面。

2. 创建表名规范规则。

单击“新建规则”，在“新建规则”对话框配置基本信息。

其中表名由 N 个前缀+表名主体+N 个后缀拼接而成。其中表名主体在维度建模中填写表名主体。前缀与后缀由拼接项组成。

参数	描述
挂载类型	下拉单选，表模型所属的挂载类型，如数据域、业务过程等。
点击新建拼接项 (前缀/后缀)	<p>每个拼接项，单选。拼接项可选固定字符串、业务分类缩写、数仓要素、正则表达式、存储策略。</p> <p>其中：</p> <ul style="list-style-type: none"> ● 固定字符：合法字符，最大字符数为 64；比如填写了 <code>dwd</code>，则表名里固定有 <code>dwd</code> 拼接项。 ● 业务分类缩写：在业务分类里建的业务分类的缩写。根据用户在维度建模时选择具体的业务分类进行拼接。 ● 数仓要素：在数仓要素建的要素，英文缩写。根据用户在维度建模时选择具体的数仓要素进行拼接。 ● 正则表达式：为数字或字母，最大字符数为 128。根据用户在维度建模建表时，修改成具体的数字或字母进行拼接。 ● 存储策略：在命名词典里建的应用范围含【存储策略】的命名词典。根据用户在维度建模时选择具体的存储策略进行拼接。主要用于数据引入层、维度层、明细层和自定义分层的表模型。 ● 统计周期：用于限定统计数据的时间范围，主要用于汇总层、应用层的表模型。

如下例子：

新建「维度层」规则

* 挂载类型: 数据域

前缀	控制逻辑	操作
p1	固定字符串	dim 3/64
p2	业务分类缩写	表名会按照用户选择的直接挂载追溯到业务分...
p3	一级数据域缩写	表名会按照用户选择的直接挂载进行替换

+ 新增拼接项

后缀	控制逻辑	操作
s1	正则表达式	[0-9a-z]{1,128}
s2	存储策略	表名会按照用户选择的存储策略进行替换

取消 确定

生成的表名规范为：

规范列表 待治理列表

表名规范 指标名规范

数据分层	序号	挂载类型	表名规则	修改人	修改时间	操作
数据引入层	1		共1条规则			新建规则
维度层	2		共4条规则			新建规则
	2.1	数据域	业务分类_表名主体_一级数据域	fr1	2024-01-28 14:20:39	查看 修改 删除 查看检测结果 禁用 <input checked="" type="checkbox"/> 启用 <input type="checkbox"/> 弱规则 <input checked="" type="checkbox"/> 强规则 <input type="checkbox"/>
	2.2	数据域	[0-9a-z]{1,128}_表名主体_cdcd_存储策略	fr1	2024-01-28 15:13:48	查看 修改 删除 查看检测结果 禁用 <input checked="" type="checkbox"/> 启用 <input type="checkbox"/> 弱规则 <input checked="" type="checkbox"/> 强规则 <input type="checkbox"/>
	2.3	数据域	[0-9a-z]{1,128}_表名主体_存储策略	fr1	2024-01-28 15:48:00	查看 修改 删除 查看检测结果 禁用 <input checked="" type="checkbox"/> 启用 <input type="checkbox"/> 弱规则 <input checked="" type="checkbox"/> 强规则 <input type="checkbox"/>
	2.4	数据域	dim_业务分类_一级数据域_表名主体_[0-9a-z]{1,128}_存储策略	zaf1	2024-04-01 16:53:08	查看 修改 删除 查看检测结果 禁用 <input checked="" type="checkbox"/> 启用 <input type="checkbox"/> 弱规则 <input checked="" type="checkbox"/> 强规则 <input type="checkbox"/>
明细数据层	3		共2条规则			新建规则
汇总数据层	4		共58条规则			新建规则
应用数据层	5		共3条规则			新建规则

表名规则管理

- 修改：可以修改填写时的规范。
- 删除：可删除规范。

- 禁用/启用：如果该规则不想被引用且不想删掉记录，可点击禁用。
- 弱规则/强规则：弱规则，在维度建模引用后，可以不按规范命名，在治理清单列出不符合规范的表名。强规则，在维度建模引用后，必须按照规范命名。


后续步骤

- 在维度建模创建和修改表模型时引用“表名规范”，详见“维度建模”。
- 点击查看检测结果，则可查看不符合规范的清单。

指标名规范

管理指标名规范

1. 进入指标名规范。

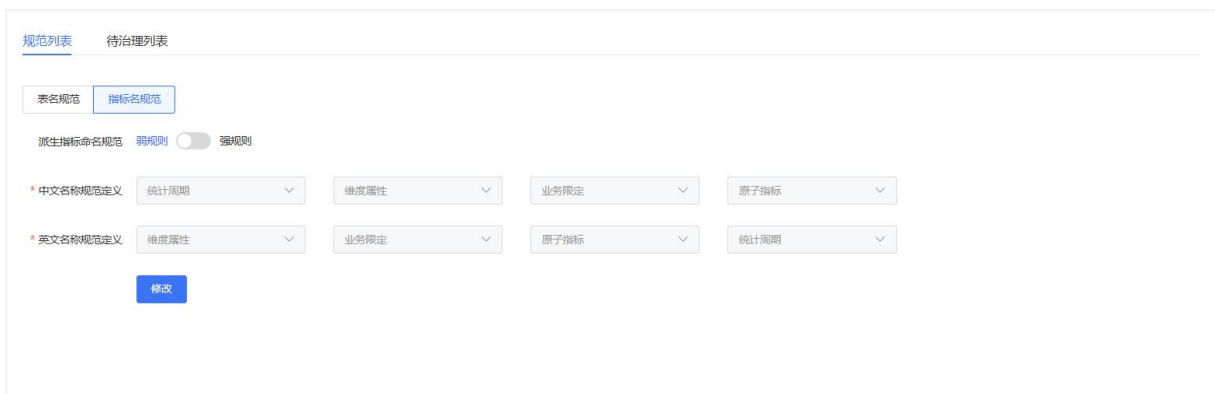
点击左上方的，选择全部产品 > 数据建模 > 数据标准；在数据标准左侧抽屉栏，单击规范管理，在规范列表，选择“指标名规范”页签。

2. 修改指标名规范。

默认的指标名规范如下：

自动生成的派生指标中文名，用下划线连接各元素：统计周期_维度属性_业务限定_原子指标；

自动生成的派生指标英文名，用下划线连接各元素：维度属性_业务限定_原子指标_统计周期。



规范列表 待治理列表

表名规范 指标名规范

派生指标命名规范 弱规则 强规则

* 中文名称规范定义 统计周期 维度属性 业务限定 原子指标

* 英文名称规范定义 维度属性 业务限定 原子指标 统计周期

修改

可点击“修改”进入编辑模式：

- 可拨动开关，选择强规则或弱规则。
- 可把已选的规范定义选项全部去除，再重新选择各元素按新的顺序拼接。

3. 点击“保存”，保存指标名规范。

后续步骤

派生指标按已配置的指标名规范自动生成英文名与中文名，详见派生指标。

4.3.4 数仓规划

4.3.4.1 业务分类


业务分类模块支持创建编辑一级、二级分类，在业务复杂且较为庞大时，可通过业务分类管理对数据进行业务区分。

新建一级分类

1. 进入业务分类。

点击左上方的，选择全部产品 > 数据建模 > 数仓规划，默认进入业务分类页面。

2. 创建业务分类。

- 在业务分类页面，点击，弹出“新建一级业务分类”弹窗。



- 配置“新建一级业务分类”对话框的各项参数。

参数	描述
英文缩写	必填，按照键入前文字规则输入，最大字符数为 128。
英文名称	必填，按照键入前文字规则输入，最大字符数为 512。
中文名称	必填，按照键入前文字规则输入，最大字符数为 512。
负责人	必填，默认为当前登录账号，可切换成其他账号，可多选。
备注	最大字符数为 2048。

方式 1：不在命名词典注册，直接按照上述参数要求，新建一级分类。

新建一级业务分类 ×

1 建议先注册命名词典，再建业务分类，规范命名！ [查看命名词典](#)

* 英文缩写

* 英文名称

* 中文名称

* 负责人

备注

0/2048

点击“确定”，生成“未注册到命名词典”的分类。

业务分类 ▲

如您所在的集团或公司业务复杂且较为庞大时，可创建业务分类，对数据进行业务区分，便于对业务数据进行更加方便的管理。您可以按需 [新建业务分类](#)

[查看后续步骤](#)


基础信息

英文缩写	education	英文名称	education
中文名称	教育分类	治理标签	未注册到命名词典
创建人	yanshi	创建时间	2023-02-22 11:25:22
修改人	system	修改时间	2023-04-06 11:15:35
负责人	yanshi(yanshi_luo管理员)	备注	

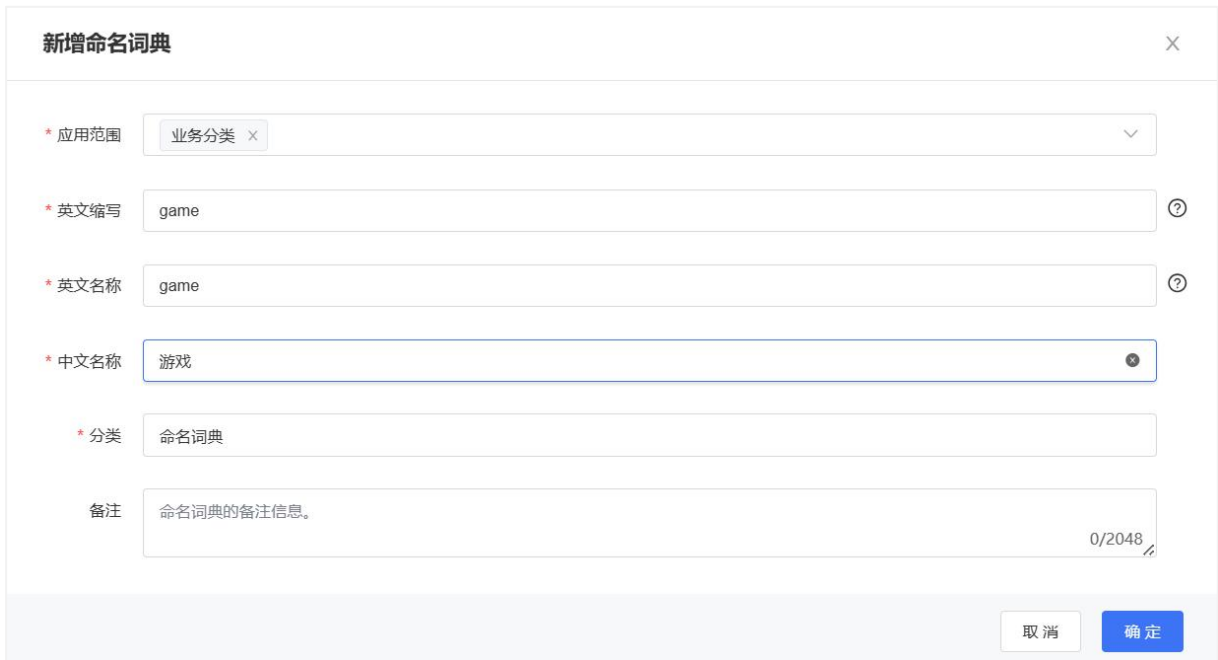
子业务分类 + 新建子业务分类

子分类英文缩写	子分类英文名称	子分类中文名	治理标签	负责人	创建人	创建时间	修改人	修改时间	操作
k12	k12	K12	未注册到命名词典	yanshi(yanshi_luo管理员)	yanshi	2023-02-22 11:25:22	system	2023-04-06 11:15:35	删除
highedu	highedu	高级教育	未注册到命名词典	yanshi(yanshi_luo管理员)	yanshi	2023-02-22 11:28:52	system	2023-04-06 11:15:35	删除

方式 2：先注册命名词典，再建业务分类，规范命名。

进入命名词典：点击左上方的，选择全部产品 > 数据建模 > 数据标准 > 命名词典，进入命名词典页面。

选择应用范围是业务分类，并填写命名词典参数。



新增命名词典

* 应用范围


* 英文缩写

* 英文名称

* 中文名称

* 分类

备注 0/2048

在业务分类页面，点击，弹出“新建一级业务分类”弹窗，输入英文缩写、英文名称、中文名称任一参数，点击“用命名词典填充”。

新建一级业务分类

建议先注册命名词典，再建业务分类，规范命名！ [查看命名词典](#)

* 英文缩写

* 英文名称

* 中文名称

* 负责人

备注 0/2048

根据推荐命名，选择其中符合的命名。

温馨提示

如下是在命名词典找到的匹配记录，默认选择第一条，您可切换选择！建议【用已选词典填充，并确认】，规范命名。

英文缩写	英文名称	中文名称	应用范围	备注	操作
game	game	游戏	业务分类		<input checked="" type="radio"/> 用此行填充

点击“用已选词典填充，并确认”，生成符合标准的分类，且默认生成 1 个默认的子分类。



业务分类

输入业务分类搜索

- 电子商务
 - 直播电商
- 教育分类
- 订单
- 汉字Ab&011_
- 游戏

业务分类

如您所在的集团或公司业务复杂且较为庞大时，可创建业务分类，对数据进行业务区分，便于对业务数据进行更加方便的管理。您可以按需 [新建业务分类](#)

[查看后续步骤](#)


基础信息

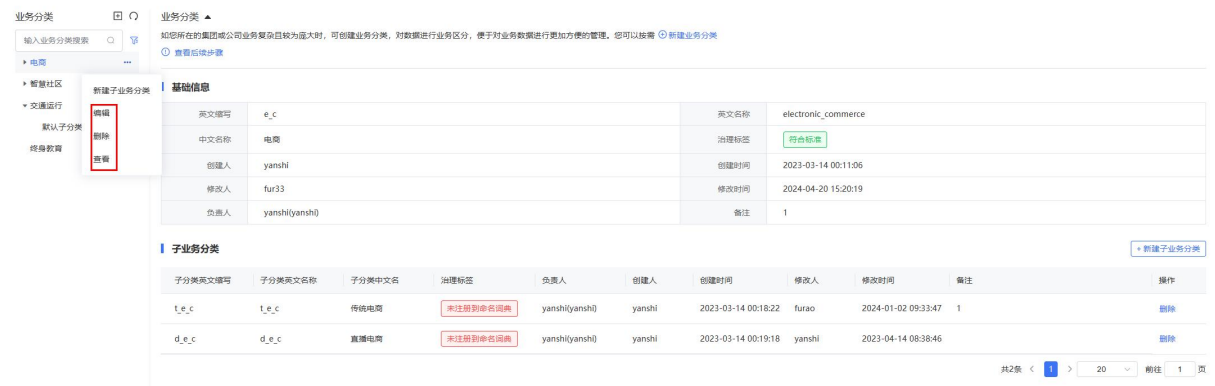
英文缩写	game	英文名称	game
中文名称	游戏	治理标签	符合标准
创建人	yanshi	创建时间	2023-08-31 20:01:34
修改人	yanshi	修改时间	2023-08-31 20:01:34
负责人	yanshi(yanshi_luo管理员)	备注	

子业务分类 [+ 新建子业务分类](#)

子分类英文缩写	子分类英文名称	子分类中文名称	治理标签	负责人	创建人	创建时间	修改人	修改时间	操作
default3	default3	默认子分...	未注册到命名词典	yanshi(yanshi_luo管理员)	yanshi	2023-08-31 20:01:34	yanshi	2023-08-31 20:01:34	删除

管理一级分类

管理一级分类：在一级业务分类下，悬停至一级分类名称后 ，点击编辑、删除或查看。



业务分类

输入业务分类搜索

- 智慧社区
- 交通出行
- 终身教育

业务分类

如您所在的集团或公司业务复杂且较为庞大时，可创建业务分类，对数据进行业务区分，便于对业务数据进行更加方便的管理。您可以按需 [新建业务分类](#)

[查看后续步骤](#)

基础信息


英文缩写	e_c	英文名称	electronic_commerce
中文名称	电商	治理标签	符合标准
创建人	yanshi	创建时间	2023-03-14 00:11:06
修改人	fur33	修改时间	2024-04-20 15:20:19
负责人	yanshi(yanshi)	备注	1

子业务分类 [+ 新建子业务分类](#)

子分类英文缩写	子分类英文名称	子分类中文名称	治理标签	负责人	创建人	创建时间	修改人	修改时间	备注	操作
t_e_c	t_e_c	传统电商	未注册到命名词典	yanshi(yanshi)	yanshi	2023-03-14 00:18:22	furao	2024-01-02 09:33:47	1	删除
d_e_c	d_e_c	直播电商	未注册到命名词典	yanshi(yanshi)	yanshi	2023-03-14 00:19:18	yanshi	2023-04-14 08:38:46		删除

共2条 < 1 > 20 前往 1 页

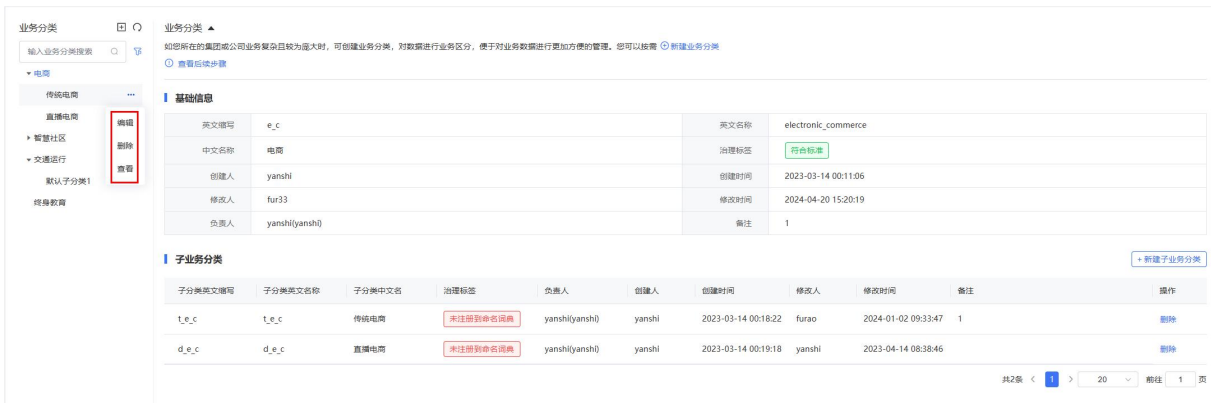
新建子分类

新建子分类：在一级业务分类下，悬停至一级分类名称后 ，点击“新建子业务分类”，创建方法与“新建一级业务分类”相同。



管理子分类

悬停至子分类名称后[⋮]，点击编辑、删除或查看。



后续步骤

在配置数仓要素：如数据域、数据集市时，可挂载到具体的业务分类。具体可看“数仓要素”。

4.3.4.2 数仓分层

本章节主要介绍数仓分层的用途及新建数仓分层的步骤。

功能概述

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物理层的详细划分。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。数仓分层还用于规范表模型的命名。

规划数仓分层

系统默认创建的数仓分层有：

- **数据引入层 ODS**

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。

- **维度层 DIM**

DIM 层使用维度构建数据模型。基于实际业务，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，构建整个企业一致性数据维表，帮助您降低数据计算口径和算法不统一的风险。

- **明细数据层 DWD**

DWD 层通过企业业务活动构建数据模型。基于具体业务事件特点，构建最细粒度明细数据表。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。

- **汇总数据层 DWS**

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。

- **应用数据层 ADS**

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。

如下方前 5 个分层是系统内置的分层；同时，针对不同业务使用场景和架构规划，也支持新建管理自定义数仓分层。如第 6 个分层。

数仓分层

数仓分层是结合业务场景、实际数据、使用系统的综合分析，对数据模型进行的具体结构设计及物理实现的详细部分。用于将不同来源的数据，按照划分不同的分层，便于更好地组织、管理、维护数据。数仓分层适用于规范类模型的命名规范。您可以[查看](#) [新建数仓分层](#)


[查看帮助文档](#)

<p>数据引入层 ods</p> <p>负责人: 系统 符合标准</p> <p>修改时间: 2023-06-20 16:52:48</p> <p>描述: ODS用于接收并处理需要存储至数据仓库系统的原始数据。其数据源的结构与原始数据所在的数据系统中的表结构一致。是数据仓库的数据源基础。</p>	<p>维度层 dim</p> <p>负责人: 系统 符合标准</p> <p>修改时间: 2023-06-20 16:52:48</p> <p>描述: DIM层使用维度构建数据模型。基于实际业务，通过定义维度，确定维度主题，添加维度属性，关联不同维度等操作，构建整个企业一致性的数据模型，解决因降低数据计算口径和算法不统一的风险。</p>	<p>明细数据层 dwd</p> <p>负责人: 系统 符合标准</p> <p>修改时间: 2023-06-20 16:52:48</p> <p>描述: DWD层通过企业业务活动构建数据模型。基于具体业务事件特点，构建明细粒度明细数据表。您可以结合企业的业务使用特点，标明数据表的某些重要维度属性字段进行冗余，即宽表化处理。</p>
<p>汇总数据层 dws</p> <p>负责人: 系统 符合标准</p> <p>修改时间: 2023-06-20 16:52:48</p> <p>描述: DWS层通过分析的主题构建数据模型。基于上层应用和产品的需求需求，构建公共粒度的汇总指标事实表。</p>	<p>应用数据层 ads</p> <p>负责人: 系统 符合标准</p> <p>修改时间: 2023-06-20 16:52:48</p> <p>描述: 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。</p>	<p>专题数据层 topic</p> <p>负责人: yemah@2023-11-01 更快捷，请及时创建个人于账号 符合标准</p> <p>修改时间: 2023-06-06 15:25:32</p> <p>描述: 专题数据层</p>

创建数仓分层

系统已默认为您创建了五个分层（ODS层、DIM层、DWD层、DWS层、ADS层），可满足绝大部分的场景和需求，如您存在部分需要个性化处理的场景，可参见如下步骤创建新的数仓分层。个性化处理的场景示例：抽象出TMP（临时表）层、TOPIC（专题）层，在创建表时严格将表归属到对应分层，对每个分层设定一些标准和校验规则，譬如命名规范等。

1. 进入数仓分层。

点击左上方的，选择全部产品 > 数据建模 > 数仓规划，在数仓规划左侧抽屉栏，单击数仓分层，进入数仓分层页面。

2. 创建数仓分层

单击“新建数仓分层”，在“新建数仓分层”对话框配置分层的基本信息。

参数	描述
英文缩写	必填，按照键入前文字规则输入，最大字符数为 128。
英文名称	必填，按照键入前文字规则输入，最大字符数为 512。
中文名称	必填，按照键入前文字规则输入，最大字符数为 512。
负责人	必填，默认为当前登录账号，可切换成其他账号，可多选。
备注	最大字符数为 2048。

方式 1：不在命名词典注册，按照上述参数要求，直接新建数仓分层。

新建数仓分层

×

① 建议先注册命名词典，再建数仓分层，规范命名！ ➔ 查看命名词典

* 英文缩写

* 英文名称

* 中文名称

* 负责人

备注 0/2048

取消
用命名词典填充
确定


点击“确定”后，在新的弹窗点击“不去注册，并确认”。



生成“未注册到命名词典”的分层。



方式 2：先注册命名词典，再建数仓分层，规范命名。

进入命名词典：点击左上方的，选择全部产品 > 数据建模 > 数据标准 > 命名词典，进入命名词典页面。

选择应用范围是数仓分层，并填写命名词典参数。

新增命名词典 ✕

* 应用范围 ✕ ▼

* 英文缩写 ✕ ?

* 英文名称 ?

* 中文名称 ✕

* 分类

备注 0/2048

在数仓分层页面，点击 [新建数仓分层](#)，弹出“新建数仓分层”弹窗，输入英文缩写、英文名称、中文名称任一参数，点击“用命名词典填充”。

新建数仓分层 ×

! 建议先注册命名词典，再建数仓分层，规范命名! ↻ 查看命名词典

* 英文缩写

* 英文名称

* 中文名称

* 负责人

备注 0/2048

根据推荐命名，选择其中符合的命名。

温馨提示 ×

i 如下是在命名词典找到的匹配记录，默认选择第一条，您可切换选择！建议【用已选词典填充，并确认】，规范命名。

英文缩写	英文名称	中文名称	应用范围	备注	操作
test	test	测试层	数仓分层		<input checked="" type="radio"/> 用此行填充
stg	stage	暂存	数仓分层		<input type="radio"/> 用此行填充

取消 用已选词典填充，并确认

点击“用已选词典填充，并确认”，生成符合标准的分层。

★ **测试层 test** ✎ 🗑

负责人 yanshi(yanshi_luo管理员) 符合标准

修改时间 2023-08-31 20:17:36

描述

后续步骤

- 数仓分层创建完成后，可创建表名规范规则，用于规范数仓分层中表的命名规范，详见“规范管理”。
- 数仓分层创建完成后，可创建不同分层的表模型，详见“可视化建模”。

4.3.4.3 规划配置

规划配置一般情况下根据系统默认设置即可，如需修改，可参照本章修改说明和步骤进行操作。

功能入口

1. 点击左上方的■，选择全部产品 > 数据建模 > 数仓规划；在数仓规划页面左侧抽屉栏，单击规划配置，进入规划配置页面。
2. 点击右上角“修改”按钮进行修改。

操作步骤

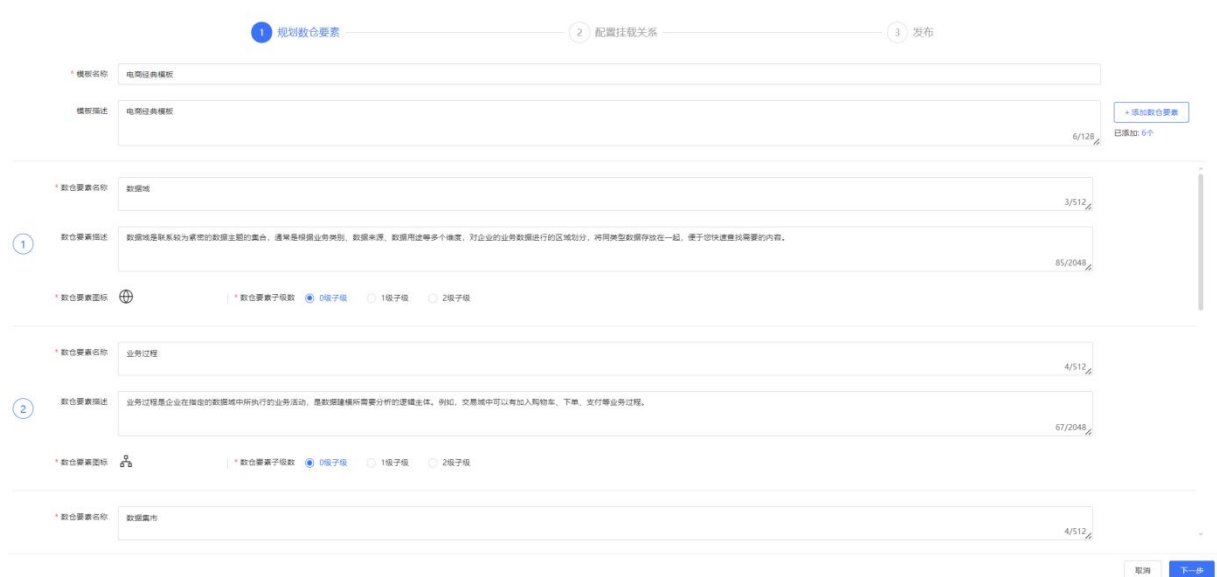
1. 规划数仓要素。

- 修改数仓要素相关信息。

说明

- 数仓子级数：“0级”子级代表没有子级；“1级子级”代表有1级的子级，包含父级共有2级；其他子级数以此类推。
- 子级数只能向大的级数修改。

- 点击“+ 添加数仓要素” 按需添加新的数仓要素。



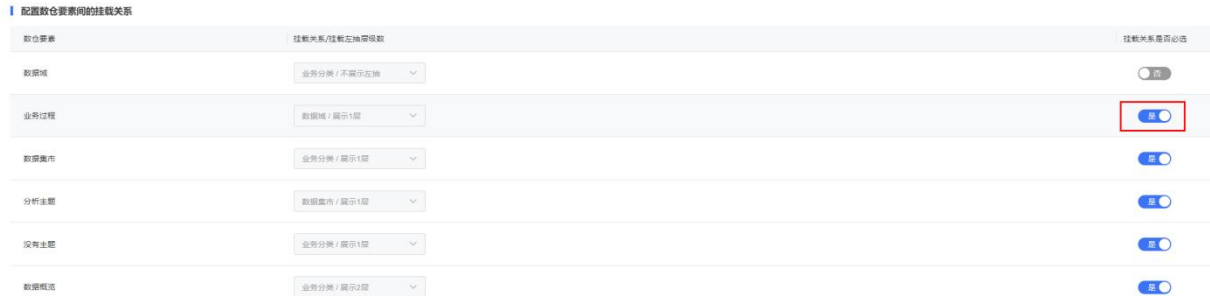
The screenshot shows the '规划数仓要素' (Plan Data Warehouse Elements) configuration page. It features a progress bar at the top with three steps: 1. 规划数仓要素 (Plan Data Warehouse Elements), 2. 配置挂载关系 (Configure Mounting Relationship), and 3. 发布 (Publish). Below the progress bar, there are three main sections, each with a form for adding or editing an element:

- Element 1:** Name: 数据域 (Data Domain), Description: 数据域是数据仓库中数据主题的集合... (Data domain is the collection of data subjects in the data warehouse...), Level: 0级子级 (0-level sub-level).
- Element 2:** Name: 业务过程 (Business Process), Description: 业务过程是企业运营中的数据主题... (Business process is the data subject in enterprise operations...), Level: 0级子级 (0-level sub-level).
- Element 3:** Name: 数据集市 (Data Mart), Description: (empty), Level: 0级子级 (0-level sub-level).

Each section includes a '添加数仓要素' (Add Data Warehouse Element) button. The page also has a '取消' (Cancel) button and a '下一步' (Next Step) button at the bottom right.

2. 配置挂载信息。

- 配置数仓要素间的挂载关系：可修改项：挂载关系是否必选；以业务过程的配置为例，如果此处选择必选，则在业务过程配置时，必须选择具体的数据域。



- 配置数据指标与数仓要素的关系。

可修改项 1：添加挂载关系；以维度为例，如果维度除了可挂载在数据域下，还想挂载在其他数仓要素下，则选择对应的数仓要素，增加挂载类型。



可修改项 2：挂载关系是否必选；以维度为例，如果此处选择必选，则在维度配置时，必须选择具体的数据域、数据集市或主题域。



- 配置数仓分层与数仓要素的关系。

可修改项 1：添加挂载关系；以汇总数据层为例，如果汇总数据层除了可挂载在数据域下，还想挂载在其他数仓要素下，则选择对应的数仓要素，增加挂载类型。



可修改项 2：挂载关系是否必选；如下截图的设置，代表所有的数仓分层的表，在维度建模时，都必须配置挂载关系。



3. 点击“提交发布”，发布成功后，该规划配置生效。

4.3.4.4 数仓要素

数仓要素：是 DataWings 用于不同类目类型的抽象叫法。如本章节以 DataWings 提供的系统模板为例：含数据域、业务过程、数据集市、分析主题等数仓要素。

- **数据域**是联系较为紧密的数据主题的集合，通常是根业务类别、数据来源、数据用途等多个维度，对企业的业务数据进行的区域划分，将同类型数据存放在一起，便于您快速查找需要的内容。
- **业务过程**是企业指定的数据域中所执行的业务活动，是数据建模所需要分析的逻辑主体。例如，交易域中可以有加入购物车、下单、支付等业务过程。
- **数据集市**是基于业务分类，面向特定应用场景或者产品的数据组织。通常位于数据应用层，依赖于公共层的整合数据。
- **分析主题**用于将数据集市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。如盘货分析、关键指标分析等。

操作以**数据域**为例，其他数仓要素类似。

创建数据域

1. 进入数据域。

点击左上方的■，选择全部产品 > 数据建模 > 数仓规划，在数仓规划左侧抽屉栏，单击数据域，进入数据域页面。

2. 创建数据域。

① 说明

- 如果“规划配置”的数仓要素（比如数据域）配置子级数为0，则单击“新建数据域”进行创建。
- 如果“规划配置”的数仓要素（比如数据域）配置子级数>0，则单击“新建一级数据域”进行创建。

如下举例的是在“规划配置”的“数据域”配置子级数>0，则单击“新建一级数据域”，在“新建一级数据域”对话框配置数据域的基本信息。

参数	描述
英文缩写	必填，按照键入前文字规则输入，最大字符数为 128。
英文名称	必填，按照键入前文字规则输入，最大字符数为 512。
中文名称	必填，按照键入前文字规则输入，最大字符数为 512。
挂载类型及具体的 挂载类型名	<p>选择具体的挂载类型后，选择具体的名称。</p> 
负责人	必填，默认为当前登录账号，可切换成其他账号，可多选。
备注	最大字符数为 2048。

方式 1：不在命名词典注册，按照上述参数要求，直接新建数据域。

新建一级数据域 ×

ⓘ 建议先注册命名词典，再建数仓元素，规范命名! [查看命名词典](#)

* 英文缩写

* 英文名称

* 中文名称

挂载类型 业务分类

* 负责人

备注 0/2048

点击“确定”，在新的弹窗点击“不去注册，并确认”。

新建一级数据域 ✕

📌 建议先注册命名词典，再建数仓元素，规范命名!
➔ 查看命名词典

* 英文缩写

* 英文名称

* 中文名称

挂载类型

* 负责人

温馨提示1 ✕

📌 该数据域scence(场景)在命名词典不存在，建议先注册命名词典，再建数据域，规范命名!

取消
去命名词典注册
不去注册，并确认

备注

0/2048

取消
用命名词典填充
确定

生成“未注册到命名词典”的数据域。

序号	英文缩写	英文名	中文名	治理标签	挂载类型	挂载类型名	负责人	创建人	创建时间	修改人
1	scence	scence	场景	未注册到命名词典	业务分类	电子商务	yanshi(yanshi_luo管 理员)	yanshi	2023-08-31 20:23:16	yanshi

方式 2：先注册命名词典，再建数据域，规范命名。

进入命名词典：点击左上方的，选择全部产品 > 数据建模 > 数据标准，进入命名词典页面。

选择应用范围是数据域，并填写命名词典参数。

新增命名词典 ×

* 应用范围

* 英文缩写 ?

* 英文名称 ?

* 中文名称

* 分类

备注 0/2048

在数据域页面，点击“新建一级数据域”，在弹窗中，输入英文缩写、英文名称、中文名称任一参数，点击“用命名词典填充”。

新建一级数据域 ✕

① 建议先注册命名词典，再建数仓元素，规范命名! [查看命名词典](#)

* 英文缩写

* 英文名称

* 中文名称

挂载类型 业务分类

* 负责人

备注 0/2048

根据推荐命名，选择其中符合的命名。

温馨提示 ×

① 如下是在命名词典找到的匹配记录，默认选择第一条，您可切换选择！建议【用已选词典填充，并确认】，规范命名。

英文缩写	英文名称	中文名称	应用范围	备注	操作
tech	technology	技术	数据域		<input checked="" type="radio"/> 用此行填充

取消
用已选词典填充，并确认

点击“用已选词典填充，并确认”，生成符合标准的数据域。

序号	英文缩写	英文名	中文名	治理标签	挂载类型	挂载类型名	负责人	创建人	创建时间	修改人	修改时间	操作
1	tech	technology	技术	符合标准			yanshi(yanshi_luo管理员)	yanshi	2023-09-01 10:13:49	yanshi	2023-09-01 10:13:49	修改 删除 新建子级

后续步骤

根据规划配置的【步骤 2：配置挂载关系】，后续建新的数仓要素、数据指标、维度建模分层时，挂载到具体的数据域。详见“数仓要素”、“数据指标”、“维度建模”。

4.3.5 数据指标

4.3.5.1 原子指标

本章节主要介绍原子指标的用途及操作步骤。

功能概述

原子指标用于明确业务的统计口径和计算逻辑，是基于用户的业务活动（即业务过程）创建的，


用于统计业务活动中某一业务状况的数值。例如，用户的业务活动为购买，则原子指标就可以指定为支付金额。

前提条件

已创建数仓要素，如业务过程，用于明确需要统计分析的具体业务活动。

新建原子指标

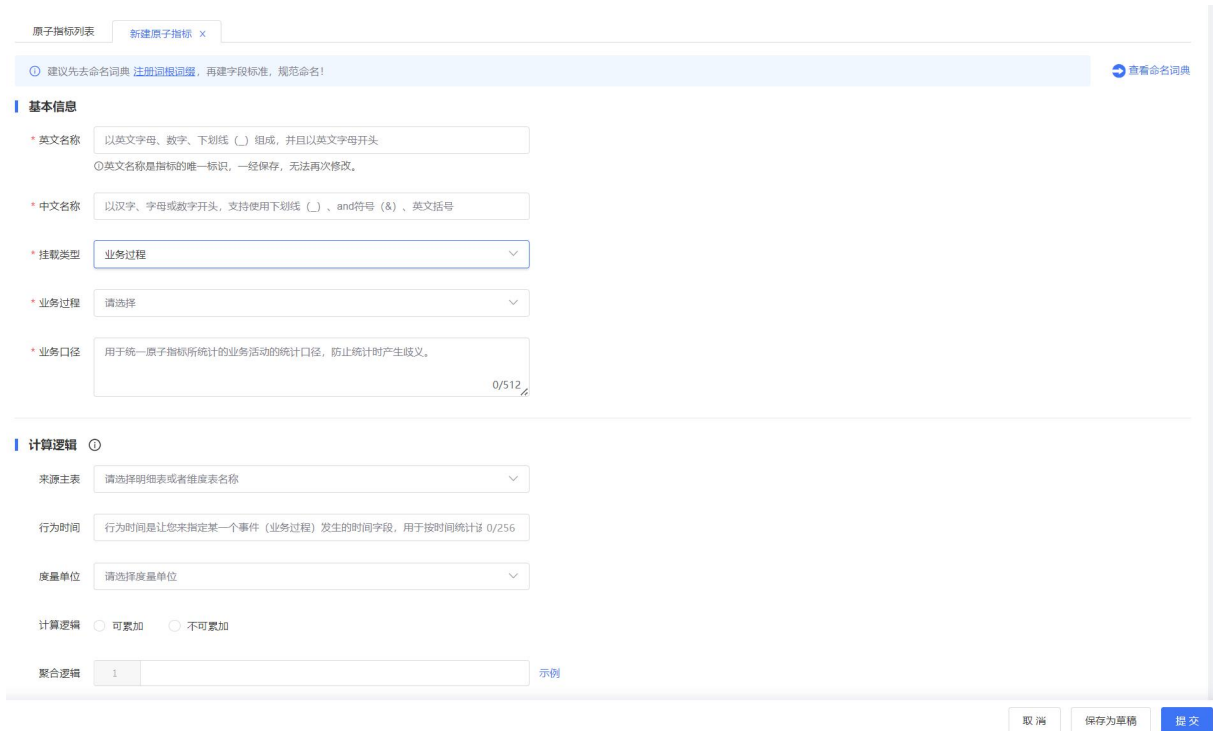
1. 进入原子指标。

点击左上方的，选择全部产品 > 数据建模 > 数据指标，在数据指标页面左侧抽屉栏，单击原子指标，进入原子指标页面。

2. 创建原子指标，单击，在“新建原子指标”页面配置基本信息。

参数	描述
基本信息	展示的分组名称，非填写参数，将原子指标的英文名称、中文名称、挂载类型及挂载类型名、业务口径归到基本信息分组。
英文名称	必填，按照键入前文字规则输入，最大字符数为 512。 此项发布成功后不可再修改。
中文名称	必填，按照键入前文字规则输入，最大字符数为 512。
挂载类型	原子指标所属的挂载类型，如业务过程。此参数“是否必选”是根据规划配置的“挂载关系是否必选”生成约束；请根据参数前是否有星号来判断即可，有星号代表必选。
挂载类型名	跟挂载类型对应，选择具体的数仓要素的名称。此参数“是否必选”是根据规划配置的“挂载关系是否必选”生成约束；请根据参数前是否有星号来判断即可，有星号代表必选。
业务口径	必填，用于统一原子指标所统计的业务活动的统计口径，防止统计时产生歧义。
计算逻辑	展示的分组名称，非填写参数，将计算逻辑相关的参数归到计算逻辑分组。
来源主表	选择来源主表。
行为时间	指定某一事件（业务过程）发生的事件字段，用于按事件统计该指标。
度量单位	选择该原子指标对应的度量单位，需先在【度量单位】定义，再在此处引用。
计算逻辑	单选：可累加，或不可累加。比如销售额是可以累加的，UV 是不可累加的。

参数	描述
聚合逻辑	该原子指标的聚合逻辑，比如 count(distinct user_id),sum(amount)。



3. 点击“保存为草稿”或“提交”；点击“保存为草稿”则可在左侧“草稿”tab，查看草稿列表；点击“提交”则生成审批单，经过审批后，系统自动发布到生产环境。

批量导入原子指标

如您有大量原子指标待创建，则可点击图标^[1]，DataWings 为您提供了导入模板，您可按照模板表格填写好待导入的原子指标后批量导入。



1. 参考上述步骤进入原子指标页面后，单击导入图标。
2. 选择导入类型并下载模板。
3. 在本地填写模板。
4. 选择文件，目前仅支持 excel 文件。

5. 点击“保存为草稿”或“导入并提交”。




点击“导入并提交”后，可进入导入单详情查看导入结果；如下是进入导入单详情的方法：

- 点击弹窗的导入单号，进入导入单详情。
- 通过全部产品 > 系统管理 > 流程中心 > 我的导入，点击导入单号，进入导入单详情。

导入成功后，系统自动生成审批单，经过审批后，系统自动发布到生产环境。

管理原子指标

- 批量导出，点击  批量导出。



导出已选：则按照用户勾选的指标导出；

导出已搜：则按搜索结果导出；

导出全部：则导出当前 tab 全部指标。

- 批量删除，选择列表上需删除的指标，点击“批量删除”。

说明

- 删除后的指标将无法再被关联引用；若当前原子指标已被模型或派生指标关联引用，则需先删除引用关系后方可成功删除指标。
- 在生产tab发起的删除需走审批流程。

批量删除

<input type="checkbox"/>	序号	英文名称	中文名称	治理标签	状态	挂载类型	挂载类型名	最后修改时间	最后修改人	操作
<input checked="" type="checkbox"/>	1	test0420	test0420	词根词缀未注册	已发布	业务过程	浏览	2024-04-20 16:23:48	pocfr	版本信息 查看 修改 删除
<input checked="" type="checkbox"/>	2	atom_test_frfr	原子指标_测试...	词根词缀未注册	已发布	业务过程	浏览	2024-04-09 10:46:18	yanshi	版本信息 查看 修改 删除
<input checked="" type="checkbox"/>	3	fr_age	fr_age	词根词缀未注册	已发布	业务过程	订单	2024-04-07 11:26:43	yanshi	版本信息 查看 修改 删除
<input checked="" type="checkbox"/>	4	BMRS	报名人数	词根词缀未注册	已发布	业务过程	学生报名	2024-04-02 11:30:39	wangyan01	版本信息 查看 修改 删除
<input type="checkbox"/>	5	user_cnt	客户数	词根词缀未注册	已发布	业务过程	订单	2024-04-02 11:19:55	yanshi	版本信息 查看 修改 删除
<input type="checkbox"/>	6	stock	库存数量	词根词缀未注册	已发布	业务过程	订单	2024-04-02 11:19:55	yanshi	版本信息 查看 修改 删除
<input type="checkbox"/>	7	order_amount	成交金额	词根词缀部分注册	已发布	业务过程	订单	2024-04-02 11:19:55	yanshi	版本信息 查看 修改 删除
<input type="checkbox"/>	8	price	单价(件)	词根词缀已注册	已发布	业务过程	订单	2024-04-02 11:19:55	yanshi	版本信息 查看 修改 删除
<input type="checkbox"/>	9	perchase_mer...	商品销售件数	词根词缀未注册	已发布	业务过程	订单	2024-04-02 11:19:55	yanshi	版本信息 查看 修改 删除
<input type="checkbox"/>	10	gmv	成交价格	词根词缀未注册	已发布	业务过程	订单	2024-04-02 11:19:55	zhengzf11	版本信息 查看 修改 删除

共11条 < 1 2 > 10 前往 1 页

- 批量提交：在开发列表、草稿列表，才露出此操作。提交是指提交发布申请，提交后走审批流程。

批量提交 **批量删除**

<input checked="" type="checkbox"/>	序号	英文名称	中文名称	状态	挂载类型	挂载类型名	最后修改时间	最后修改人	操作
<input checked="" type="checkbox"/>	1	user_cnt	user_cnt	草稿	业务过程	订单	2024-04-07 11:26:20	yanshi	查看 修改 删除 提交
<input checked="" type="checkbox"/>	2	fgr_age	库存数量	草稿	业务过程	订单	2024-04-07 11:26:20	yanshi	查看 修改 删除 提交

共2条 < 1 > 10 前往 1 页

- 版本信息：只有曾经发布成功的指标才有版本信息；点击查看版本列表，选择 2 个版本进行对比。

进入“版本列表”的方法：生产环境的列表，点击“版本信息”进入版本列表；曾经发布过的版本，在修改过程中，可从开发环境的列表，点击“版本信息”进入版本列表。

如下举例是从生产环境的列表进入，点击“版本信息”：

原子指标 🔍 🔍 🔍 🔍 🔍

生产 开发 草稿

挂载类型 业务过程

请输入内容

- 流量域(2)
- 交易域(8)
- 用户域(0)
- 招录域(1)
- 教务域(0)

原子指标列表

原子指标用于明确业务的统计口径和计算逻辑，是基于用户的业务活动（即业务过程）创建的，用于统计业务活动中某一业务状态的数值。例如，用户的业务活动为购买，则原子指标就可以指定为支付金额。您可以[新建原子指标](#)

[🔍 查看前提条件、后续步骤](#)

[批量删除](#) [查询](#) [重置](#)

<input type="checkbox"/>	序号	英文名称	中文名称	治理标签	状态	挂载类型	挂载类型名	最后修改时间	最后修改人	操作
<input type="checkbox"/>	1	test0420	test0420	词根词缀未注册	已发布	业务过程	浏览	2024-04-20 16:23:48	pocfr	版本信息 查看 修改 删除
<input type="checkbox"/>	2	atom_test_frfr	原子指标_测试...	词根词缀未注册	已发布	业务过程	浏览	2024-04-09 10:46:18	yanshi	版本信息 查看 修改 删除
<input type="checkbox"/>	3	fr_age	fr_age	词根词缀未注册	已发布	业务过程	订单	2024-04-07 11:26:43	yanshi	版本信息 查看 修改 删除
<input type="checkbox"/>	4	BMRS	报名人数	词根词缀未注册	已发布	业务过程	学生报名	2024-04-02 11:30:39	wangyan01	版本信息 查看 修改 删除
<input type="checkbox"/>	5	user_cnt	客户数	词根词缀未注册	已发布	业务过程	订单	2024-04-02 11:19:55	yanshi	版本信息 查看 修改 删除
<input type="checkbox"/>	6	stock	库存数量	词根词缀未注册	已发布	业务过程	订单	2024-04-02 11:19:55	yanshi	版本信息 查看 修改 删除
<input type="checkbox"/>	7	order_amount	成交金额	词根词缀部分注册	已发布	业务过程	订单	2024-04-02 11:19:55	yanshi	版本信息 查看 修改 删除
<input type="checkbox"/>	8	price	单价(件)	词根词缀已注册	已发布	业务过程	订单	2024-04-02 11:19:55	yanshi	版本信息 查看 修改 删除
<input type="checkbox"/>	9	perchase_mer...	商品销售件数	词根词缀未注册	已发布	业务过程	订单	2024-04-02 11:19:55	yanshi	版本信息 查看 修改 删除
<input type="checkbox"/>	10	gmv	成交价格	词根词缀未注册	已发布	业务过程	订单	2024-04-02 11:19:55	zhengzf11	版本信息 查看 修改 删除

共11条 < 1 2 > 10 前往 1 页

进入版本列表：

原子指标列表

[← 返回](#) | **版本列表** [🔍 版本对比](#)

<input type="checkbox"/>	版本	发布时间	提交人	审批人	操作
<input type="checkbox"/>	当前(v.10)	2024-01-24 10:58:03	zhengzf11	zhengzf11	查看
<input type="checkbox"/>	v.9	2023-09-26 15:54:44	zhengzf11	zhengzf11	查看
<input type="checkbox"/>	v.8	2023-08-07 16:30:15	yanshi	yanshi	查看
<input type="checkbox"/>	v.7	2023-07-14 11:02:39	yanshi	yanshi	查看
<input type="checkbox"/>	v.6	2023-06-02 16:12:21	yanshi	yanshi	查看
<input type="checkbox"/>	v.5	2023-06-02 16:11:46	yanshi	yanshi	查看
<input type="checkbox"/>	v.4	2023-05-30 15:04:31	yanshi	yanshi	查看
<input type="checkbox"/>	v.2	2023-05-17 18:09:58	yanshi	yanshi	查看
<input type="checkbox"/>	v.1	2023-05-17 18:06:22	yanshi	yanshi	查看

选择两个版本进行对比，点击“版本对比”控件进行对比。

原子指标列表

< 返回 | 版本列表 版本对比

版本	发布时间	提交人	审批人	操作
<input checked="" type="checkbox"/> 当前(v.10)	2024-01-24 10:58:03	zhengzf11	zhengzf11	查看
<input checked="" type="checkbox"/> v.9	2023-09-26 15:54:44	zhengzf11	zhengzf11	查看
<input type="checkbox"/> v.8	2023-08-07 16:30:15	yanshi	yanshi	查看
<input type="checkbox"/> v.7	2023-07-14 11:02:39	yanshi	yanshi	查看
<input type="checkbox"/> v.6	2023-06-02 16:12:21	yanshi	yanshi	查看
<input type="checkbox"/> v.5	2023-06-02 16:11:46	yanshi	yanshi	查看
<input type="checkbox"/> v.4	2023-05-30 15:04:31	yanshi	yanshi	查看
<input type="checkbox"/> v.2	2023-05-17 18:09:58	yanshi	yanshi	查看
<input type="checkbox"/> v.1	2023-05-17 18:06:22	yanshi	yanshi	查看

如下例子：中文名、业务口径有修改，用蓝色虚线圈出。

原子指标列表

< 返回 | 版本对比 新增 修改 删除

v.9	VS	v.10
gmv	*英文名称	gmv
成交价格1	*中文名称	成交价格
业务过程	*挂载类型	业务过程
订单	*业务过程	订单
商品最终成交价格0714	*业务口径	商品最终成交价格
--	来源主表	--
--	行为时间	--
--	度量单位	--
<input type="radio"/> 可累加 <input checked="" type="radio"/> 不可累加	计算逻辑	<input type="radio"/> 可累加 <input checked="" type="radio"/> 不可累加
--	聚合逻辑	--

后续步骤

- 审批人员在全部产品 > 系统管理 > 流程中心 > 我的审批中进行审批，详见我的审批。
- 审批通过后，系统自动执行，执行成功后生成原子指标新的版本，可在原子指标“生产”tab，查看到该指标的最新版本。
- 生产版的原子指标才可被事实表和派生指标引用。详见新建事实表、派生指标。

4.3.5.2 业务限定

本章节主要介绍业务限定的用途及操作步骤。

功能概述


业务限定用于限定统计数据的数据范围。

前提条件

已创建数仓要素，如数据域，用于确定业务限定所属的数据范围。

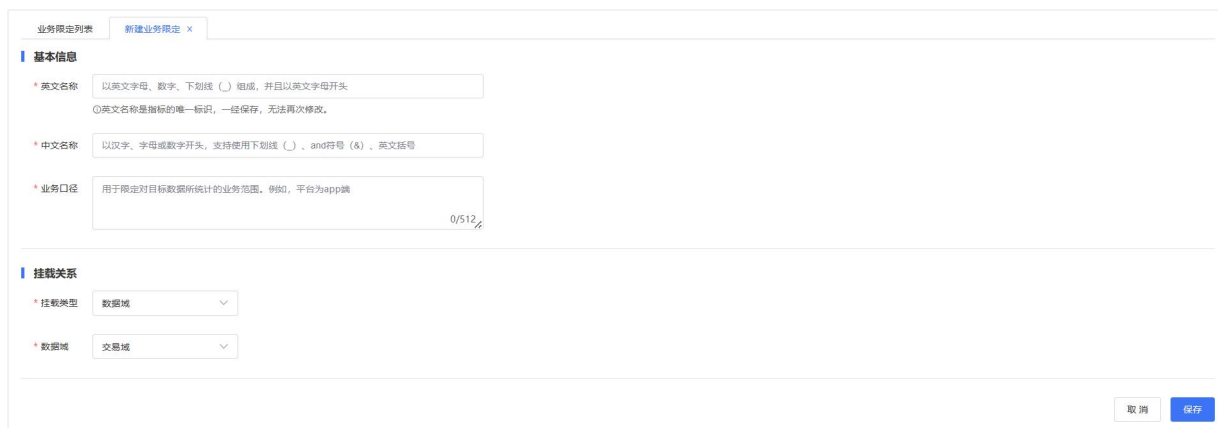
新建业务限定

1. 进入业务限定。

点击左上方的，选择全部产品 > 数据建模 > 数据指标，在数据指标页面左侧抽屉栏，单击业务限定，进入业务限定页面。


2. 创建业务限定，单击，在“新建业务限定”对话框配置基本信息。

参数	描述
英文名称	必填，按照键入前文字规则输入，最大字符数为 512。
中文名称	必填，按照键入前文字规则输入，最大字符数为 512。
业务口径	必填，用于限定对目标数据所统计的业务范围，最大字符数为 512。
挂载类型	业务限定所属的挂载类型，如数据域。此参数“是否必选”是根据规划配置的“挂载关系是否必选”生成约束；请根据参数前是否有星号来判断即可，有星号代表必选。
挂载类型名	跟挂载类型对应，选择具体的数仓要素的名称。此参数“是否必选”是根据规划配置的“挂载关系是否必选”生成约束；请根据参数前是否有星号来判断即可，有星号代表必选。



3. 点击“保存”，完成新建。

批量导入业务限定

如您有大量业务限定待创建，则点击 ，DataWings 为您提供了导入模板，您可按照模板表格填写好待导入的业务限定后批量导入。

1. 参考上述步骤进入业务限定页面后，点击
2. 选择导入类型并下载模板。
3. 在本地填写模板。
4. 选择文件，目前仅支持 excel 文件。
5. 点击“确定”。



查看导入结果：

- 点击弹窗的导入单号，进入导入单详情。
- 通过全部产品 > 系统管理 > 流程中心 > 我的导入，点击导入单号，进入导入单详情。
- 点击“完成”，刷新当前页面查看结果。

管理业务限定

- 查看：点击“查看”可查看业务限定详情。
- 修改：修改时，英文名称不可修改。
- 删除：进行二次确认，则可删除。
- 导出业务限定：有 3 种导出模式。可按需选择。



导出全部：导出全部记录；

导出已搜：按搜索结果导出；

按选中导出：按照用户勾选的记录导出。

后续步骤

创建的业务限定与原子指标、统计周期、维度，可以被派生指标引用，用于统计某一业务活动统计目标指标在具体时间、维度、业务条件下的数值表现。如果业务限定已经被派生指标引用，则需要删除所有关联的派生指标后，才可以删除业务限定。详见派生指标。

4.3.5.3 统计周期

本章节主要介绍统计周期的用途及操作步骤。

功能概述


统计周期即统计数据的时间范围，例如近 7 天、近 30 天等。用于在统计派生指标时，限定业务统计的时间范围。

新建统计周期

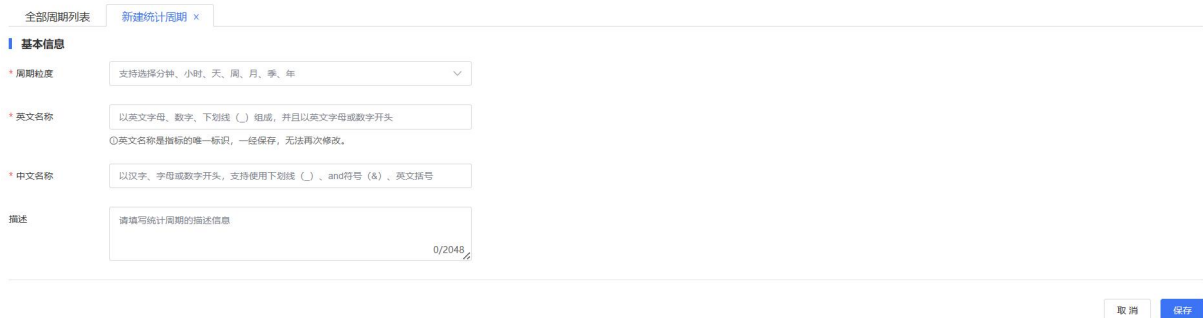
1. 进入统计周期。

点击左上方的■，选择全部产品 > 数据建模 > 数据指标，在数据指标页面左侧抽屉栏，单击

统计周期，进入统计周期页面。


2. 创建统计周期，单击 ，在“新建统计周期”对话框配置基本信息。








参数	描述
周期粒度	必选，单选，支持选择分钟、小时、天、周、月、季、年。
英文名称	必填，按照键入前文字规则输入，最大字符数为 512。 此项保存后不可修改。
中文名称	必填，按照键入前文字规则输入，最大字符数为 512。
描述	非必填，统计周期的描述信息，最大字符数为 2048。



3. 点击“保存”，完成新建。

批量导入统计周期

如您有大量统计周期待创建，则单击 ，DataWings 为您提供了导入模板，您可按照模板表格填写好待导入的统计周期后批量导入。

1. 参考上述步骤进入统计周期页面后，单击       。

2. 选择导入类型并下载模板。

3. 在本地填写模板。
4. 选择文件，目前仅支持 excel 文件。
5. 点击“确定”。



查看导入结果：

- 点击弹窗的导入单号，进入导入单详情。
- 通过全部产品 > 系统管理 > 流程中心 > 我的导入，点击导入单号，进入导入单详情。
- 点击“完成”，刷新当前页面查看结果。

管理统计周期

- 查看：点击“查看”可查看统计周期详情。
- 修改：修改时，英文名称不可修改。
- 删除：进行二次确认，则可删除。
- 导出统计周期：有 3 种导出模式。可按需选择。



导出全部：导出全部记录；

导出已搜：按搜索结果导出；

按选中导出：按照用户勾选的记录导出。

后续步骤

- 创建的统计周期与原子指标、业务限定、维度，可被派生指标引用，用于统计某一业务活动统计目标指标在具体时间、维度、业务条件下的数值表现。如果统计周期已经被派生指标引用，则需要删除所有关联的派生指标后，才可以删除统计周期。详见派生指标。
- 创建的统计周期，可被维度建模的汇总表、应用表引用。用于说明汇总表和应用表是基于哪个统计周期进行统计。详见新建汇总表、新建应用表。

4.3.5.4 维度


本章节主要介绍维度的用途及操作步骤。

功能概述

维度是人们观察事物的角度，进行统计的对象。例如客户、货品、场景等具体的业务对象；地域、组织关系和产品类目等抽象的业务对象。维度属性：该维度下的属性字段。

新建维度

1. 进入维度。

点击左上方的, 选择全部产品 > 数据建模 > 数据指标, 在数据指标页面左侧抽屉栏, 单击维度, 进入维度页面。

2. 创建维度, 单击, 在“新建维度”对话框配置维度及其属性的基本信息。其中, 在对话框的维度属性栏, 点击“添加”按钮配置维度属性的基本信息。

参数	描述
维度英文名	必填，按照键入前文字规则输入，最大字符数为 512。此项保存后不可修改。
维度中文名	必填，按照键入前文字规则输入，最大字符数为 512。
属性英文名	必填，以英文字母、数字、下划线(_)组成，并以英文字母开头，最大字符数为 512。
属性中文名	必填，以汉字、字母或数字开头，支持使用下划线(_)、and 符号(&)、英文括号，最大字符数为 512。
属性是否主键	该维度下的属性字段是否设置为主键，以便后续在维度建模时规范主键设计。
挂载类型	维度所属的挂载类型，如数据域。此参数“是否必选”是根据规划配置的“挂载关系是否必选”生成约束；请根据参数前是否有星号来判断即可，有星号代表必选。
挂载类型名	跟挂载类型对应，选择具体的数仓要素的名称。此参数“是否必选”是根据规划配置的“挂载关系是否必选”生成约束；请根据参数前是否有星号来判断即可，有星号代表必选。

维度列表
新建维度 x

① 建议先去命名词典 [注册词根词组](#)，再建字段标准，规范命名！

[查看命名词典](#)

基本信息

* 维度英文名 ① 英文名称是指标的唯一标识，一经保存，无法再次修改。

* 维度中文名

*属性英文名	*属性中文名	主键	操作
<input type="text"/>	<input type="text"/>	<input type="checkbox"/>	删除

添加
清空

挂载关系


* 挂载类型




* 数据域

取消
保存

3. 点击“保存”，完成新建。

批量导入维度

如您有大量维度待创建，则可点击 ，DataWings 为您提供了导入模板，您可按照模板表格填写好待导入的维度后批量导入。

1. 参考上述步骤进入统计周期页面后，点击 维度   。
2. 下载模板。
3. 在本地填写模板。
4. 选择文件，目前仅支持 excel 文件。
5. 点击“确定”。



查看导入结果：

- 点击弹窗的导入单号，进入导入单详情。
- 通过全部产品 > 系统管理 > 流程中心 > 我的导入，点击导入单号，进入导入单详情。
- 点击“完成”，刷新当前页面查看结果。

管理维度

- 查看：点击“查看”可查看维度详情。
- 修改：修改时，维度及属性英文名称不可修改。
- 删除：进行二次确认，则可删除。
- 导出统计周期：有 3 种导出模式。可按需选择。



导出全部：导出全部记录；

导出已搜：按搜索结果导出；

按选中导出：按照用户勾选的记录导出。

后续步骤

- 创建的维度与原子指标、业务限定、统计周期，可被派生指标引用，用于统计某一业务活动统计目标指标在具体时间、维度、业务条件下的数值表现。如果维度已经被派生指标引用，则需要删除所有关联的派生指标后，才可以删除维度。详见派生指标。
- 创建的维度，可被维度建模的维度表引用。用于说明维度表基于哪个维度（统计粒度）进行统计。详见新建维度表。

4.3.5.5 派生指标

本章节主要介绍如何创建与管理派生指标。

功能概述

派生指标是由原子指标、业务限定、统计周期、维度属性(统计粒度)组成，用于反映某个粒度某一业务活动在指定时间周期及目标范围中的业务状况。例如，品牌 ID 粒度近 1 周华南地域的销


售金额。其中业务限定、维度属性为非必选项。

前提条件

准备项	说明
原子指标	用于确定业务活动数值统计的业务口径及计算逻辑。
业务限定	用于限定统计数据业务范围，非必选项。
统计周期	用于限定统计数据的时间范围。
维度属性	用于确定进行统计的对象，非必选项。
挂载类型	派生指标所属的挂载类型，如数据集市。此参数根据“规划配置”挂载关系生成。
挂载类型名	跟挂载类型对应，选择具体的数仓要素的名称。

新建派生指标

1. 进入派生指标。

点击左上方的，选择全部产品 > 数据建模 > 数据指标，在数据指标页面左侧抽屉栏，单击派生指标，进入派生指标页面。

2. 创建派生指标，单击，在“新建派生指标”对话框配置基本信息。

参数	描述
基本信息	展示的分组名称，非填写参数，将派生指标的组成元素归到基本信息分组。
原子指标	必选，单选，用于确定业务活动数值统计的业务口径及计算逻辑。
统计周期	必选，单选，用于限定统计数据的时间范围。
业务限定	非必选，可多选，用于限定统计数据的业务范围。最多可选 3 个业务限定。
维度属性	非必选，可多选，跟挂载类型对应，用于确定进行统计的对象。
派生指标	展示的分组名称，非填写参数，将派生指标的挂载信息及中英文名、描述归到派生指标分组。
挂载类型	派生指标所属的挂载类型，如数据集市。此参数“是否必选”根据规划配置的“挂载关系是否必选”生成约束；请根据参数前是否有星号来判断即可，有星号代表必选。
挂载类型名	跟挂载类型对应，选择具体的数仓要素的名称。此参数“是否必选”是根据规划配置的“挂载关系是否必选”生成约束；请根据参数前是否有星号来判断即可，有星号代表必选。
英文名称	派生指标的英文名称，系统根据规范管理的“指标规范”自动拼接。
中文名称	派生指标的中文名称，系统根据规范管理的“指标规范”自动拼接。
描述	派生指标的描述，最大字符数为 2048。

派生指标列表 新建派生指标 ×

基本信息

* 原子指标

* 统计周期

业务限定

维度属性

派生指标

* 挂载类型

* 英文名称 ⓘ

* 中文名称 ⓘ






描述 0/2048

取消 保存为草稿 提交

3. 点击“保存为草稿”或“提交”；点击“保存为草稿”则可在左侧“草稿”tab，查看草稿列表；点击“提交”则生成审批单，经过审批后，系统自动发布到生产环境。

批量导入派生指标

如您有大量派生指标待创建，则可点击图标^[4]，DataWings 为您提供了导入模板，您可按照模板表格填写好待导入的派生指标后批量导入。

1. 参考上述步骤进入派生指标页面后，单击导入图标  派生指标    。
2. 下载模板。
3. 在本地填写模板。

4. 选择文件，目前仅支持 excel 文件。
5. 点击“保存为草稿”或“导入并提交”。



点击“导入并提交”后，可进入导入单详情查看导入结果；如下是进入导入单详情的方法：

- 点击弹窗的导入单号，进入导入单详情。
- 通过全部产品 > 系统管理 > 流程中心 > 我的导入，点击导入单号，进入导入单详情。

导入成功后，系统自动生成审批单，经过审批后，系统自动发布到生产环境。

管理派生指标

- 批量导出，点击  批量导出。



导出已选：则按照用户勾选的指标导出；

导出已搜：则按搜索结果导出；

导出全部：则导出当前 tab 全部指标。

- 批量删除，选择列表上需删除的指标，点击“批量删除”。

说明

- 删除后的指标将无法再被关联引用；若当前派生指标已被复合指标或模型关联引用，则需先删除引用关系后才可成功删除指标。
- 在生产tab发起的删除需走审批流程。

批量删除

<input type="checkbox"/>	序号	派生指标英文名	派生指标中文名	挂载类型	挂载类型名	状态	最后修改人	最后修改时间	操作
<input type="checkbox"/>	1	fr_age	fr_age						
<input type="checkbox"/>	2	BMRS	报名人数						
<input type="checkbox"/>	3	stock	库存数量						
<input checked="" type="checkbox"/>	4	buy_count	购买商品数量						
<input checked="" type="checkbox"/>	4.1	buy_count_7d	近7天_购买商品数量1	数据域	交易域	已发布	yanshi	2024-04-01 09:31:02	版本信息 查看 修改 删除
<input checked="" type="checkbox"/>	4.2	buy_count_bra...	下单时间_分钟内_品牌内_购买商品数量	数据集市	用户运营	已发布	yanshi	2024-04-07 15:09:49	版本信息 查看 修改 删除
<input type="checkbox"/>	5	order_amount	成交金额						
<input type="checkbox"/>	5.1	order_amount_...	近7天_成交金额	数据域	交易域	已发布	yanshi	2024-04-01 09:31:02	版本信息 查看 修改 删除
<input type="checkbox"/>	5.2	order_amount_...	属性四_品牌id_品牌名称_近7天_品牌...	分析主题	商品分析	已发布	yanshi	2024-04-01 09:31:02	版本信息 查看 修改 删除
<input type="checkbox"/>	6	price	单价(件)						
<input type="checkbox"/>	6.1	price_brand_ap...	属性四_1天_品牌内_app端_单价(件)	数据集市	商品运营	已发布	yanshi	2024-04-01 09:31:02	版本信息 查看 修改 删除
<input type="checkbox"/>	7	gmv	成交价格						

共7条 < 1 > 10 前往 1 页

- 批量提交：在开发列表、草稿列表，才露出此操作。提交是指提交发布申请，提交后走审批流程。

批量提交 **批量删除**

<input type="checkbox"/>	序号	派生指标英文名	派生指标中文名	挂载类型	挂载类型名	状态	最后修改人	最后修改时间	操作
<input type="checkbox"/>	1	user_cnt	客户数						
<input checked="" type="checkbox"/>	1.1	user_cnt_7d	近7天_客户数	数据集市	用户运营	草稿	yanshi	2024-05-07 18:12:29	查看 修改 删除 提交
<input checked="" type="checkbox"/>	1.2	app_user_cnt_7d	近7天_app端_客户数	数据集市	用户运营	草稿	yanshi	2024-05-07 18:13:02	查看 修改 删除 提交

共1条 < 1 > 10 前往 1 页

- 版本信息：点击查看版本列表、选择 2 个版本进行版本对比。只有曾经发布成功的指标才

有版本信息。



派生指标列表

派生指标是由原子指标、业务限定、统计周期、维度(统计粒度)组成,用于反映某个粒度某一业务活动在指定时间周期及目标范围内的业务状况。例如,品牌ID粒度一周华南地域的销售金额。您可以按 [新建派生指标](#)。

[查看前提条件、后续步骤](#)

序号	派生指标英文名	派生指标中文名	挂载类型	挂载类型名	状态	最后修改人	最后修改时间	操作
1	user_cnt	客户数						
1.1	user_cnt_7d	近7天_客户数	数据集市	用户运营	已发布	yanshi	2024-05-07 18:16:26	版本信息 查看 修改 删除
2	fr_age	fr_age						
2.1	fr_age_7d	近7天_fr_age	数据集市	商品运营	已发布	pocfr	2024-04-20 16:42:35	版本信息 查看 修改 删除
2.2	dt_age_app_bran...	近7天_天分区_年...	分析主题	品牌分析	已发布	pocfr	2024-04-20 16:43:41	版本信息 查看 修改 删除
3	buy_count	购买商品数量						
3.1	buy_count_brand...	下单时间_分钟_品...	数据集市	用户运营	已发布	yanshi	2024-04-07 15:09:49	版本信息 查看 修改 删除
4	order_amount	成交金额						
4.1	order_amount_br...	属性四_品牌id_品...	分析主题	商品分析	已发布	yanshi	2024-04-01 09:31:02	版本信息 查看 修改 删除
5	price	单价(件)						
6	gmv	成交价格						

共6条 < 1 > 10 前往 1 页

从生产环境的列表进入, 点击“版本信息”进入版本列表; 如果曾经发布过的版本, 在修改过程中, 也可以开发环境的列表进入, 点击“版本信息”进入版本列表, 选择 1 个版本跟当前修改过程的版本进行对比。

如下举例是从生产环境的列表进入, 点击“版本信息”进入版本列表。



← 返回 | 版本列表 [版本对比](#)

版本	发布时间	提交人	审批人	操作
当前(v.2)	2024-05-07 18:16:26	yanshi	yanshi	查看
v.1	2024-05-07 18:15:14	yanshi	yanshi	查看

选择两个版本进行对比, 点击“版本对比”控件。



← 返回 | 版本列表 [版本对比](#)

版本	发布时间	提交人	审批人	操作
<input checked="" type="checkbox"/> 当前(v.2)	2024-05-07 18:16:26	yanshi	yanshi	查看
<input checked="" type="checkbox"/> v.1	2024-05-07 18:15:14	yanshi	yanshi	查看

如下例子: 描述有修改, 用蓝色虚线圈出。

v.1	VS	v.2
user_cnt	*原子指标	user_cnt
7d	*统计周期	7d
--	业务限定	--
--	维度属性	--
数据集市	*挂载类型	数据集市
用户运营	*数据集市	用户运营
user_cnt_7d	*英文名称	user_cnt_7d
近7天_客户数	*中文名称	近7天_客户数
--	描述	时间周期为近7天，下单的客户数

后续步骤

- 审批人员在审批进行审批，审批通过后，系统自动执行，执行成功则生成派生指标新的版本，可在派生指标“生产”tab，查看到该指标的最新版本。
- 生产版的派生指标才可被复合指标或模型引用。详见新建汇总表、新建应用表。

4.3.5.6 复合指标

本章节主要介绍如何创建与管理复合指标。

功能概述


复合指标是通过指标的加减乘除等运算构建而成的指标，比如比例、比率等等。

前提条件

准备项	说明
派生指标	用于确定复合指标需基于哪些派生指标进行计算。
挂载类型	复合指标所属的挂载类型，如数据集市。此参数根据“规划配置”挂载关系生成。
挂载类型名	跟挂载类型对应，选择具体的数仓要素的名称。

新建复合指标

1. 进入复合指标。

点击左上方的，选择全部产品 > 数据建模 > 数据指标，在数据指标页面左侧抽屉栏，单击复合指标，进入复合指标页面。

2. 创建复合指标，单击，在“新建复合指标”对话框配置基本信息。

参数	描述
基本信息	展示的分组名称，非填写参数，将复合指标的英文名称、中文名称、挂载类型、挂载类型名、描述说明归到基本信息分组。
英文名称	必填，按照键入前文字规则输入，最大字符数为 512。 此项发布成功后不可再修改。
中文名称	必填，按照键入前文字规则输入，最大字符数为 512。
挂载类型	复合指标所属的挂载类型，如数据集市。此参数“是否必选”根据规划配置的“挂载关系是否必选”生成约束；请根据参数前是否有星号来判断即可，有星号代表必选。
挂载类型名	跟挂载类型对应，选择具体的数仓要素的名称。此参数“是否必选”是根据规划配置的“挂载关系是否必选”生成约束；请根据参数前是否有星号来判断即可，有星号代表必选。
描述说明	非必填，复合指标的描述说明。
计算逻辑	展示的分组名称，非填写参数，将计算逻辑相关的参数归到计算逻辑分组。
派生指标	用于确定复合指标需基于哪些派生指标进行计算。与下方的复合指标，至少选择其中一项。
复合指标	新建的复合指标可依赖已发布的复合指标。与上方的派生指标，至少选择其中一项。

复合指标列表
新建复合指标 x

基本信息

* 英文名称 以英文字母、数字、下划线 () 组成, 并且以英文字母开头
Ⓞ 英文名称是指标的唯一标识, 一经保存, 无法再次修改。

* 中文名称 以汉字、字母或数字开头, 支持使用下划线 ()、and符号 (&)、英文括号

* 挂载类型

- 数据集市
- 分析主题


描述说明 0/512





计算逻辑 (必填)

派生指标	复合指标	表达式
<input type="text" value="请选择派生指标"/>	<input type="text" value="请选择复合指标"/>	<div style="border: 1px solid #ccc; padding: 5px;"> <p style="text-align: center;">+ - * / ()</p> <p style="font-size: 8px; text-align: center;">请点击指标与运算符创建表达式</p> </div> <div style="text-align: right; margin-top: 5px;"> <input type="button" value="清除"/> </div>

3. 点击“保存为草稿”或“提交”；点击“保存为草稿”则可在左侧“草稿”tab，查看草稿列表；点击“提交”则生成审批单，经过审批后，系统自动发布到生产环境。

批量导入复合指标

如您有大量复合指标待创建，则可点击图标 ，DataWings 为您提供了导入模板，您可按照模板表格填写好待导入的复合指标后批量导入。

1. 参考上述步骤进入复合指标页面后，单击导入图标    。
2. 下载模板。
3. 在本地填写模板。
4. 选择文件，目前仅支持 excel 文件。
5. 点击“保存为草稿”或“导入并提交”。



点击“导入并提交”后，可进入导入单详情查看导入结果；如下是进入导入单详情的方法：

- 点击弹窗的导入单号，进入导入单详情。
- 通过全部产品 > 系统管理 > 流程中心 > 我的导入，点击导入单号，进入导入单详情。

导入成功后，系统自动生成审批单，经过审批后，系统自动发布到生产环境。

管理复合指标

- 批量导出，点击  批量导出。



导出已选：则按照用户勾选的指标导出；

导出已搜：则按搜索结果导出；

导出全部：则导出当前 tab 全部指标。

- 批量删除，选择列表上需删除的指标，点击“批量删除”。

说明

- 删除后的指标将无法再被关联引用；若当前复合指标已被关联引用，则需先删除引用关系后方可成功删除指标。
- 在生产tab发起的删除需走审批流程。

复合指标列表

复合指标 ▲
复合指标通过指标的加减乘除等运算构建而成的指标，比如比例、比率等等。您可以按需 [新建复合指标](#)
[查看前提条件、后续步骤](#)

批量删除 [查询](#) [重置](#)

<input type="checkbox"/>	序号	英文名称	中文名称	状态	挂载类型	挂载类型名	最后修改时间	最后修改人	操作
<input checked="" type="checkbox"/>	1	fhzb_0420	fhzb_0420	已发布	分析主题	品牌分析	2024-04-20 17:11:05	pocfr	查看 修改 删除
<input checked="" type="checkbox"/>	2	composite_index	composite_index	已发布	分析主题	品牌分析	2024-04-07 16:53:12	yanshi	查看 修改 删除
<input type="checkbox"/>	3	ywmcfhzc	复合指标修改中文名	已发布	数据集市	低级别流量运营	2024-04-07 15:59:23	yanshi	查看 修改 删除
<input type="checkbox"/>	4	fuhezhib	复合指标2	已发布	数据集市	低级别流量运营	2024-04-03 17:06:55	yanshi	查看 修改 删除

共4条 < 1 > 10 前往 1 页

- 批量提交：在开发列表、草稿列表，才露出此操作。提交是指提交发布申请，提交后走审批流程。

复合指标列表

复合指标 ▲
复合指标通过指标的加减乘除等运算构建而成的指标，比如比例、比率等等。您可以按需 [新建复合指标](#)
[查看前提条件、后续步骤](#)

批量提交 **批量删除** [查询](#) [重置](#)

<input type="checkbox"/>	序号	英文名称	中文名称	状态	挂载类型	挂载类型名	最后修改时间	最后修改人	操作
<input checked="" type="checkbox"/>	1	zzf_fox_test_8	复合zzf8	草稿	数据集市	数据集市2	2024-05-09 16:04:55	zzf1	查看 修改 删除 提交
<input checked="" type="checkbox"/>	2	zzf_fox_test_9	复合zzf9	草稿	数据集市	数据集市2	2024-05-09 16:04:55	zzf1	查看 修改 删除 提交
<input type="checkbox"/>	3	zzf_fox_test_2	复合zzf2	草稿	数据集市	数据集市2	2024-05-09 16:04:54	zzf1	查看 修改 删除 提交
<input type="checkbox"/>	4	zzf_fox_test_3	复合zzf3	草稿	数据集市	数据集市2	2024-05-09 16:04:54	zzf1	查看 修改 删除 提交
<input type="checkbox"/>	5	zzf_fox_test_4	复合zzf4	草稿	数据集市	数据集市2	2024-05-09 16:04:54	zzf1	查看 修改 删除 提交
<input type="checkbox"/>	6	zzf_fox_test_5	复合zzf5	草稿	数据集市	数据集市2	2024-05-09 16:04:54	zzf1	查看 修改 删除 提交
<input type="checkbox"/>	7	zzf_fox_test_6	复合zzf6	草稿	数据集市	数据集市2	2024-05-09 16:04:54	zzf1	查看 修改 删除 提交
<input type="checkbox"/>	8	zzf_fox_test_1	复合zzf1	草稿	数据集市	数据集市2	2024-05-09 16:04:53	zzf1	查看 修改 删除 提交
<input type="checkbox"/>	9	test123213123123	7天_用户	草稿	数据集市	数据集市3	2024-05-09 16:02:34	zzf1	查看 修改 删除 提交

共9条 < 1 > 10 前往 1 页

- 版本信息：点击查看版本列表、选择 2 个版本进行版本对比。只有曾经发布成功的指标才有版本信息。

复合指标 复合指标列表

生产 开发 草稿

挂载类型 数据集市

请输入内容

批量删除

请输入复合指标英文名称、中文名称搜索

查询 重置

序号	英文名称	中文名称	状态	挂载类型	挂载类型名	最后修改时间	最后修改人	操作
1	average_sales	客均支付金额	已发布	数据集市	数据集市3	2024-05-20 14:39:36	zzf1	引用分析 版本信息 查看 修改 删除
2	lql_aaa_fuhe	lql_aaa_fuhe	已发布	数据集市	数据集市2	2024-05-10 16:20:23	zzf1	引用分析 版本信息 查看 修改 删除
3	zzf_fox_test_5	复合zzf5	已发布	数据集市	数据集市2	2024-05-09 17:42:45	zzf1	引用分析 版本信息 查看 修改 删除
4	zzf_fox_test_4	复合zzf4	已发布	数据集市	数据集市2	2024-05-09 16:29:18	zzf1	引用分析 版本信息 查看 修改 删除
5	fhzbauto_update	fhzbauto_update	已发布	数据集市	数据集市2	2024-05-09 08:08:42	autotestuser	引用分析 版本信息 查看 修改 删除
6	zzf05060013	复合指标13	已发布	数据集市	数据集市3	2024-05-06 17:41:26	zzf1	引用分析 版本信息 查看 修改 删除
7	gp1	gp1test	已发布	数据集市	数据集市3	2024-05-06 17:22:23	zzf1	引用分析 版本信息 查看 修改 删除
8	gp2	gp2测试	已发布	数据集市	数据集市3	2024-05-06 17:22:11	zzf1	引用分析 版本信息 查看 修改 删除
9	gp3	gp3测试	已发布	数据集市	数据集市3	2024-05-06 17:22:05	zzf1	引用分析 版本信息 查看 修改 删除
10	test11123123123...	test11123123123...	已发布	数据集市	数据集市3	2024-04-17 19:43:38	zzf1	引用分析 版本信息 查看 修改 删除

共16条 < 2 > 10 前往 1 页

从生产环境的列表进入，点击“版本信息”进入版本列表；如果曾经发布过的版本，在修改过程中，也可以开发环境的列表进入，点击“版本信息”进入版本列表，选择1个版本跟当前修改过程的版本进行对比。

如下举例是从生产环境的列表进入，点击“版本信息”进入版本列表。

复合指标列表

< 返回 | 版本列表

版本对比

版本	发布时间	提交人	审批人	操作
当前(v.2)	2024-05-20 14:39:36	zzf1	zzf1	查看
v.1	2024-05-20 14:38:43	zzf1	zzf1	查看

选择两个版本进行对比，点击“版本对比”控件。

< 返回 | 版本列表

版本对比

版本	发布时间	提交人	审批人	操作
当前(v.2)	2024-05-20 14:39:36	zzf1	zzf1	查看
v.1	2024-05-20 14:38:43	zzf1	zzf1	查看

如下例子：描述说明有修改，用蓝色虚线圈出。

← 返回 | 版本对比 ● 新增 ○ 修改 ⋯ 删除

v.1	VS	v.2
average_sales	*英文名称	average_sales
客均支付金额	*中文名称	客均支付金额
数据集市	*挂载类型	数据集市
数据集市3	*数据集市	数据集市3
--	描述说明	客均支付金额=近7天_支付金额/近7天_交易客户数
近7天_支付金额/近7天_交易客户数	依赖派生指标	近7天_支付金额,近7天_交易客户数
--	依赖复合指标	--
近7天_支付金额/近7天_交易客户数	*表达式	近7天_支付金额/近7天_交易客户数

4.3.6 维度建模

根据业务情况设计并创建接入表、维度表、明细表、汇总表、应用表等，并将模型快速发布到相应的数据源。同时还可以进行逆向建模，把存量的物理表反向生成模型，后续可以进行统一管理。

4.3.6.1 正向建模

4.3.6.1.1 可视化建模

4.3.6.1.1.1 可视化建模概述

本章节主要介绍可视化建模的概述及前提条件。

功能概述

根据业务情况，通过可视化的方式设计并创建接入表、维度表、明细表、汇总表、应用表等，并将模型快速发布到相应的数据源。

前提条件

- 集群：集群同步/注册成功，且含有 Hive 组件；
- 引擎管理：将集群与调度资源组进行关联，且引擎实例状态为成功；
- 数据源管理：Hive 内置数据源，成功自动生成；
- 数据库：Hive 内置数据源下，已创建数据库；

- 规划模板：已规划好各数仓要素，且配置好维度模型与数仓要素的关系；
- 选择模板：在首次进入数据建模时，租户管理员选定某个规划模板；
- 业务分类：已配置好数仓要素要挂载的业务分类；
- 数仓分层：已创建对应的数仓分层；
- 数仓要素：配置维度建模需要挂载的数仓要素；
- 表名规范（非必须）：已定义表名规范；
- 存储策略（非必须）：已在命名词典定义存储策略，如日增量（di）、日全量（df）；
- 维度（非必须）：已定义维度，用于表达该为表是什么维度；
- 数据标准（非必须）：已定义数据标准，在字段管理时引用标准；
- 数据指标（非必须）：已定义数据指标，在字段管理时关联指标。



4.3.6.1.1.2 新建接入表

本章节主要介绍新建接入表的步骤。

背景信息

接入表：主要是用于数据库、日志等基础源数据的引入而落地的表。

功能入口

- 单击左上方的图标，选择全部产品 > 数据建模 > 维度建模 > 正向建模，进入正向建模页面；
- 新建接入表：在正向建模页面，鼠标悬停至图标，单击“新建接入表”。

详细操作步骤

1. 表信息填写。

1 表信息填写
2 字段管理
3 设置主键/分区键
4 发布申请

* 数据源

* 表物理类型

* 表属性 内部表 外部表

* 数仓分层

* 挂载类型

* 表名主体

表名规范

存储策略

描述

* 表英文名 使用推荐表名

①若拼接项有正则表达式，「使用推荐表名」后，正则表达式对应内容需手工替换成具体值

表别名

* 表中文名

* 负责人

* 数据库名

* 数据域

取消 保存为草稿 下一步

主要参数说明如下：

填写参数	参数说明
数据源	需通过集群注册 Hive 组件，且在引擎管理中将集群与调度资源组绑定，系统自动生成内置的 Hive 数据源，如果有多个，下拉单选。
数据库名	该数据源下已创建的数据库，下拉单选。
表物理类型	表的物理类型。

填写参数	参数说明
表属性	内部表：默认为内部表；外部表：外部表需要填写外部表的 Location。
数仓分层	跟所选的分层对应。
挂载类型	表所属的挂载类型，如数据域。
挂载类型名	跟挂载类型对应，选择具体的数仓要素的名称。
表名主体	用于表达该表的作用，是表英文名的组成部分；由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。一般情况下表英文名包含：固定字符(如 ods)_数据域_表名主体_存储策略。
表名规范	通过已配置的表名规范约束表名。选择表名规范后，可点击“使用推荐表名”，系统生成推荐表名。
存储策略	接入表基于什么策略（即存储数据的时间周期及数据量范围）存储数据。
描述	表描述。

填写参数	参数说明
表英文名	接入表的英文名，由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。 只要选择了表名规范，不管是强规则还是弱规则，都可以点击“使用推荐表名”，系统生成推荐表名。如果选择的表名规范为强规则，则需遵循表名规范进行建设。
表别名	接入表的别名，由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。一般用于逆向建模，当表英文名不符合规范时，此参数可填写规范的名称，作为辅助表英文名用。
表中文名	接入表的中文名，由中文、英文字母、下划线（_）、and（&）、英文括号组成。
负责人	当前接入表的负责人，默认为创建接入表的用户，可选多个。

完成表信息填写，点击“保存为草稿”或“下一步”；点击“保存为草稿”则可在左侧“草稿”tab，查看草稿列表；点击“下一步”，则进入步骤2“字段管理”。

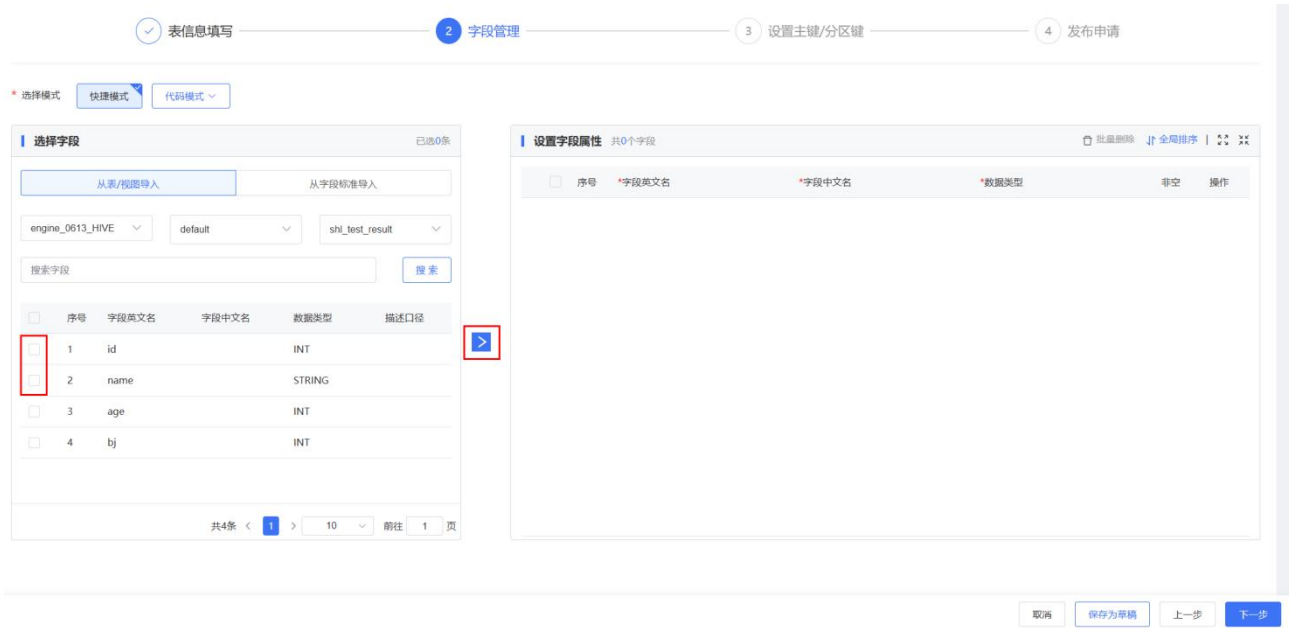
2. 字段管理。

● 选择字段。

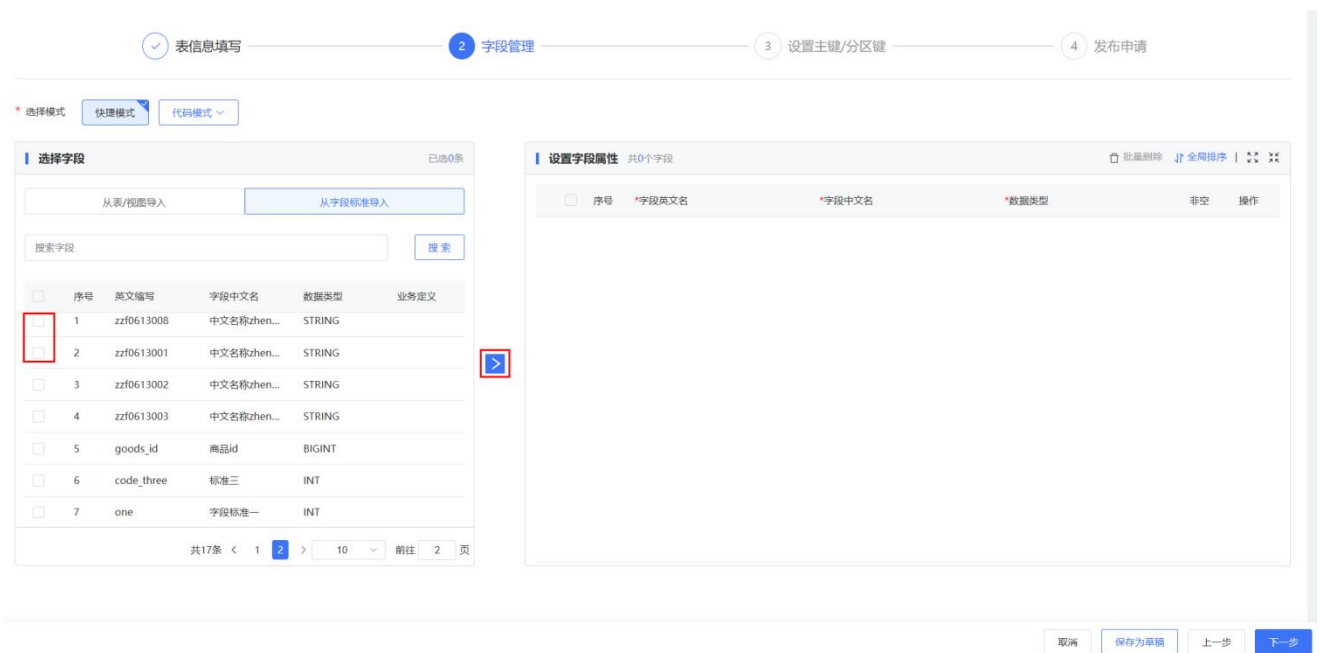
字段管理支持快捷模式与代码模式。

A. 快捷模式：支持从表/视图导入、从字段标准导入。

如下截图为从表/视图导入：



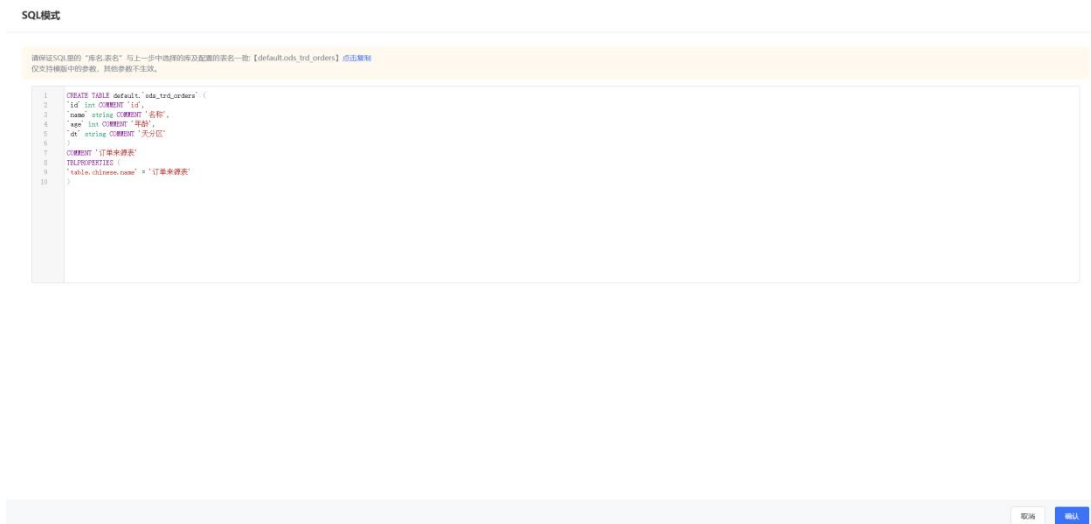
如下截图为从字段标准导入：



代码模式导入：支持解析 sql 内容导入字段。有两种方式：



其一：保留快捷模式字段，将快捷模式所选的字段自动生成 sql 带到代码模式的输入框；其二，清空快捷模式字段，不将快捷模式所选的字段带到代码模式的输入框。



点击“确认”，则进入“设置字段属性”。

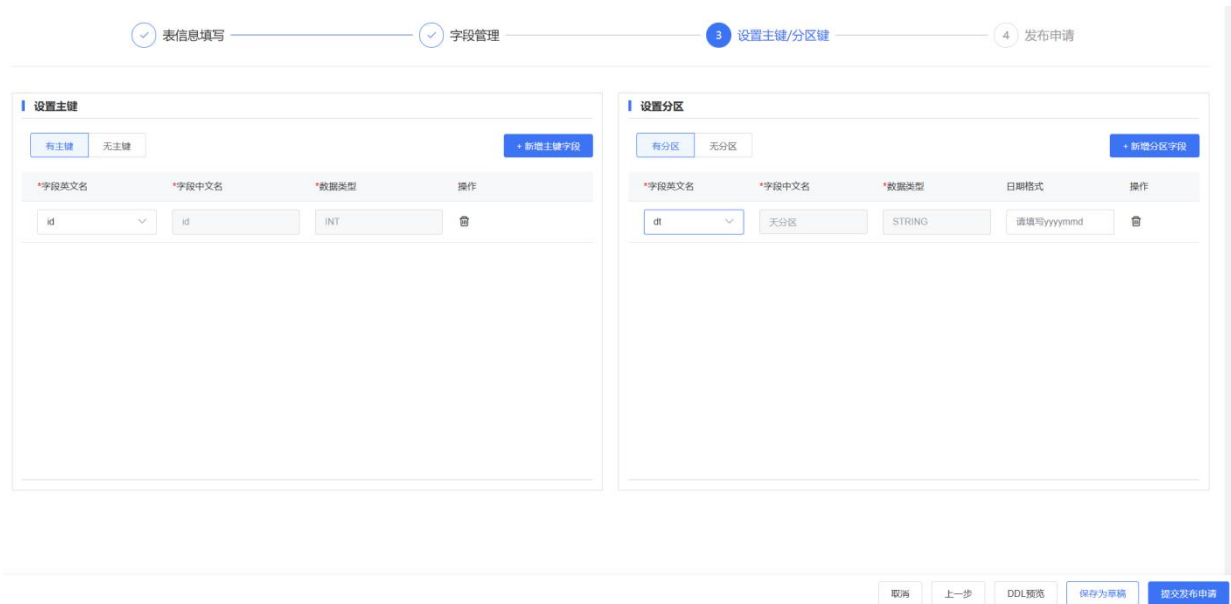
● 设置字段属性。

- 字段英文名、字段中文名、数据类型、描述口径。
- 设置非空属性、关联字段标准、标准值域。
- 关联类型选维度属性，并选择所关联的内容。

3. 设置主键和分区键。

字段配置完成后，可以根据实际场景配置 1 个或多个主键字段；针对 Hive 表，建议您配置分区

字段并确定分区字段格式。



4. 发布申请。

设置完成后，即可提交发布申请。

后续步骤

- 发布申请生成审批单，需审批人员在“我的审批”进行审批，审批通过后，系统自动发布生成物理表。系统生成后，可在建模“生产”tab，查看到该表的最新版本。
- 发布成功后，可在“任务创建”，进行相关的数据集成任务操作，将数据接入至该表。

4.3.6.1.1.3 新建维度表



本章节主要介绍新建维度表的步骤。

背景信息

维度表：结合业务的数据域规划，提取出各业务数据域中进行数据分析时可能存在的维度，并将维度及其属性通过维度表的方式存储下来。例如，在进行电商业务数据分析时，可用的维度

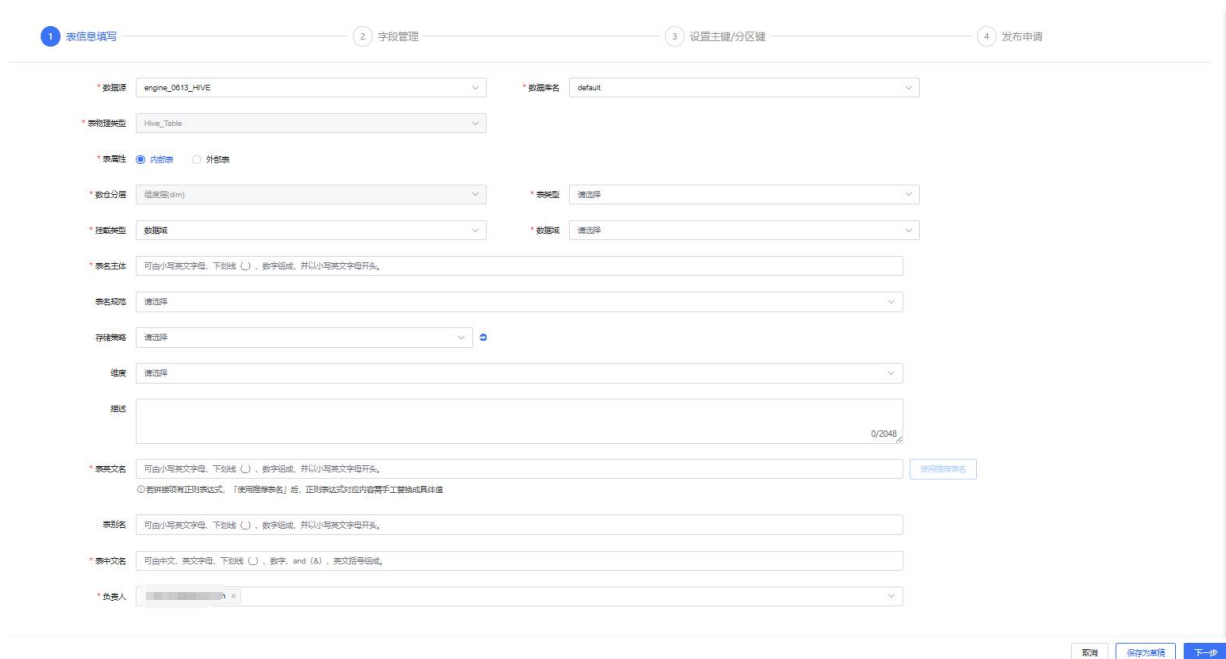
及其属性有：订单维度（属性包括订单 ID、订单创建时间、买家 ID、卖家 ID 等）、用户维度（性别、出生日期等）、商品维度（包括商品 ID、商品名称、商品上架时间）等，此时您就可以将这些维度和属性创建为订单维度表、用户维度表、商品维度表等，将维度属性记录作为维度表的字段。后续您可将这些维度表发布到数仓中，通过 ETL 将实际维度数据按照维度表定义的方式进行存储，方便业务人员在后续的数据分析进行取用。

功能入口

- 单击左上方的  图标，选择全部产品 > 数据建模 > 维度建模 > 正向建模，进入正向建模页面；
- 新建维度表：在正向建模页面，鼠标悬停至  图标，单击“新建维度表”。

详细操作步骤

1. 表信息填写。



新建维度表配置界面，包含以下配置项：

- 1 表信息填写
- 2 字段管理
- 3 设置主键/分区键
- 4 发布申请

配置项包括：

- * 数据源: engine_0013_HIVE
- * 数据库名: default
- * 物理表类型: Hive_Table
- * 表属性: 内部表 外部表
- * 数仓分区: 维度层(dim)
- * 表类型: 请选择
- * 表名: 数据源
- * 表名主体: 可由小写英文字母、下划线 (_)、数字组成，并以小写英文字母开头。
- * 表名称: 请选择
- * 存储策略: 请选择
- * 描述: 请选择
- * 表英文名: 可由小写英文字母、下划线 (_)、数字组成，并以小写英文字母开头。
① 若拼接原有正则表达式，「使用理解命名」后，正则表达式对应内容需手工编辑成具体值
- * 表别名: 可由小写英文字母、下划线 (_)、数字组成，并以小写英文字母开头。
- * 表中文名: 可由中文、英文字母、下划线 (_)、数字、and (&)、英文逗号组成。
- * 负责人: 请选择

底部操作按钮：取消、保存为草稿、下一步

主要参数说明如下:

填写参数	参数说明
数据源	需通过集群注册 Hive 组件，且在引擎管理中将集群与调度资源组绑定，系统自动生成内置的 Hive 数据源，如果有多个，下拉单选。
数据库	该数据源下已创建的数据库，下拉单选。
表物理类型	表的物理类型。
表属性	内部表：默认为内部表； 外部表：外部表需要填写外部表的 Location。
数仓分层	跟所选的分层对应。
表类型	维度表的表类型： <ul style="list-style-type: none"> ● 普通维度表：无特殊需求可选择此项； ● 枚举维度表：可进行枚举的维度表，如部门 id 与部门名称； ● 层级维度表：需要设置层级关系的维度表，支持指标做上卷和下钻。 例如国家、省份、市县等。
挂载类型	为表所属的挂载类型，如数据域。
挂载类型名	跟挂载类型对应，选择具体的数仓要素的名称。

填写参数	参数说明
表名主体	<p>用于表达该表的作用，是表英文名的组成部分；由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。</p> <p>一般情况下表英文名包含：固定字符(如 dim)_数据域_表名主体_存储策略。</p>
表名规范	通过已配置的表名规范约束表名。选择表名规范后，可点击“使用推荐表名”，系统生成推荐表名。
存储策略	维度表基于什么策略（即存储数据的时间周期及数据量范围）存储数据。
维度	维度表是基于哪个维度建设的。
描述	表描述。
表英文名	<p>维度表的英文名，由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。只要选择了表名规范，不管是强规则还是弱规则，都可以点击“使用推荐表名”，系统生成推荐表名。如果选择的表名规范为强规则，则需遵循表名规范进行建设。</p>
表别名	<p>维度表的别名，由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。一般用于逆向建模，当表英文名不符合规范时，此参数可填写规范的名称，作为辅助表英文名用。</p>
表中文名	维度表的中文名，由中文、英文字母、下划线（_）、and（&）、英文

填写参数	参数说明
	括号组成。
负责人	当前维度表的责任人，默认为创建维度表的用户。可选多个。

完成表信息填写，点击“保存为草稿”或“下一步”；点击“保存为草稿”则可在左侧“草稿”tab，查看草稿列表；点击“下一步”，则进入步骤2“字段管理”。

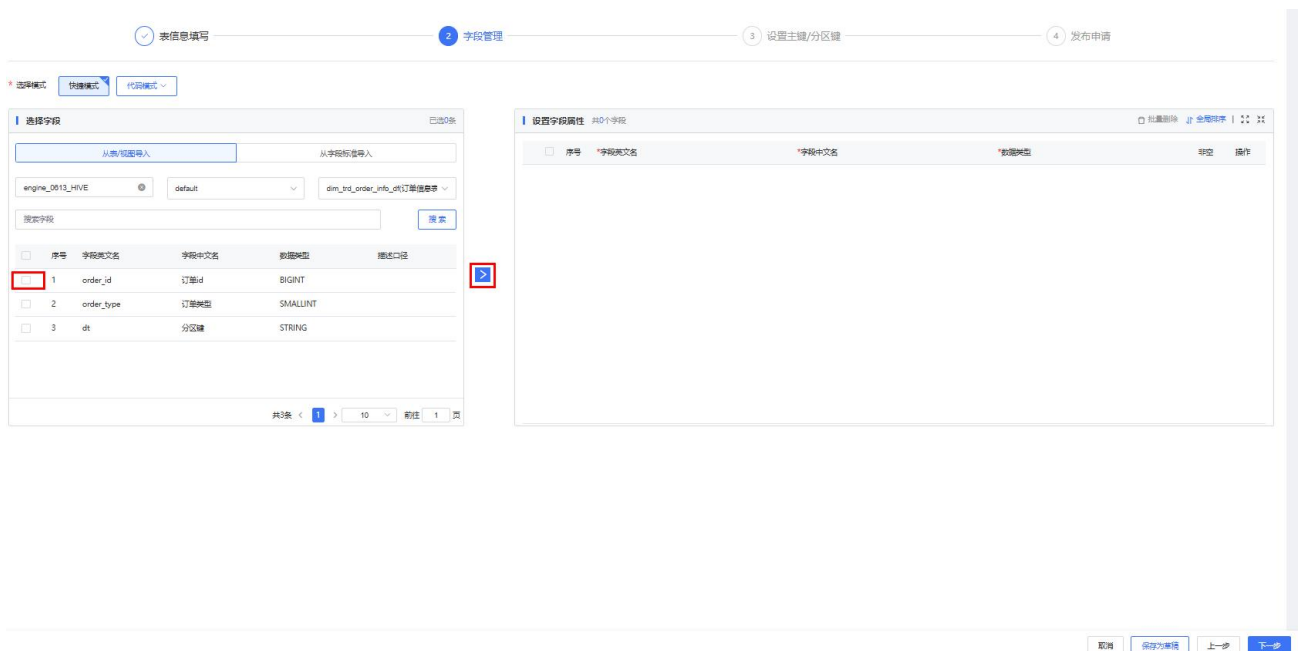
2. 字段管理。

● 选择字段。

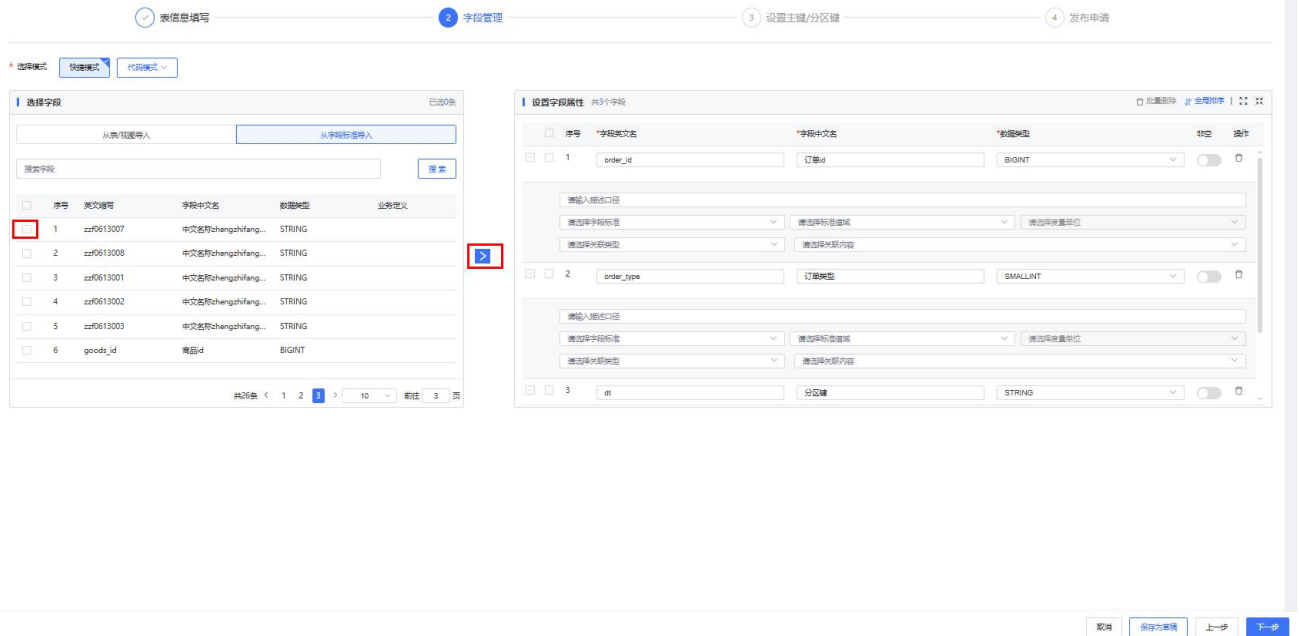
字段管理支持快捷模式与代码模式。

A. 快捷模式：支持从表/视图导入、从字段标准导入。

如下截图为从表/视图导入：



如下截图为从字段标准导入：



B. 代码模式导入：支持解析 sql 内容导入字段。有两种方式：



其一：保留快捷模式字段，将快捷模式所选的字段自动生成 sql 带到代码模式的输入框；其二，清空快捷模式字段，不将快捷模式所选的字段带到代码模式的输入框。

SQL语句

```
请通过SQL语句“新建字段”在上一级中创建新的字段及配置属性。 [ default_charset=utf8mb4_unicode_ci ] 帮助文档  
在您的脚本中设置时，其语法如下所示。  
  
1 CREATE TABLE `table` ( `field` `type` `comment` )  
2 `field` `type` `comment` ( `comment` )  
3 `field` `type` `comment` ( `comment` )  
4 `field` `type` `comment` ( `comment` )  
5 `field` `type` `comment` ( `comment` )  
6  
7 COMMENT: ( `comment` )  
8  
9 TABLESPACE: `name` ( `comment` )  
10  
11
```

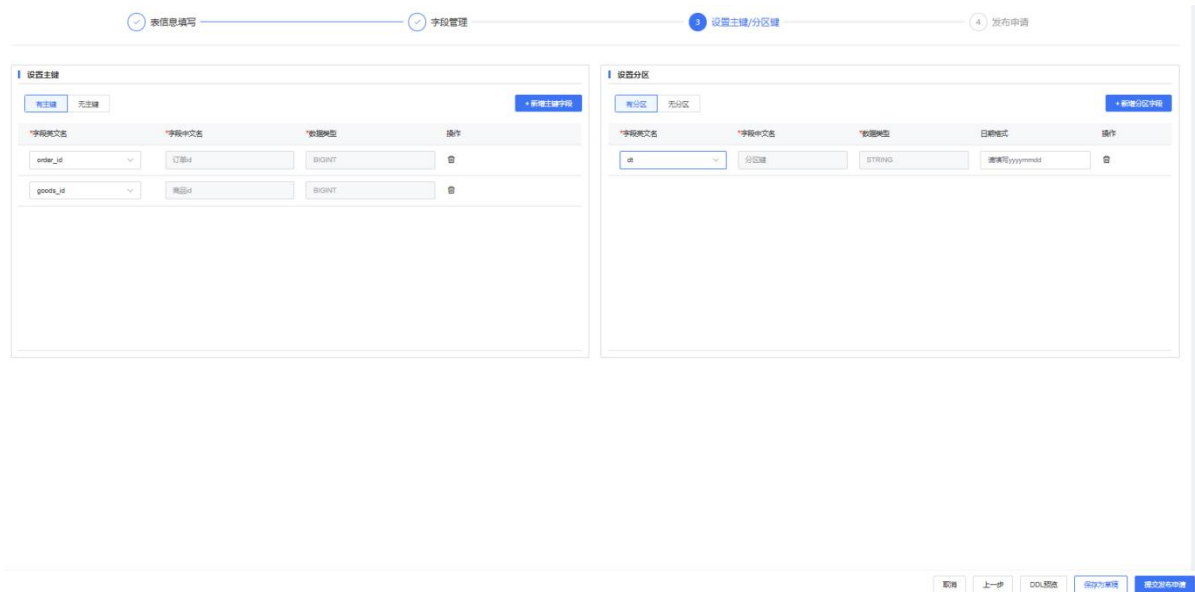
点击“确认”，则进入“设置字段属性”。

● **设置字段属性。**

- 字段英文名、字段中文名、数据类型、描述口径。
- 设置非空属性、关联字段标准、标准值域。
- 关联类型选维度属性，并选择关联的内容。

3. 设置主键和分区键。

字段配置完成后，可以根据实际场景配置 1 个或多个主键字段；针对 Hive 表，建议您配置分区字段并确定分区字段格式。



4. 发布申请。

设置完成后，即可提交发布申请。

后续步骤

- 发布申请生成审批单，需审批人员在“我的审批”进行审批，审批通过后，系统自动发布生成物理表。系统生成后，可在建模“生产”tab，查看到该表的最新版本。
- 发布成功后，可在“ workflow 开发”，基于该表进行相关的数据开发操作。

4.3.6.1.1.4 新建事实表



本章节主要介绍新建事实表的步骤。

背景信息

事实表：又叫明细数据表，结合业务过程的规划，梳理分析各业务过程中可能产生的实际数据，将这些实际数据字段通过明细数据表的方式存储下来。例如下单这一业务过程中，您可以创建下单这一明细数据表，用于记录下单过程可能产生实际数据字段，例如订单 ID、订单创建时间、

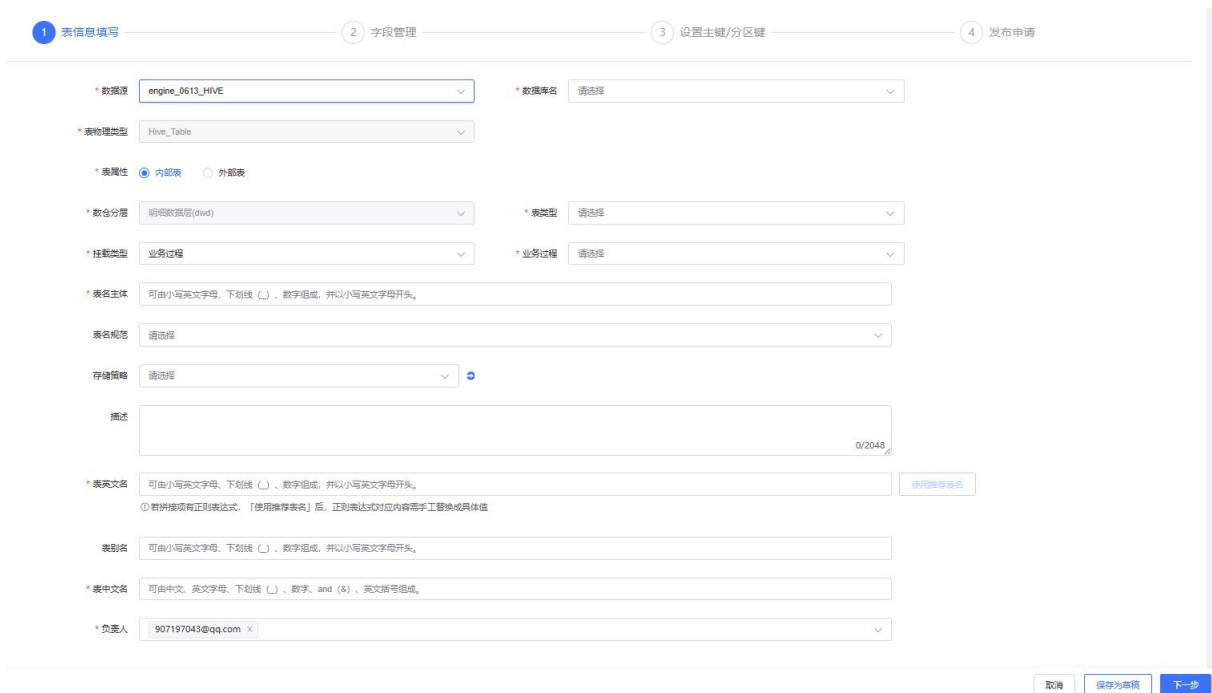
商品 ID、数量、金额等。后续您可将明细数据表发布到数仓中，通过 ETL 将真实的数据按照明细数据表的定义方式进行存储，便于业务分析时取用。

功能入口

- 单击左上方的  图标，选择全部产品 > 数据建模 > 维度建模 > 正向建模，进入正向建模页面；
- 新建事实表：在正向建模页面，鼠标悬停至  图标，单击“新建事实表”。

详细操作步骤

1. 表信息填写。



1 表信息填写 2 字段管理 3 设置主键/分区键 4 发布申请

* 数据源 engine_0613_HIVE * 数据库名 请选择

* 表物理类型 Hive_Table

* 表属性 内部表 外部表

* 数仓分层 明细数据层(dwd) * 表类型 请选择

* 挂载类型 业务过程 * 业务过程 请选择

* 表名主体 可由小写英文字母、下划线(_)、数字组成，并以小写英文字母开头。

表名规范 请选择

存储策略 请选择

描述 0/2048

* 表英文名 可由小写英文字母、下划线(_)、数字组成，并以小写英文字母开头。
① 若拼接项有正则表达式，「使用推荐表名」后，正则表达式对应内容需手工替换成具体值 [使用推荐表名](#)

表别名 可由小写英文字母、下划线(_)、数字组成，并以小写英文字母开头。

* 表中文名 可由中文、英文字母、下划线(_)、数字、and (&)、英文符号组成。

* 负责人 907197043@qq.com X

[取消](#) [保存为草稿](#) [下一步](#)

主要参数说明如下：

填写参数	参数说明
数据源	需通过集群注册 Hive 组件，且在引擎管理中将集群与调度资源组绑定，系统自动生成内置的 Hive 数据源，如果有多个，下拉单选。
数据库	该数据源下已创建的数据库，下拉单选。
表物理类型	表的物理类型。
表属性	内部表：默认为内部表； 外部表：外部表需要填写外部表的 Location。
数仓分层	跟所选的分层对应。
表类型	根据实际情况，选择事务事实表、周期快照表或累加表。
挂载类型	为表所属的挂载类型，如业务过程。
挂载类型名	跟挂载类型对应，选择具体的数仓要素的名称。
表名主体	用于表达该表的作用，是表英文名的组成部分；由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。 一般情况下表英文名包含：固定字符(如 dwd)_业务过程_表名主体_存储策略。
表名规范	通过已配置的表名规范约束表名。选择表名规范后，可点击“使用推荐表名”，系统生成推荐表名。

填写参数	参数说明
存储策略	事实表基于什么策略（即存储数据的时间周期及数据量范围）存储数据。
描述	表描述。
表英文名	事实表的英文名，由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。只要选择了表名规范，不管是强规则还是弱规则，都可以点击“使用推荐表名”，系统生成推荐表名供使用。如果选择的表名规范为强规则，则需遵循表名规范进行建设。
表别名	事实表的别名，由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。一般用于逆向建模表英文名不规范时，此参数可填写规范的名称，作为辅助表英文名用。
表中文名	事实表的中文名，由中文、英文字母、下划线（_）、and（&）、英文括号组成。
负责人	当前事实表的责任人，默认为创建事实表的用户。可选多个。

点击“保存为草稿”或“下一步”；点击“保存为草稿”则可在左侧“草稿”tab，查看草稿列表；点击“下一步”，则进入步骤2“字段管理”。

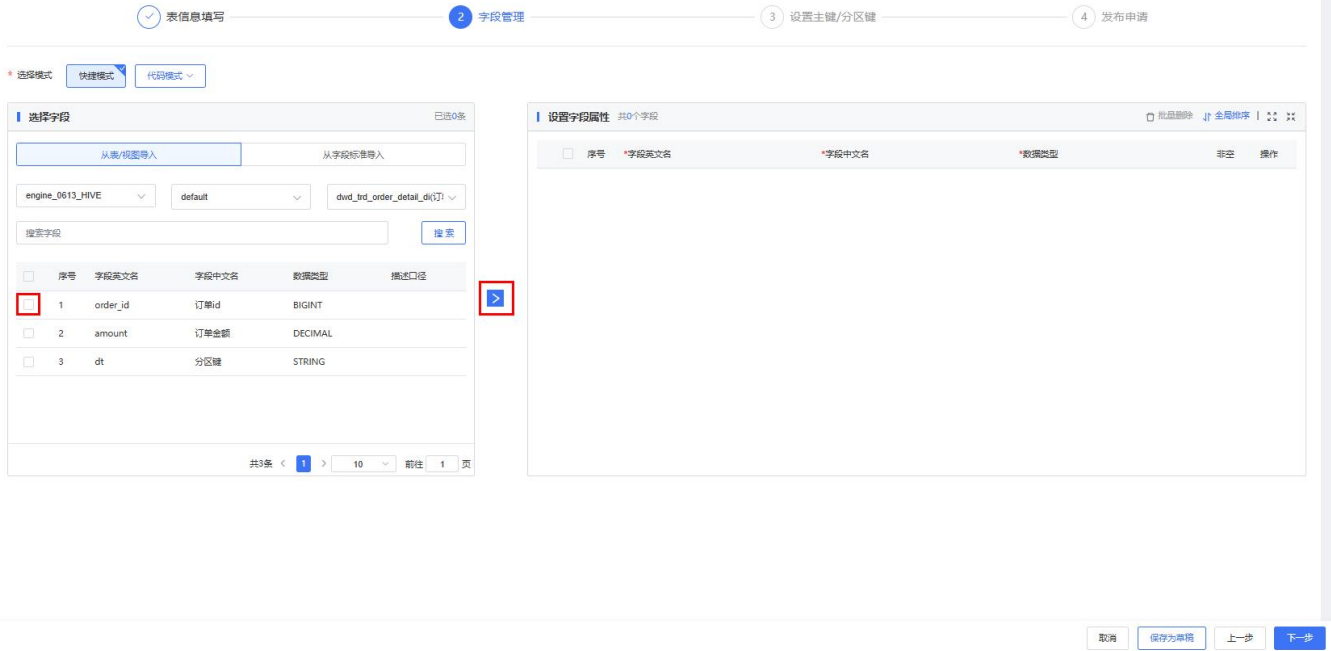
2. 字段管理。

● 选择字段。

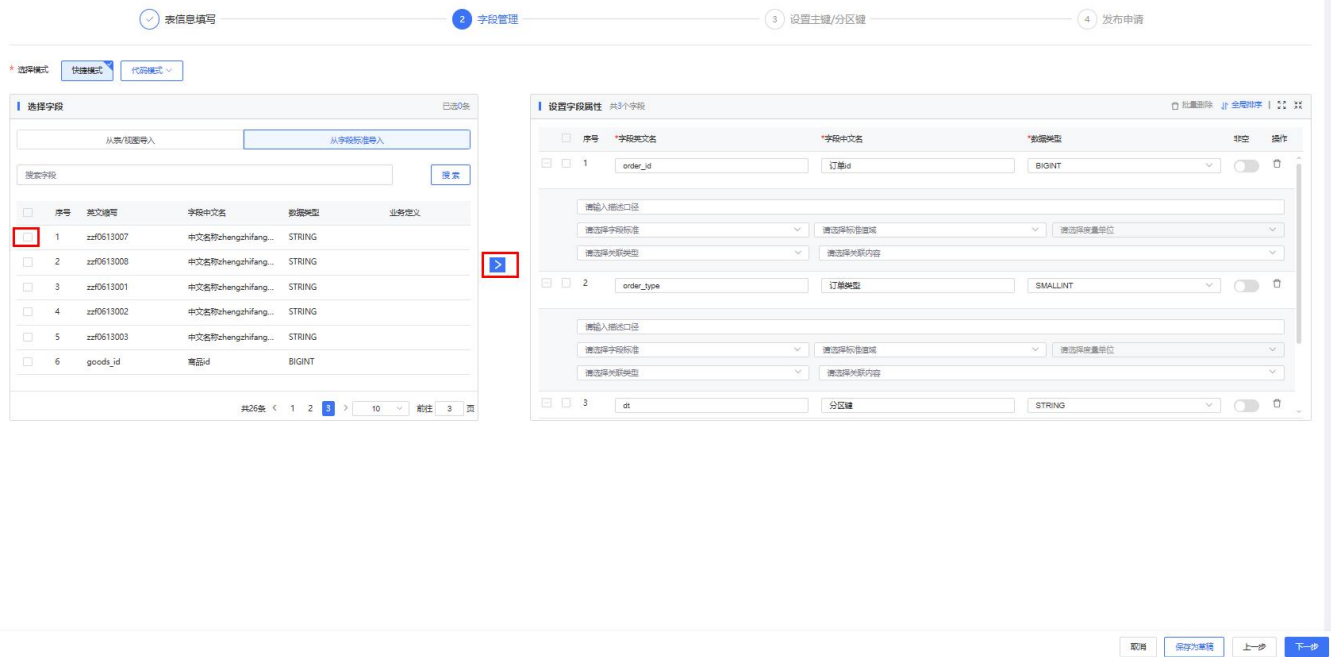
字段管理支持快捷模式与代码模式。

A. 快捷模式：支持从表/视图导入、从字段标准导入。

如下截图为从表/视图导入：



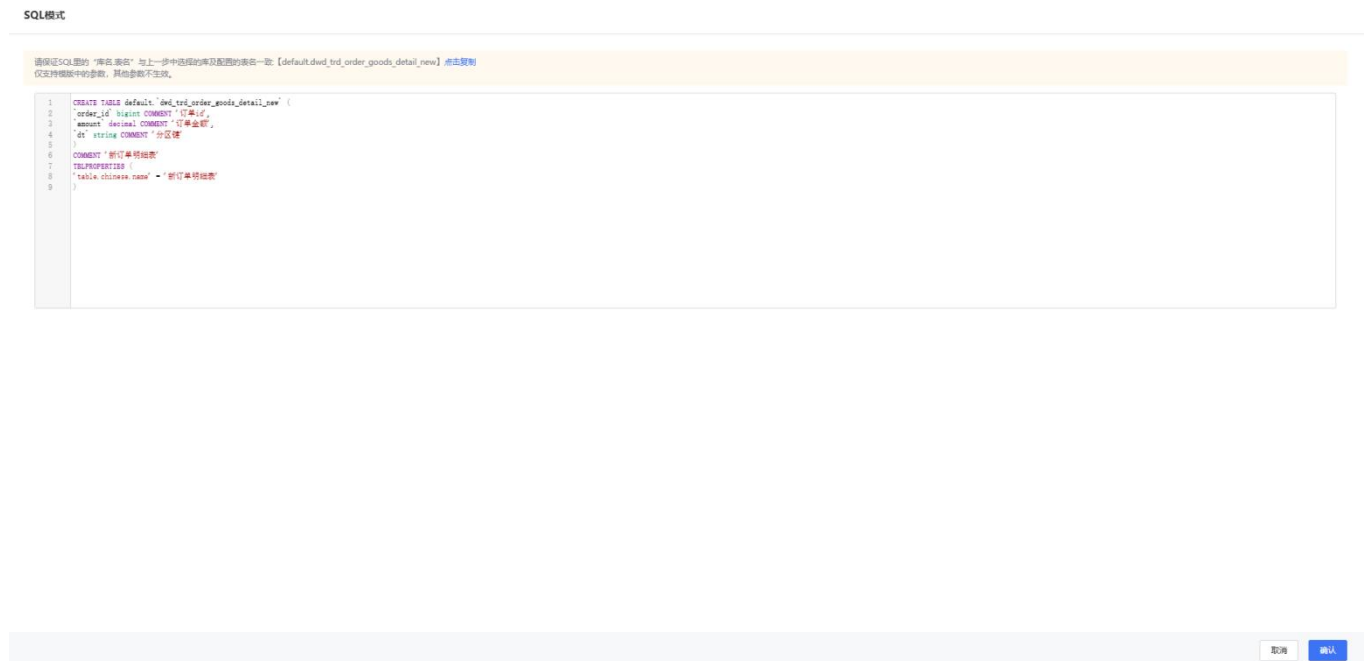
如下截图为从字段标准导入：



B. 代码模式导入：支持解析 sql 内容导入字段。有两种方式：



其一：保留快捷模式字段，将快捷模式所选的字段自动生成 sql 带到代码模式的输入框；其二，清空快捷模式字段，不将快捷模式所选的字段带到代码模式的输入框。



点击“确认”，则进入“设置字段属性”。

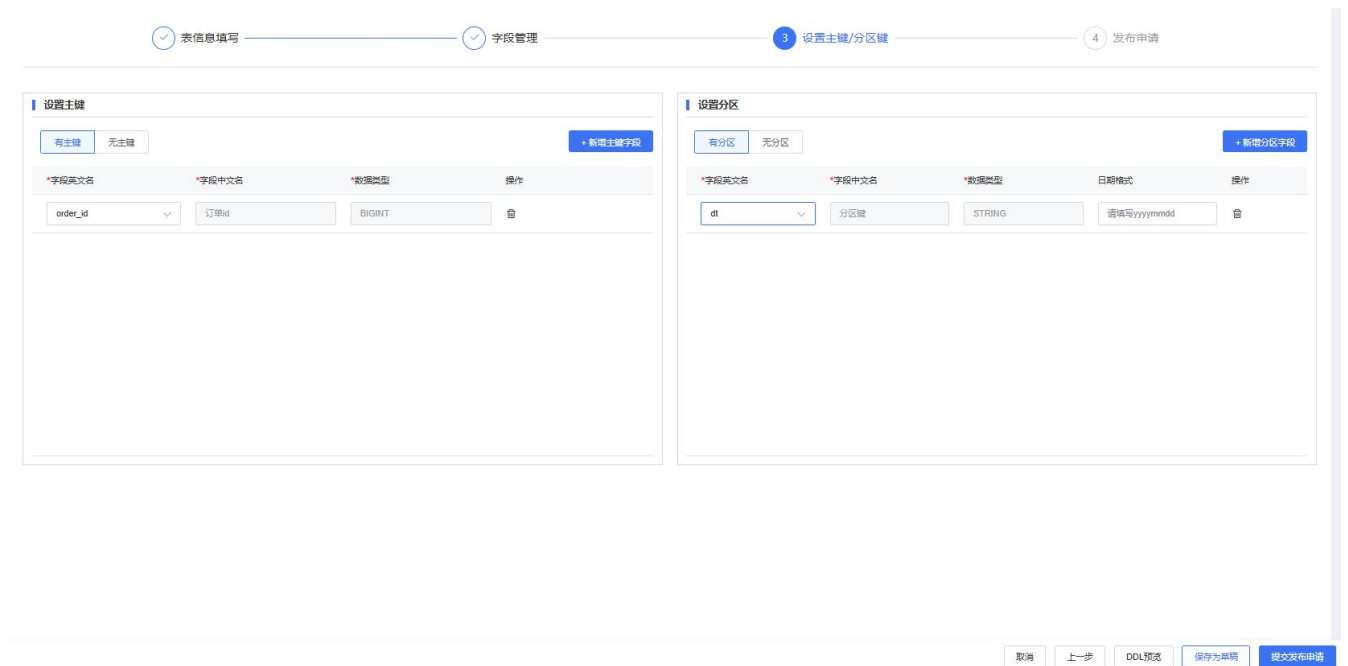
● 设置字段属性。

- 字段英文名、字段中文名、数据类型、描述口径。
- 设置非空属性、关联字段标准、标准值域、设置度量单位。
- 关联类型可选择维度属性或原子指标，并选择所关联的内容。

3. 设置主键和分区键。

字段配置完成后，可以根据实际场景配置 1 个或多个主键字段；针对 Hive 表，建议您配置分区

字段并确定分区字段格式。



4. 发布申请。

设置完成后，即可提交发布申请。

后续步骤

- 发布申请生成审批单，需审批人员在“我的审批”中进行审批，审批通过后，系统自动发布生成物理表。系统生成后，可在建模“生产”tab，查看到该表的最新版本。
- 发布成功后，可在“ workflow 开发”，基于该表进行相关的数据开发操作。

4.3.6.1.1.5 新建汇总表



本章节主要介绍新建汇总表的步骤。

背景信息

汇总表：结合业务数据分析和数仓分层，将一些明细的事实数据和维度数据先进行汇总分析，

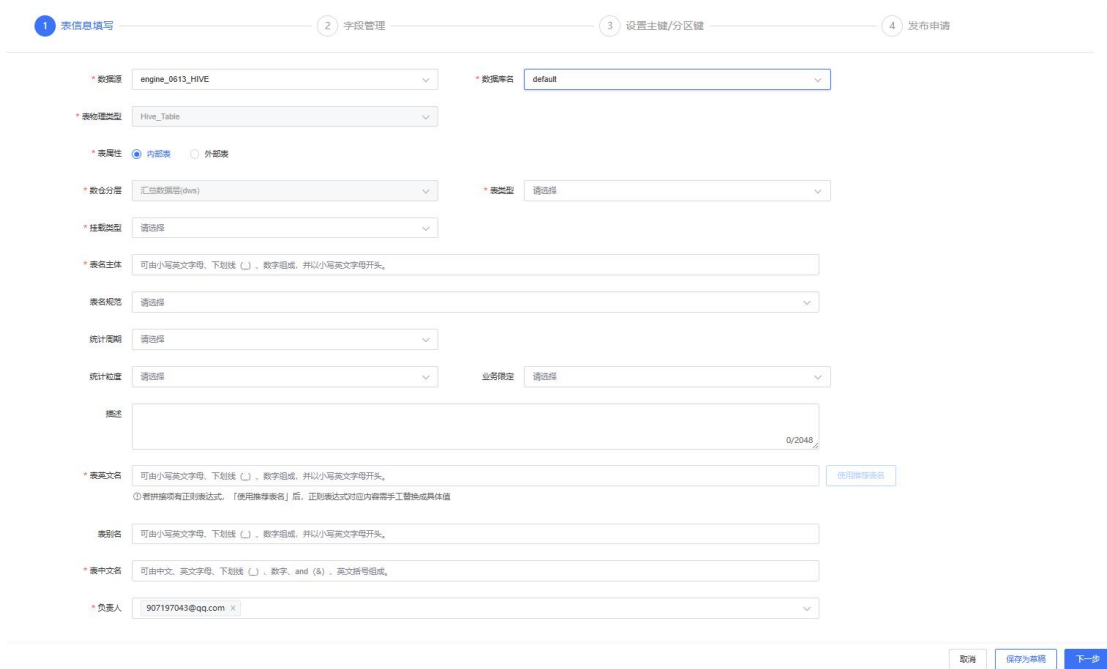
创建汇总表，后续数据分析时直接取用汇总表中的数据即可，无需再取用明细表和维度表中的数据。

功能入口

- 单击左上方的  图标，选择全部产品 > 数据建模 > 维度建模 > 正向建模，进入正向建模页面；
- 新建汇总表：在正向建模页面，鼠标悬停至  图标，单击“新建汇总表”。

详细操作步骤

1. 表信息填写。



1 表信息填写 | 2 字段管理 | 3 设置主键/分区键 | 4 发布申请

* 数据源: engine_0613_HIVE | * 数据库名: default

* 表物理类型: Hive_Table

* 表属性: 内部表 外部表

* 聚合维度: 汇总数据(agg) | * 表类型: 请选择

* 挂载类型: 请选择

* 表名主体: 可由小写英文字母、下划线 (_)、数字组成，并以小写英文字母开头。

表名后缀: 请选择

统计周期: 请选择

统计粒度: 请选择 | 业务绑定: 请选择

描述: 0/2048

* 表英文名: 可由小写英文字母、下划线 (_)、数字组成，并以小写英文字母开头。
① 若拼接带有正则表达式，「使用推荐表名」后，正则表达式对应内容需手工替换成具体值

表别名: 可由小写英文字母、下划线 (_)、数字组成，并以小写英文字母开头。

* 表中文名: 可由中文、英文字母、下划线 (_)、数字、and (&)、英文符号组成。

* 负责人: 907197043@qq.com x

取消 | 保存为草稿 | 下一步

主要参数说明如下：

填写参数	参数说明
数据源	需通过集群注册 Hive 组件，且在引擎管理中将集群与调度资源组绑定，系统自动生成内置的 Hive 数据源，如果有多个，下拉单选。
数据库名	该数据源下已创建的数据库，下拉单选。
表物理类型	表的物理类型。
表属性	内部表：默认为内部表； 外部表：外部表需要填写外部表的 Location。
数仓分层	跟所选的分层对应。
表类型	汇总表的表类型： <ul style="list-style-type: none"> ● 普通汇总表：只能统计单粒度的数据，不会做多个粒度数据的再次汇总； ● 轻度汇总表：可以统计多个粒度的数据，并将不同粒度的数据进行再次汇总。
挂载类型	表所属的挂载类型，如数据域。
挂载类型名	跟挂载类型对应，选择具体的数仓要素的名称。
表名主体	用于表达该表的作用，是表英文名的组成部分；由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。一般情况下表英文名包含：固定字符(如 dws)_数据域_表名主体_统计周期。
表名规范	通过已配置的表名规范约束表名。选择表名规范后，可点击“使用推荐表名”，系统生成推荐表名。

填写参数	参数说明
统计周期	用于确定汇总表需要整合的统计数值的时间范围。例如，近一天，近一周。
统计粒度	汇总表是基于哪个维度，也即是基于哪个统计粒度建设的。
业务限定	汇总表是基于哪个业务范围，也即是基于哪个业务限定建设的。
描述	表描述。
表英文名	汇总表的英文名，由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。只要选择了表名规范，不管是强规则还是弱规则，都可以点击“使用推荐表名”，系统生成推荐表名供使用。如果选择的表名规范为强规则，则需遵循表名规范进行建设。
表别名	汇总表的别名，由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。一般用于逆向建模表英文名不规范时，此参数可填写规范的名称，作为辅助表英文名用。
表中文名	汇总表的中文名，由中文、英文字母、下划线（_）、and（&）、英文括号组成。
负责人	当前汇总表的责任人，默认为创建汇总表的用户，可选多个。

点击“保存为草稿”或“下一步”；点击“保存为草稿”则可在左侧“草稿”tab，查看草稿列表；点击“下一步”，则进入步骤2“字段管理”。

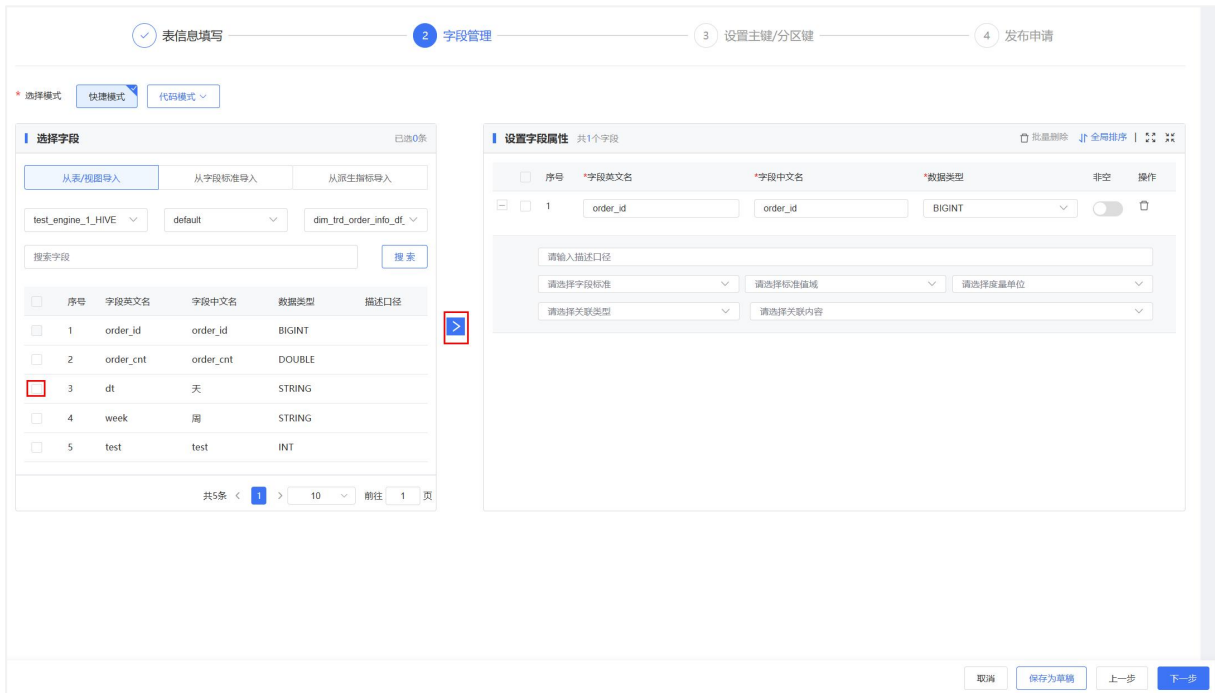
2. 字段管理。

- **选择字段。**

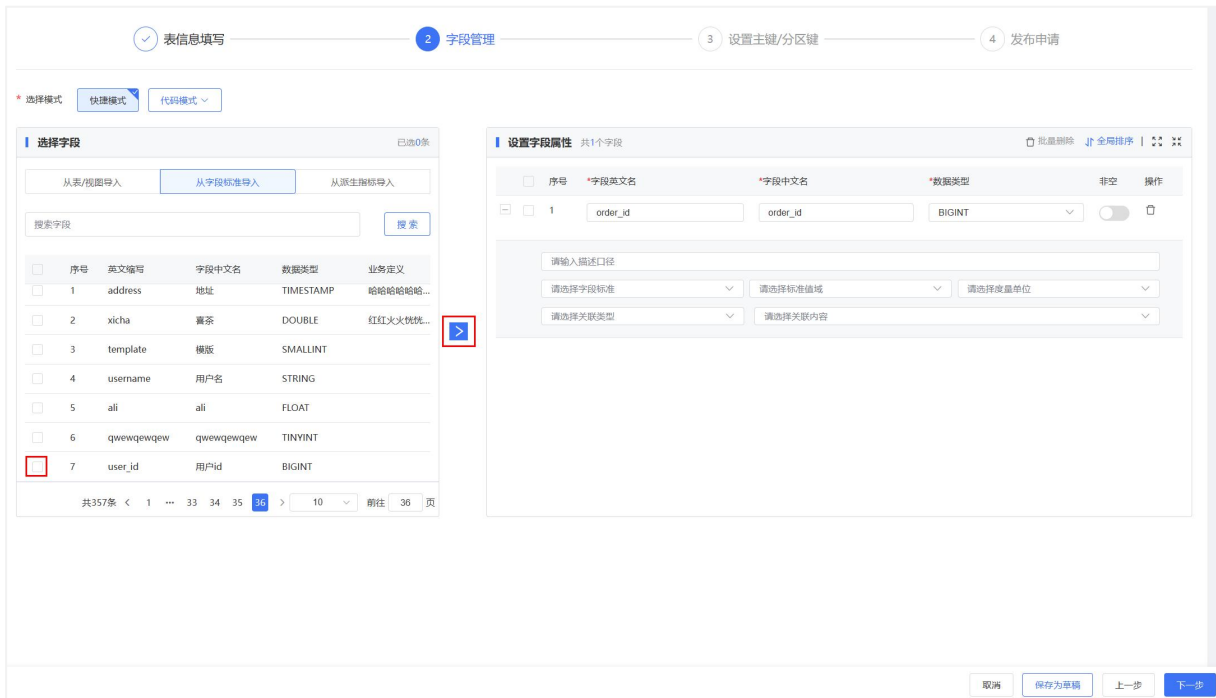
字段管理支持快捷模式与代码模式。

A. 快捷模式：支持从表/视图导入、从字段标准导入、从派生指标导入。

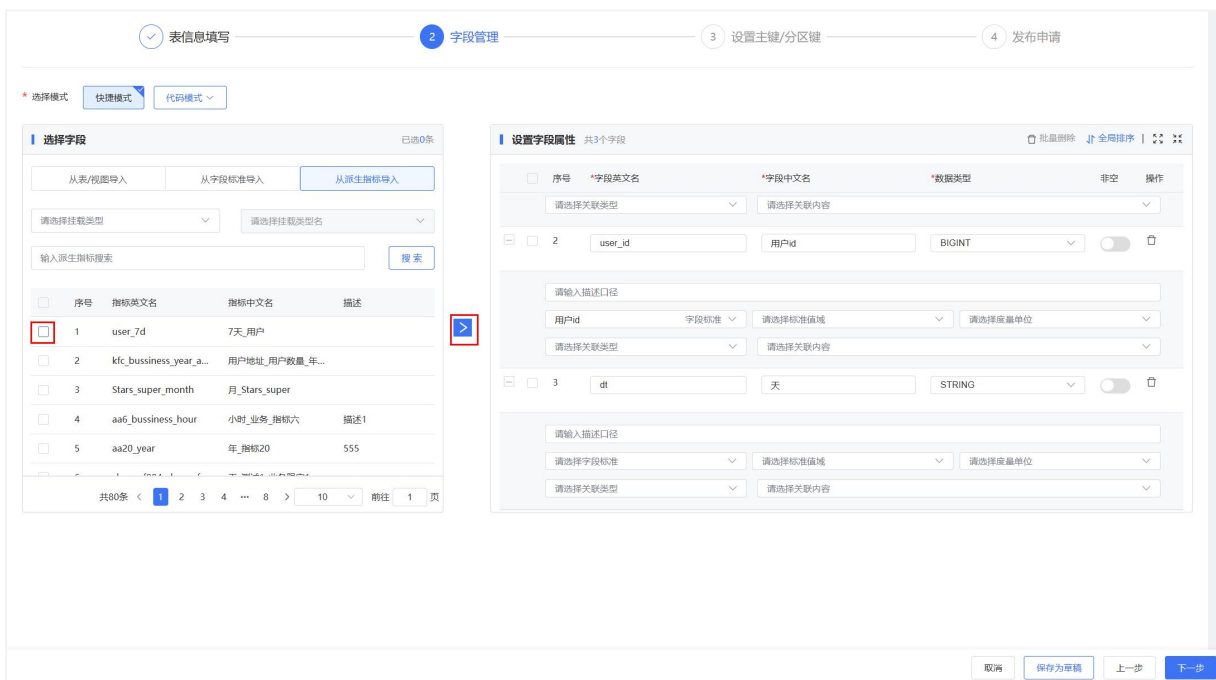
如下截图为从表/视图导入：



如下截图为从字段标准导入：



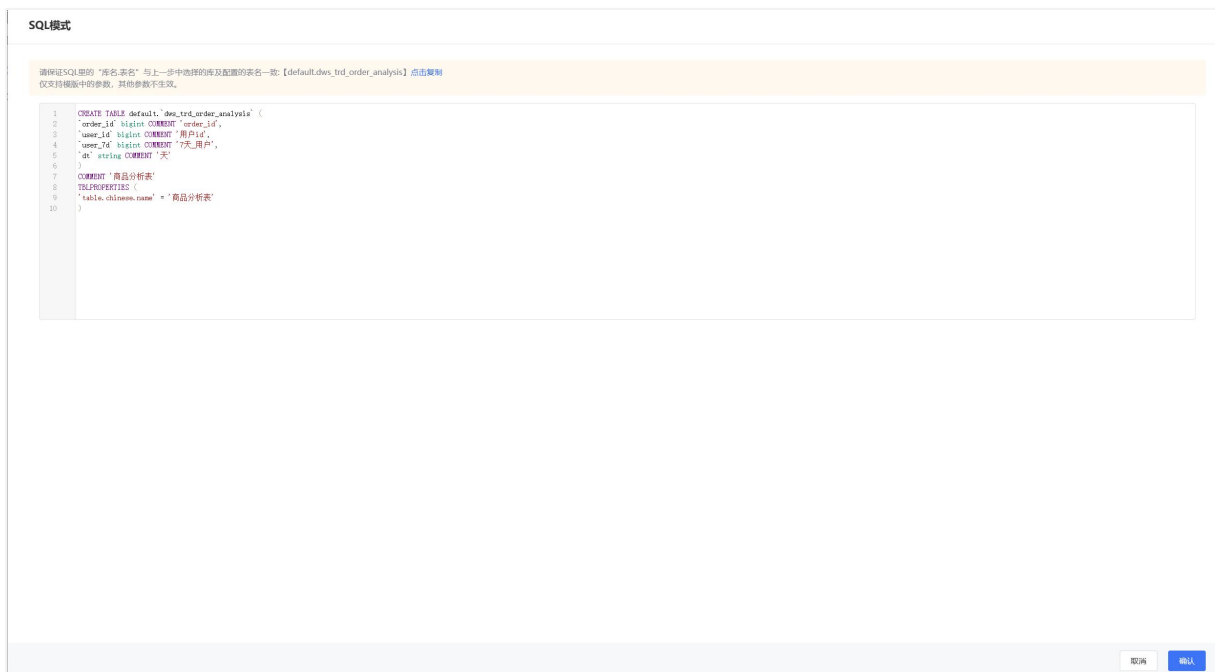
如下截图为从派生指标导入：



B. 代码模式导入：支持解析 sql 内容导入字段。有两种方式：



其一：保留快捷模式字段，将快捷模式所选的字段自动生成 sql 带到代码模式的输入框；其二，清空快捷模式字段，不将快捷模式所选的字段带到代码模式的输入框。选择其中一种代码模式，进入 SQL 模式弹窗。



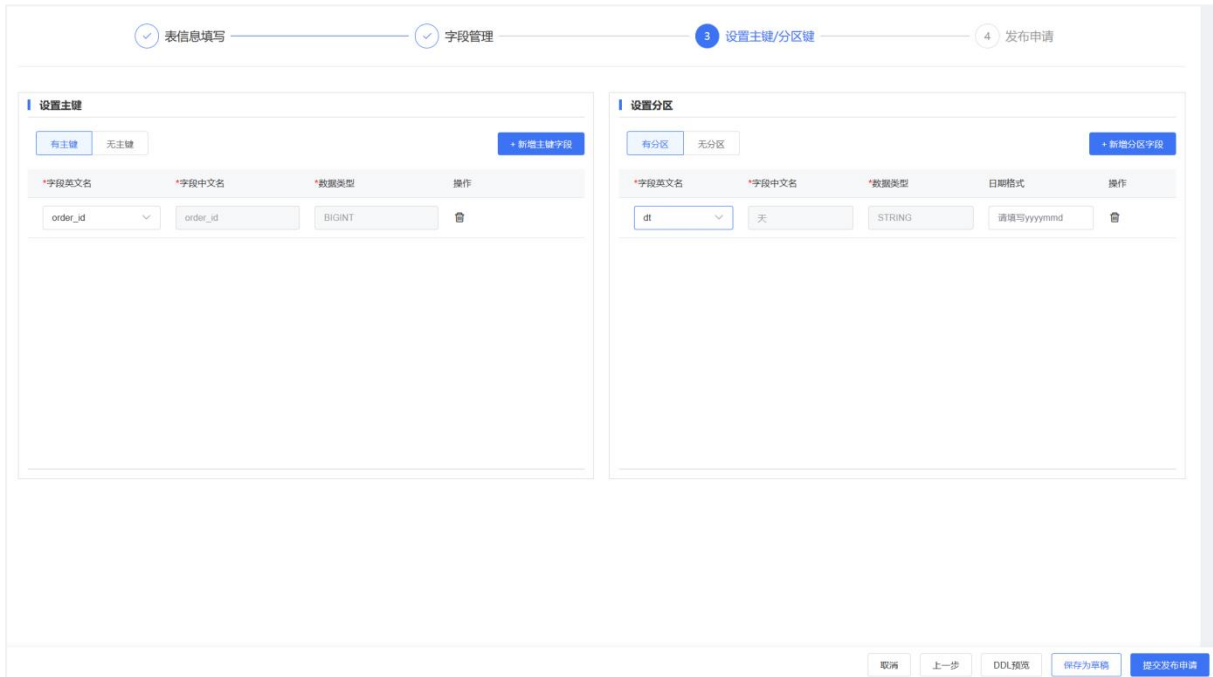
点击“确认”，则进入“设置字段属性”。

● 设置字段属性。

- 字段英文名、字段中文名、数据类型、描述口径。
- 设置非空属性、关联字段标准、标准值域、度量单位。
- 关联类型可选择维度属性或派生指标，并选择所关联的内容。

3. 设置主键和分区键。

字段配置完成后，可以根据实际场景配置 1 个或多个主键字段；针对 Hive 表，建议您配置分区字段并确定分区字段格式。



4. 发布申请。

设置完成后，即可提交发布申请。

后续步骤

- 发布申请生成审批单，需审批人员在“我的审批”中进行审批，审批通过后，系统自动发布生成物理表。系统生成后，可在建模“生产”tab，查看到该表的最新版本。
- 发布成功后，可在“ workflow 开发”，基于该表进行相关的数据开发操作。



4.3.6.1.1.6 新建应用表

本章节主要介绍新建应用表的步骤。

背景信息

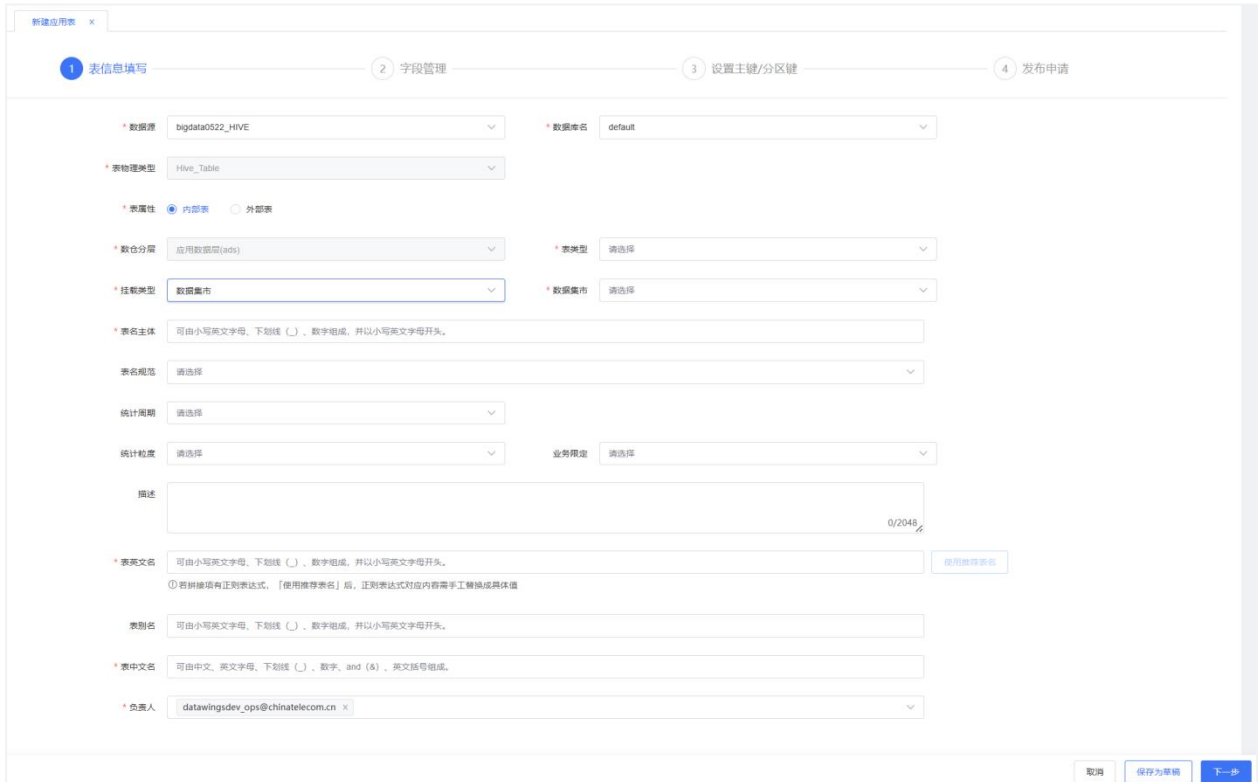
应用表是面向具体业务场景，用于组织相同时间周期、相同维度的指标或统计粒度的统计数据，为后续的业务查询、OLAP 分析、数据分发等提供基础。

功能入口

- 单击左上方的  图标，选择全部产品 > 数据建模 > 维度建模 > 正向建模，进入正向建模页面；
- 新建应用表：在正向建模页面，鼠标悬停至  图标，单击“新建应用表”。

详细操作步骤

1. 表信息填写。



新建应用表配置界面，包含以下配置项：

- 1 表信息填写
- 2 字段管理
- 3 设置主键/分区键
- 4 发布申请

配置项包括：

- 数据源: bigdata0622_HIVE
- 数据命名: default
- 表物理类型: Hive_Table
- 表属性: 内部表 外部表
- 数仓分库: 应用数据层(ads)
- 表类型: 请选择
- 挂载类型: 数据集市
- 数据集市: 请选择
- 表名主体: 可由小写英文字母、下划线 (_)、数字组成，并以小写英文字母开头。
- 表名规范: 请选择
- 统计周期: 请选择
- 统计粒度: 请选择
- 业务限定: 请选择
- 描述: 0/2048
- 表英文名: 可由小写英文字母、下划线 (_)、数字组成，并以小写英文字母开头。
① 若拼接项有正则表达式，「使用推荐表名」后，正则表达式对应内容需手工替换成具体值
- 表别名: 可由小写英文字母、下划线 (_)、数字组成，并以小写英文字母开头。
- 表中文名: 可由中文、英文字母、下划线 (_)、数字、and (&)、英文括号组成。
- 负责人: datawingsdev_ops@chinatelecom.cn

底部按钮: 取消 保存为草稿 下一步

主要参数说明如下：

参数	参数说明
数据源	需通过集群注册 Hive 组件，且在引擎管理中将集群与调度资源组绑定，系统自动生成内置的 Hive 数据源，如果有多个，下拉单选。
数据库名	该数据源下已创建的数据库，下拉单选。
表物理类型	表的物理类型。
表属性	内部表：默认为内部表； 外部表：外部表需要填写外部表的 Location。
数仓分层	跟所选的分层对应。
表类型	应用表的表类型： <ul style="list-style-type: none"> ● 普通应用表：只能统计单粒度的数据，不会做多个粒度数据的再次汇总； ● 轻度应用表：可以统计多个粒度的数据，并将不同粒度的数据进行再次汇总。
挂载类型	表所属的挂载类型，如数据集市或分析主题。
挂载类型名	跟挂载类型对应，选择具体的数仓要素的名称。
表名主体	用于表达该表的作用，是表英文名的组成部分； 由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。 一般情况下表英文名包含：固定字符(如 ads)_数据集市_表名主体_统计周期。
表名规范	通过已配置的表名规范约束表名。选择表名规范后，可点击“使用推荐表名”，

参数	参数说明
	系统生成推荐表名。
统计周期	用于确定应用表需要整合的统计数值的时间范围。例如，近一天，近一周。
统计粒度	应用表是基于哪个维度，也即是基于哪个统计粒度建设的。
业务限定	应用表是基于哪个业务范围，也即是基于哪个业务限定建设的。
描述	表描述。
表英文名	应用表的英文名，由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。只要选择了表名规范，不管是强规则还是弱规则，都可以点击“使用推荐表名”，系统生成推荐表名供使用。如果选择的表名规范为强规则，则需遵循表名规范进行建设。
表别名	应用表的别名，一般用于逆向建模表英文名不规范时，此参数可填写规范的名称，作为辅助表英文名用。由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。
表中文名	应用表的中文名，由中文、英文字母、下划线（_）、and（&）、英文括号组成。
负责人	当前应用表的责任人，默认为创建应用表的用户，可选多个。

点击“保存为草稿”或“下一步”；点击“保存为草稿”则可在左侧“草稿”tab，查看草稿列表；点击“下一步”，则进入步骤2“字段管理”。

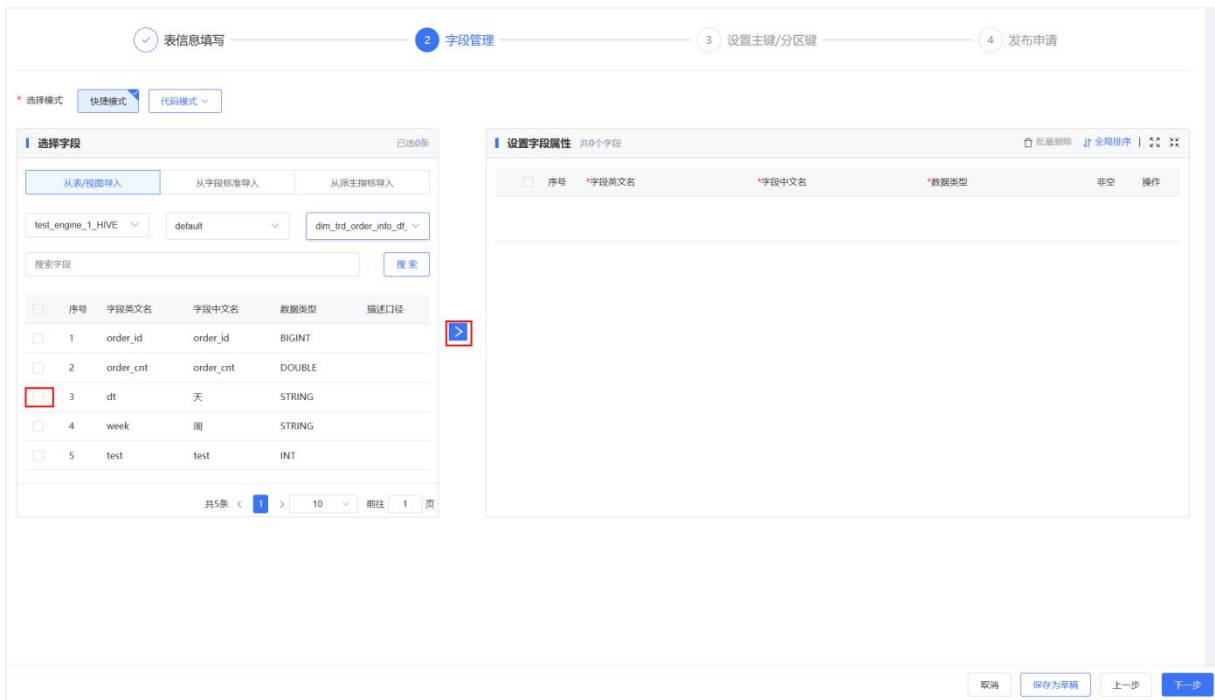
2. 字段管理。

- **选择字段。**

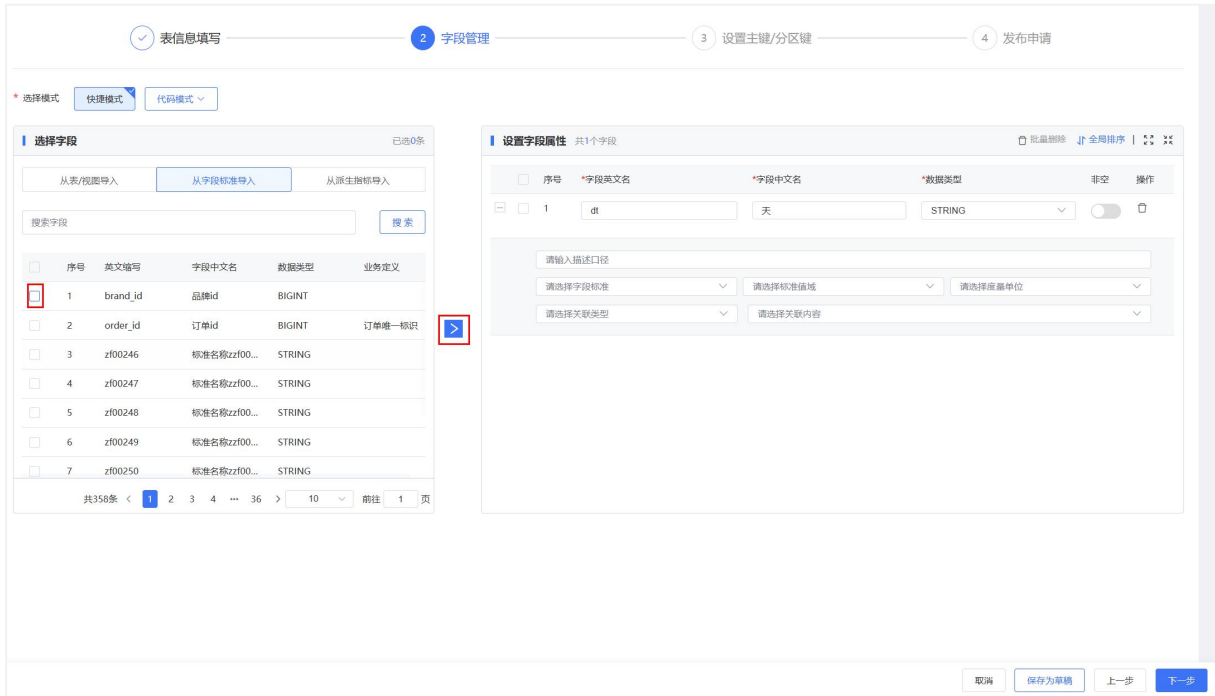
字段管理支持快捷模式与代码模式。

A. 快捷模式：支持从表/视图、从字段标准导入、从派生指标导入。

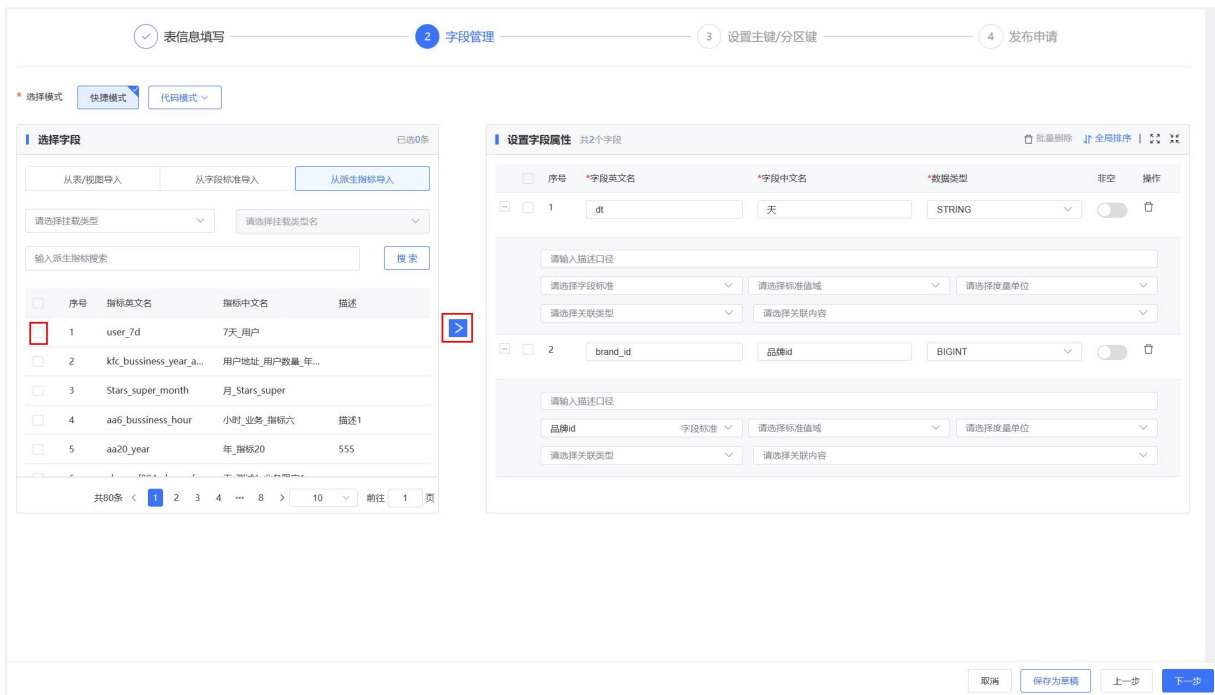
如下截图为从表/视图导入：



如下截图为从字段标准导入：



如下截图为从派生指标导入：



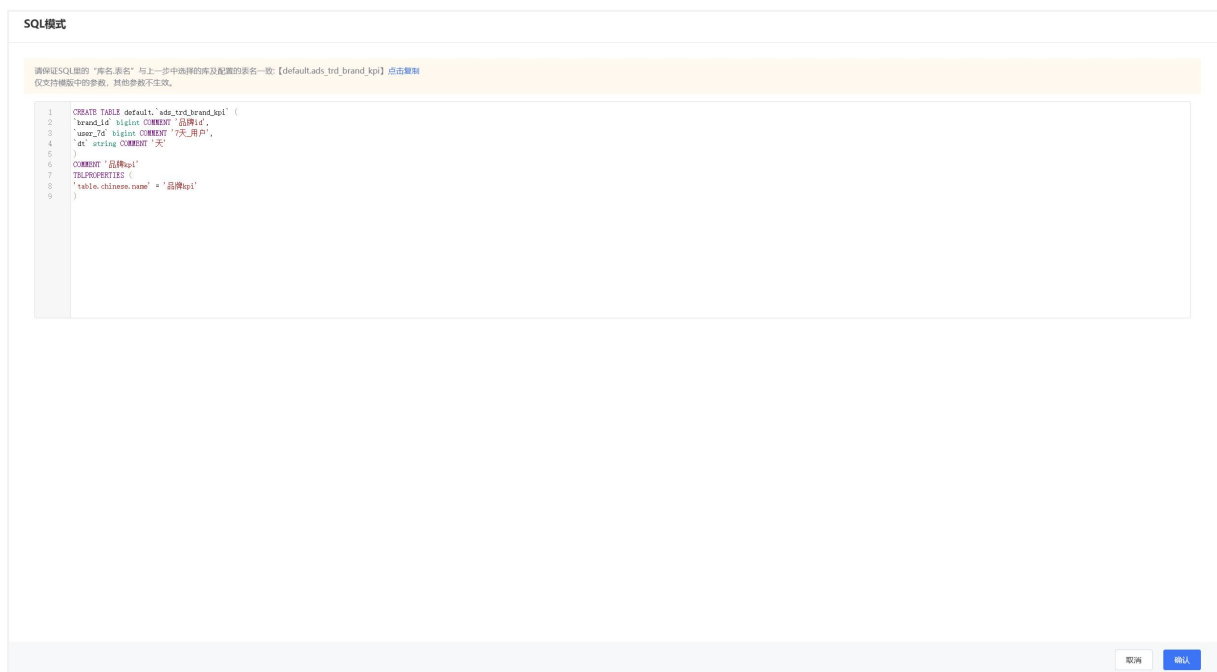
B. 代码模式导入：支持解析 sql 内容导入字段。有两种方式：



其一：保留快捷模式字段，将快捷模式所选的字段自动生成 sql 带到代码模式的输入框；

其二，清空快捷模式字段，不将快捷模式所选的字段带到代码模式的输入框。

选择其中一种代码模式，进入 SQL 模式弹窗。

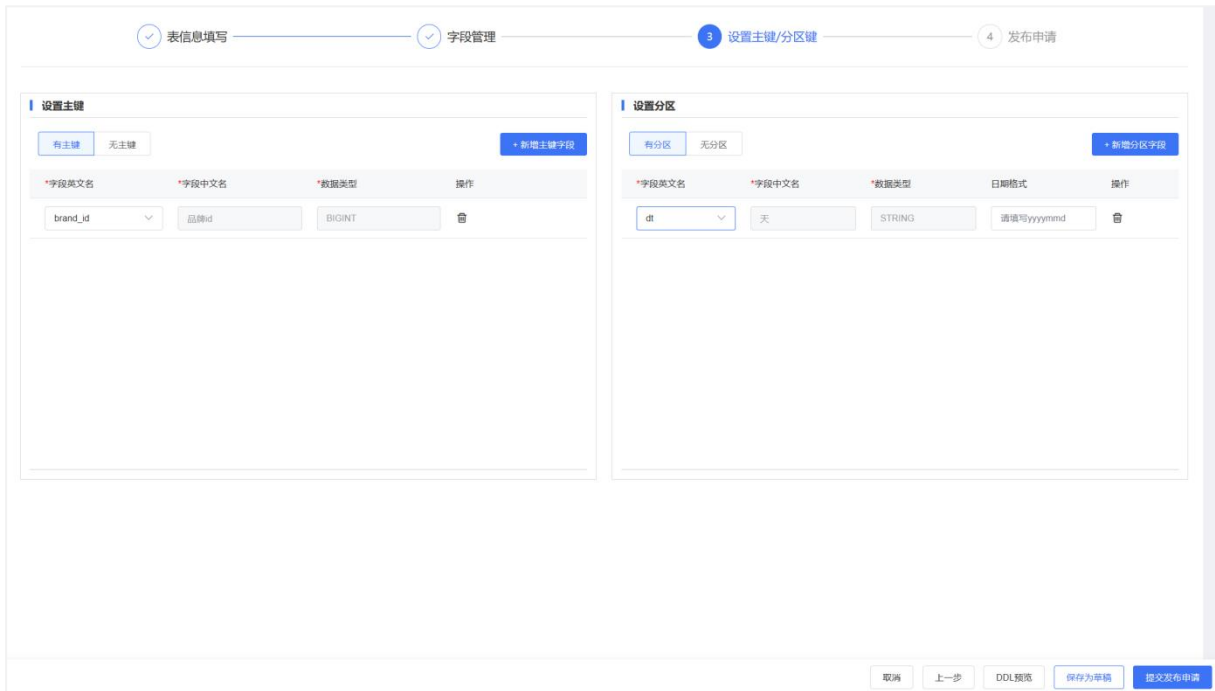


点击“确认”，则进入“设置字段属性”。

- 设置字段属性。
 - 字段英文名、字段中文名、数据类型、描述口径。
 - 设置非空属性、关联字段标准、标准值域、度量单位。
 - 关联类型可选择维度属性或派生指标，并选择所关联的内容。

3. 设置主键和分区键。

字段配置完成后，可以根据实际场景配置 1 个或多个主键字段；针对 Hive 表，建议您配置分区字段并确定分区字段格式。



4. 发布申请。

设置完成后，即可提交发布申请。

后续步骤

- 发布申请生成审批单，需审批人员在“我的审批”中进行审批，审批通过后，系统自动发布生成物理表。系统生成后，可在建模“生产”tab，查看到该表的最新版本。
- 发布成功后，可在“ workflow 开发”，基于该表进行相关的数据开发操作。



4.3.6.1.1.7 新建其他表

本章节主要介绍新建其他表的步骤。

背景信息

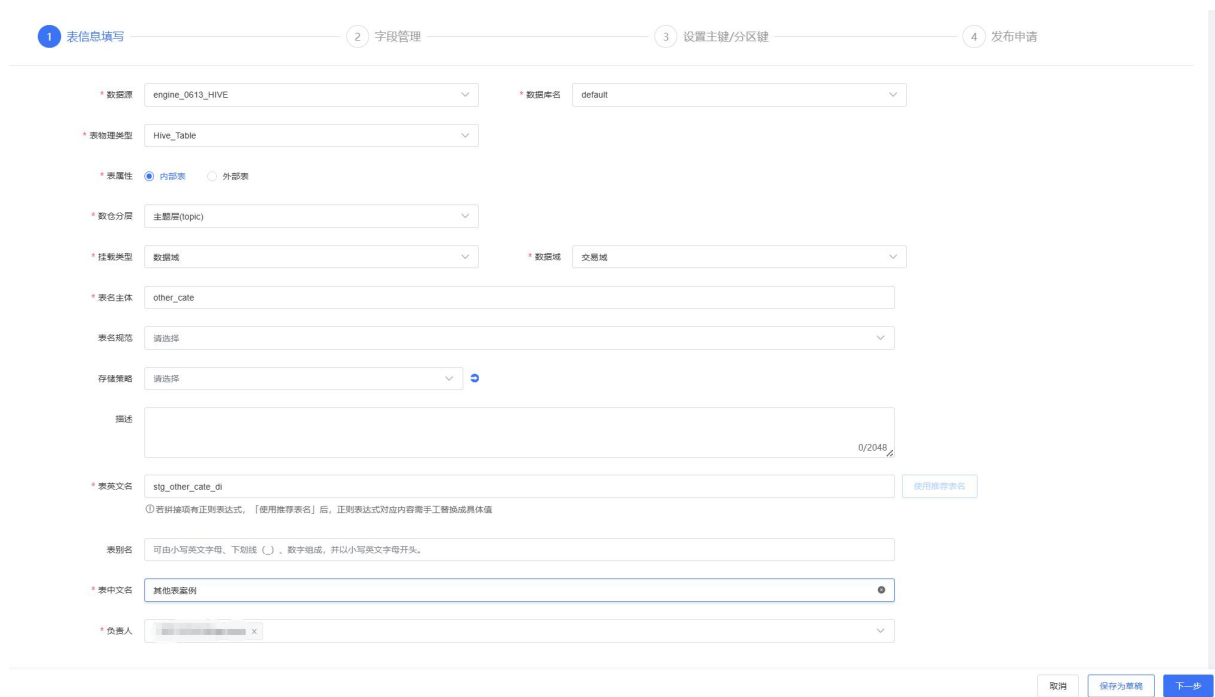
其他表：主要是用于承接自定义数仓分层的表。

功能入口

- 单击左上方的  图标，选择全部产品 > 数据建模 > 维度建模 > 正向建模，进入正向建模页面；
- 新建其他表：在正向建模页面，鼠标悬停至  图标，单击“新建其他表”。

详细操作步骤

1. 表信息填写。



正向建模-表信息填写界面截图。顶部有四个步骤指示器：1. 表信息填写 (当前选中), 2. 字段管理, 3. 设置主键/分区键, 4. 发布申请。

表单内容如下：

- * 数据源: engine_0613_HIVE
- * 数据库名: default
- * 表物理类型: Hive_Table
- * 表属性: 内部表 外部表
- * 数仓分层: 主题层(topic)
- * 挂载类型: 数据域
- * 数据域: 交易域
- * 表名主体: other_cate
- 表名规范: 请选择
- 存储策略: 请选择
- 描述: (输入框, 0/2048)
- * 表英文名: stg_other_cate_di (右侧有“使用推荐表名”按钮)
- ① 若拼接项有正则表达式, 「使用推荐表名」后, 正则表达式对应内容需手工替换成具体值
- 表别名: 可由小写英文字母、下划线(_)、数字组成, 并以小写英文字母开头。
- * 表中文名: 其他表类别
- * 负责人: (输入框)

底部操作按钮: 取消, 保存为草稿, 下一步

主要参数说明如下：

填写参数	参数说明
数据源	需通过集群注册 Hive 组件，且在引擎管理中将集群与调度资源组绑定，系统自动生成内置的 Hive 数据源，如果有多个，下拉单选。
数据库名	该数据源下已创建的数据库，下拉单选。
表物理类型	表的物理类型。
表属性	内部表：默认为内部表；外部表：外部表需要填写外部表的 Location。
数仓分层	自定义的数仓分层。
挂载类型	表所属的挂载类型，如数据域。
挂载类型名	跟挂载类型对应，选择具体的数仓要素的名称。
表名主体	用于表达该表的作用，是表英文名的组成部分；由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。一般情况下表英文名包含：固定字符(如 stg)_数据域_表名主体_存储策略。
表名规范	通过已配置的表名规范约束表名。选择表名规范后，可点击“使用推荐表名”，系统生成推荐表名。

填写参数	参数说明
存储策略	该表基于什么策略（即存储数据的时间周期及数据量范围）存储数据。
描述	表描述。
表英文名	由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。只要选择了表名规范，不管是强规则还是弱规则，都可以点击“使用推荐表名”，系统生成推荐表名。如果选择的表名规范为强规则，则需遵循表名规范进行建设。
表别名	由小写英文字母、下划线（_）、数字组成，并以小写英文字母开头。一般用于逆向建模，当表英文名不符合规范时，此参数可填写规范的名称，作为辅助表英文名用。
表中文名	由中文、英文字母、下划线（_）、and（&）、英文括号组成。
负责人	表的责任人，默认为创建表的用户，可选多个。

完成表信息填写，点击“保存为草稿”或“下一步”；点击“保存为草稿”则可在左侧“草稿”tab，查看草稿列表；点击“下一步”，则进入步骤2“字段管理”。

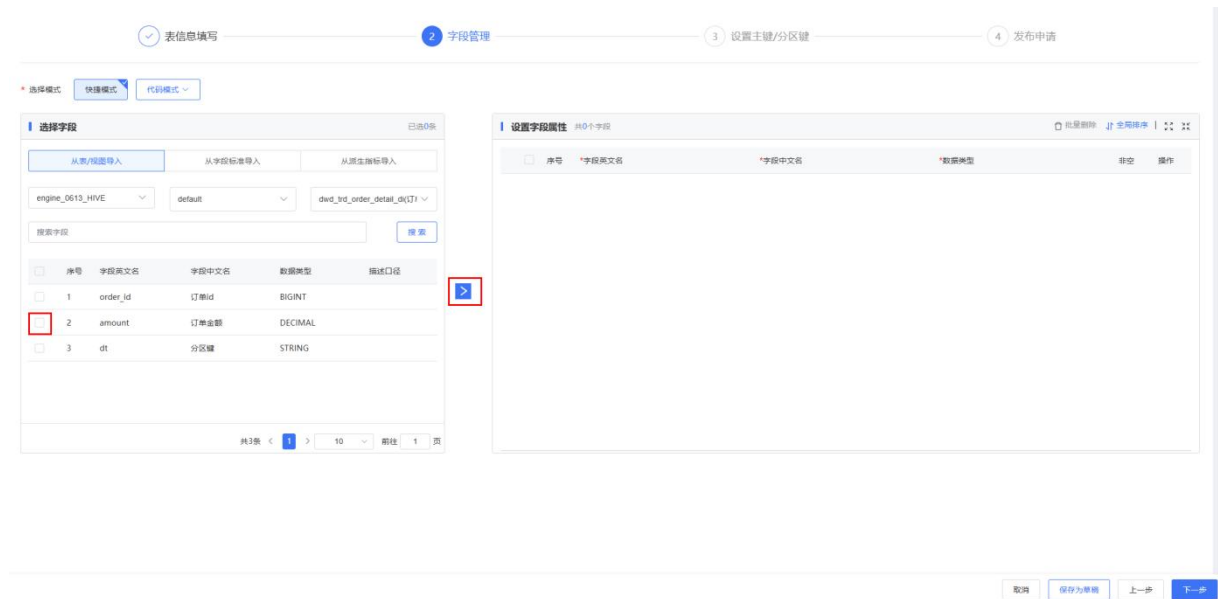
2. 字段管理。

● 选择字段。

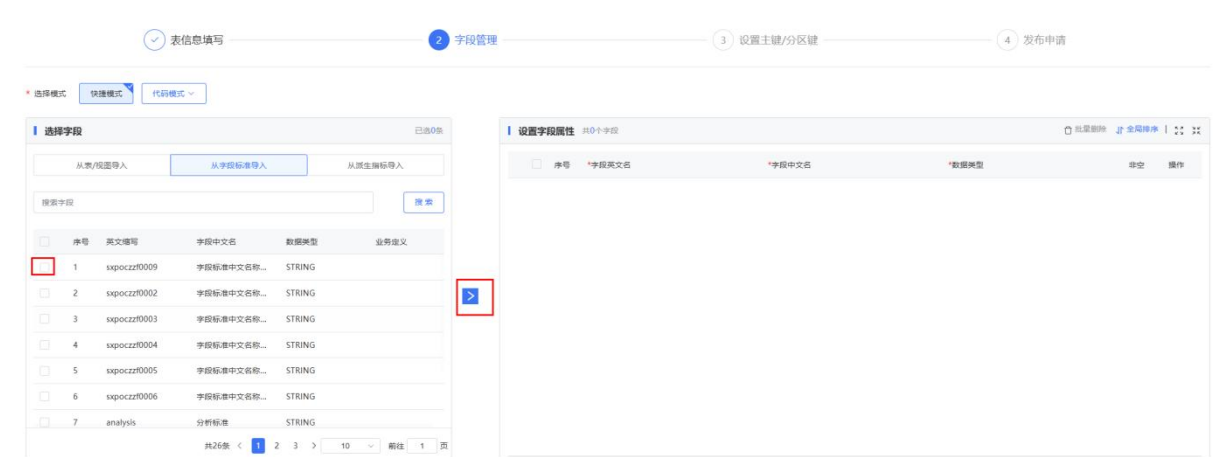
字段管理支持快捷模式与代码模式。

A. 快捷模式：支持从表/视图导入、从字段标准导入、从派生指标导入。

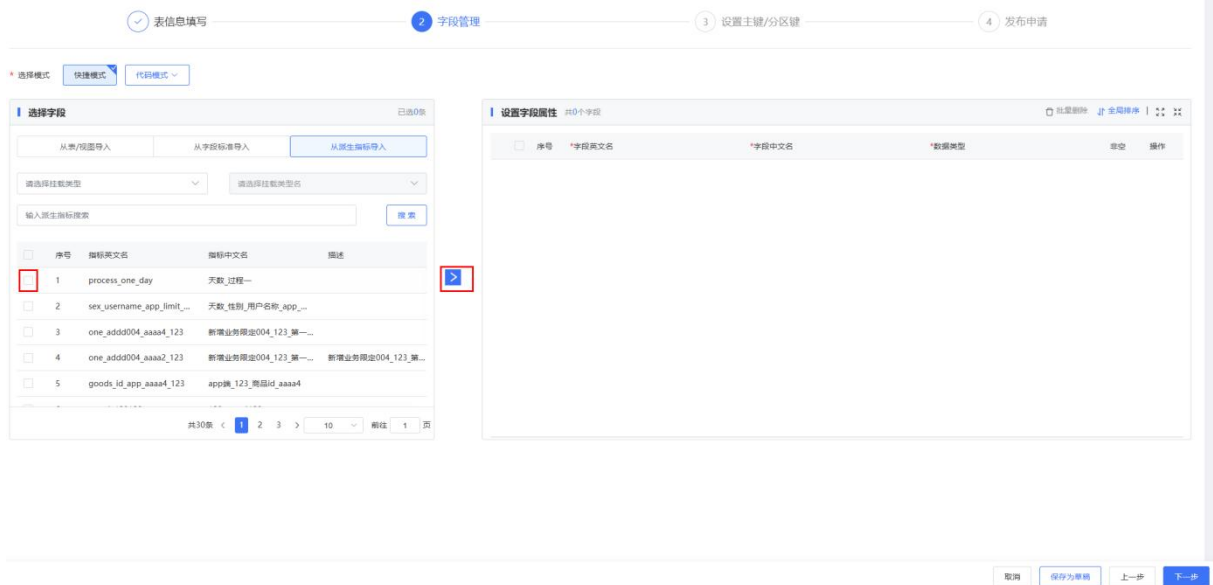
如下截图为从表/视图导入：



如下截图为从字段标准导入：



如下截图为从派生指标导入：



B. 代码模式导入：支持解析 sql 内容导入字段。有两种方式：



其一：保留快捷模式字段，将快捷模式所选的字段自动生成 sql 带到代码模式的输入框；

其二，清空快捷模式字段，不将快捷模式所选的字段带到代码模式的输入框。

选择其中一种代码模式，进入 SQL 模式弹窗。

SQL模式

请保证SQL里的“商品名称”与上一步中选择的库及配置的名称一致：【default.stg_other_cate_d】 请复制
页面上编辑中的参数，其他参数不修改。

```
1 CREATE TABLE default.`stg_other_cate_d` (  
2   `code_id` bigint NOT NULL COMMENT '商品id',  
3   `order_id` bigint COMMENT '订单id',  
4   `dt` string COMMENT '分区键',  
5 )  
6 COMMENT '其他表属性'  
7 TABLEPROPERTIES (  
8   'table.chinese.name' = '其他表属性'  
9 )
```

取消 确认

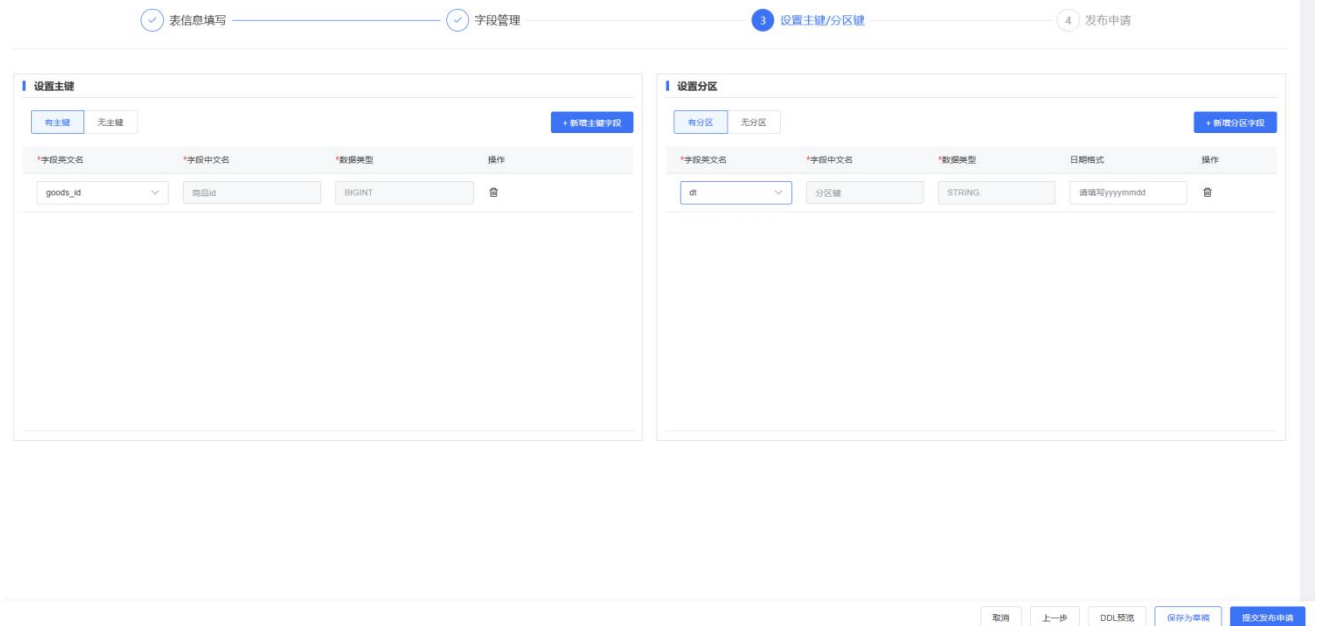
点击“确认”，则进入“设置字段属性”。

● 设置字段属性。

- 字段英文名、字段中文名、数据类型、描述口径。
- 设置非空属性、关联字段标准、标准值域、度量单位。
- 关联类型可选择维度属性、原子指标、派生指标，并选择所关联的内容。

3. 设置主键和分区键。

字段配置完成后，可以根据实际场景配置 1 个或多个主键字段；针对 Hive 表，建议您配置分区字段并确定分区字段格式。



4. 发布申请。

设置完成后，即可提交发布申请。


后续步骤

- 发布申请生成审批单，需审批人员在“我的审批”进行审批，审批通过后，系统自动发布生成物理表。系统生成后，可在建模“生产”tab，查看到该表的最新版本。
- 发布成功后，可在“ workflow 开发”，基于该表进行相关的数据开发操作。

4.3.6.1.2 查看表详情

本章节主要介绍如何查看表详情。

功能入口

- 单击左上方的  图标，选择全部产品 > 数据建模 > 维度建模 > 正向建模，进入正向建模页面；
- 点击表模型列表里蓝色的表英文名，可进入表详情。

表模型列表

请选择表物理类型 | 请选择数据源 | 请输入表英文名,表中文名,表描述搜索

default.dim_trd_order_info_1029_df 订单信息表1029

数据源: engine 0613 HIVE | 数仓分层: 维度层(dim) | 数据域: 交易域 | 表类型: 普通维度表 | 负责人: 907197043@qq.com

描述:

default.hive_dim_zzf_0923_001 新建hive_dim_zzf_0923_001

数据源: engine 0613 HIVE | 数仓分层: 维度层(dim) | 数据域: 数据一 | 表类型: 普通维度表 | 负责人: zhengf0613@chinateleco...

描述: 新建hive_dim_zzf_0613_002

查看表详情

表详情主要包括表信息、字段信息、主键和分区信息。

表模型列表 dim_trd_o... x

表信息 已发布 版本信息

* 英文名	dim_trd_order_info_1029_df	* 表中文名	订单信息表1029
* 表别名(非显示名)	--	* 数据中名	default
* 数据源	engine_0613_HIVE	* 表物理类型	Hive Table
* 表属性	内嵌表	* 表位置	Location
* 数仓分层	维度层(dim)	* 表类型	普通维度表
* 挂载类型	数据域	* 挂载别名	交易域
* 表名主键	order_info_1029	* 表名规范	dim_数据源缩写_表名主键_存储策略 (保留0)
存储策略	日总量(df)	* 表描述	--
维度名称	--	* 负责人	--

字段信息

序号	字段英文名	数据类型	字段中文名	描述或口径	是否必填	字段标准	标准描述	度量单位	关联类型	关联内容
1	order_id	BIGINT	订单id		否					
2	order_type	SMALLINT	订单类型		否					
3	dt	STRING	分区键		否					

主键

序号	字段英文名	数据类型	字段中文名	描述或口径
1	order_id	BIGINT	订单id	


分区键

序号	字段英文名	数据类型	字段中文名	日期格式	描述或口径
1	dt	STRING	分区键		

4.3.6.1.3 管理表模型

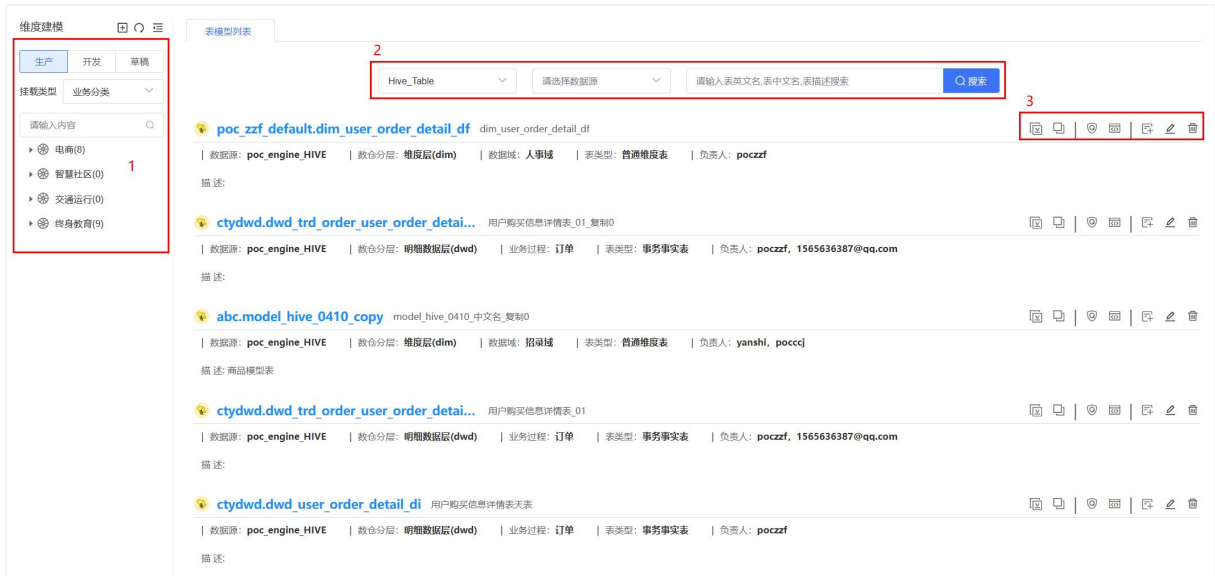
管理表模型主要是包含筛选、搜索和对于表模型的各种操作。

功能入口

- 单击左上方的  图标, 选择全部产品 > 数据建模 > 维度建模 > 正向建模, 进入正向建模页面;
- 进入表模型列表: 点击表模型列表进入。

功能说明

如下方截图红色框圈出的操作。



左抽：筛选与搜索。

- 生产、开发、草稿 tab 可切换查看。
 - 生产：代表生产环境；展示新增或修改的表模型，提交发布申请，通过申请且系统执行，新增或修改的信息发布到生产环境的表模型。
 - 开发：代表开发环境；新增表模型时，提交发布申请，但未发布到生产环境；或提交发布申请，被驳回修改过程中的列表；或已存在生产版本，修改表模型时，选择暂存，可以在开发环境查看开发过程中的表模型。
 - 草稿：新增表模型时，选择保存到草稿的表模型。
- 挂载类型切换：比如业务分类、数据域、业务过程、数据集市、分析主题
- 挂载类型名与数仓分层搜索：搜索具体的业务分类名、数据域名、业务过程名、数据集市名、分析主题名；具体的数仓分层。
- 挂载类型名树形筛选：可选择具体的挂载类型名进行筛选。


- 数仓分层筛选：可选择具体的数仓分层进行筛选。

横栏：数据源及数据库筛选、搜索。

列表操作。

- 版本信息：点击查看版本列表、选择 2 个版本进行版本对比；只有曾经发布成功的表模型才有版本信息。

进入方式 1：从生产环境的列表，点击  图标进入版本列表；进入列表后，选择两个版本进行对比；

进入方式 2：如果曾经发布过的版本，在修改过程中，也可以从开发环境的列表，点击  图标进入版本列表；进入列表后，选择 1 个版本跟当前修改过程的版本进行对比。

如下举例是从生产环境的列表，点击  图标进入版本列表。

← 返回 | 版本列表 版本对比

<input type="checkbox"/>	版本	发布时间	提交人	审批人	操作
<input type="checkbox"/>	当前(v.3)	2024-05-27 14:22:19	datawingsdev_ops@chinatelecom.cn	datawingsdev_ops@chinatelecom.cn	查看
<input type="checkbox"/>	v.2	2024-05-27 14:20:40	datawingsdev_ops@chinatelecom.cn	datawingsdev_ops@chinatelecom.cn	查看
<input type="checkbox"/>	v.1	2024-05-27 14:12:28	datawingsdev_ops@chinatelecom.cn	datawingsdev_ops@chinatelecom.cn	查看

选择两个版本进行对比，点击“版本对比”控件。

← 返回 | 版本列表 版本对比

<input type="checkbox"/>	版本	发布时间	提交人	审批人	操作
<input checked="" type="checkbox"/>	当前(v.3)	2024-05-27 14:22:19	datawingsdev_ops@chinatelecom.cn	datawingsdev_ops@chinatelecom.cn	查看
<input type="checkbox"/>	v.2	2024-05-27 14:20:40	datawingsdev_ops@chinatelecom.cn	datawingsdev_ops@chinatelecom.cn	查看
<input checked="" type="checkbox"/>	v.1	2024-05-27 14:12:28	datawingsdev_ops@chinatelecom.cn	datawingsdev_ops@chinatelecom.cn	查看

查看对比情况，含表信息、字段信息、主键与分区键。如下例子：表信息的描述有修改，用蓝色虚线圈出。

v.1	VS	v.3
dim_trd_order_info_df	*表英文名	dim_trd_order_info_df
订单信息表	*表中文名	订单信息表
...	表别名(表显示名)	...
bigdata0522_HIVE	*数据源	bigdata0522_HIVE
default	*数据库名	default
HIVE	*表物理类型	HIVE
内部表	*表属性	内部表
...	Location	...
维度层(dim)	*数仓分层	维度层(dim)
普通维度表	表类型	普通维度表
数据域	*挂载类型	数据域
业务域	*挂载类型名	业务域
order_info	*表名主体	order_info
...	表名规范	...
...	存储策略	...
...
...	维度名称	...
...	描述	订单信息表, 包含订单id等信息
datawingsdf	*负责人	datawingsdf

字段信息：order_id 数据类型有修改，用蓝色虚线圈出，新增字段 order_type，用绿色点标注。

序号	字段英文名	数据类型	字段中文名	描述或口径	是否非空	字段标准	标准值域	度量单位	关联类型	关联内容
1	order_id	BIGINT	订单ID		否				维度属性	order_id(订单ID)
		STRING	订单ID		否				维度属性	order_id(订单ID)
2	order_sn	STRING	订单SN		否				维度属性	order_sn(订单SN)
● 3	order_type	TINYINT	订单类型		否				维度属性	order_type(订单...
4	dt	STRING	日分区		否				维度属性	dt(日分区)

主键与分区键：主键删除 order_id，用红线划掉，新增主键 order_sn，用绿色点标注。

序号	字段英文名	数据类型	字段中文名	描述或口径
● 1	order_sn	STRING	订单SN	
2	order_id	BIGINT	订单ID	

序号	字段英文名	数据类型	字段中文名	日期格式	描述或口径
1	dt	STRING	日分区		

- 克隆表：快速复制已有表模型，如果您已创建完成部分表，后续希望创建内容一致的表，可使用克隆表快速创建。表参数配置与新建表参数配置一致。表英文名自动增加_copy，可

修改表英文名。

- 跳转到数据质量：跳转到数据质量，并自动带上相关参数。



- 跳转到数据开发：跳转到数据开发界面。
- 新增字段：新增字段，可通过此入口。
- 编辑：修改全部、修改表信息、新增/修改字段、设置主键；根据所需修改内容，选择对应选项。
- 删除：二次确认，进行删除。注意此处是会删除逻辑模型和物理模型，需慎重操作。

4.3.6.2 逆向建模

本章节主要介绍逆向建模操作步骤。

功能概述

当用户的数仓中存在大量的物理表，没有进行统一的分层划域管理，希望通过 DataWings 的数据建模进行统一管理，则可通过“逆向建模”功能，将已有物理表反向建模至 DataWings 的维度建模中。该功能帮助用户无需再次执行建模操作即可快速创建模型，节省了大量的时间成本。

前提条件

- 用户的数仓中存在大量的物理表，需进行统一的分层划域管理；
- 集群：集群同步/注册成功，且含有 Hive 组件；

- 引擎管理：将集群与调度资源组进行关联，且引擎实例状态为成功；
- 数据源管理：Hive 内置数据源，成功自动生成；
- 数据库：Hive 内置数据源下，已创建数据库；
- 规划模板：已规划好各数仓要素，且配置好维度层模型与数仓要素的关系；
- 选择模板：在首次进入数据建模时，租户管理员选定某个规划模板；
- 业务分类：已配置好数仓要素要挂载的业务分类；
- 数仓分层：已创建对应的数仓分层；
- 数仓要素：配置维度建模需要挂载的数仓要素；
- 元数据：数据源、数据库和表模型等元数据采集成功；
- 表名规范（非必须）：已定义表名规范。

使用限制

仅支持对 DataWings 内置 Hive 数据源的表进行逆向建模。

操作步骤

逆向建模使用流程如下：

① 逆向策略

确定逆向建模范围：选择所属数据源、数据库、通过【表名匹配规则】匹配的结果

预设逆向建模后的模型所属的数仓分层、所属的数仓要素、所属的表名规范

确定逆向建模后的执行方式：本期刊支持增量更新（也即是匹配到的模型，如果之前已经存在于维度建模的生产环境，则跳过。）

② 系统解析并匹配模型

根据配置的建模策略进行解析，匹配出待创建的模型。

③ 确认模型

初步生成的表模型可能会存在偏差，需要根据业务需求进行模型信息的调整。例如调整表类型，以及所属的数仓分层、数据域等信息。

④ 提交发布申请

提交发布申请；

- 1) 生成导入单，导入成功的记录，可通过站内信、逆向建模列表了解导入结果；
- 2) 自动提交到审批单，审批成功且系统执行（可通过全部产品 > 系统管理 > 流程中心 > 我的申请 了解审批和执行结果）成功后，在维度建模列表显示。

1. 逆向策略。

< 逆向建模列表

1 逆向策略 ————— 2 确定模型 ————— 3 发布申请

确定建模范围 ①数据源、数据库和表模型列表小时级别更新，存在一定延迟。

* 数据源 test_engine_reconfig_3_HIVE

* 数据库 default

表名匹配规则 模糊匹配 精确匹配

test

预设逆向建模后所属

* 数仓分层 维度层(dim)

* 挂载类型 数据域

* 数据域 交易

表名规范 请选择

确定执行方式 增量更新

取消 预创建模型

配置说明：

- 选择所属数据源、数据库，通过【表名匹配规则】匹配的结果。

模糊匹配：输入关键词，即可匹配到名称中包含该关键词的所有表。

精准匹配：输入的表名称为表的精准全称。多个表名称之间使用英文分号 (;) 分隔，并且英文分号 (;) 后不能添加空格。

若未匹配到任何满足条件的表，则此次逆向建模失败，即不会生成任何模型。

- 预设逆向建模后的模型所属的数仓分层、所属的数仓要素、所属的表名规范；因物理表已存在，逆向建模不可修改表英文名，在此处的表名规范下拉框仅展示启用的弱规则。

- 确定逆向建模后的执行方式：本期仅支持增量更新（也即是匹配到的模型，如果之前已经存在于维度建模的生产环境，则跳过。）

2. 确定模型。

逆向建模列表

逆向策略 2 确定模型 3 发布申请

确定建模范围 初步生成的表模型可能会存在偏差, 可根据业务需求进行模型信息的调整

序号	*数据源名	*表英文名	表名规范	表别名(表显示名) ①	*表中文名 ②	操作
1	test_engine_1_HIVE	default.test0109	请选择	可按表名规范, 填写规范表名, 用于显示	可由中文、英文字母、下划线 C...	删除
		*数仓分层 维度...	*表类型 请选择	*挂载类型 数据域	*数据域 交易	*负责人 zzf1 x
2	test_engine_1_HIVE	default.test0116	请选择	可按表名规范, 填写规范表名, 用于显示	可由中文、英文字母、下划线 C...	删除
		*数仓分层 维度...	*表类型 请选择	*挂载类型 数据域	*数据域 交易	*负责人 zzf1 x

共2条 < 1 > 10 前往 1 页

上一步 下一步

配置说明:


- 表英文名: 是存量物理表名, 不可修改。
- 表名规范: 可按需选择。因物理表已存在, 逆向建模不可修改表英文名, 在此处的表名规范下拉框仅展示启用的弱规则。
- 表别名(表显示名): 可点击 ① 进行查看填写规范。
- 表中文名: 必填, 默认获取建表语句的表 comment; 可修改, 点击 ② 查看填写规范。
- 数仓分层、挂载类型、挂载类型名按照“步骤 1”批量预设展示, 可以在“步骤 2”进行二次调整。
- 表类型: 必选, 根据所选的数仓分层不同, 有不同的选项, 按需选择。
- 负责人: 默认为本次逆向建模的创建者。

3. 发布申请。

生成导入单。

< 逆向建模列表

逆向策略 ———— 确定模型 ———— 3 发布申请



导入单生成

导入单: I1-20230908155900-191

- 「逆向建模」使用异步导入方式，导入所需时长根据导入量大小而定，请关注「站内信」或者逆向建模列表了解执行结果。
- 由于逆向建模需审核后才可以发布，故导入成功后需走审批流程。与本导入单相关的审批列表可点击站内信或逆向建模列表“跳转到我的申请”查看。

[确认](#)

后续步骤

- 在维度建模 > 逆向建模的“逆向建模列表”筛选和查看导入情况。

数据建模 / 逆向建模

逆向建模列表

请输入导入单号 请选择当前状态 选择操作时间

请输入操作人 [刷新查询](#)

[+ 逆向建模](#)

注意事项： 逆向建模需要审批；可点击“跳转到我的申请-跳转”查看审批与执行结果。

导入单号	跳转到我的申请	操作时间	操作人	执行完成时间	当前导入状态
1 I1-20230908155900-191	跳转	2023-09-08 15:59:00	gzw_tenantry_2	2023-09-08 15:59:00	全部失败
2 I1-20230829162529-190	跳转	2023-08-29 16:25:29	zzf0828	2023-08-29 16:25:32	全部成功
3 I1-20230825112512-188	跳转	2023-08-25 11:25:13	zzf0821	2023-08-25 11:25:18	全部成功
4 I1-20230824192131-187	跳转	2023-08-24 19:21:32	zzf0821	2023-08-24 19:21:35	全部成功
5 I1-20230824191231-186	跳转	2023-08-24 19:12:31	zzf0821	2023-08-24 19:12:35	全部成功
6 I1-20230824191113-185	跳转	2023-08-24 19:11:14	zzf0821	2023-08-24 19:11:14	全部失败
7 I1-20230824190934-184	跳转	2023-08-24 19:09:34	zzf0821	2023-08-24 19:09:37	全部成功
8 I1-20230824160921-183	跳转	2023-08-24 16:09:21	zzf0821	2023-08-24 16:09:22	全部失败

- 导入失败的表，点击蓝色导入单号，可查看导入失败原因。

逆向建模列... I1-202309... ×

I1-20230908155900-191:2023-09-08_逆向建模 刷新查询

导入结果: 总计1条, 导入成功0条, 导入失败1条; 导入失败的记录, 请查看原因, 重新调整后, 再导入。

表名	表别名(表显示名)	表类型	导入结果	导入失败原因	审批单号
1	engine0622.test_database_1.shl_zj_dim_test	普通维度表	失败	表中文名称重复	

共1条 < 1 > 前往 1 页

- 导入成功的记录, 生成审批单, 点击“跳转”则可跳转到“我的申请”查看审批和执行情况。
- 审批人员在“我的审批”中进行审批, 审批通过并执行成功后, 可在维度建模“生产”tab, 按照分层划域管理表模型, 详见“管理表模型”。

4.4 数据集成

4.4.1 数据集成概述

提供简单易用的迁移能力和多种数据源到数据仓库的集成能力, 降低了客户数据迁移和集成的复杂性, 有效地提高数据迁移和集成的效率。

支持的同步类型

同步类型	来源	目标
离线同步	MySQL	Hive
离线同步	MySQL	Doris
离线同步	Oracle	Hive

同步类型	来源	目标
离线同步	Oracle	Doris
离线同步	Oracle	HDFS
离线同步	SQL Server	Hive
离线同步	PostgreSQL	Hive
离线同步	PostgreSQL	Doris
离线同步	Greenplum	MySQL
离线同步	Greenplum	Hive
离线同步	Hive	MySQL
离线同步	Hive	Hive
离线同步	TeleDB	MySQL
离线同步	TeleDB	Hive
离线同步	TeleDB	HDFS

同步类型	来源	目标
离线同步	TeleDB	Doris
离线同步	TelePG	MySQL
离线同步	TelePG	Hive
实时同步	MySQL	Doris
实时同步	MySQL	Hudi
实时同步	MySQL	Iceberg
实时同步	PostgreSQL	Doris
实时同步	PostgreSQL	Hudi
实时同步	PostgreSQL	Iceberg
实时同步	TeleDB	Doris
实时同步	TeleDB	Hudi
实时同步	TeleDB	Iceberg

同步类型	来源	目标
实时同步	TelePG	Doris
实时同步	TelePG	Hudi
实时同步	TelePG	Iceberg

4.4.2 集成集群管理

集成集群管理功能模块提供集群资源注册以及资源组注册能力，支持对集群以及资源组的状态、启停、连接测试等进行管理。

4.4.2.1 资源组

资源组用于添加已购买的翼 MapReduce 集群资源，该资源用于运行数据同步任务。

进入页面

单击左上方的图标，选择全部产品 > 数据集成 > 集成集群管理 > 资源组

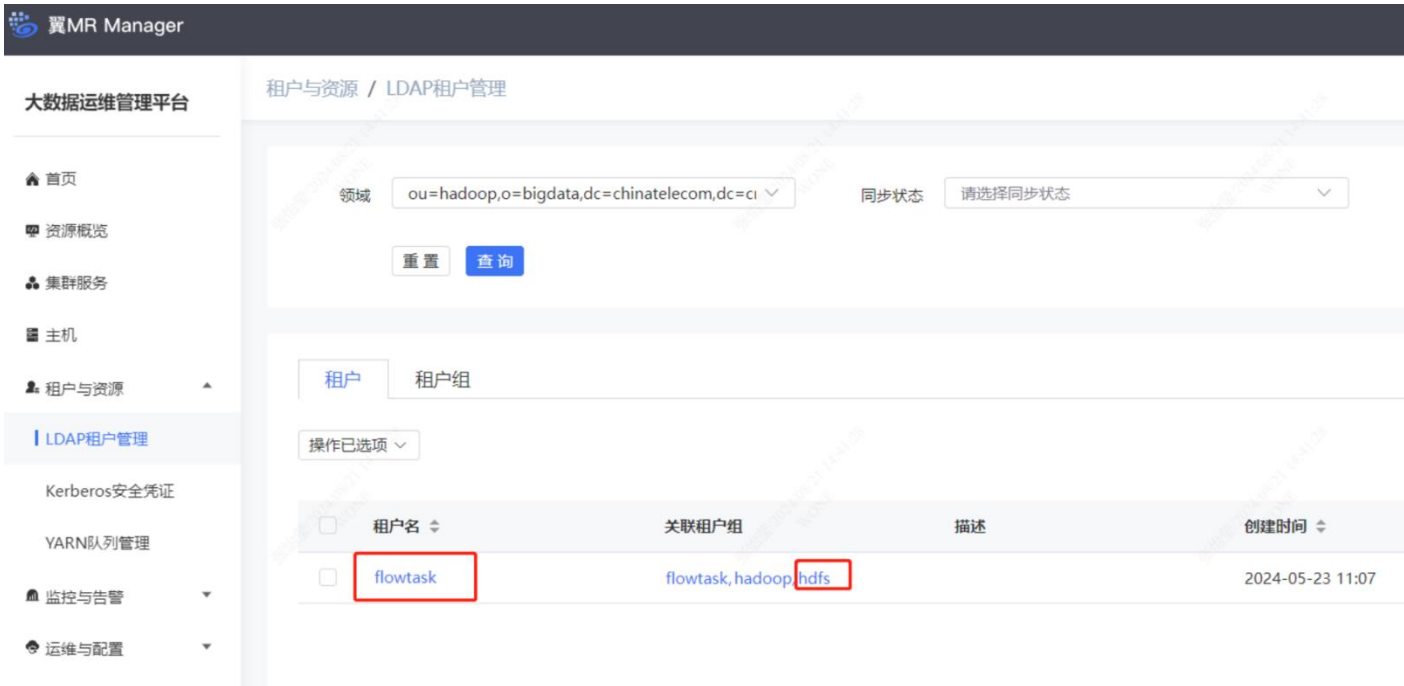
新增资源组

获取翼 MR 集群 Keytab 配置文件。

登录翼 MR 控制台，进入用户管理，若用户管理为空，则需要先进行 IAM 同步，保证用户管理中有当前天翼云账号对应的用户。



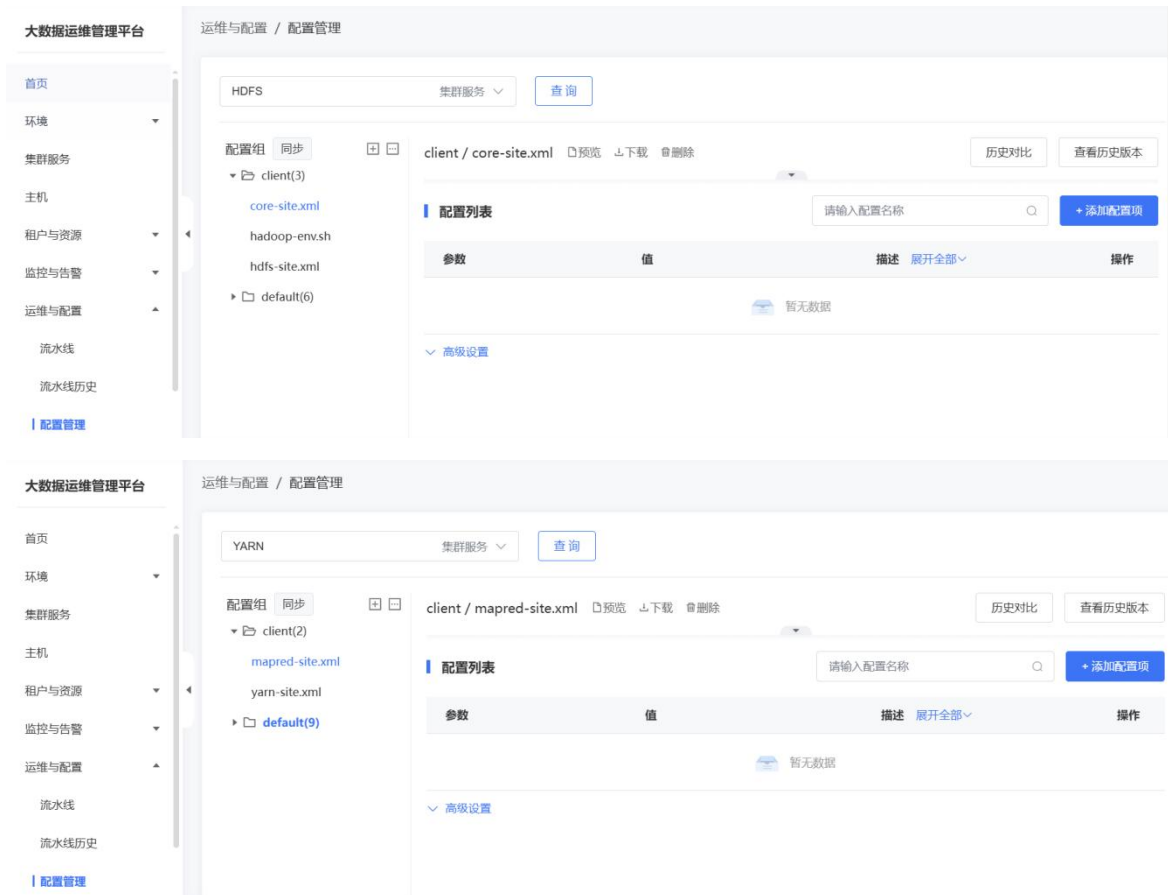
进入到翼 MR Manager 以后, 点击菜单 “租户与资源 > LDAP 租户管理 > 租户与资源概述”。点击"+新增", 新增租户并设置该租户在 hdfs 组中。



点击菜单 “租户与资源 > Kerberos 安全凭证 > 租户与资源概述”，点击"+新增" 新增 Principal, 下载 keytab 文件。



点击菜单“运维与配置 > 配置管理-配置管理概述”，选择 HDFS 和 YARN 集群，下载 yarn-site.xml、core-site.xml、hdfs-site.xml 以及 krb5.conf，加上 keytab 文件，总共有 5 个配置文件。



登录 DataWings，选择全部产品> 数据集成 > 集成集群管理 > 资源组，点击【新增资源组】按钮。



填写基本信息，信息填写完毕后点击【确认】按钮即可创建资源组。

* 资源名称

* 资源类型

* Hadoop 配置文件

① 指定配置文件路径(末尾无/), 需包含以下文件: core-site.xml,hdfs-site.xml,yarn-site.xml

* Kerberos账号

* keytab配置文件

① 指定配置文件路径(末尾无/), 需包含以下文件: krb5.conf,keytab

* 工作空间

指定HDFS工作空间路径(末尾无/),用于存储数据集成用到的组件或代理服务。

取消

确定

基本信息	说明
资源名称	必填项，输入资源名称。
资源类型	必填项，选择资源类型，支持 Hadoop On Yarn。
Hadoop 配置文件	必填项，上传 Hadoop 配置文件，包含 core-site.xml,hdfs-site.xml,yarn-site.xml。
Kerberos 账号	必填项，输入 Kerberos 账号，需包含工作空间写入读取权限。
keytab 配置文件	必填项，上传 keytab 配置文件，包含 krb5.conf,keytab。
工作空间	必填项，输入 HDFS 工作空间地址，用于上传数据集成服务运行同步任务的相关 jar 包和配置文件。输入格式为 hdfs://ctyunns/xxx。

查看资源组

注册资源组后，用户可以查看现有资源组的详情信息。

1. 选择注册好的资源组。
2. 点击列表右侧"详情"按钮。



查看资源组

资源名称	hadoop
资源类型	Hadoop On Yarn
Hadoop 配置文件	core-site.xml,hdfs-site.xml,yarn-site.xml
Kerberos 账号	hive@BIGDATA.CHINATELECOM.CN
keytab 配置文件	krb5.conf,hive.keytab
工作空间	hdfs://ctyunns/user/flowtask/workspace

关闭

编辑资源组

注册资源组后，您可以编辑现有资源组的基本信息。

1. 选择注册好的资源组。
2. 点击列表右侧"编辑"按钮。

编辑资源组

* 资源名称	<input type="text" value="hadoop"/>
* 资源类型	<input type="text" value="Hadoop On Yarn"/>
* Hadoop 配置文件	<input type="button" value="上传"/>
	<input type="checkbox"/> core-site.xml <input checked="" type="checkbox"/>
	<input type="checkbox"/> hdfs-site.xml <input checked="" type="checkbox"/>
	<input type="checkbox"/> yarn-site.xml <input checked="" type="checkbox"/>
	<small>① 指定配置文件路径（末尾无/），需包含以下文件：core-site.xml,hdfs-site.xml,yarn-site.xml</small>
* Kerberos账号	<input type="text" value="hive@BIGDATA.CHINATELECOM.CN"/>
* keytab配置文件	<input type="button" value="上传"/>
	<input type="checkbox"/> krb5.conf <input checked="" type="checkbox"/>
	<input type="checkbox"/> hive.keytab <input checked="" type="checkbox"/>

删除资源组


当该资源组下没有关联的集群时，您可以删除不再需要的资源组。

1. 选择需要删除的资源组卡片。
2. 点击列表右侧“删除”按钮。

4.4.2.2 集群注册

集群注册用于注册数据集成服务集群。

进入页面

单击左上方的图标，选择全部产品 > 数据集成 > 集成集群管理 > 集群注册

集成集群管理

集群注册

资源组

数据集成 / 集群管理 / 集群注册

flinkcdc		测试连接	详情	禁用	编辑	删除
集群类型	FLINK_ON_YARN	执行模式	Yarn Application	创建时间	2024-04-20 18:39:27	
集群状态	有效	所属资源	emr_poc	更新时间	2024-04-20 18:55:38	
		测试连接状态	● 连接成功			

+ 新增集群

新增集群

在集群注册页面，单击新增集群。

flink		测试连接	详情	禁用	编辑	删除
集群类型	FLINK_ON_YARN	执行模式	Yarn Application	创建时间	2024-07-25 17:57:26	
集群状态	有效	所属资源	hadoop	更新时间	2024-08-08 12:50:16	
		测试连接状态	● 连接成功			

+ 新增集群

填写基本信息。

基本信息

* 集群类型

* 组件版本

* 集群名称

* 所属资源 +

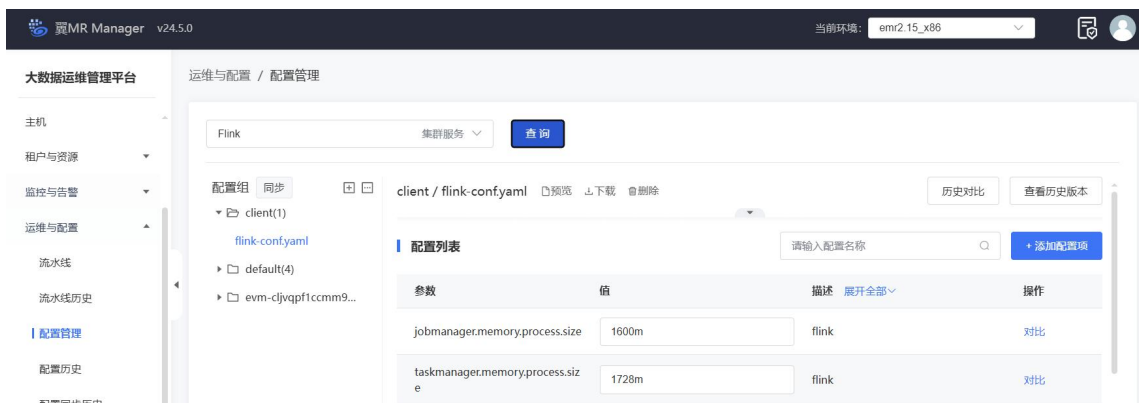
* 执行模式

基本信息	说明
集群类型	必填项，支持 FLINK ON YARN 类型。
组件版本	必填项，选择部署的组件版本，当前支持 Flink1.16.0。
集群名称	必填项，输入集群名称。

基本信息	说明
所属资源	必填项，选择已配置的翼 MR 资源。
执行模式	必填项，支持 Yarn Application 类型。

上传 Flink 配置文件。

登录翼 MR Manager>运维与配置 >配置管理-配置管理概述, 选择 Flink 集群, 下载 Flink-conf.yaml 文件。



将配置文件上传到集群注册。

Flink

* flink-conf.yaml

yarn代理地址

填写 yarn 代理地址，可以通过该地址访问在 yarn 上运行的任务运行详情页面。

登录翼 MR 控制台>节点管理，找到有安装 ResourceManager 的节点，点击更多绑定弹性 IP。

< flowtask-emr3

基础信息 节点管理 翼MR Manager 用户权限 操作日志 访问链接与端口

请选择节点状态 查询 重置 刷新

节点组名称	节点组类型	付费类型	角色实例		规格	操作	
▼ master	MASTER	包年包月	<ul style="list-style-type: none"> NameNode JournalNode ZKFC HDFSClient ResourceManager YARNClient ZKServer KDCSlave Slapd HiveServer2 HiveClient Spark-HistoryServer SparkClient Doris-FE 		(s7.2xlarge.4) 8 vCPU, 32GB	配置升级	
节点名称	节点状态	内网IP	外网IP	角	规格	磁盘信息	操作
bigdata-vm-1716428813-x	运行中	192.168.0.48	-	8	(s7.2xlarge.4) 8 vCPU, 32...	普通IO云硬盘 200GB * 1 普通IO云硬盘 200GB * 2	远程连接 更多
bigdata-vm-1716428813-l	运行中	192.168.0.52	-	1	(s7.2xlarge.4) 8 vCPU, 32...	普通IO云硬盘 200GB * 1 普通IO云硬盘 200GB * 2	远程连接 更多
bigdata-vm-1716428813-s	运行中	192.168.0.53	121.237.178.130	14个角色实例	(s7.2xlarge.4) 8 vCPU, 32...	普通IO云硬盘 200GB * 1 普通IO云硬盘 200GB * 2	远程连接 更多

地址格式为：`http://{绑定的弹性 ip}:{ResourceManager 端口}/proxy`。

填写安装路径，点击"安装"部署集成所需组件（Flink cdc 和 Seatunnel）到指定的 HDFS 路径下，注意您需有该路径的写入权限。

Flink CDC

* 版本选择

SeaTunnel

* 版本选择

* 安装路径 安装

填写高级配置，若不填将使用默认配置。信息填写完毕后点击【确认】按钮即可新增集群。

高级配置

高级配置

* 执行队列

Task Slot并发数

1

JobManager内存配置

MB



TaskManager内存配置

MB



其他配置

key=value,多个配置以逗号分隔

高级配置	说明
执行队列	yarn.application.queue
Task Slot 并发数	taskmanager.numberOfTaskSlots
jobmanager 内存配置	jobmanager.memory.process.size
taskmanager 内存配置	taskmanager.memory.process.size
其他配置	其他 Flink 运行配置，多个配置以逗号分隔。

编辑集群

注册集群后，您可以编辑现有集群的基本信息。

1. 选择注册好的集群。
2. 点击集群右侧"编辑"按钮。

编辑集群 ×

基本信息

* 集群类型	FLINK_ON_YARN	▼
* 组件版本	1.16.0	▼
* 集群名称	flink	
* 所属资源	hadoop	▼ +
* 执行模式	Yarn Application	▼

Flink

* flink-conf.yaml	上传
flink-conf.yaml	👍

删除集群

当该集群下没有关联的同步任务时，您可以删除不再需要的集群。

1. 选择需要删除的集群卡片。
2. 点击集群右侧"删除"按钮。


4.4.3 同步任务

任务配置模块用于创建、管理数据同步相关的任务，支持离线同步、实时同步和自定义任务三类数据同步任务。

4.4.3.1 任务创建

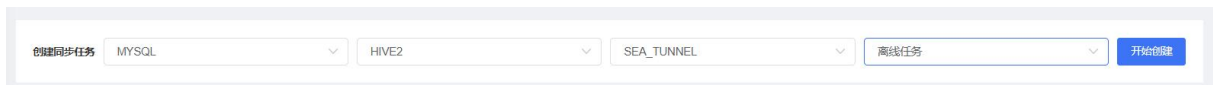
任务创建功能用于创建离线、实时或自定义同步任务，将数据从来源数据源同步到目标数据源中。

进入页面

单击左上方的图标，选择全部产品 > 数据集成 > 任务配置 > 任务管理

新建离线任务

在任务管理页面，选择数据来源类型、目标源类型、任务引擎（SEA_TUNNEL）、同步方式(离线任务)，点击"开始创建"配置同步任务。



创建同步任务

MYSQL HIVE2 SEA_TUNNEL 离线任务 开始创建

填写基本信息

基本信息

*任务名称 *任务分组

任务描述 0/300

基本信息	说明
任务名称	数据接入任务的名称。不可重复，支持中文、英文、数字、下划线（_），且只能以英文或中文开头，1~64 个字符。
任务分组	任务所属分组。
描述	对任务的说明。

选择数据来源

完成基本信息填写后，首先需要在数据来源配置离线同步任务节点的读取端数据源，以及需要同步的表等信息。

此处以 MySQL 为例：

任务表单

数据来源-Mysql

* 来源数据源

* 来源数据库

* 来源表名

来源数据筛选条件

字段名	字段类型	过滤规则	过滤参数	操作
-----	------	------	------	----

暂无数据

+ 添加一行

* 一次从结果集中提取的结果行数

基本信息	说明
来源数据源	必填项，选择来源数据源名称
来源数据库	必填项，内容为所选数据源配置的数据库。
来源表名	必填项，所选数据库的已有数据表。
过滤条件	非必填项，按列设置过滤规则，可添加一或多组条件，条件默认为 AND 关系。
一次从结果集中提取的结果行数	必填项，一次性批量提交的记录数大小，该值可以极大减少数据同步系统与 Hive 的网络交互次数，并提升整体吞吐量。如果该值设置过大，会导致数据同步运行进程 OOM 异常。单次默认 2000 行。

选择数据去向

完成数据来源的配置后，可以在下方配置数据去向的数据源，以及需要写入的表信息等。

此处以 Hive 为例：

数据去向-Hive

* 目标数据源

* 目标数据库

* 目标表名

* 元数据存储URI

基本信息	说明
目标数据源	必填项，选择数据源名称。
目标数据库	必填项，选择数据库。
目标表名	必填项，选择数据表。
元数据存储 URI	必填项，Hive 元数据连接地址。

Hive 元数据 URI 可登录翼 MR Manager 获取。

登录翼 MR Manager>运维与配置 > 配置管理，选择 Hive 集群、点击 hive-site.xml 获取 hive.metastore.uris 参数值。

Hive 集群服务 ▾ 查询

配置组 同步 + -

- ▶ client(2)
- ▼ default(5)
 - Hive_vars.yaml
 - hive-env.sh
 - hive-log4j2.properties
 - [hive-site.xml](#)
 - *krb5.conf

le	<input type="text" value="/etc/security/keytabs/hive."/>	Kerberos Keytab文件的路径	对比
hive.metastore.kerberos.principal	<input type="text" value="hive/_HOST@BIGDATA.C"/>	MetaStore服务器的服务主体	对比 恢复默认值
hive.metastore.uris	<input type="text" value="thrift://evm-cljvqpf1ccmm9"/>		对比 恢复默认值
hive.server2.authentication.kerberos.keytab	<input type="text" value="/etc/security/keytabs/hive."/>	服务器主体的Kerberos密...	对比
hive.server2.authentication.kerberos.principal	<input type="text" value="hive/_HOST@BIGDATA.C"/>	Kerberos服务器主体	对比 恢复默认值

配置字段映射

在完成数据来源和数据去向的配置后，需要指定数据来源端和去向端的映射关系。支持同名映

射、同行映射以及自定义映射。

字段映射

字段映射

同名映射 同行映射 取消映射

来源表字段	字段类型	目标表字段	字段类型	是否必填	分区字段
ID	BIGINT	id	BIGINT	否	否
COLLATION_NAME	VARCHAR	name	VARCHAR	否	否

共 2 条 < 1 > 10条/页 前往 1 页

seatunnel与linkcdc任务不允许使用新增字段，不允许使用空变量，不允许将源表字段映射至分区字段，否则将会失败

来源表字段信息	说明
同名映射	根据字段名称建立映射关系，首次显示字段时，默认采取同名映射。无法映射的部分，目标表字段按照表中字段顺序依次填入选择框中，来源表字段处保留选择框为空，用户可自行匹配字段。
同行映射	来源表字段和目标表字段均按照表中的字段顺序填入选择框中。显示的行数量以来源表和目標表中字段数量少的为准。
取消映射	取消映射后，用户可自行选择源表字段来对应目标表字段。

填写高级配置

任务并行度是数据同步任务中，最大并行读取或并行写入数据数。并发数影响数据同步的效率，并发设置越高对应资源消耗也越多，由于资源原因或者任务本身特性等原因，实际执行时并发数可能小于等于此值。默认 1，取值范围 1-50。

高级配置

* 任务并行度

选择任务配置

配置任务的调度策略。支持手动执行或设置 cron 表达式调度离线任务。

任务配置

* 调度策略 cron表达式调度 手动执行

设置 cron 表达式

秒分时天月

每一秒钟

每隔 秒钟执行, 从 秒钟开始

具体秒钟数 (可多选)

周期从 到 秒钟

新建实时任务

在任务管理页面，选择数据来源类型、目标源类型、任务引擎（FLINK_CDC）、同步方式(实时任务)，点击"开始创建"配置同步任务。

创建同步任务

填写基本信息

基本信息

* 任务名称 * 任务分组

任务描述 0/300

基本信息	说明
任务名称	数据接入任务的名称。不可重复，支持中文、英文、数字、下划线（_），且只能以英文或中文开头，1~64 个字符。
任务分组	任务所属分组。

基本信息	说明
描述	对任务的说明。

选择数据来源

完成基本信息填写后，首先需要在数据来源侧配置实时同步任务的读取端数据源，以及需要同步的表等信息。

此处以 TeleDB 为例：

任务表单

数据来源-Mysql

* 来源数据源

* 来源数据库

* 接入表名

* serverId

连接到MySQL集群的客户端实际上是一个简化的副本(server)，并且Server ID值在整个复制组中必须是唯一的(即不同于任何主服务器或副本使用的任何其他Server ID)。因此每个同步任务的实例在整个Mysql集群中都必须具有唯一的Server ID。

基本信息	说明
来源数据源	必填项，选择来源数据源名称
来源数据库	必填项，内容为所选数据源配置的数据库。
来源表名	必填项，所选数据库的已有数据表。
serverId	必填项，用于连接到 MySQL 集群的客户端，不可与其他任务的 serverId 重复。

选择数据去向

完成数据来源的配置后，可以在下方选择数据去向的数据源。

此处以 Hudi 为例：

数据去向-Hudi

* 目标数据源

* 目标数据库

* 目标表

基本信息	说明
目标数据源	必填项，选择数据源名称。
目标数据库	必填项，选择目标数据库名称。
目标表	必填项，选择目标表名称。

填写高级配置

任务并行度是指在数据同步任务中的最大并行读取或并行写入的算子数。并发数会影响数据同步的效率；并发设置越高，对应的资源消耗也就越多。但由于资源限制或任务本身的特性等原因，实际执行时的并发数可能会小于或等于设定的值。默认情况下，任务并行度为 1，其取值范围为 1-50。


高级配置

* 任务并行度

4.4.3.2 任务管理

任务管理为您展示创建成功的同步任务，可对任务进行基本的运维管理。

进入页面

单击左上方的图标，选择全部产品 > 数据集成 > 任务配置 > 任务管理

任务列表

任务列表展示您创建成功的同步任务，您可执行任务发布、查看详情、查看任务详情、启停任务以及删除任务等操作。



任务发布

您创建好的同步任务初始状态为“未发布”，需要点击操作栏对应的“发布”按钮，将任务提交至已注册的集群。

任务启停

对于发布状态为“已发布”、任务状态为“已停止”的任务，点击操作栏对应的“启动”按钮，可启动对应任务。

序号	任务名称	任务分组	同步方式	执行组件	同步方向	任务状态	发布状态	修改时间	操作
1	jctest_ods	test	离线任务	SEA_TUNNEL	↔ HIVE2 -> ↔ HIVE2	● 结束	已发布	2024-08-09 14:22:22	▲ 发布 ● 启动 @ 任务详情 更多 ▾
2	process_test	test	离线任务	SEA_TUNNEL	↔ HIVE2 -> ↔ HIVE2	● 结束	已发布	2024-08-15 14:14:59	▲ 发布 ● 启动 @ 任务详情 更多 ▾
3	test	test0808	离线任务	SEA_TUNNEL	↖、MYSQL -> ↔ HIVE2	● 结束	已发布	2024-08-09 10:33:49	▲ 发布 ● 启动 @ 任务详情 更多 ▾
4	mysql2dori...	test	实时任务	FLINK_CDC	↖、MYSQL -> ▶ DORIS	● 运行中	已发布	2024-08-12 14:55:46	▲ 发布 ● 停止 @ 任务详情 更多 ▾

对于任务状态为“运行中”的任务，点击操作栏对应的“停止”按钮，将弹出停止提示弹窗，点击“确定”按钮将停止对应任务。

提示

×

 操作将停止 mysql2doris0730_2, 是否确认停止?

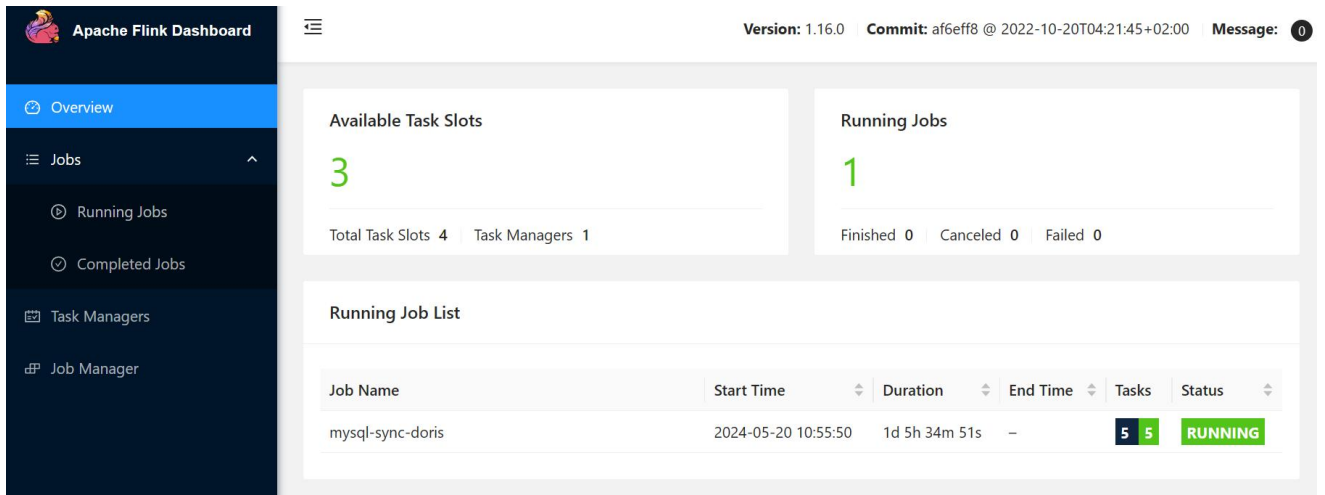
取消

确定

运行详情

对于任务状态为“运行中”的任务，点击操作栏对应的“任务详情”按钮，可查看 Flink Dashboard

任务执行详情。



The screenshot shows the Apache Flink Dashboard interface. The top navigation bar includes the Apache Flink logo, the text 'Apache Flink Dashboard', and a hamburger menu icon. On the right side of the top bar, it displays 'Version: 1.16.0', 'Commit: af6eff8 @ 2022-10-20T04:21:45+02:00', and 'Message: 0'. The left sidebar contains a menu with 'Overview' (selected), 'Jobs', 'Running Jobs', 'Completed Jobs', 'Task Managers', and 'Job Manager'. The main content area is divided into three sections: 'Available Task Slots' showing 3 slots (Total Task Slots 4, Task Managers 1), 'Running Jobs' showing 1 job (Finished 0, Canceled 0, Failed 0), and 'Running Job List'. The 'Running Job List' table has columns for Job Name, Start Time, Duration, End Time, Tasks, and Status. A single job 'mysql-sync-doris' is listed with a start time of 2024-05-20 10:55:50, a duration of 1d 5h 34m 51s, and a status of 'RUNNING' with 5 tasks.

Job Name	Start Time	Duration	End Time	Tasks	Status
mysql-sync-doris	2024-05-20 10:55:50	1d 5h 34m 51s	-	5 / 5	RUNNING

注意事项

- 1.仅状态为“运行中”的任务可查看“任务详情”。
- 2.使用该功能需在新增/编辑集群时，填写 yarn 代理地址。

编辑集群



* 集群名称

* 所属资源

* 执行模式

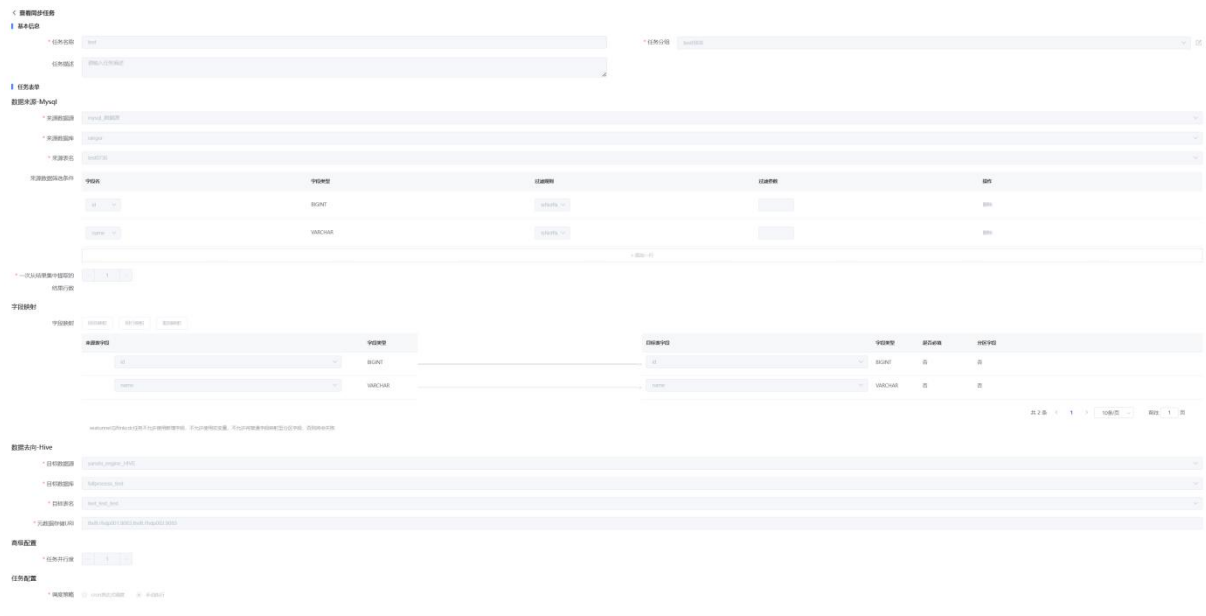
Flink

* flink-conf.yaml

yarn代理地址

任务详情

点击任务列表操作栏对应的"详情"按钮，可查看任务配置的基本信息、数据来源、数据去向、高级配置等详细信息。



任务编辑/删除

对于任务状态为"初始化/结束/失败"的任务，点击操作栏对应的"编辑"或"删除"按钮，可编辑或

删除对应任务。

序号	任务名称	任务分组	同步方式	执行组件	同步方向	任务状态	发布状态	修改时间	操作
1	jctest_ods	test	离线任务	SEA_TUNNEL	HIVE2 -> HIVE2	● 结束	已发布	2024-08-09 14:22:22	发布 启动 任务详情 更多
2	process_test	test	离线任务	SEA_TUNNEL	HIVE2 -> HIVE2	● 结束	已发布	2024-08-15 14:14:59	发布 启动 任务详情 更多
3	test	test0808	离线任务	SEA_TUNNEL	MYSQL -> HIVE2	● 结束	已发布	2024-08-09 10:33:49	发布 启动 任务详情 更多
4	mysql2dori...	test	实时任务	FLINK_CDC	MYSQL -> DORIS	● 运行中	已发布	2024-08-12 14:55:46	发布 停止 任务详情 日志
5	mysql2dori...	test	实时任务	FLINK_CDC	MYSQL -> DORIS	● 已停止	已发布	2024-07-30 14:16:16	发布 启动 任务详情 编辑 详情
6	MySQLtoD...	测试验证	实时任务	FLINK_CDC	MYSQL -> DORIS	● 已停止	已发布	2024-07-30 14:24:28	发布 启动 任务详情 删除

< 编辑同步任务

基本信息

* 任务名称 * 任务分组

任务描述 12/300

任务表单

数据来源-Hive

* 来源数据源

* 来源hive数据库

* 来源hive表名

* 元数据存储URI

字段映射

字段映射

来源表字段	字段类型	目标表字段	字段类型	是否必填	分区字段
<input type="text" value="gmt_modified"/>	VARCHAR	<input type="text" value="gmt_modified"/>	VARCHAR	否	否
<input type="text" value="gmt_create"/>	VARCHAR	<input type="text" value="gmt_create"/>	VARCHAR	否	否

任务日志

点击任务列表操作栏对应的"日志"按钮，可查看任务发布、启动或停止过程中的运行异常日志，便于定位和排查问题。

点击下载图标，支持默认编码 UTF8 打开文件。点击刷新图标，刷新日志信息。点击全屏图标，支持全屏展示或缩放日志页面。

序号	任务名称	任务分组	同步方式	执行组件	同步方向	任务状态	发布状态	修改时间	操作
1	jctest_ods	test	离线任务	SEA_TUNNEL	HIVE2 -> HIVE2	● 结束	已发布	2024-08-09 14:22:22	发布 启动 任务详情 更多
2	process_test	test	离线任务	SEA_TUNNEL	HIVE2 -> HIVE2	● 结束	已发布	2024-08-15 14:14:59	发布 启动 任务详情 更多
3	test	test0808	离线任务	SEA_TUNNEL	MYSQL -> HIVE2	● 结束	已发布	2024-08-09 10:33:49	发布 启动 任务详情 更多
4	mysql2dori...	test	实时任务	FLINK_CDC	MYSQL -> DORIS	● 运行中	已发布	2024-08-12 14:55:46	发布 停止 任务详情 更多
5	mysql2dori...	test	实时任务	FLINK_CDC	MYSQL -> DORIS	● 已停止	已发布	2024-07-30 14:16:16	发布 启动 任务详情 日志
6	MySQLtoD...	测试验证	实时任务	FLINK_CDC	MYSQL -> DORIS	● 已停止	已发布	2024-07-30 14:24:28	发布 启动 任务详情 编辑 详情
7	MySQLtoHI...	测试验证	离线任务	SEA_TUNNEL	MYSQL -> HIVE2	● 初始化	已发布	2024-08-07 17:13:27	发布 启动 任务详情 删除

查看日志



Web应用日志

任务停止失败 | 停止Flink作业失败 | The given configuration directory name 'task-submit-conf/tenant-2/task-group-3/task-10/flink-home' (/soft/datawings/task-submit-conf/tenant-2/task-group-3/task-10/flink-home) does not describe an existing directory.

关闭

4.5 数据处理

4.5.1 数据开发


4.5.1.1 项目管理

项目是数据处理管理任务、成员、分配角色和权限的基本单元，所有开发工作都在项目内完成，您需要在数据开发前先创建项目。

前提条件

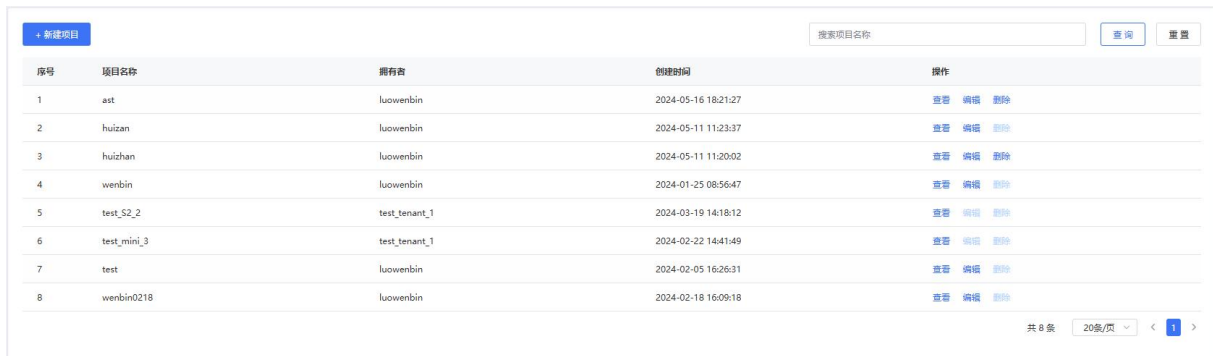
创建项目前请确保您已完成引擎实例的创建，否则无法正常执行引擎任务。创建引擎请前往引擎管理。

操作步骤

单击左上方的图标，选择全部产品 > 数据处理 > 数据开发 > 项目管理。

添加项目

1、点击【新建项目】按钮。

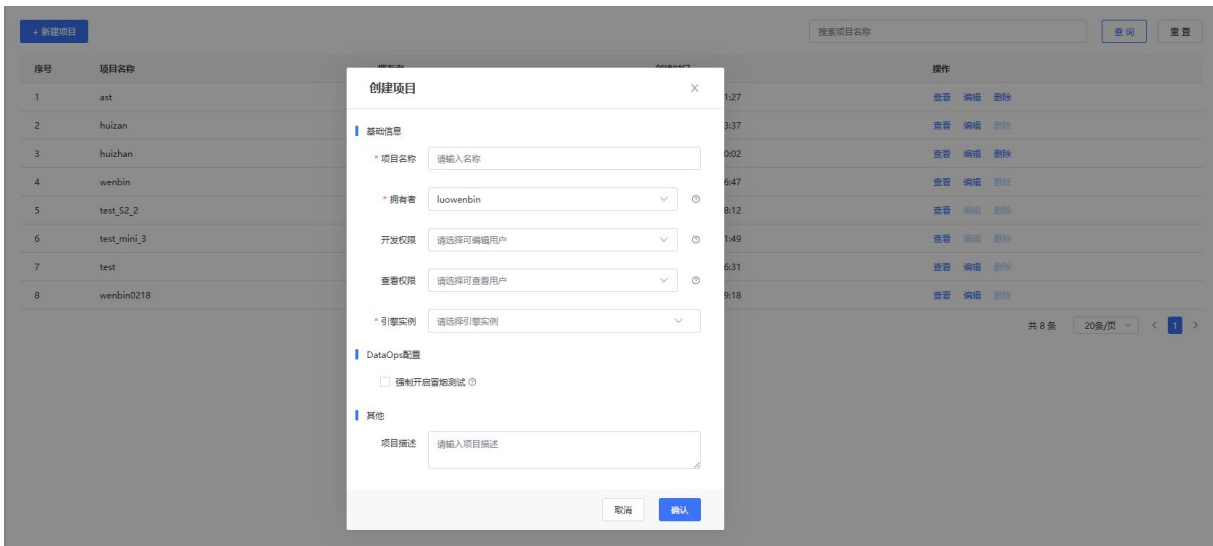


序号	项目名称	拥有者	创建时间	操作
1	ast	luowenbin	2024-05-16 18:21:27	查看 编辑 删除
2	huizan	luowenbin	2024-05-11 11:23:37	查看 编辑 删除
3	huizhan	luowenbin	2024-05-11 11:20:02	查看 编辑 删除
4	wenbin	luowenbin	2024-01-25 08:56:47	查看 编辑 删除
5	test_S2_2	test_tenant_1	2024-03-19 14:18:12	查看 编辑 删除
6	test_mini_3	test_tenant_1	2024-02-22 14:41:49	查看 编辑 删除
7	test	luowenbin	2024-02-05 16:26:31	查看 编辑 删除
8	wenbin0218	luowenbin	2024-02-18 16:09:18	查看 编辑 删除

共 8 条 20条/页 < 1 >

填写项目的基本信息并配置权限。信息填写完毕后点击【确认】按钮即可创建项目。

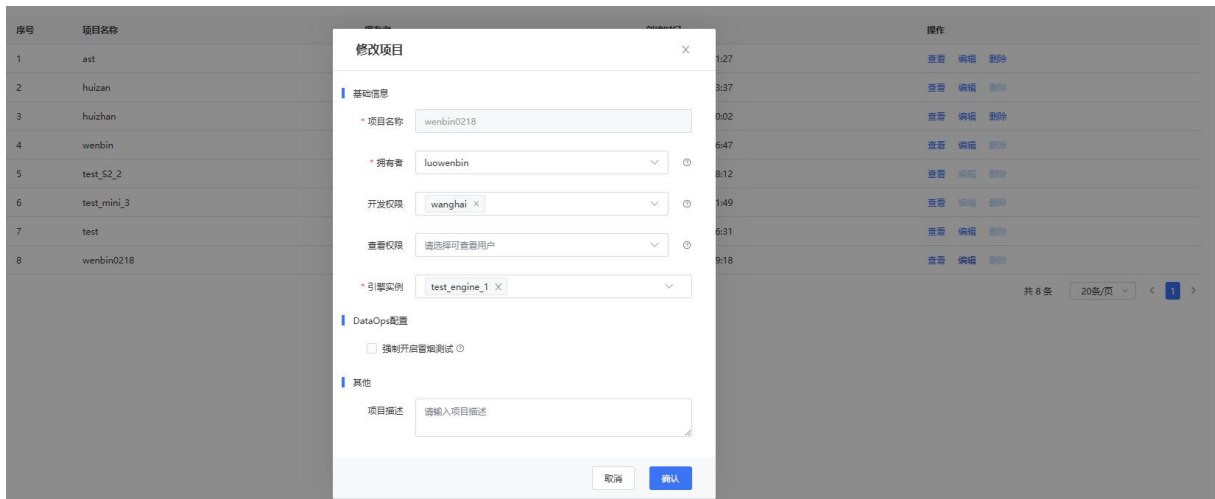
参数	说明
拥有者	可配置项目权限，对项目进行删除，转移项目拥有者。拥有者同时有开发和查看权限。
开发权限	可进行项目内工作流的创建、配置、发布、删除，可查看、配置当前项目工作流在任务调度中的调度情况，同时拥有查看权限。
查看权限	可查看工作流内容，不可配置与发布工作流，不可在任务调度中查看当前项目的调度情况。
引擎实例	配置该项目下关联的引擎。
强制开启冒烟测试	开启后，工作流内的代码类节点需通过冒烟测试才可以发布。



编辑项目

创建项目后，用户可以编辑当前项目的信息。

- 1、仅当前项目的拥有者可以编辑项目信息。
- 2、编辑当前项目的信息后，点击确认即可保存。



删除项目

您可以删除不再需要的项目。

- 1、仅项目拥有者可以删除项目。
- 2、点击【删除】按钮即可删除项目。

4.5.1.2 workflow 开发

DataWings 支持您进行可视化的 workflow DAG 开发，并且支持多种类型大数据开发节点以及版本管理能力。

前提条件

- 已创建项目，项目已配置引擎实例。

- 您需要拥有项目的开发权限才可以新建和开发工作流。

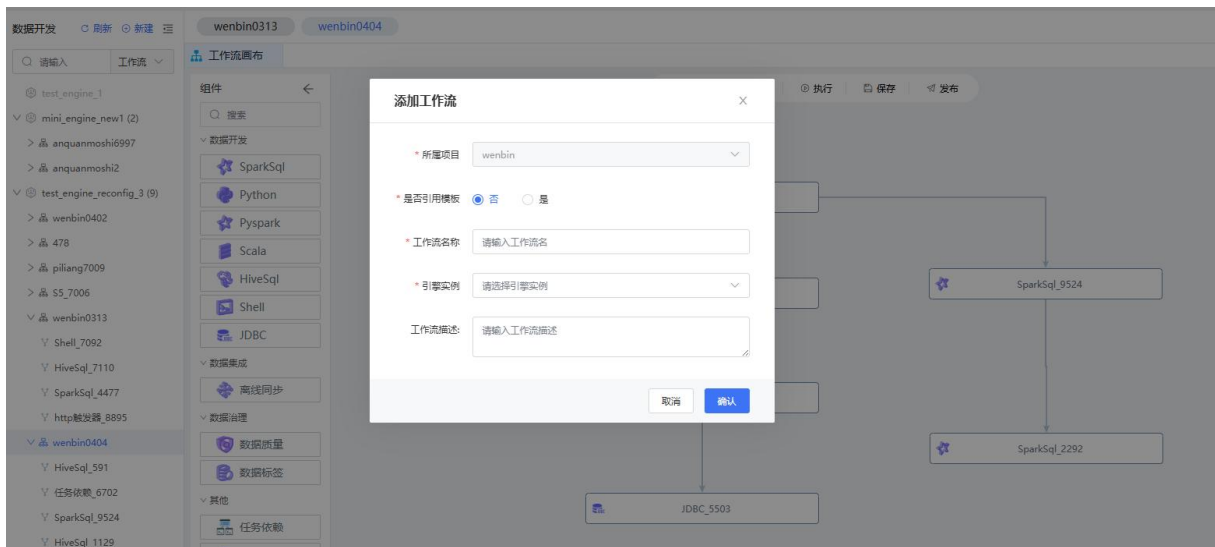
生成 DAG 工作流

新建工作流项目后，点击进入工作流开发页面，您可以通过拖拽组件以搭建工作流。

操作步骤

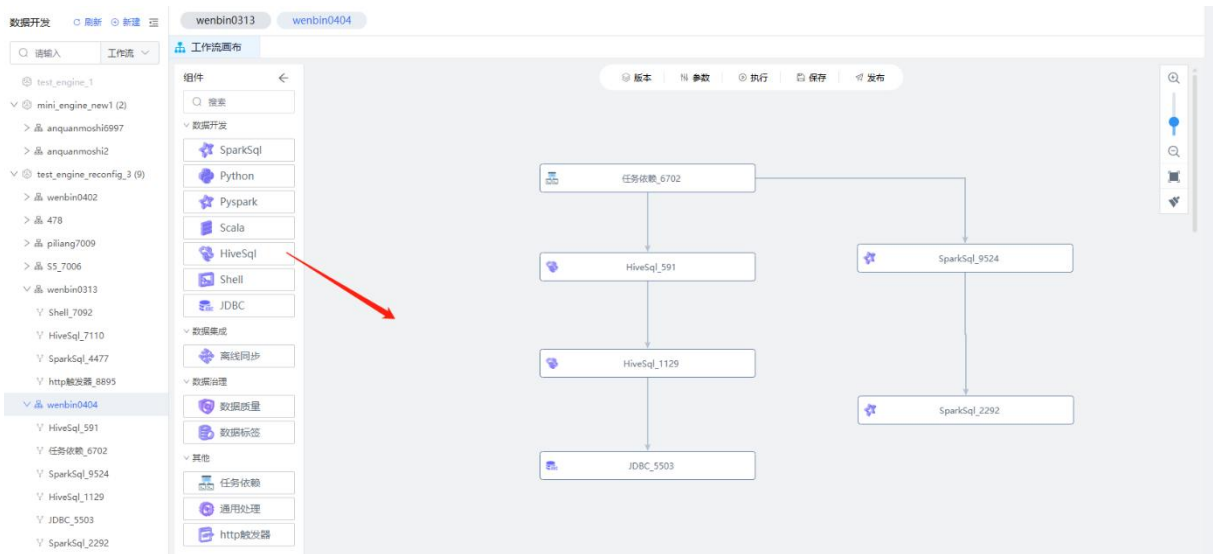
1、新建工作流。

点击【新建工作流】按钮，配置工作流的基本信息。



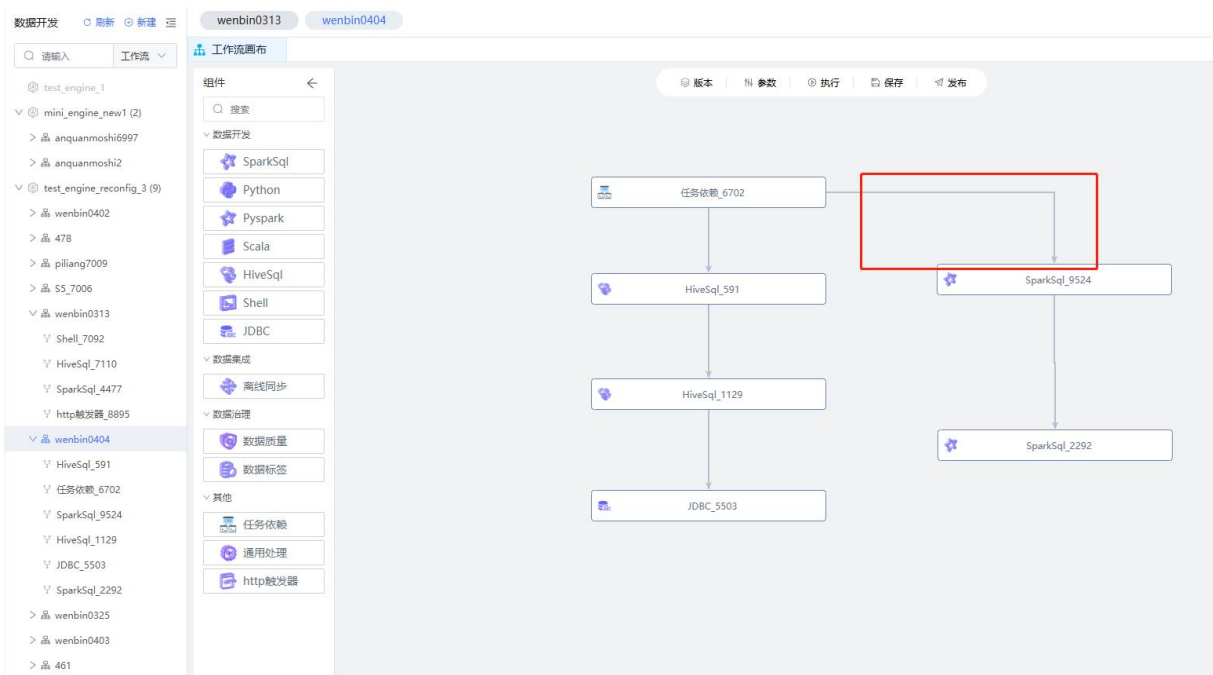
2、添加数据开发节点。

在组件栏中选择所需的组件类型，点击鼠标并拖拽到画布中，即可在工作流中添加一个任务节点。



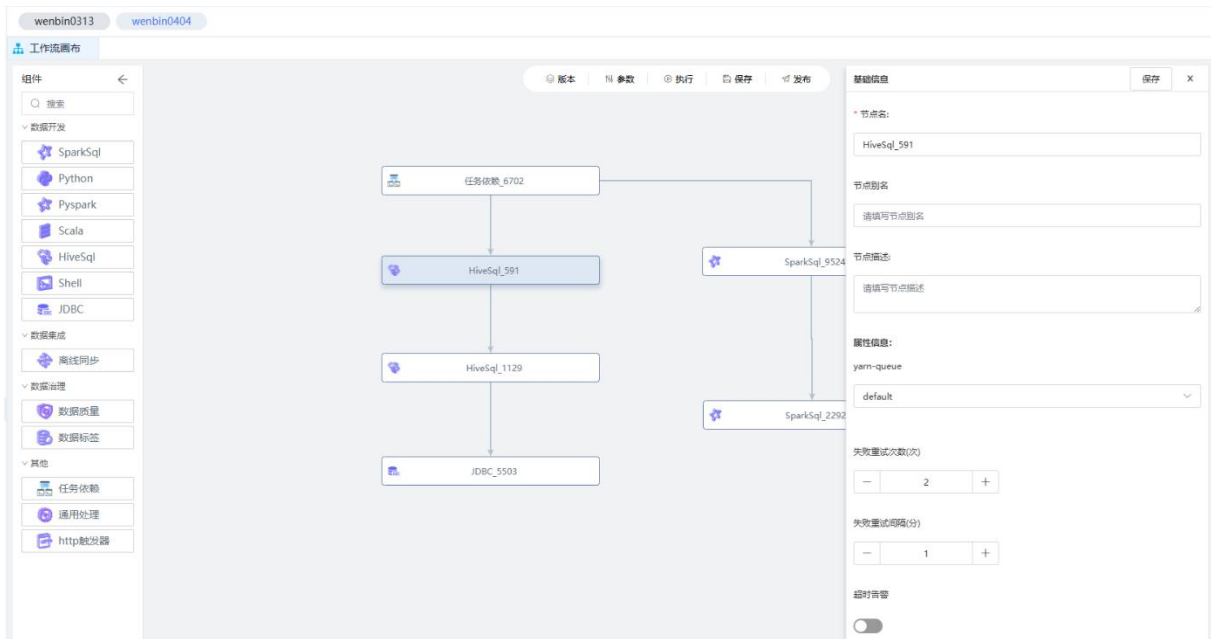
3、连接 workflow。

将鼠标悬停在工作流的节点上，可以显示节点连接处，拖动鼠标可以连接两个节点形成 DAG 工作流，连接后前后节点形成节点依赖关系。



4、配置节点属性。

鼠标单击 workflow 节点，可以配置节点的基础信息，例如节点名称、标签、属性、资源等信息。



编辑 workflow 节点

DataWings 平台内置脚本编辑器，用户可按自身需求编辑 workflow 节点的脚本代码。

操作步骤

1、打开 workflow 节点页面。

鼠标双击 workflow 画布中的节点，即可进入节点编辑页面（部分节点进入节点编辑弹窗，例如数据质量节点、数据标签节点、数据集成节点等等），本文以 SparkSql 节点举例。



2、 workflow 节点操作。

操作	说明
保存	可保存当前节点脚本。
运行	可运行当前脚本以便调试测试，选中部分代码点击运行也可调试部分代码。
配置	可添加脚本所需的自定义参数。
版本	可查看当前节点历史版本信息，进行版本对比。
UDF	节点类型为 HiveSql 或 SparkSql 时可配置节点需要的 udf 函数。
冒烟测试	开发完成后可针对当前版本进行冒烟测试，确保节点运行正常。
测试记录	可查看冒烟测试记录。

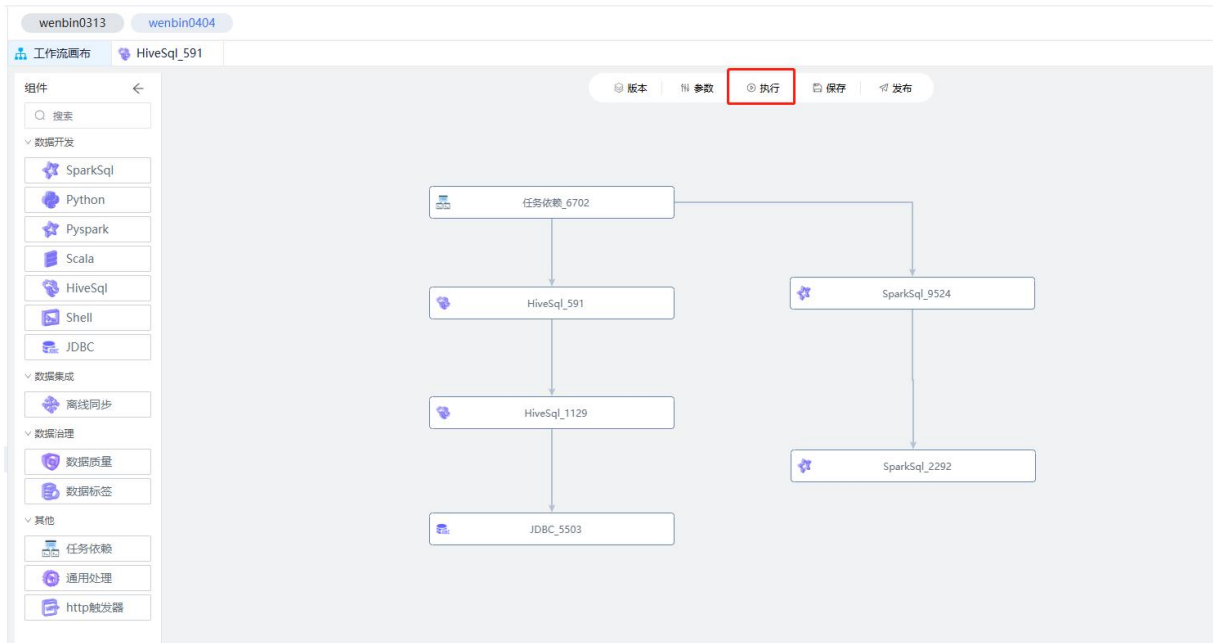


执行 workflow

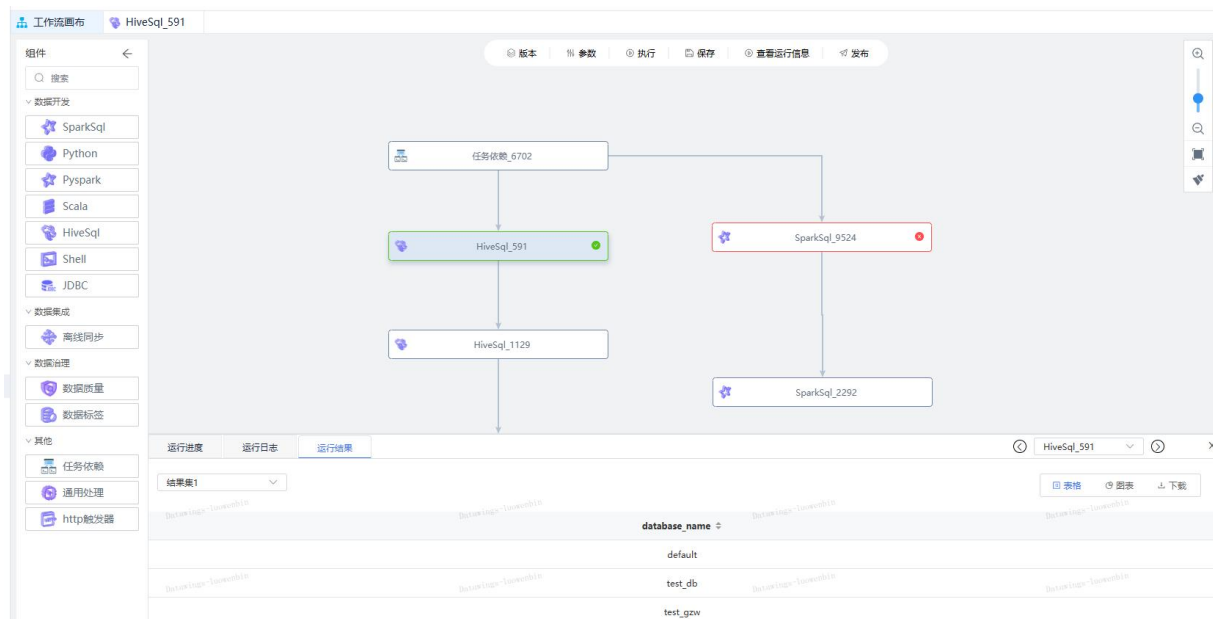
完成 workflow 编排和节点开发后，您可以手动执行当前 workflow，点击 workflow 执行将会自动保存一次。

操作步骤

- 1、点击执行。

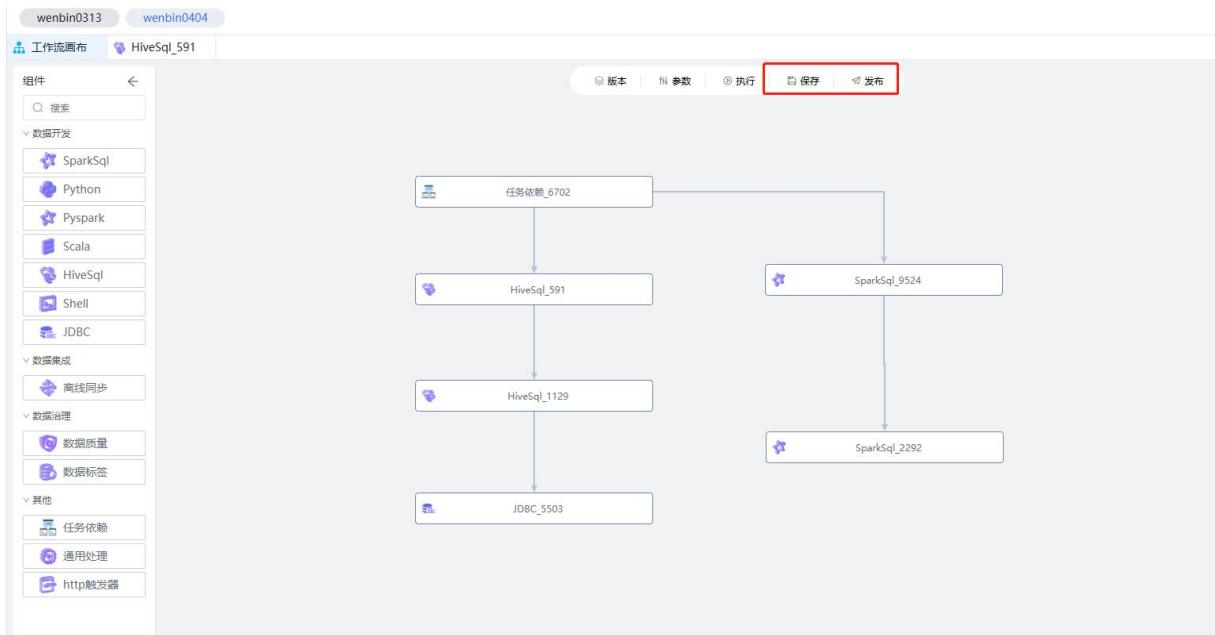


2、在画布中查看 workflow 各节点的运行状态。



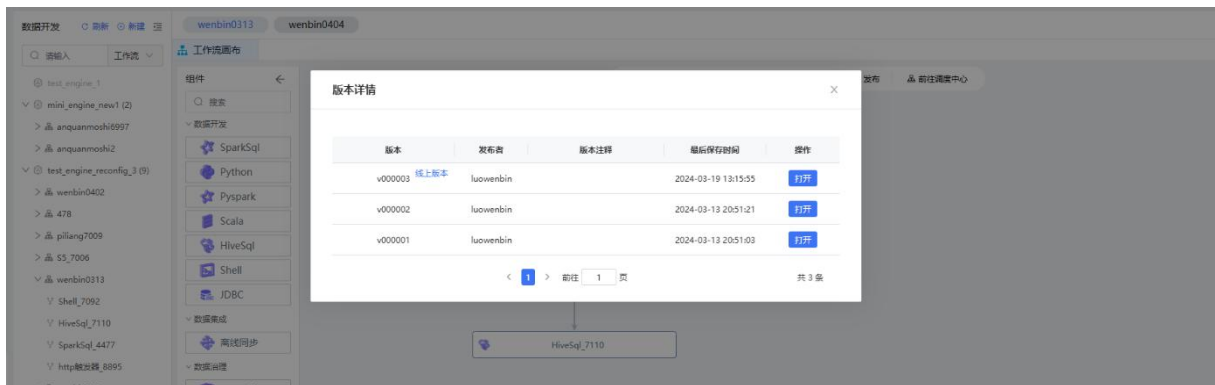
保存和发布 workflow

您可以保存编辑完成的工作流，还可以将工作流发布到任务调度，可以对已发布的工作流进行调度。



查看工作流版本

工具栏点击【版本】，即可查看工作流已发布版本信息，在版本列表中，点击【打开】，可查看历史已发布版本信息。



4.5.1.3 工作流模板

工作流模板会保存工作流的配置信息、节点信息、代码信息，方便您复用提升开发效率。


前提条件

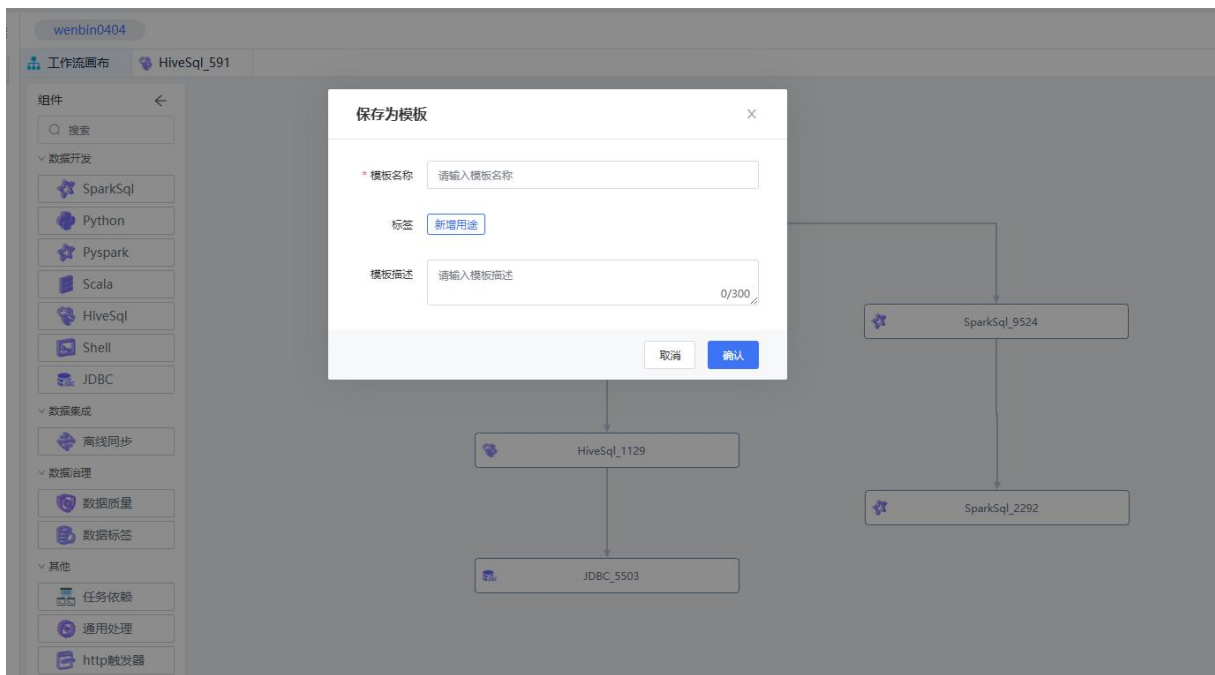
- 已创建项目、配置引擎实例信息。

- 已创建工作流。

创建工作流模板

操作步骤

- 1、单击左上方的图标，选择全部产品 > 数据处理 > 数据开发 > 项目与工作流。
- 2、选择已创建的工作流，点击... > 保存为模板。



- 3、创建后你可以在工作流模板页面查看已保存的工作流模板。


序号	模板名称	模板描述	标签	发布用户	创建时间	操作
1	中文模板		sh11		2024-05-10 16:36:52	查看 使用 下载
2	保存模板		sh11		2024-04-12 14:07:16	查看 使用 下载
3	test_temple		sh11		2024-04-10 16:26:39	查看 使用 下载
4	para		test_tenant_1		2024-04-09 14:30:16	查看 使用 下载
5	test_mb		sh11		2024-04-08 15:27:11	查看 使用 下载
6	t_from_1		test_tenant_1		2024-04-08 10:27:37	查看 使用 下载
7	t3333		test_tenant_1		2024-04-08 10:19:44	查看 使用 下载
8	tmp2		test_tenant_1		2024-04-02 15:50:36	查看 使用 下载
9	tem1		test_tenant_1		2024-04-02 15:39:39	查看 使用 下载
10	t_gov_1	fasqf	test_tenant_1		2024-03-26 14:26:43	查看 使用 下载
11	tmp1		test_tenant_1		2024-02-29 17:07:00	查看 使用 下载

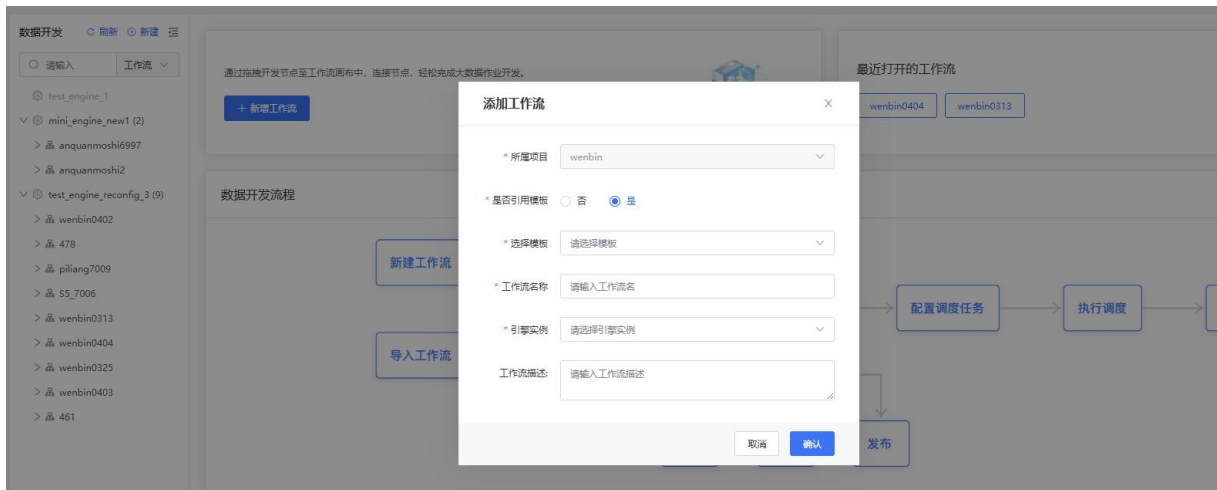
使用工作流模板

创建模板后，您可以在新建工作流时引用模板或者基于模板新建工作流。

引用 workflow 模板


操作步骤

- 1、单击左上方的图标，选择全部产品 > 数据处理 > 数据开发 > 项目与 workflow。
- 2、点击新建 workflow。
- 3、选择引用模板。



基于模板新建 workflow

操作步骤

- 1、单击左上方的图标，选择全部产品 > 数据处理 > 数据开发 > workflow 模板。
- 2、选择制定模板，点击【使用】。

序号	模板名称	模板描述	标签	发布用户	创建时间	操作
1	中文模板			sh1	2024-05-10 16:36:52	查看 使用 下载
2	保存模板			sh1	2024-04-12 14:07:16	查看 使用 下载
3	test_template			sh1	2024-04-10 16:26:39	查看 使用 下载
4	para			test_tenant_1	2024-04-09 14:30:16	查看 使用 下载
5	test_mb			sh1	2024-04-08 15:27:11	查看 使用 下载
6	t_from_1			test_tenant_1	2024-04-08 10:27:37	查看 使用 下载
7	t3333			test_tenant_1	2024-04-08 10:19:44	查看 使用 下载
8	tmp2			test_tenant_1	2024-04-02 15:50:36	查看 使用 下载
9	tem1			test_tenant_1	2024-04-02 15:39:39	查看 使用 下载
10	t_gzw_1	fasfqf		test_tenant_1	2024-03-26 14:26:43	查看 使用 下载
11	tmp1			test_tenant_1	2024-02-29 17:07:00	查看 使用 下载

4.5.1.4 数据开发节点介绍

4.5.1.4.1 SparkSql 节点

您可以在工作流开发中创建 SparkSql 节点，构建大型、低延迟的大数据分析应用。

前提条件

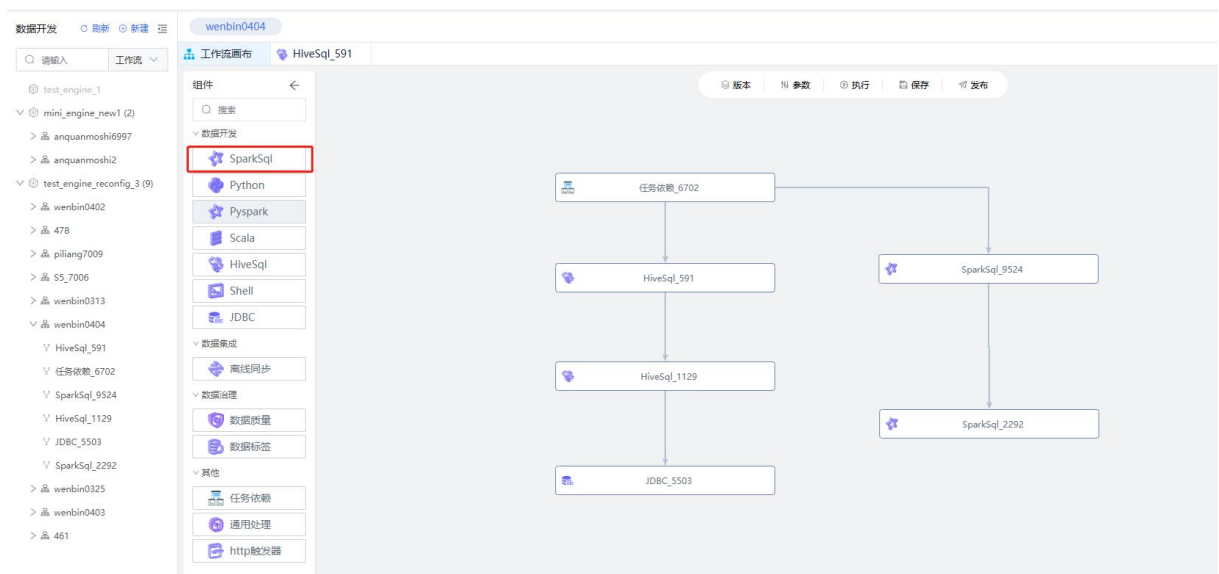
- 已创建项目并绑定至少一个翼 MR 引擎。
- 已创建工作流并绑定翼 MR 引擎。

使用限制

- 执行 SparkSql 节点任务仅支持独享调度资源组运行。

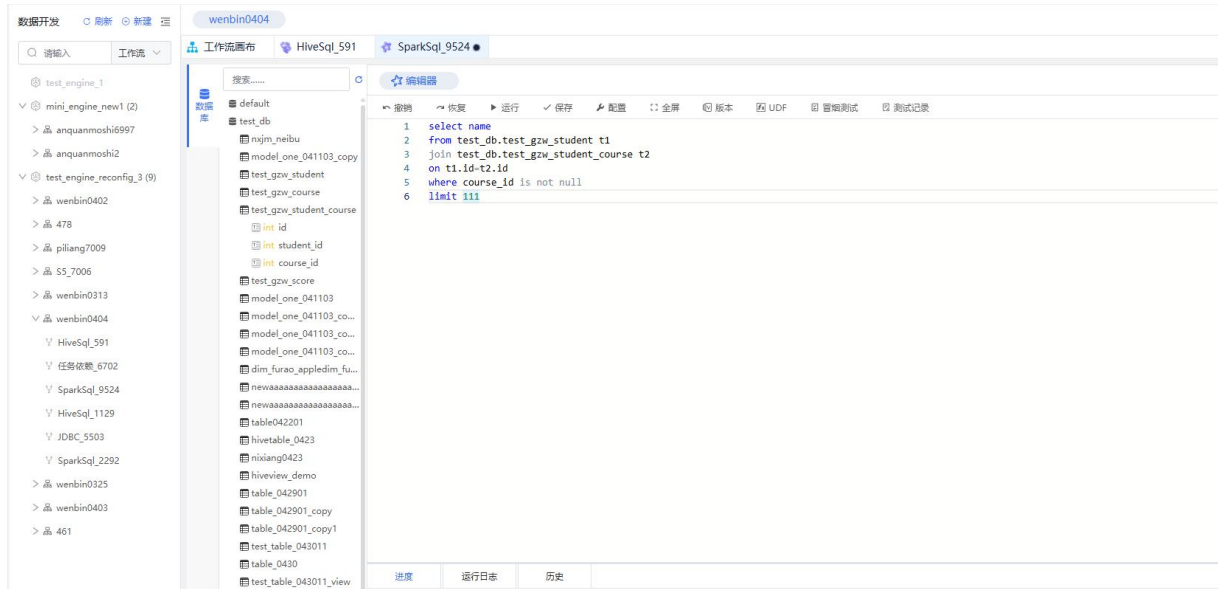
创建 sparksql 节点

- 在工作流开发页面，您可以通过拖拽【SparkSql 节点】进入画布完成节点新增。



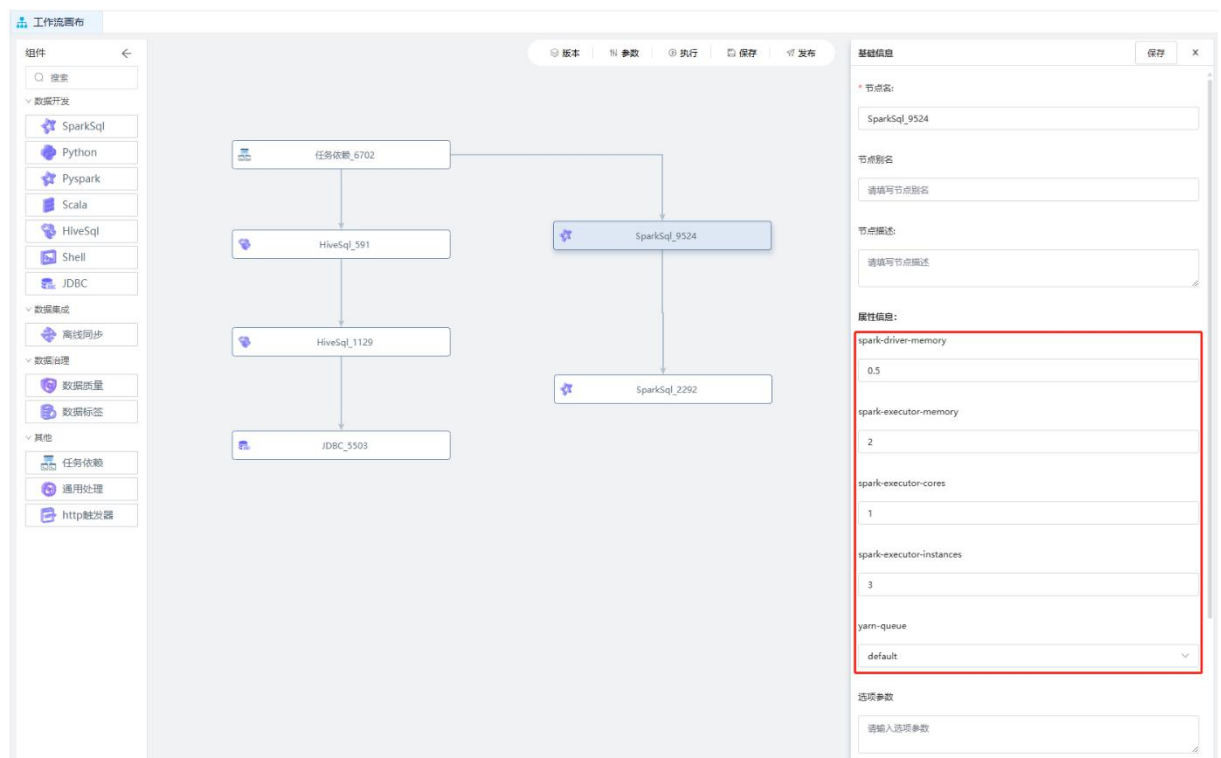
编辑 sparksql 脚本

- 双击 SparkSql 节点可开始编辑 SparkSql 代码，在编辑页面，您可以快速查看内置 hive 元数据，以及调试、保存开发代码。



配置 SparkSql 节点参数

- 单击 SparkSql 节点，您除了可以配置节点基本信息外，还可以编辑当前节点的 Spark 运行参数。



4.5.1.4.2 数据质量节点

您可以在数据开发工作流中增加数据质量节点并配置定时策略，通过周期调起质量节点运行数据质检规则，质检规则（强规则）不符合预期时将自动拦截问题任务，有效阻断脏数据向下游蔓延。

前提条件

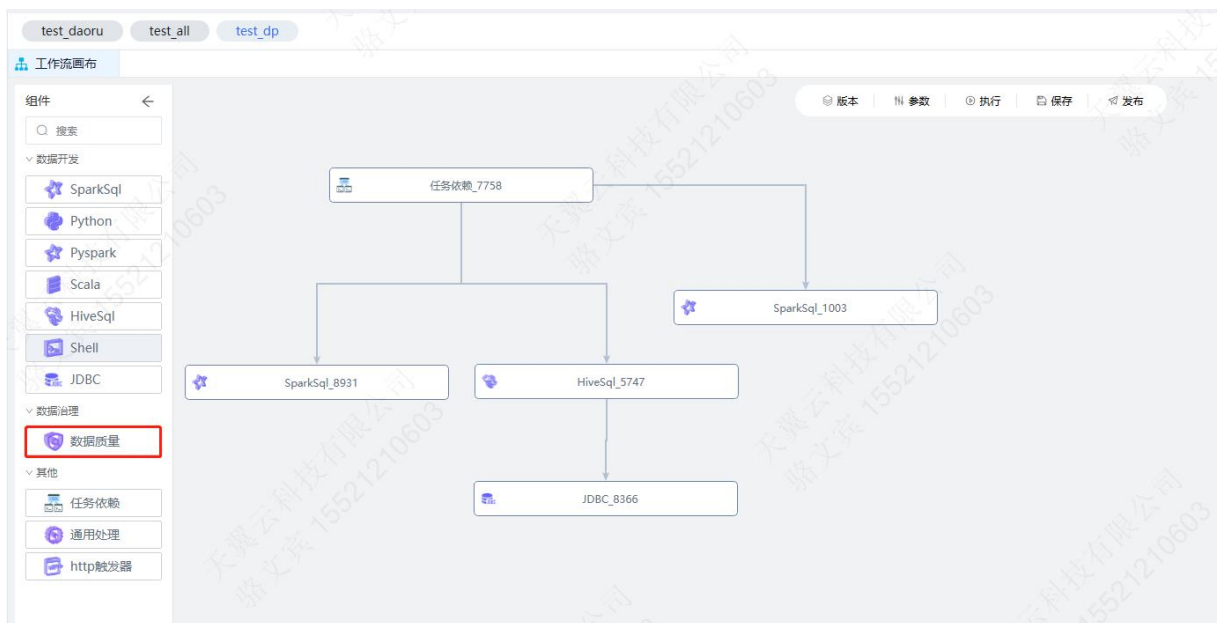
- 您已在数据质量模块创建质检规则并上线。
- 您已绑定至少一个翼 MR 引擎，质检任务以 SparkSql 的方式执行。

使用限制

- 质检节点仅支持当前工作流已绑定的引擎实例的内置 hive 数据源，其他数据源的支持敬请期待。

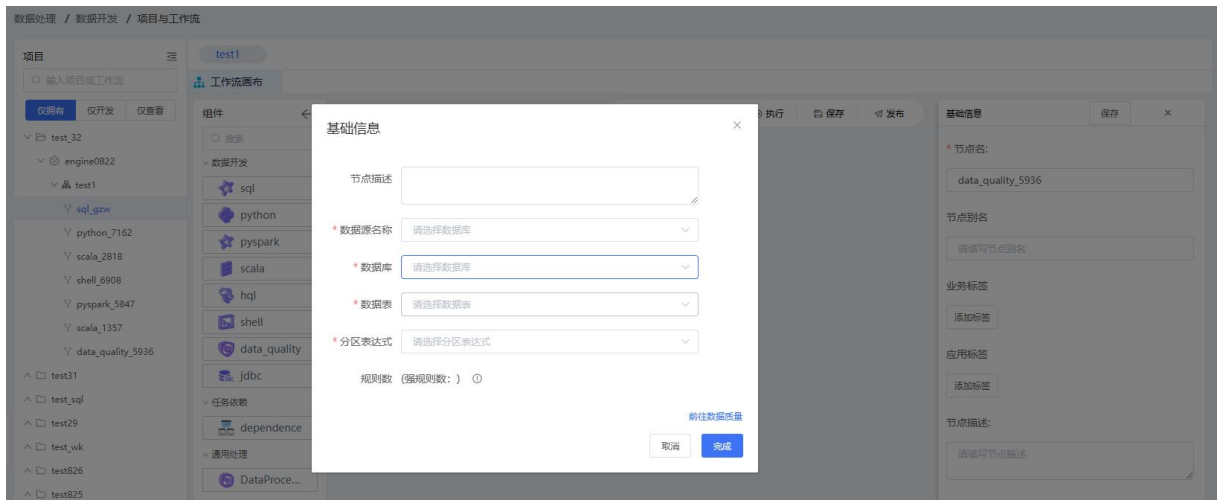
创建数据质量节点

- 在工作流开发页面，您可以通过拖拽【数据质量节点】进入画布完成节点新增。



配置数据质量节点

- 双击数据质量节点即可打开节点配置弹窗，您可以快速选择数据源、库、表信息，并指定当前节点需要运行的分区表达式规则，完成配置后，您可以即时查看当前分区表达式下已配置的规则数及强规则数量。
- 强规则：强规则数据质检规则失败或者不符合预期时，将阻断下游任务执行，避免脏数据蔓延。



4.5.1.4.3 通用处理节点

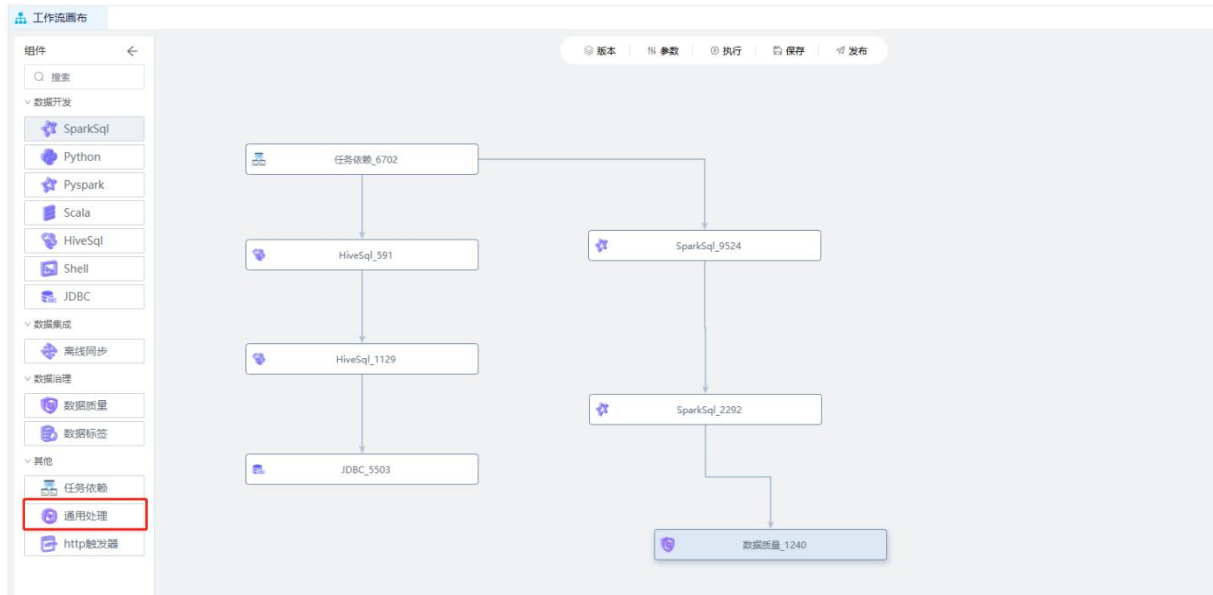
创建通用处理节点，您可以通过简单的操作实现过滤、去重、函数转换、增删字段等复杂 SQL 的构建和调试。

使用限制

- 当前通用处理节点仅支持写入内置 hive 数据源，其他数据源敬请期待。

创建通用处理节点

在工作流开发页面，您可以通过拖拽【通用处理节点】进入画布完成节点新增。

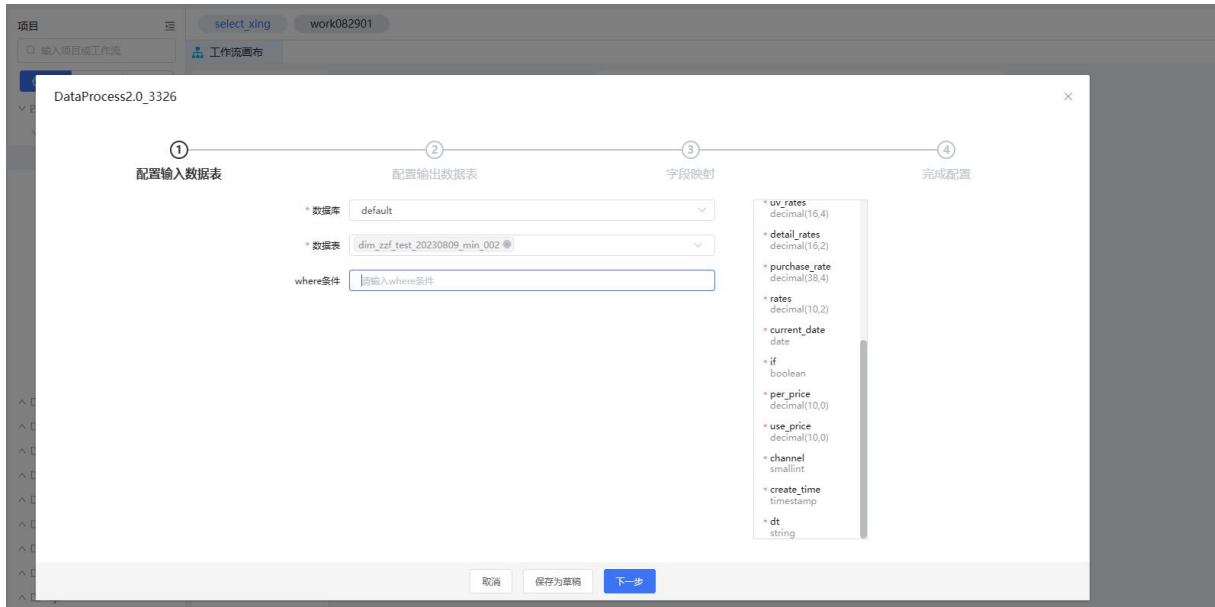


配置通用处理节点

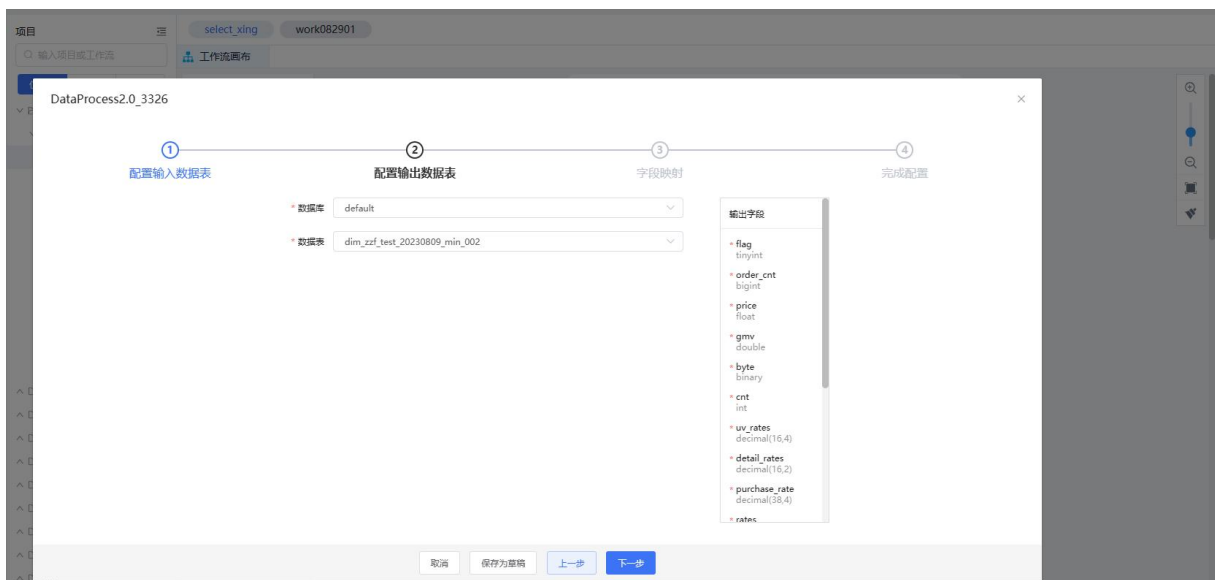
操作步骤

1、配置输入表。

参数	说明
数据库	选择当前工作流绑定引擎的内置 hive 数据源下的数据库。
数据表	选择当前数据库下的数据表。
where 条件	支持输入 where 条件，对单表或多表组合后的数据信息进行进一步的筛选。
输入字段	展示所选表的字段信息。
保存为草稿	点击后，为节点保存当前填写的配置信息，返回工作流画布页。



2、配置输出表。

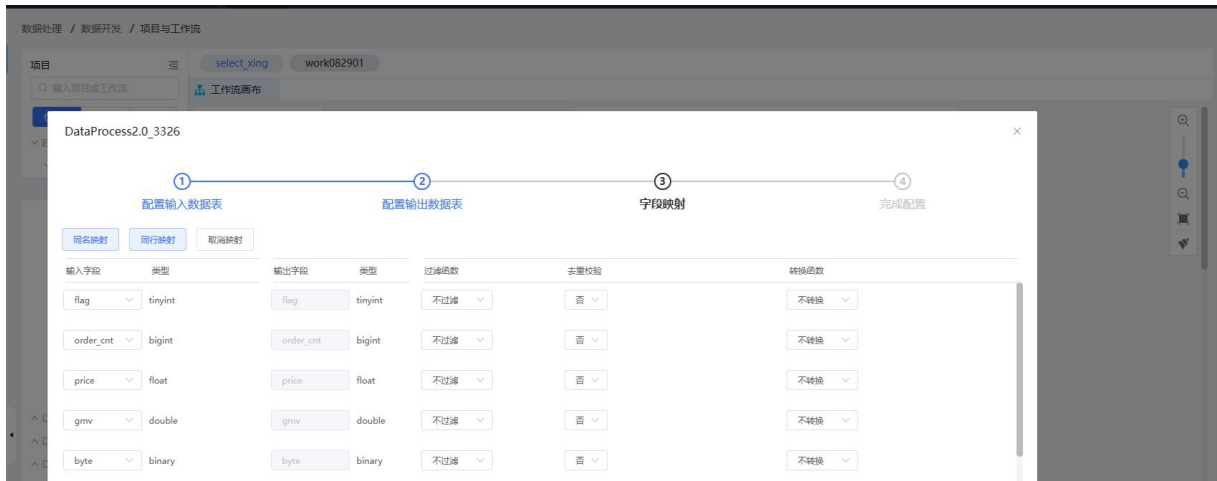


3、配置字段映射。

映射类型描述：

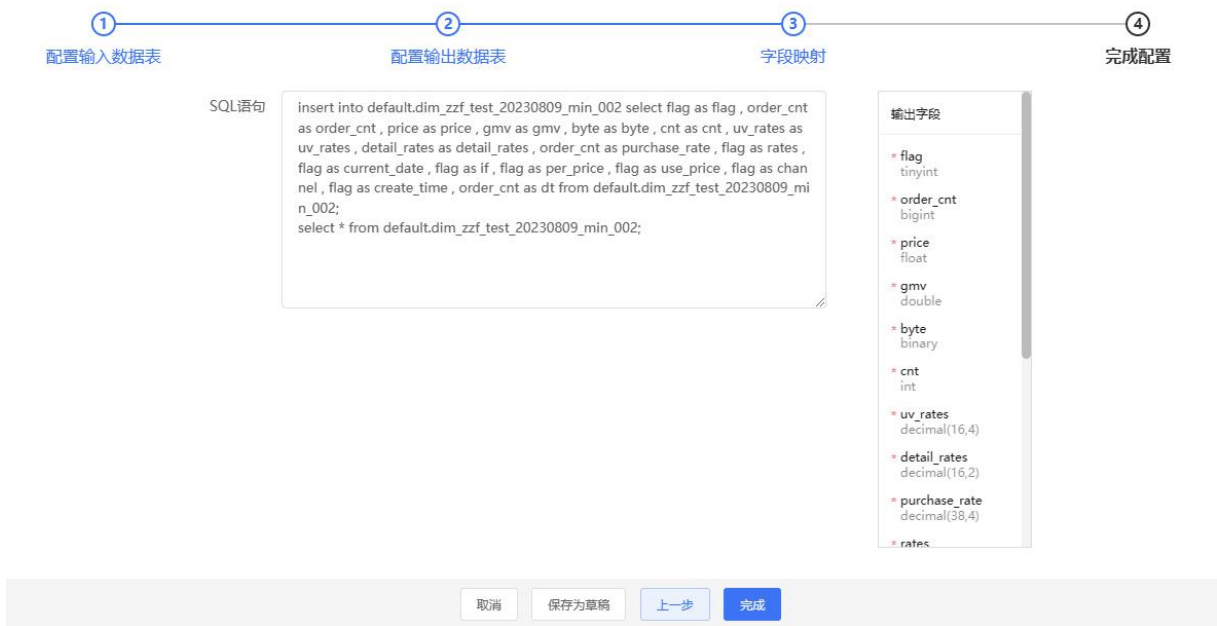
- 同名映射：根据输出字段名来匹配输入字段，进行映射。
- 同行映射：根据输入字段顺序，依次匹配输出字段进行映射。
- 取消映射：取消当前的映射规则。

字段	说明
输入字段	<p>1) 选择当前输入表的字段。</p> <p>2) 自定义输入：您可以输入常量和参数。</p> <p>常量：可输入任意数字或者字符串。</p> <p>参数：可输入基于\${变量名}格式的参数。</p>
过滤函数	<p>您可以选择多种过滤规则，包括 isnull, isnotnull, filterInt, 逻辑运算符, 模糊匹配, 正则匹配等。</p>
去重校验	<p>选择去重后，您可以配置【字段选择】和【保留方式】。</p> <p>字段选择：用于判定数据记录是否重复的依据。</p> <p>保留方式：保留规则用于判定该保留哪个数据记录，可保留较大值/保留较小值。</p>
转换函数	<p>1) 自定义函数：选择此类型后转换规则中的选择项包括用户自定义的 UDF 函数，弹窗中出现“自定义参数”输入项。</p> <p>2) SQL 内置函数：选择此类型，转换规则中的选择项包括常见的 SQL 内置函数。</p>



4、预览 SQL。

完成前述步骤后，您可以预览和校验自动生成的 sql。



4.5.1.4.4 任务依赖节点

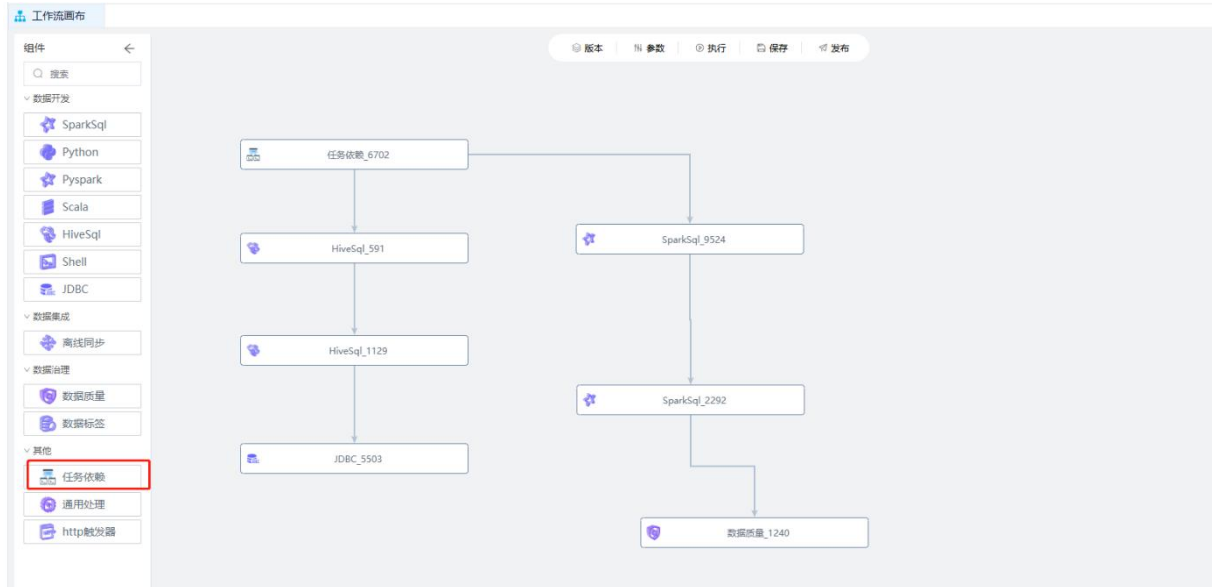
您可以在 workflow 开发中创建依赖节点，通过依赖节点，您可以轻松实现跨 workflow、跨项目依赖，构建全方位的工作流网络。

前提条件

- 已创建项目及至少一个工作流。

创建依赖节点

在工作流开发页面，您可以通过拖拽【任务依赖】节点进入画布完成节点新增。



配置依赖节点

双击依赖节点，可以编辑当前节点的依赖规则。

依赖推荐

智能推荐说明：

1. 依赖推荐功能会基于 SparkSql、HiveSql 节点的历史血缘关系进行节点依赖推荐，帮您尽可能完整地填写依赖关系，依赖数据来源仅包含已调度上线的工作流任务。
2. 非 sql 类型的节点暂不支持依赖推荐，您可以自行配置以来规则。

下游信息：

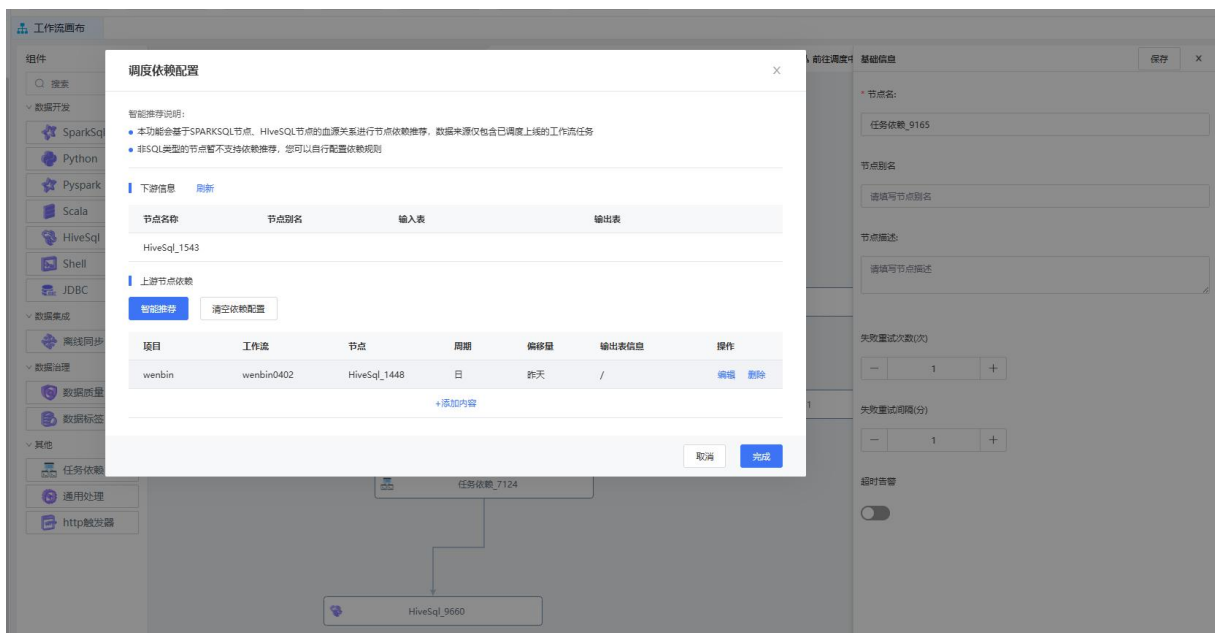
1. 根据工作流当前上下游关系获取下游节点的信息，并根据下游节点的 sql 内容展示输入、输出表信息。

上游节点依赖：

1. 点击智能推荐将会根据下游节点的输入表信息，推荐依赖信息。
2. 您可以基于推荐的依赖信息再次修改满足业务需求。

手动创建依赖关系

1. 您可以按照项目- workflow-任务粒度选择上游依赖，并配置依赖的偏移量和周期。
2. 依赖节点将根据上游节点在偏移量周期内的最后一次调度时间的实例状态来判断是否依赖成功，例如：依赖了 A 节点，周期为日，偏移量为昨天。依赖检查时，将会根据判断调度日期为昨日的 A 节点的最后一次实例是否成功来判断是否依赖成功。
3. 依赖节点仅发布后才会生效，在数据开发画布里执行时将不会做依赖检查。




4.5.2 任务调度

4.5.2.1 调度大屏

调度大屏用以展示用户可见视图下（指用户作为拥有者或开发权限），整个租户内不同时间段

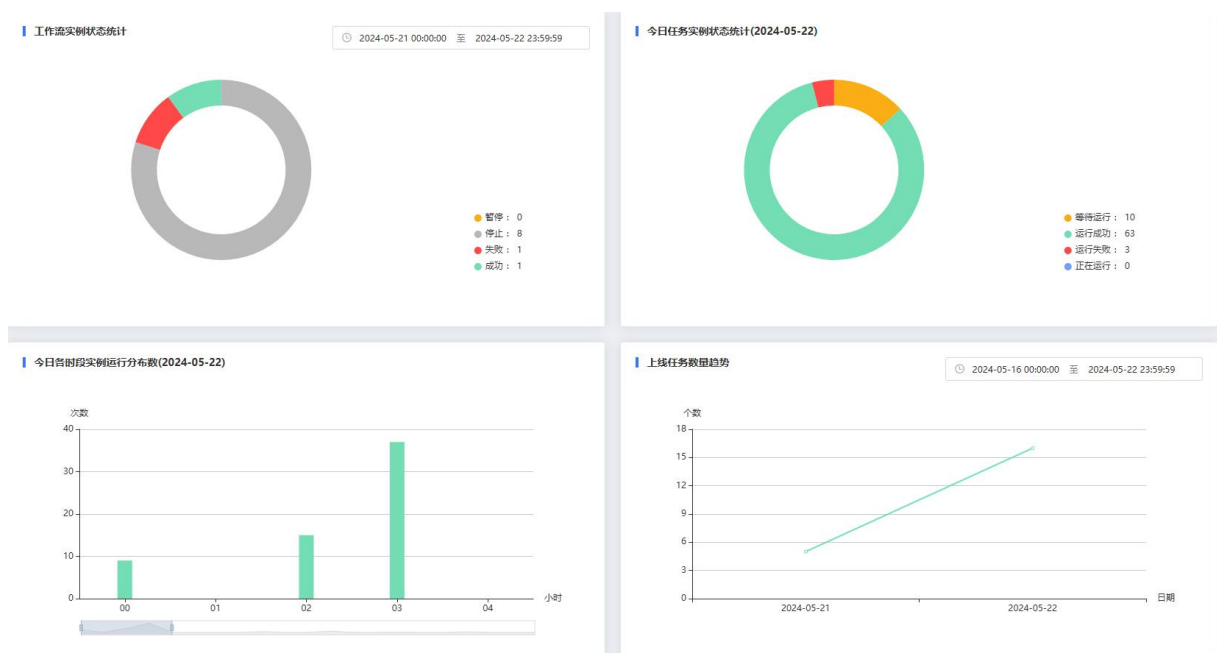
内已发布工作流的统计图表监控信息，分别展示了流程状态统计信息、周期实例完成情况、工作流实例与成功率统计和流程定义统计。

操作步骤

单击左上方的图标，选择全部产品 > 数据处理 > 任务调度 > 调度大屏。

查看大屏指标

工作流及实例执行统计



图表内容解释如下：

图表名称	解释
工作流实例状态统计	统计当前时段内所有工作流实例的状态分布。
今日任务实例状态统计	统计当天已执行任务实例的状态分布。

图表名称	解释
今日各时段实例运行分布数	统计当天已执行任务实例在每个小时运行的数量。
上线任务数量趋势	统计所有已上线任务的日数量趋势。

任务运行耗时排行 top10

统计当前时段内最大运行耗时的 top10 任务。

任务运行耗时排行top10						
排行	任务/节点id	任务名称	任务模块	任务负责人	运行最大耗时	运行平均耗时
1	13678084360540	DEBUG_13678084360540	数据处理	datawingsdev_ops@chinatelecom.cn	52.00s	52.00s
2	13678084360284	DEBUG_13678084360284	数据处理	datawingsdev_ops@chinatelecom.cn	48.00s	48.00s
3	13677423800284	DEBUG_13677423800284	数据处理	datawingsdev_ops@chinatelecom.cn	47.00s	47.00s
4	13678084359470	DEBUG_13678084359470	数据处理	datawingsdev_ops@chinatelecom.cn	42.00s	42.00s
5	13677423800495	DEBUG_13677423800495	数据处理	datawingsdev_ops@chinatelecom.cn	41.00s	41.00s
6	13677423801052	DEBUG_13677423801052	数据处理	datawingsdev_ops@chinatelecom.cn	36.00s	36.00s
7	13678427525340	DEBUG_13678427525340	数据处理	datawingsdev_ops@chinatelecom.cn	31.00s	31.00s
8	13677409900846	DEBUG_13677409900846	数据处理	datawingsdev_ops@chinatelecom.cn	30.00s	30.00s
9	13677416238812	DEBUG_13677416238812	数据处理	datawingsdev_ops@chinatelecom.cn	29.00s	29.00s
10	13678438098012	DEBUG_13678438098012	数据处理	datawingsdev_ops@chinatelecom.cn	29.00s	29.00s

任务运行失败次数排行 top10


统计当前时段内失败次数最多的 top10 任务。

任务运行失败次数排行top10						
排行	任务/节点id	任务名称	任务模块	任务负责人	运行失败次数	
1	1367727722030	DEBUG_1367727722030	数据处理	datawingsdev_ops@chinatelecom.cn	1	
2	1367727722286	DEBUG_1367727722286	数据处理	datawingsdev_ops@chinatelecom.cn	1	
3	1367727725486	DEBUG_1367727725486	数据处理	datawingsdev_ops@chinatelecom.cn	1	
4	13677284725980	DEBUG_13677284725980	数据处理	datawingsdev_ops@chinatelecom.cn	1	
5	13677317521582	DEBUG_13677317521582	数据处理	datawingsdev_ops@chinatelecom.cn	1	
6	13677332390574	DEBUG_13677332390574	数据处理	datawingsdev_ops@chinatelecom.cn	1	
7	13677342215086	DEBUG_13677342215086	数据处理	datawingsdev_ops@chinatelecom.cn	1	
8	13677352150190	DEBUG_13677352150190	数据处理	datawingsdev_ops@chinatelecom.cn	1	
9	13676809401180	DEBUG_13676809401180	数据处理	datawingsdev_ops@chinatelecom.cn	1	
10	13676817417564	DEBUG_13676817417564	数据处理	datawingsdev_ops@chinatelecom.cn	1	

4.5.2.2 任务运维

任务运维页面为您提供了 workflow 关系、workflow 定义和实例、节点定义和实例的运维管理能力，提升数据开发的运维效率。

操作步骤

单击左上方的图标，选择全部产品 > 数据处理 > 任务调度 > 任务运维。

workflow 关系

workflow 关系页面展示了该项目下 workflow 之间的依赖关系，及 workflow 的状态信息。

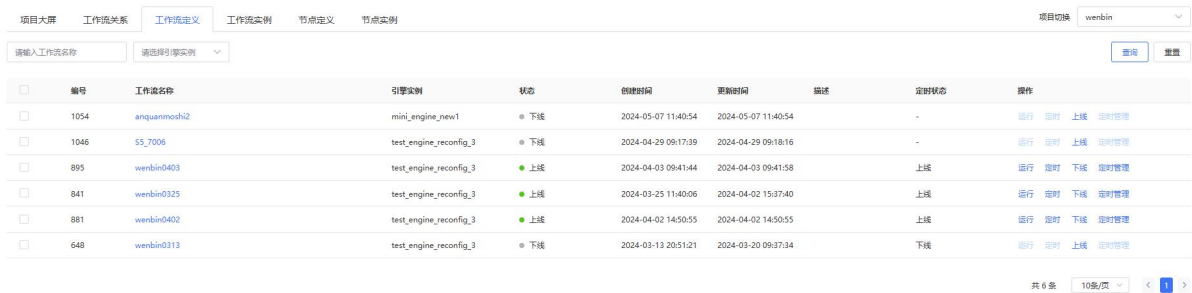


workflow 定义

workflow 定义展示了当前选中项目下定义的所有 workflow 的信息，可以在操作栏对单个 workflow 进行运行、定时、上线/下线和定时管理等操作。

功能	说明
上线	workflow 状态为"下线"时，上线 workflow，只有"上线"状态的 workflow 能运行，但不能编辑。
下线	workflow 状态为"上线"时，下线 workflow，下线状态的 workflow 可以编辑，但不能运行。

功能	说明
运行	只有上线的工作流能运行。运行操作步骤见运行工作流。
定时	只有上线的工作流能设置定时，系统自动定时调度工作流运行。创建定时后的状态为"下线"，需在定时管理页面上线定时才生效。定时操作步骤见工作流定时。
定时管理	定时管理页面可编辑、上线/下线、删除定时。



编号	工作流名称	引擎实例	状态	创建时间	更新时间	描述	定时状态	操作
1054	anquanmoshi2	mini_engine_new1	下线	2024-05-07 11:40:54	2024-05-07 11:40:54	-	-	运行 定时 上线 定时管理
1046	55_7006	test_engine_reconfig_3	下线	2024-04-29 09:17:39	2024-04-29 09:18:16	-	-	运行 定时 上线 定时管理
895	wenbin0403	test_engine_reconfig_3	上线	2024-04-03 09:41:44	2024-04-03 09:41:58	-	上线	运行 定时 下线 定时管理
841	wenbin0325	test_engine_reconfig_3	上线	2024-03-25 11:40:06	2024-04-02 15:37:40	-	上线	运行 定时 下线 定时管理
881	wenbin0402	test_engine_reconfig_3	上线	2024-04-02 14:50:55	2024-04-02 14:50:55	-	上线	运行 定时 下线 定时管理
648	wenbin0313	test_engine_reconfig_3	下线	2024-03-13 20:51:21	2024-03-20 09:37:34	-	下线	运行 定时 上线 定时管理

运行工作流

- 1、选择指定工作流，点击运行按钮。
- 2、在运行弹窗中配置运行参数。

功能	说明
失败策略	当某一个任务节点执行失败时，其他并行的任务节点需要执行的策略。 继续：其他任务节点正常执行。

功能	说明
	结束：终止所有正在执行的任务，并终止整个流程。
通知策略	当流程结束，根据流程状态发送流程执行信息通知邮件，包含任何状态都不发，成功发，失败发，成功或失败都发。
流程优先级	流程运行的优先级，分五个等级：最高（HIGHEST），高（HIGH），中（MEDIUM），低（LOW），最低（LOWEST）。当 master 线程数不足时，级别高的流程在执行队列中会优先执行，相同优先级的流程按照先进先出的顺序执行。
告警组	触发告警时，将会发送告警消息到指定的通知组的所有成员。
补数	支持串行补数和并行补数两种方式。 串行补数：指定时间范围内，从开始日期至结束日期依次执行补数，依次生成 N 条流程实例。 并行补数：指定时间范围内，多天同时进行补数，同时生成 N 条流程实例。



创建工作流定时

- 1、选择指定 workflow，点击定时按钮。
- 2、在定时弹窗中配置定时信息。

失败策略、通知策略、流程优先级、告警组和工作流运行操作的参数相同，定时策略创建后默认为下线状态。



- 3、管理定时策略。

选择指定 workflow，点击定时管理按钮。

功能	说明
编辑	编辑当前定时策略，仅下线的定时策略支持编辑。
上线	上线当前定时策略， workflow 将按照定时策略周期触发。
下线	下线当前定时策略，定时失效。
删除	删除当前定时策略，仅下线的定时策略支持删除。



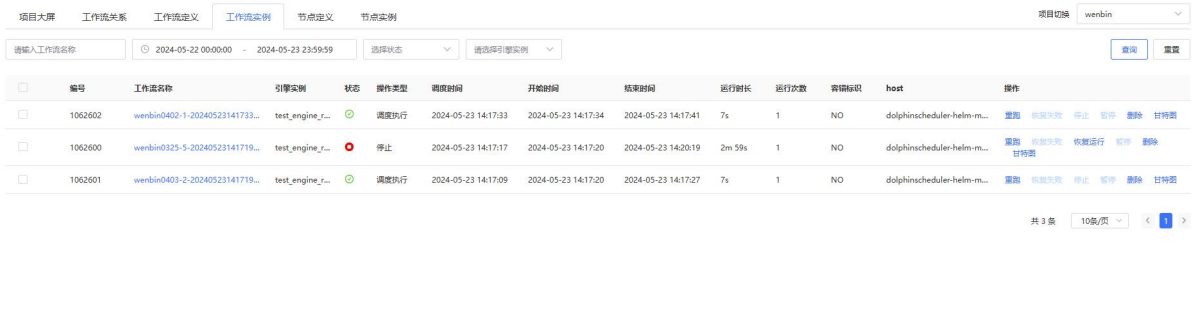
编号	工作流名称	开始时间	结束时间	crontab	失败原因	状态	创建时间	更新时间	操作
51	wenbin0325	2024-05-23 00:00:00	2124-05-23 00:00:00	每5分钟	CONTINUE	下线	2024-05-23 14:17:15	2024-05-23 14:17:15	编辑 上线 删除

工作流实例

工作流实例页面以列表形式展示了工作流的状态、运行类型、开始时间、运行时长等信息，并可对工作流的运行状态进行操作。

功能	说明
查看详情	点击工作流名称，您可以查看当前工作流实例运行的实时 DAG 状态图。
重跑	重新执行已经结束的流程。

功能	说明
停止	对正在运行的流程进行中止操作。
暂停	对正在运行的流程进行暂停操作，系统状态变为等待执行，会等待正在执行的任务结束，暂停下一个要执行的任务。
恢复运行	对暂停的流程恢复，直接从暂停的节点开始运行。
删除	删除 workflow 实例及 workflow 实例下的任务实例。
甘特图	Gantt 图纵轴是某个 workflow 实例下的任务实例的拓扑排序，横轴是任务实例的运行时间。



编号	工作流名称	引擎实例	状态	操作类型	调度时间	开始时间	结束时间	运行时长	运行次数	资源标识	host	操作
1062602	wenbin0402-1-20240523141733...	test_engine_f...	🟢	调度执行	2024-05-23 14:17:33	2024-05-23 14:17:34	2024-05-23 14:17:41	7s	1	NO	dolphinscheduler-helm-m...	重新 恢复实例 停止 暂停 删除 甘特图
1062600	wenbin0325-5-20240523141719...	test_engine_f...	🔴	停止	2024-05-23 14:17:17	2024-05-23 14:17:20	2024-05-23 14:20:19	2m 59s	1	NO	dolphinscheduler-helm-m...	重新 恢复实例 恢复运行 暂停 删除 甘特图
1062601	wenbin0403-2-20240523141719...	test_engine_f...	🟢	调度执行	2024-05-23 14:17:09	2024-05-23 14:17:20	2024-05-23 14:17:27	7s	1	NO	dolphinscheduler-helm-m...	重新 恢复实例 停止 暂停 删除 甘特图

节点定义

节点定义页面以列表形式展示 workflow 中各个节点任务的名称、所属 workflow、workflow 状态、任务类型、上游任务、创建时间等信息。

项目大屏 工作流关系 工作流定义 工作流实例 节点定义 节点实例 项目切换 wenbin

请输入任务名称 请输入工作流名称 请选择引擎实例 查询 重置

编号	任务名称	引擎实例	工作流名称	工作流状态	任务类型	上游任务	创建时间
1	HiveSql_1543	test_engine_reconfig_3	wenbin0325	ONLINE	HIVE_SQL	[13134456999008]任务依赖_9165	2024-04-02 15:37:40
2	任务依赖_7124	test_engine_reconfig_3	wenbin0325	ONLINE	DEPENDENT	[13134455811552]HiveSql_1543	2024-04-02 15:37:40
3	HiveSql_8213	test_engine_reconfig_3	S5_7006	OFFLINE	HIVE_SQL		2024-04-29 09:17:39
4	任务依赖_9165	test_engine_reconfig_3	wenbin0325	ONLINE	DEPENDENT		2024-04-02 15:37:40
5	HiveSql_1448	test_engine_reconfig_3	wenbin0402	OFFLINE	HIVE_SQL		2024-04-02 14:50:55
6	HiveSql_9660	test_engine_reconfig_3	wenbin0325	ONLINE	HIVE_SQL	[13134463080672]任务依赖_7124	2024-04-02 15:37:40
7	HiveSql_2321			OFFLINE	HIVE_SQL		2024-03-25 11:40:06
8	Shell_7092	test_engine_reconfig_3	wenbin0313	OFFLINE	SHELL		2024-03-19 13:15:55
9	HiveSql_591	test_engine_reconfig_3	wenbin0403	OFFLINE	HIVE_SQL		2024-04-03 09:41:44
10	HiveSql_1146	mini_engine_new1	anquanmoshi2	OFFLINE	HIVE_SQL		2024-05-07 11:40:54

共 15 条 10条/页 1

节点实例

节点实例页面以列表形式展示工作流中各个节点任务的基本信息和运行情况，并可查看任务实例的运行日志。

项目大屏 工作流关系 工作流定义 工作流实例 节点定义 节点实例 项目切换 wenbin

请输入实例名称 2024-05-22 00:00:00 - 2024-05-23 23:59:59 选择状态 请选择引擎实例 查询 重置

编号	名称	工作流实例	节点类型	状态	提交时间	开始时间	结束时间	运行时长	重试次数	host	操作
90971	HiveSql_1448	wenbin0402-1-20240523141733604	HIVE_SQL	成功	2024-05-23 14:17:34	2024-05-23 14:17:33	2024-05-23 14:17:40	7s	0	192.168.61.62:1234	查看日志
90969	任务依赖_9165	wenbin0325-5-20240523141719530	DEPENDENT	失败	2024-05-23 14:17:20	2024-05-23 14:17:20	2024-05-23 14:20:19	2m 59s	0	dolphinsheduler-helm-master-1.dolph...	查看日志
90970	HiveSql_591	wenbin0403-2-20240523141719531	HIVE_SQL	成功	2024-05-23 14:17:20	2024-05-23 14:17:19	2024-05-23 14:17:26	7s	0	192.168.61.62:1234	查看日志

共 3 条 10条/页 1

4.5.3 即席查询


4.5.3.1 用户空间

用户空间是一个文件目录，用户对该目录拥有所有的权限可以进行文件管理操作等。

前提条件

- 使用即席查询模块前请至少保证已创建 1 个引擎实例。

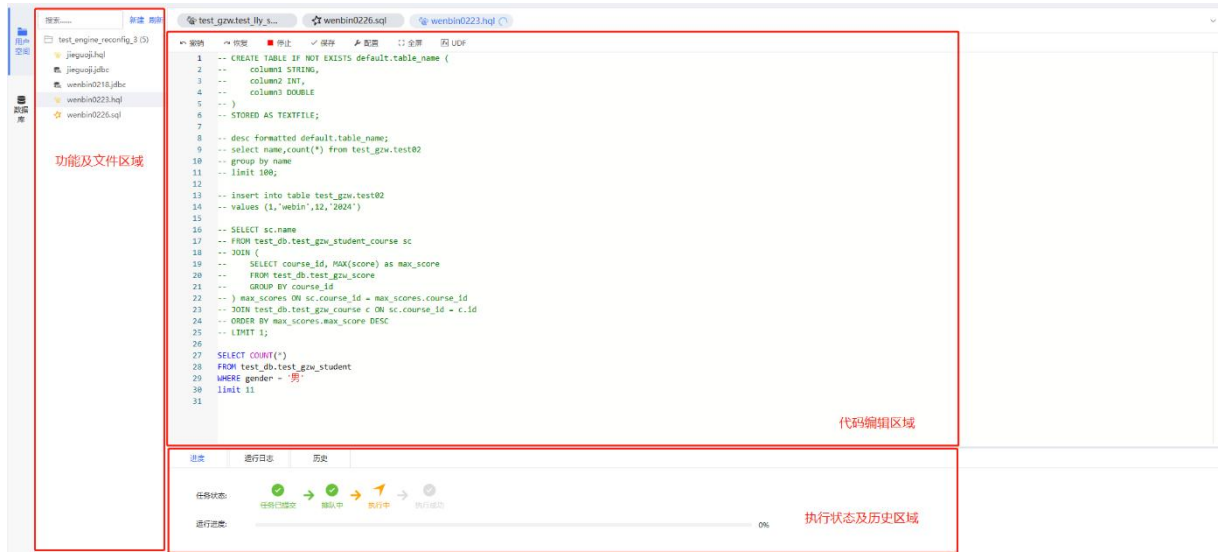
页面进入

- 单击左上方的图标，选择全部产品 > 数据处理 > 即席查询 > 用户空间。

页面功能说明

- 1) 页面区域划分：功能与文件目录区域、操作及脚本编辑区域、执行与历史状态区域。
- 2) 具体功能说明。

功能	说明
新建脚本	您可以创建 SparkSql、HiveSql、Python、JDBC、Hudi 等脚本。
上传脚本	除了手动创建外，您还可以从本地上传已有脚本。
运行	您可以选择部分代码进行执行，也可以直接点击执行运行整个脚本。
配置	配置定义自定义变量，仅该脚本内生效。
撤销	撤销上一步编辑内容。
恢复	恢复上一步编辑内容。
进度	当前脚本执行的进度详情。
运行日志	当前脚本运行的实时日志。
历史	当前脚本的历史运行日志和结果。



4.5.3.2 数据库

数据库模块为您清晰地展示数据库表数据，支持快速查询数据和查询表结构等。

使用限制

当前数据库仅支持展示内置 hive 数据源的元数据，其他数据源的元数据敬请期待。

页面进入

单击左上方的图标，选择全部产品 > 数据处理 > 即席查询 > 数据库。

页面说明

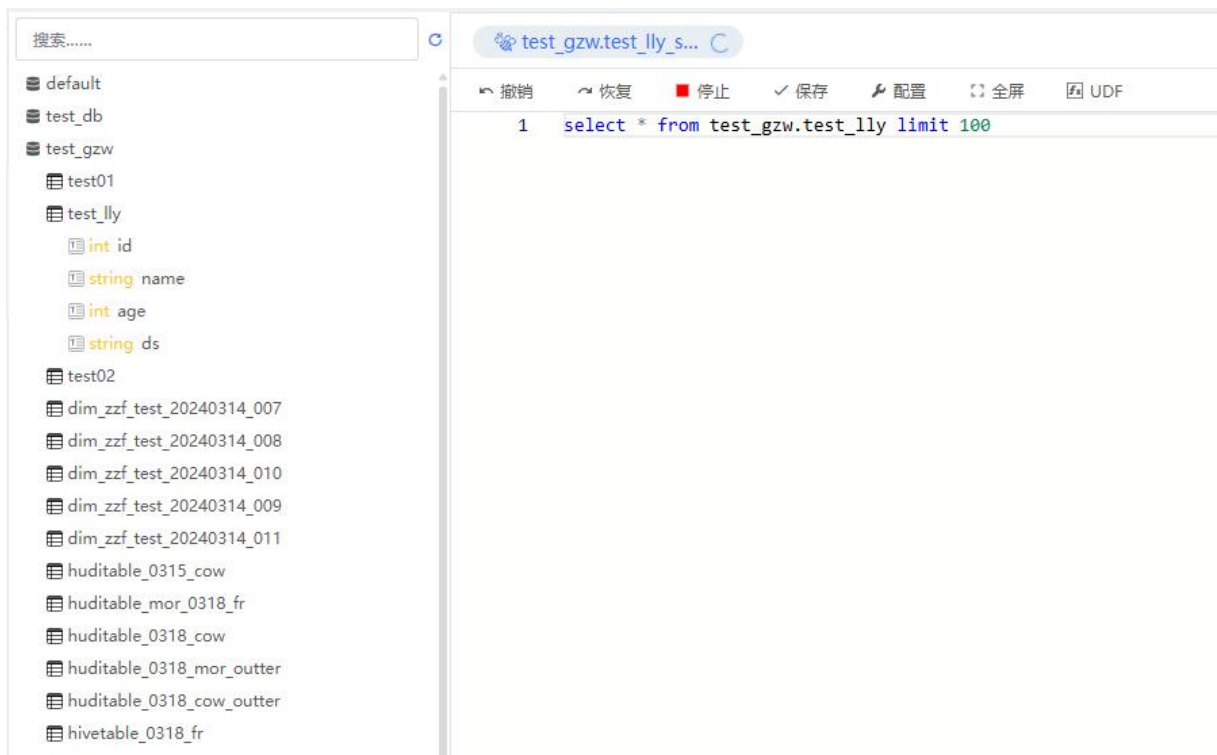
1、查看数据库表字段信息

您可以按库、表、字段逐级查看元数据信息。



2、查询表数据

选择指定表, 右键选择查询表, 即可快速生成查询表数据的 SQL 并执行。



3、查看表结构

选择指定表，右键选择查看表结构，即可跳转到数据地图模块，查看当前表的详细元数据信息。

4.5.4 配置中心

您可以对数据开发内涉及的文件资源、UDF 函数、内置变量进行统一的管理和查看。

内置变量

您可以查看可在数据开发/即席查询节点或者脚本呢使用的内置变量，及其对应的描述。

使用方式如下：`select "${yyyy-MM-dd-1}"`

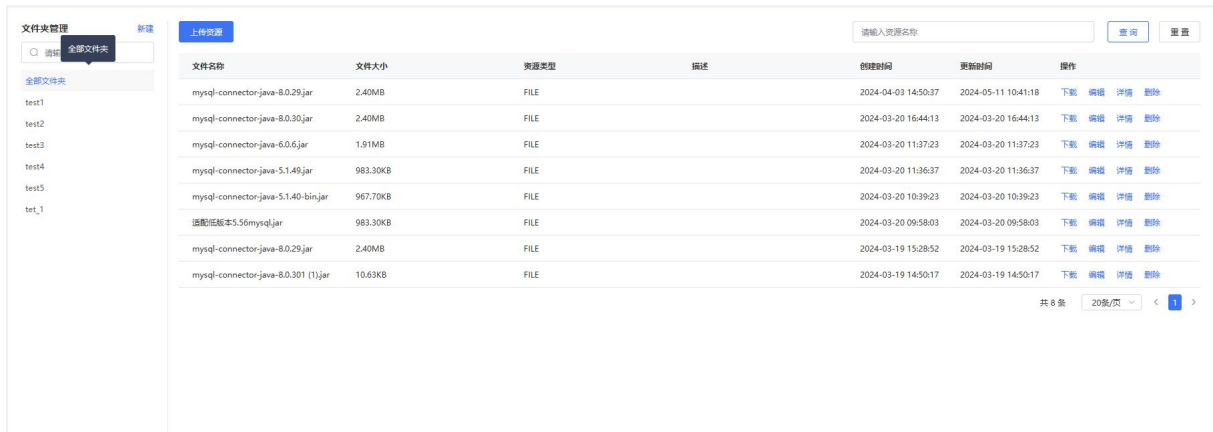
特别的，针对 `yyyyMMdd` 格式的内置变量，您可以在日期格式间使用或者不使用 `/`，`-` 符号。



内置变量	SQL 表达式	描述
内置变量1	<code>\${add_months(yyyyMMdd,12*N)}</code>	后 N 年
内置变量2	<code>\${add_months(yyyyMMdd,-12*N)}</code>	前 N 年
内置变量3	<code>\${add_months(yyyyMMdd,N)}</code>	后 N 月
内置变量4	<code>\${add_months(yyyyMMdd,-N)}</code>	前 N 月
内置变量5	<code>\${yyyyMMdd+7*N}</code>	后 N 周
内置变量6	<code>\${yyyyMMdd-7*N}</code>	前 N 周
内置变量7	<code>\${yyyyMMdd+N}</code>	后 N 天
内置变量8	<code>\${yyyyMMdd-N}</code>	前 N 天
内置变量9	<code>\$(H:mm:ss+N/24)</code>	后 N 小时
内置变量10	<code>\$(H:mm:ss-N/24)</code>	前 N 小时
内置变量11	<code>\$(H:mm:ss+N/24/60)</code>	后 N 分钟
内置变量12	<code>\$(H:mm:ss-N/24/60)</code>	前 N 分钟

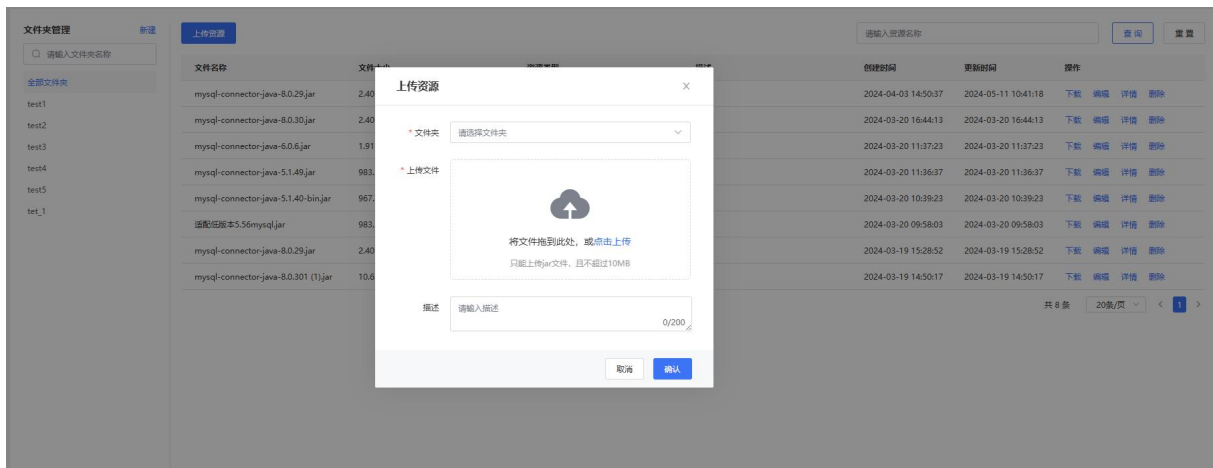
文件管理

您可以在文件管理中上传、管理文件资源。



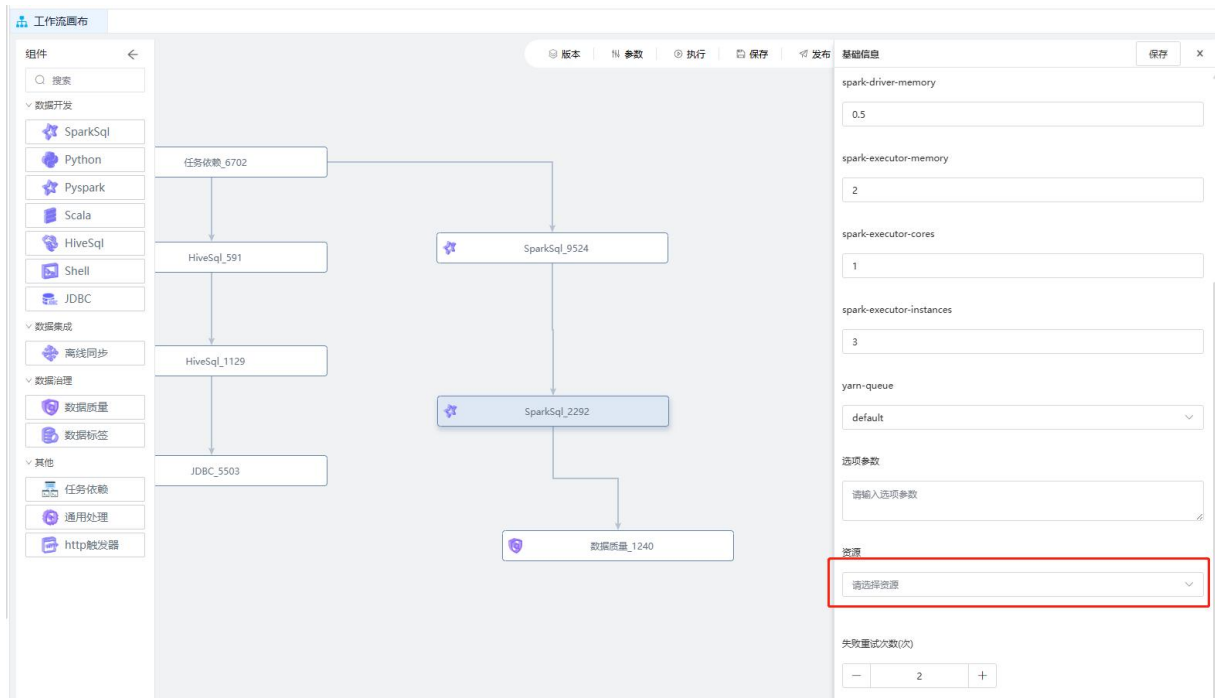
上传资源

当前仅支持上传 jar 包资源。



文件资源使用

在数据开发 workflow 画布中，单击 SparkSql/Scala 节点，您可以选择加载文件资源，保存后，执行当前任务时，会引用已上传的资源执行大数据计算任务。



UDF/函数管理

资源管理

资源管理专门用来存放 UDF 的资源 jar 包，您可以在资源管理里上传、下载、编辑、删除 UDF 资源 jar 包。UDF 资源上传操作和文件管理资源上传操作一致。



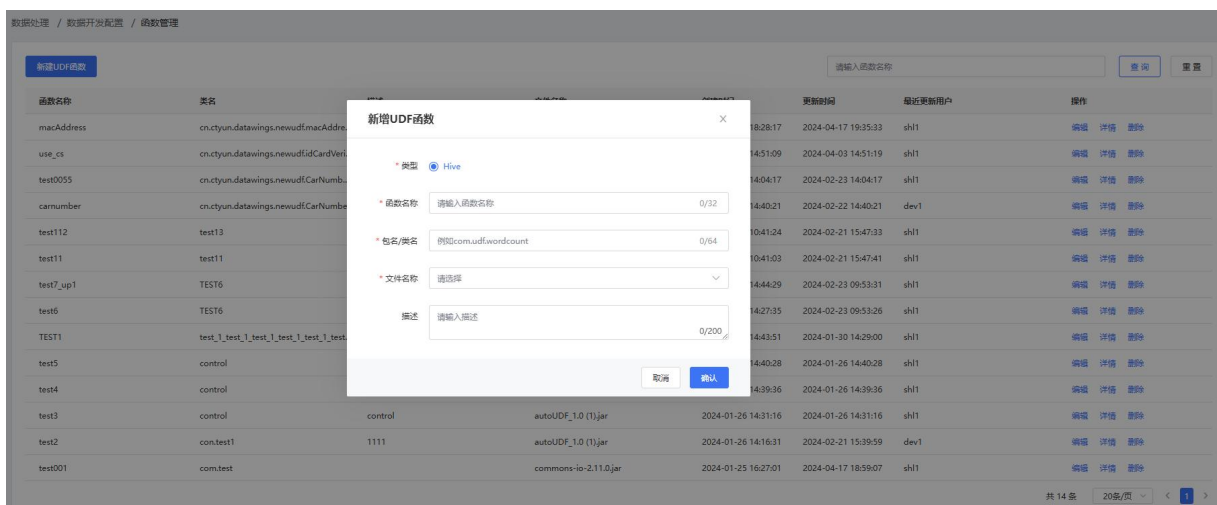
函数管理

函数管理用于已创建的 UDF 函数，包括新建、编辑、查看、删除 UDF 函数。

函数名称	类名	描述	文件名称	创建时间	更新时间	最近更新用户	操作
macAddress	cn.ctyun.datawings.newudf.macAddre...		datawings.jar	2024-04-17 18:28:17	2024-04-17 19:35:33	sh11	编辑 详情 删除
use_cs	cn.ctyun.datawings.newudf.CardVeri...	测试专用	datawings.jar	2024-04-03 14:51:09	2024-04-03 14:51:19	sh11	编辑 详情 删除
test0055	cn.ctyun.datawings.newudf.CarNumb...	111	datawings-udf01.jar	2024-02-23 14:04:17	2024-02-23 14:04:17	sh11	编辑 详情 删除
carnumber	cn.ctyun.datawings.newudf.CarNumber		datawings.jar	2024-02-22 14:40:21	2024-02-22 14:40:21	dev1	编辑 详情 删除
test112	test13		datawings.jar	2024-02-21 10:41:24	2024-02-21 15:47:33	sh11	编辑 详情 删除
test11	test11		datawings.jar	2024-02-21 10:41:03	2024-02-21 15:47:41	sh11	编辑 详情 删除
test7_up1	TEST6	test6	commons-io-2.11.0.jar	2024-01-30 14:44:29	2024-02-23 09:53:31	sh11	编辑 详情 删除
test6	TEST6	TEST6TEST6TEST61	autoUDF_1.0 (1).jar	2024-01-30 14:27:35	2024-02-23 09:53:26	sh11	编辑 详情 删除
TEST1	test_1_test_1_test_1_test_1_test...		autoUDF_1.0 (1).jar	2024-01-26 14:43:51	2024-01-30 14:29:00	sh11	编辑 详情 删除
test5	control		autoUDF_1.0 (1).jar	2024-01-26 14:40:28	2024-01-26 14:40:28	sh11	编辑 详情 删除
test4	control	control	autoUDF_1.0 (1).jar	2024-01-26 14:39:36	2024-01-26 14:39:36	sh11	编辑 详情 删除

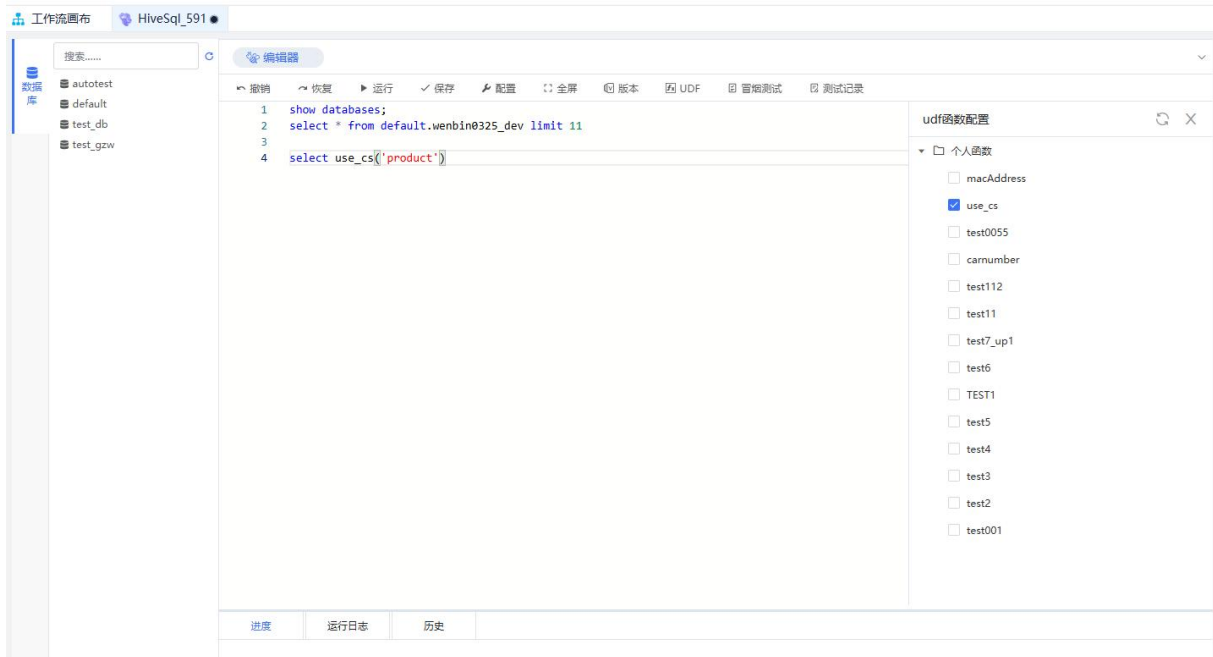
新建 UDF 函数

新建 UDF 函数需要您填写自定义函数名、包名/类名，选择引用的 UDF 资源名称信息。



使用 UDF 函数

UDF 函数创建后，您可以在数据开发 workflow 节点/即席查询脚本中使用 UDF 函数，使用前您需要在工作栏点击配置中先选中需要的 UDF 函数。




4.6 数据治理

4.6.1 数据调研

数据调研支持在数据接入前进行全方位的数据资产盘点，提供资产目录配置、模板配置、录入数据等功能。

操作步骤


1. 进入数据调研。

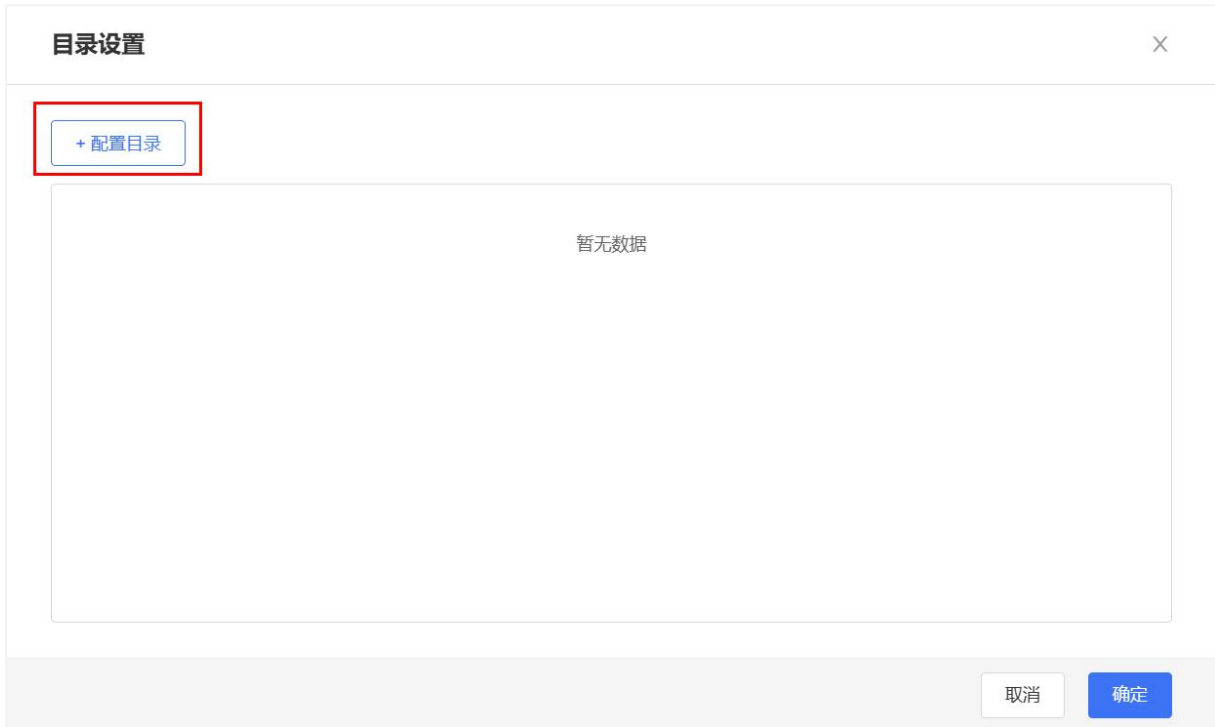
点击左上方的，选择全部产品 > 数据治理 > 数据调研，进入数据调研。

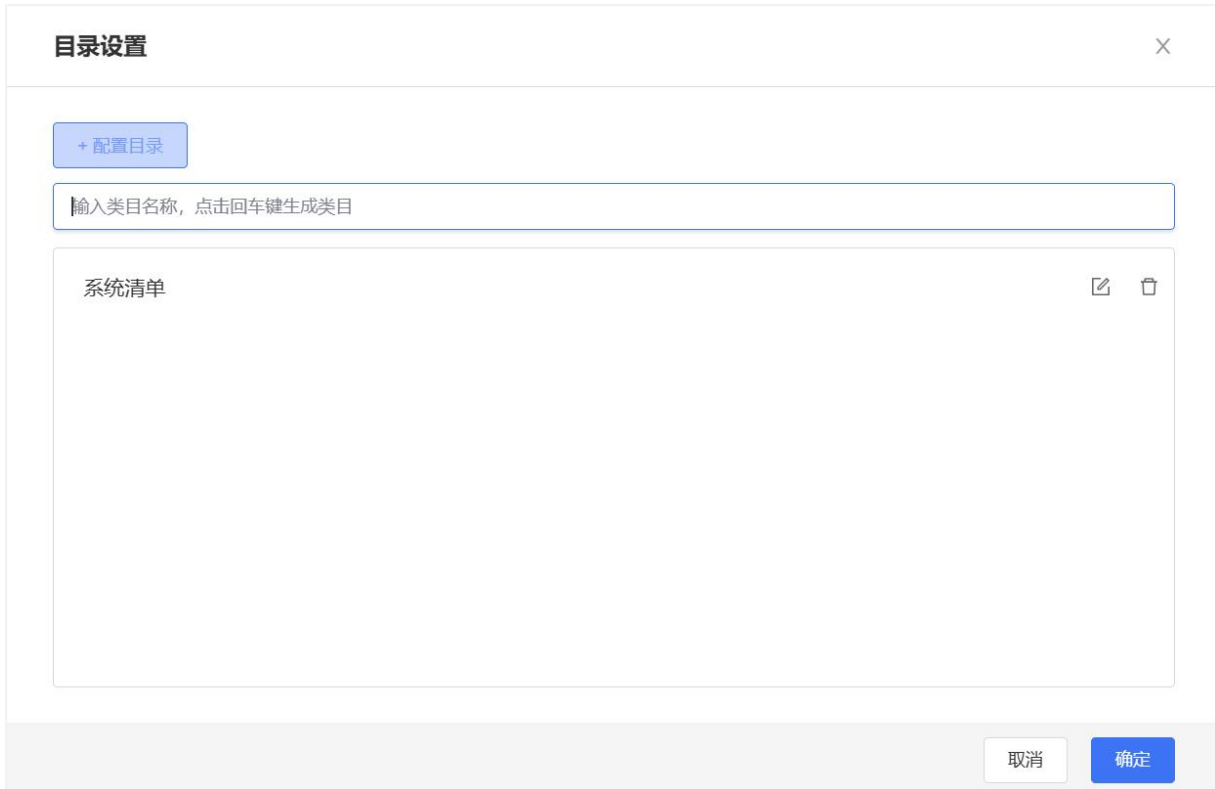
2. 添加目录。

a. 点击 ，进行目录配置。



b. 点击  ，在输入框输入目录名称，点击回车键生成类目预览。点击“确定”保存目录配置。





3. 设置模板。点击 **设置模版** 进入设置模板页面。



a. 点击 **新建** 按钮，进入新建参数弹窗。



b. 填写参数。

设置
✕

* 显示名称

* 输入类型

默认值

* 编辑表单 权限

数据校验

* 英文名称

提示文本

* 新建表单 权限

展示与筛选设置

是否在数据列表展示 展示顺序

是否作为筛选条件 筛选顺序

是否作为必选条件 ?

跳转页面

跳转字段

参数	描述
显示名称	用于在列表中显示的列名。
英文名称	用于列名的唯一性判断的参数。
输入类型	<p>单选项，可按实际情况选择。</p> <ul style="list-style-type: none"> 下拉框：跟“数据来源”联动，可选择“自定义”或“关联已有数据”作为选项值。 <p>数据来源有两种方式：</p> <p>自定义：通过“添加新选项”按钮添加选项作为下拉框选项值。选项内容</p>

参数	描述
	<p>手动输入合法字符即可。</p> <p>关联已有数据：通过选择“来源页面”、“来源字段”作为下拉框的选项值。</p> <ul style="list-style-type: none"> ● 输入框：如果参数的类型是短文本，可选择此类型。 ● 多行输入框：如果参数的类型是长文本，可选择此类型。
编辑表单权限	单选项。可选项为：可编辑、只读。默认为可编辑。
新建表单权限	单选项。可选项为：可编辑、只读。默认为可编辑。
数据校验开关	开关默认为关闭。有如下校验项，按需勾选。
展示与筛选设置	<ul style="list-style-type: none"> ● 是否在数据列表展示，默认为勾选，可按需配置。 ● 是否作为筛选项，默认为不勾选，可按需配置。 ● 是否作为必选条件：是指查询或筛选时，必须输入或选择具体的内容，否则页面不展示内容。例如按表名查询表结构，则必须输入表名。
跳转页面	可跳转的页面。例如从表清单页面跳转到单表的表结构页面。需先配置好表结构的目录，在表清单页面选择表结构页面。
跳转字段	通过设置该字段，可点击该列名，即可跳转到对应页面，并按照该字段值进行筛选。例如从表清单的表中文名跳转到该表的表结构页面。

c. 点击  按钮，保存模板。

← 返回 | 设置模版 批量删除 新建

<input type="checkbox"/>	序号	显示名称	英文名称	输入类型	是否在数据列表展示	展示顺序	是否作为筛选条件	筛选顺序	是否数据校验	编辑表单权限	新建表单权限	操作
<input type="checkbox"/>	1	业务系统名称	ywxtmc	输入框	是	1	否	1	否	可编辑	可编辑	删除 设置
<input type="checkbox"/>	2	系统功能描述	gnms	输入框	是	2	否	1	否	可编辑	可编辑	删除 设置
<input type="checkbox"/>	3	部门联系方式	bmlxfs	输入框	是	3	否	1	否	可编辑	可编辑	删除 设置
<input type="checkbox"/>	4	数据存储类型	sjcck	输入框	是	4	否	1	否	可编辑	可编辑	删除 设置

取消 保存

4. 录入数据。

a. 点击 录入数据 按钮进行录入数据。



← 返回 | 录入数据

业务系统名称

系统功能描述

部门联系方式

数据存储类型

提交

b. 点击 提交 按钮进行数据提交。

4.6.2 数据质量

4.5.2.1 数据质量概述

数据质量管理是指通过建立数据质量评估标准和管理规范，及时发现、监测定位、跟踪解决各类数据质量问题，形成数据质量问题的闭环处理，以保证数据质量的稳定可靠。平台支持定义数据全生命周期的质量监测、实时质量分析、数据质量报告等相关策略。

使用前提

1. 您需购买全生命周期数据自治及以上版本才可以使用数据质量相关功能。
2. 在进行数据质量规则配置前，请确保数据源已在数据源管理创建成功。
3. 创建数据质量规则后，您需要前往数据开发-工作流中配置数据质量节点才可以周期调度执行数据质量规则。

使用场景

在离线数据质量稽核场景下，通过创建的表-分区表达式来匹配数据开发每天产出的表分区数据，通过在数据开发的工作流中嵌入数据质量节点，当上游任务执行结束后会自动触发数据质量稽核校验，您可以设置数据质量规则的强弱属性来控制节点的成功或失败，从而减少和避免脏数据进一步污染下游，您还可以通过规则配置告警信息，第一时间接收质量告警并处理。

配置规则

1. 创建规则：您可以在数据质量模块创建单表的质检规则，也可以通过模板列表批量创建质检规则。
2. 配置告警策略：您可以在配置质检规则时配置告警信息，支持发送告警信息至邮件、站内信、短信、企微机器人、飞书机器人、钉钉机器人等渠道。

质检结果

您可以在运维管理中查看已执行的质检规则的稽核检验结果。

- 查看质检日志

质检任务开始运行后，支持实时查看质检规则运行的日志信息。

实例列表
实例ID:3872日志 x

数据源	test_engine_reconfig_3_HIVE	库名	default
表名	zzf_test_stardand_20240318_001	分区表达式	zhidiaobaozun005_dt='20240521'
执行时间	2024-05-21 15:25:50	调度时间	2024-05-21 15:25:48

```

[INFO] 2024-05-21 15:25:59.701 +0800 - TenantCode:test_tenant_1 check success
[INFO] 2024-05-21 15:25:59.701 +0800 - ProcessExecDir/tmp/dolphinscheduler/exec/process/test_tenant_1/12348965832704/13676282766000_0/1062009/89859 check success
[INFO] 2024-05-21 15:25:59.701 +0800 - Resources:{} check success
[INFO] 2024-05-21 15:25:59.701 +0800 - Task plugin: SHELL create success
[INFO] 2024-05-21 15:25:59.702 +0800 - shell task params ["rawScript": "\n /tmp/test_tenant_1/2024-05-21\n /tmp/test_tenant_1/2024-05-21/790d8235-baae-4b06-bb5e-e9277040a791.result\nkeytab=/etc/security/keytabs/hive.keytab\nproxyUser=hive/mozi-metastore-2@datawings.newtest.cn\n #[(keytab) #[(proxyUser)\nspark-sql --master yarn --hiveconf hive.cli.print.header=true --executor-memory 2G --num-executors 3 --executor-cores 1 --driver-memory 512M --driver-cores 1 --proxy-user test_ldap_3 --queue default --name test_ldap_3_spark-sql -e "\n ;select count(*) from default.zzf_test_stardand_20240318_001 where zhidiaobaozun005_dt='20240521' and 1=1;;select count(*) from default.zzf_test_stardand_20240318_001 where zhidiaobaozun005_dt='20240521' and 1=1;]\n05-21/790d8235-baae-4b06-bb5e-e9277040a791.result","localParams":{"}]]]
[INFO] 2024-05-21 15:25:59.702 +0800 - Success initialized task plugin instance success
[INFO] 2024-05-21 15:25:59.702 +0800 - Success set taskVarPool: null
[INFO] 2024-05-21 15:25:59.703 +0800 - raw script:

```

- 查看质检报告

质检任务运行结束后，您可以查看每个规则的质检结果报告，包括执行状态描述任务是否执行成功，质检状态描述结果是否符合您的预期。

实例列表
实例ID:3872质检报告 x

数据源	test_engine_reconfig_3_HIVE	库名	default
表名	zzf_test_stardand_20240318_001	分区表达式	zhidiaobaozun005_dt='20240521'
执行时间	2024-05-21 15:25:50	调度时间	2024-05-21 15:25:48

规则名称	检测类型	检测方式	检查数	波动值 (橙色预警)	波动值 (红色预警)	比较方式	期望值	采样结果	历史结果	质检结果	是否告警	是否命中抑制规则	执行状态	质检状态
zzf_test_stardand_20240318_001_表行数, 固定值_2024/04/26 16:56:15	数值型	与固定值比较	0	-	-	等于	58985.0	0.00	-	0.00	是	否	成功	红色异常

共1条 | 20 | < | 1 | > | 前往 | 1 | 页

- 查看告警记录

质检任务运行结束后，您可以查看每个规则的告警情况，以及告警异常信息。

实例列表 实例ID:3864告警记录 x

规则名称 zzf_test_standard_20240

基本总结

告警时间	2024-05-21 10:27:36	业务来源	data_quality
数据来源	zzf_test_standard_20240318_002_表行数, 固定值_2024/04/26 16:56:15	告警结果	成功
告警通道	企业机器人, 站内信	告警等级	警告

日志查询

告警通道 请选择告警通道 告警对象 请输入告警对象 日志内容 请输入日志内容 查询 重置

仅显示异常

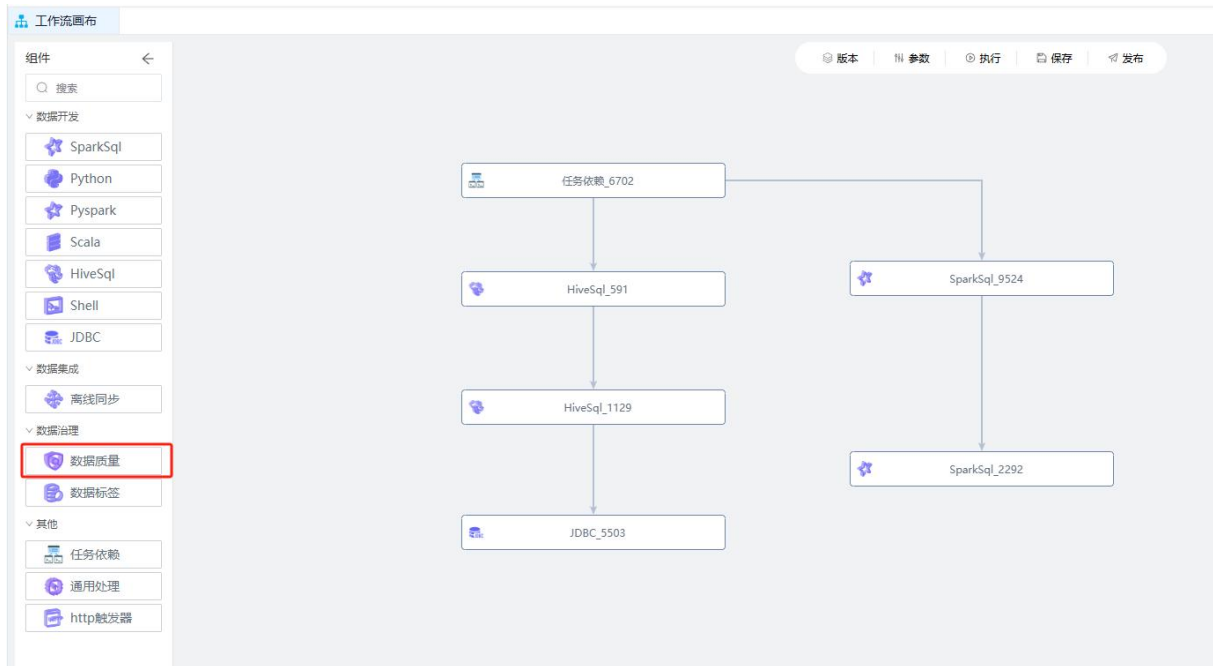
- > 告警通知: 通过企业机器人通知了测试环境-数据服务告警 企业机器人
- > 告警通知: 通过站内信通知了zzf1 (zzf1) 站内信

数据开发-数据质量节点

数据质量模块内支持创建和管理数据质量规则，如果您需要周期性跟随数据开发任务执行数据质量任务，您需要在数据开发的工作流中创建数据质量节点。

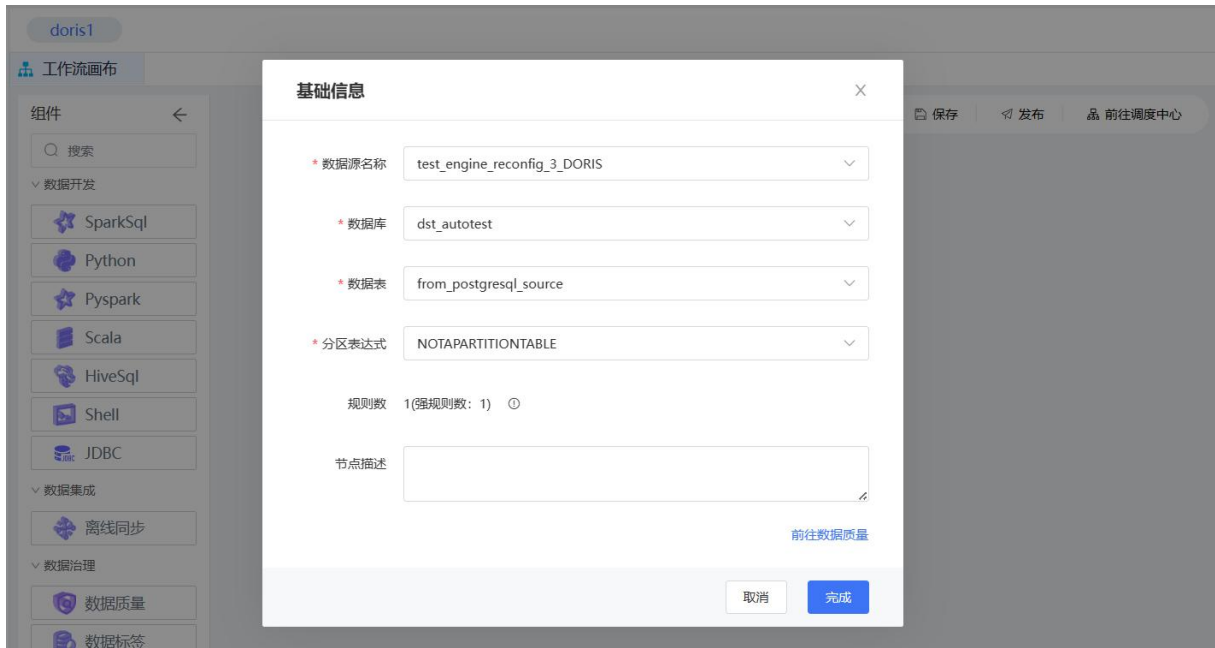
操作步骤

1. 打开数据处理-数据开发，进入数据开发页面。
2. 创建或选择一个项目。
3. 创建或选择一个工作流，并进入工作流画布页面。
4. 拖拽一个数据质量节点。



双击节点，可配置数据质量节点信息。

功能	说明
数据源名称	选择需要周期执行任务的数据源，当前仅支持内置数据源。
数据库	选择当前数据源下，已配置规则的数据库。
数据表	选择当前数据库下，已配置规则的数据表。
分区表达式	选择当前数据表已创建的分区表达式。
规则数	当前分区表达式下已上线的规则数。



4.5.2.2 数据质量大屏

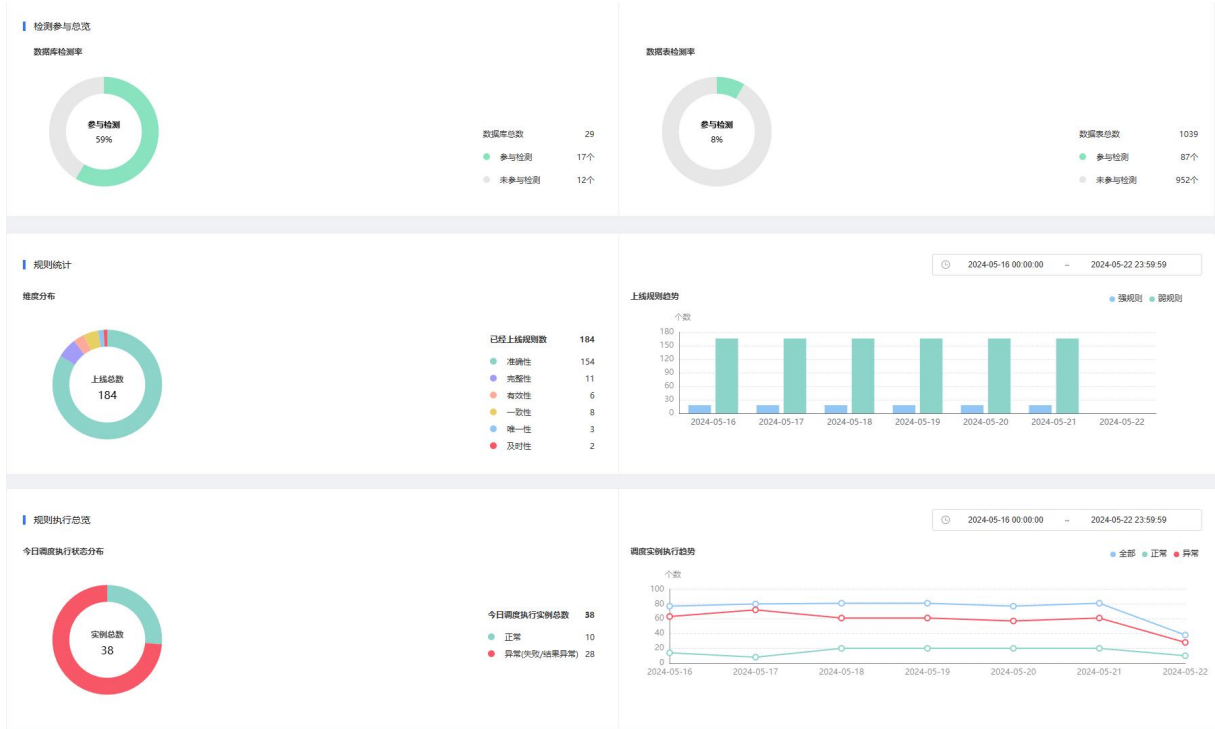
数据质量大屏为您展示数据质量的总览指标和趋势信息，包括资产配置质检稽核规则的覆盖情况、质检任务执行后告警统计及排行信息等。

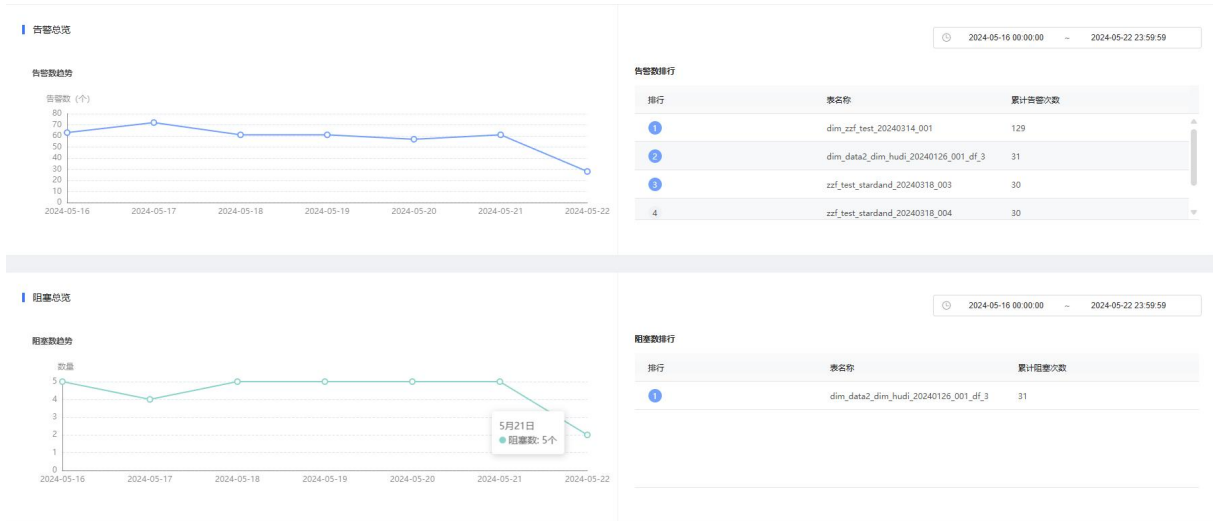
查看数据质量大屏

1. 点击数据治理-数据质量-质量大屏即可进入。
2. 以下是质量大屏统计信息描述。

标签	描述
检测参与总览	统计当前数据资产下数据库、数据表的已配置质检规则的数量和比例。
规则统计	统计上线规则维度分布及上线规则数日趋势。
规则执行总览	统计当日调度执行状态分布、按日统计调度执行实例数趋势。

标签	描述
告警总览	统计告警数趋势及表告警数排行。
阻塞总览	统计阻塞趋势及表阻塞排行。





4.5.2.3 规则管理

4.5.2.3.1 数据质量规则列表

规则列表页面按库表、质检规则颗粒度层级展示数据质量规则，用于快速检索所需查看库表的质检规则，支持新增、修改、停用、上下线切换和查询功能。

查看规则列表

- 点击数据治理-数据质量进入数据质量模块。
- 点击规则管理即可查看数据质量规则列表。
- 以下是规则管理列表页面的功能点及其描述。

功能点	描述
查看	点击进入该规则的查看页面。
修改	点击进入该规则的编辑页面，仅下线的规则支持编辑。

功能点	描述
下线	下线该规则，下线后当前规则无法使用和调度运行，可编辑和查看。
上线	上线该规则，上线后当前规则不可编辑和删除。
删除	点击更多>删除，可删除该规则。
管理分区表达式	点击打开当前数据表的分区表达式列表管理。
试运行	点击更多>试运行，打开数据质量规则试运行的配置弹窗。
配置 workflow	点击后跳转到数据处理-数据开发页面，此页面可配置数据质量节点。

test_engine_reconfiq_3_HIVE
库名
表名

分区表达式
校验类型
状态

规则名称
luowenbin
强规则

查询 重置

+ 新建规则

数据库	表名	规则数量	最后质检时间	操作
test_gzw	dim_zzf_test_20240314_007	1		管理分区表达式 更多
分区表达式	规则名称	规则字段	规则模板	状态
	dt=\${yyyyMMdd-1d]	dim_zzf_test_20240314_007_...	order_amount	内置模板
				汇总值、1天滚动率
				● 上线
				查看 修改 下线 删除
default	dim_zzf_test_20240314_001	1		管理分区表达式 更多
test_gzw	test01	1		管理分区表达式 更多

共3条 20 < > 前往 1 页

4.5.2.3.2 创建数据质量规则

本节为您介绍如何创建数据质量规则。

使用限制

- 当前数据质量规则仅支持内置 Hive 和内置 Doris 数据源。

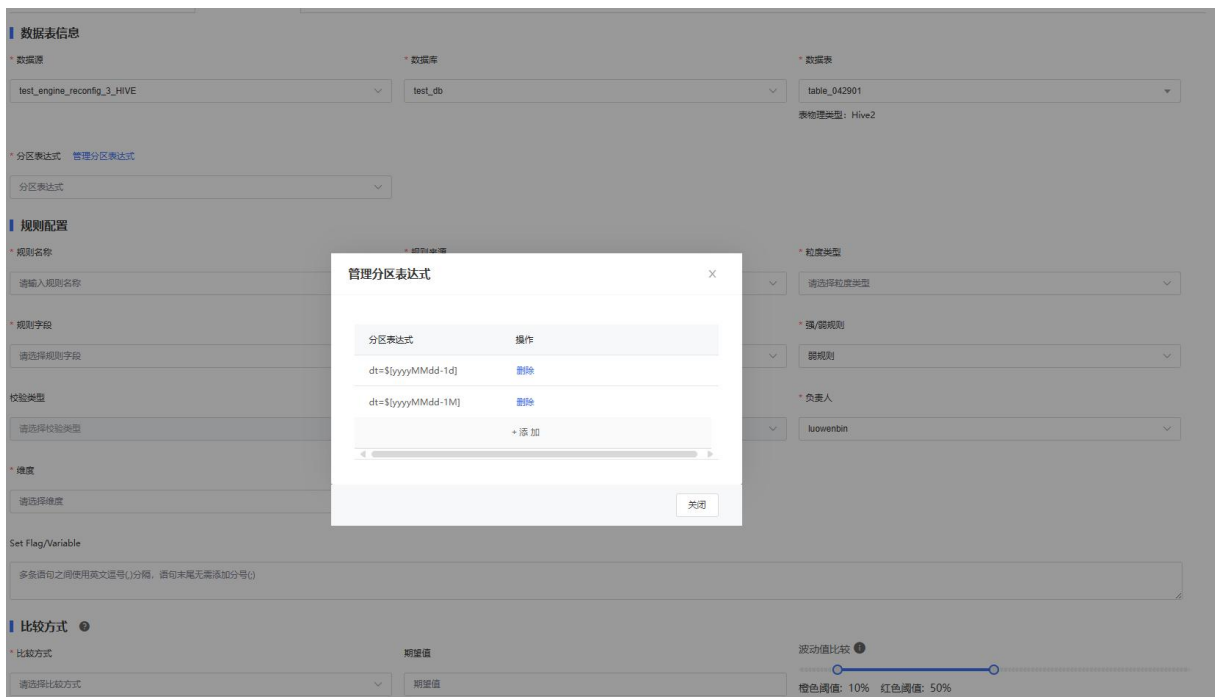
- 创建数据质量规则后，您需要在数据开发-工作流中创建数据质量节点才可以周期调度。
- 部分数据质量规则模板需要先维护数据标准（标准值域、数据格式等），若您购买的版本不包含数据建模模块，对应的规则模板也无法使用。

步骤

1. 点击数据治理-数据质量-规则管理进入数据质量规则列表页。
2. 点击【新建规则】，即可创建配置数据质量规则。

创建分区表达式

大数据场景下，为避免全表扫描，提升任务执行效率，数据质量基于分区表达式来创建管理数据质量规则，用于稽核校验指定分区的数据质量，配置规则前，您可以点击管理分区表达式来创建或者删除分区表达式。



DataWings 为您提供了一些内置的分区表达式，可直接使用，若内置表达式无法满足您的需要，则可根据需要输入自定义表达式。

配置分区表达式后，点击【计算】，数据质量会以当前时间作为调度时间，计算出分区表达式的实际结果，用来验证分区表达式是否符合您的预期。

内置分区表达式及其解释

分区表达式	解释
dt=\${yyyyMMdd}	表示分区时间取调度时间当天
dt=\${yyyyMMdd-1d}	表示分区时间取调度时间前一天
dt=\${yyyyMMdd-1M}	表示分区时间取调度时间前一月
dt=\${yyyyMMddHHmmss}	表示分区时间取调度时间(秒)
dt=\${yyyyMMddHHmmss-1d}	表示分区时间取调度时间前一天(秒)
dt=\${yyyyMMddHHmmss-1M}	表示分区时间取调度时间前一月(秒)
dt=\${yyyyMMddHHmmss-1H}	表示分区时间取调度时间前一小时(秒)
dt=\${yyyyMMddHHmmss-30m}	表示分区时间取调度时间前 30 分钟(秒)
dt=\${HHmmss-1H}	表示分区时间取调度时间前一小时(仅时分秒)
dt=\${HHmmss-30m}	表示分区时间取调度时间前 30 分钟(仅时分秒)

创建数据质量规则

创建数据质量规则需要分别填写数据表信息、规则配置、比较方式、告警配置等信息。

离线规则列表
流式规则列表
新建离线规则 x

数据表信息

* 数据源
test_engine_reconfiq_3_HIVE

* 数据库
test_db

* 数据表
table_042901
表物理类型: Hive2

* 分区表达式 [管理分区表达式](#)
dt=\${yyyyMMdd-1d}

规则配置

* 规则名称
请输入规则名称

* 规则来源
内置模板

* 粒度类型
字段粒度

* 规则字段
id(int) x

* 规则模板
重复值个数/总行数, 固定值

* 强/弱规则
弱规则

校验类型
数值型

校验方式
与固定值比较

* 负责人
luowenbin

* 维度
准确性

Set Flag/Variable
多条语句之间使用英文逗号(,)分隔, 语句末尾无需添加分号()

比较方式

* 比较方式
请选择比较方式

期望值
期望值

告警配置

联系人
请搜索联系人并选择通知方式

机器人
请选择群机器人

备注

请输入备注

配置信息解释如下：

配置项	解释说明
数据源	选择数据源名称。
数据库	选择已选择数据源下的数据库名称。
数据表	选择已选择数据库下的数据表名称。

配置项	解释说明
分区表达式	选择已创建的分区表达式，分区表达式文档参考上一节。
规则名称	输入当前规则的规则名称。
规则来源	选择内置模板或自定义。
粒度类型	如果规则来源为内置模板，则包括表粒度和字段粒度；如果规则来源为自定义，则包括表粒度、字段粒度、sql 解析。
规则字段	选择字段粒度时需要选择此信息，可单选或多选字段加入质量稽核规则。
规则模板	支持下拉选择系统内置的数据质量模板，详细模板信息请参考 模板管理 。
采样方式	-粒度类型为表级：可选 count 和 count/table -粒度类型为字段级：可选 sum、max、min、avg。
强 / 弱	强规则在数据开发的质量节点运用中可阻塞数据处理流程，弱规则不会阻塞 workflow。

配置项	解释说明
规则	
校验类型	支持选择数值型、波动型，不同类型的规则比较方式不同。
校验方式	<p>下拉单选，按校验类型的数值型和波动型显示不同内容，详情如下：</p> <p>数值型：与固定值比较、上一周期差值、1天差值。</p> <p>波动型：7天平均值波动、30天平均值波动、1天波动、7天波动、30天波动、7天标准差波动、1,7,30天波动、上一周期波动率、1,7,30，本月1号波动率。</p>
负责人	规则负责人，提交质检规则时，会校验规则负责人是否拥有当前数据表的数据权限。
维度	当前数据质量规则评估类型(准确性、完整性、一致性、唯一性、有效性、及时性)。
过滤条件	您可以输入自定义 sql 过滤条件，多条语句之间使用英文逗号分隔，语句末尾无需添加分号。
Set Flag	您可以输入自定义 SetFlag 语句，多条语句之间使用英文逗号分隔，语句末尾无需添加分号。
自定义 sql	当粒度类型是自定义 sql 时，支持输入 sql 创建数据质量规则。

配置项	解释说明
比较方式	根据校验类型选择不同，比较方式的选择范围不同，详情如下-数值型：选择范围为（大于、大于等于、等于、不等于、小于等于、小于）-波动型：选择范围为（绝对值、上升、下降）。
期望值	当校验方式为数值型时显示，仅支持输入正数和零。
波动值	当校验方式为波动型时显示，可设报警阈值并在右侧可视化显示；（波动率=（样本-基准值）/基准值）。
联系人	选择告警通知的联系人及告警通道（可在运维中心配置）。
机器人	选择告警通知的机器人（可在运维中心配置）。

后续步骤

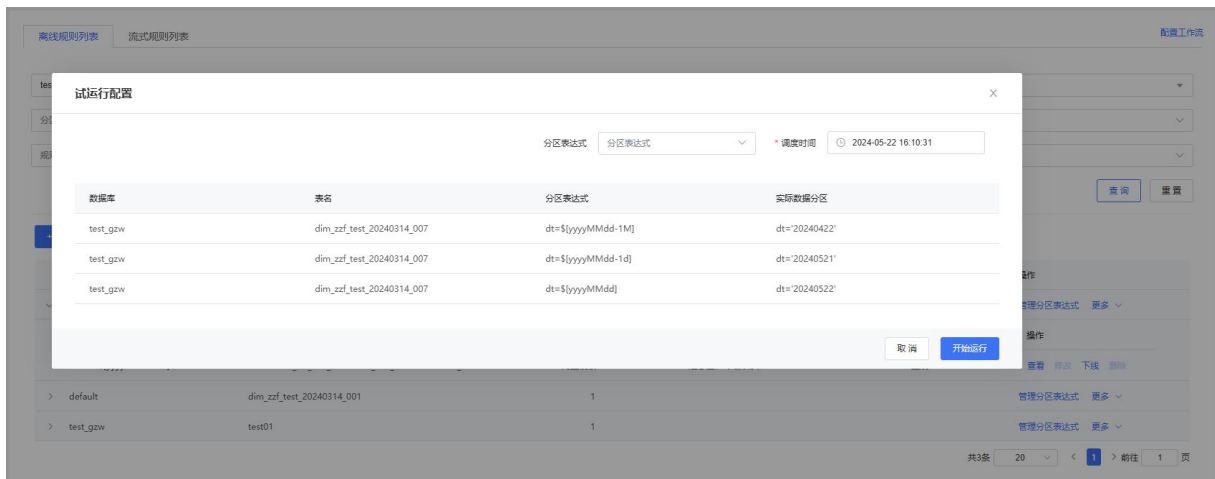
- 创建数据质量规则后，您需要前往数据开发-工作流中创建数据质量节点才可以周期调度。
- 如果您希望不符合预期的数据质量任务不阻塞数据开发工作流，可以将规则设定为弱规则或者期望抑制管理增加抑制策略，命中抑制策略的规则将不触发告警和阻塞逻辑。

4.5.2.3.3 试运行质量规则

数据质量规则完成创建后，您可以试运行数据质量规则并查看质检结果，确认试运行的结果是否符合您的预期，如果不符合您的预期，您可以针对当前规则再做适当的调整。

操作步骤

1. 点击数据治理-数据质量-规则管理进入数据质量规则列表页。
2. 选择您想试运行的数据表，点击更多-试运行。
3. 选择要试运行的分区表达式，支持选择多个。
4. 选择指定的调度时间，默认是当前时间，选择不同调度时间后，会实时显示实际数据分区。



4.5.2.4 模板管理

DataWings 内置提供了 40+ 字段粒度和表粒度的数据质量稽核规则模板，您可以在模板管理页面查看内置模板及其说明，并且支持批量按模板创建质检规则。

进入模板列表页面

1. 点击数据治理-数据质量，进入数据质量模块。
2. 点击模板管理，打开内置数据质量模板列表页面。

数据治理 / 数据质量 / 模板管理

模板列表

粒度类型: 粒度类型 查询 重置

模板名称	维度	粒度类型	模板说明	操作
1 汇总值, 1、7、30天波动率	准确性	字段粒度	取该字段的sum值, 同1天、7天和1个月前的样本(字段平均值)进行比较, 计算波动率, 再与阈值进行比较, 只要有一个不符合规则即可触发报警。	前往配置
2 汇总值, 1天波动率	准确性	字段粒度	取该字段的sum值, 与前1天进行比较, 计算出波动率后, 再与阈值进行比较。	前往配置
3 汇总值, 上周期的波动率	准确性	字段粒度	取该字段的sum值, 与上一周期进行比较, 计算出波动率后, 再与阈值进行比较, 只要有一个不符合规则即可触发报警。	前往配置
4 汇总值, 7天标准差波动率	准确性	字段粒度	取该字段的汇总值, 计算最近7天的平均值和标准差, 再计算(当前值-最近7天平均值)/标准差, 仅数值类型的字段可以使用	前往配置
5 最大值, 1、7、30天波动率	准确性	字段粒度	取该字段的最大值, 同1天、7天和1个月前的样本(字段平均值)进行比较, 计算波动率, 再与阈值进行比较, 只要有一个不符合规则即可触发报警。	前往配置
6 最大值, 1天波动率	准确性	字段粒度	取该字段的最大值, 与前1天进行比较, 计算出波动率后, 再与阈值进行比较。	前往配置
7 最大值, 上周期的波动率	准确性	字段粒度	取该字段的最大值, 与上一周期进行比较, 计算出波动率后, 再与阈值进行比较, 只要有一个不符合规则即可触发报警。	前往配置
8 最小值, 1、7、30天波动率	准确性	字段粒度	取该字段的最小值, 同1天、7天和1个月前的样本(字段平均值)进行比较, 计算波动率, 再与阈值进行比较, 只要有一个不符合规则即可触发报警。	前往配置
9 最小值, 1天波动率	准确性	字段粒度	取该字段的最小值, 与前1天进行比较, 计算出波动率后, 再与阈值进行比较。	前往配置
10 最小值, 上周期的波动率	准确性	字段粒度	取该字段的最小值, 与上一周期进行比较, 计算出波动率后, 再与阈值进行比较, 只要有一个不符合规则即可触发报警。	前往配置
11 平均值, 1、7、30天波动率	准确性	字段粒度	取该字段的平均值, 同1天、7天和1个月前的样本(字段平均值)进行比较, 计算波动率, 再与阈值进行比较, 只要有一个不符合规则即可触发报警。说明: 该字段的平均值, 分别与昨天该字段平均值、7天前该字段平均值、30天前该字段平均值进行比较。	前往配置
12 平均值, 1天波动率	准确性	字段粒度	取该字段的平均值, 与前1天进行比较, 计算出波动率后, 再与阈值进行比较。	前往配置
13 表行数, 固定值	准确性	表粒度	表行数	前往配置
14 表行数, 1、7、30天波动率	准确性	表粒度	同1天、1周和1个月前采集的表行数进行比较, 对比波动率, 说明: 表的行数, 分别与昨天的样本、7天前的样本和30天前的样本来进行数对比, 计算波动率, 再与阈值进行比较, 只要其中一个波动率超过阈值就会报警	前往配置
15 表行数, 7天平均值波动率	准确性	表粒度	该模板用于监控表行数的波动情况, 基准值是最近7天的表行数的平均值, 即(7天内每天表行数之和)/7	前往配置
16 表行数, 30天平均值波动率	准确性	表粒度	该模板用于监控表行数的波动情况, 基准值是最近30天的表行数的平均值, 即(30天内每天表行数之和)/30。	前往配置
17 表行数, 1天波动率	准确性	表粒度	基准值为昨天的样本(表行数), 比较当天采集的表行数, 对比波动率, 再与阈值进行比较, 只要有一个不符合规则即可触发报警。	前往配置
18 表行数, 7天波动率	准确性	表粒度	基准值为7天前样本(表行数), 比较当天采集的表行数, 对比波动率, 再与阈值进行比较, 只要有一个不符合规则即可触发报警。	前往配置
19 表行数, 30天波动率	准确性	表粒度	基准值为30天前的样本(表行数), 比较当天采集的表行数, 对比波动率, 再与阈值进行比较, 只要有一个不符合规则即可触发报警。	前往配置
20 表行数, 1、7、30天, 本月1号, 波动率	准确性	表粒度	表行数, 与1天前的样本、7天前的样本、30天前的样本和本月1号采集样本(表行数), 进行比较, 对比波动率, 再与阈值进行比较, 只要有一个不符合规则即可触发报警。	前往配置

共 41 条 | 20 条/页 | < 1 2 3 > 前往 1 页

按模板批量创建质检规则

选择指定模板, 点击【前往配置】按钮, 即可开始批量按模板创建质检规则, 批量创建质量规则的参数详细说明可参考[创建数据质量规则](#)。

- 配置基础信息

数据源: 选择配置的数据质量规则所应用到的数据源。

规则名称: 批量创建质检规则后将按照模板生成单独的规则名称, 名称后缀默认为当前时间, 您可以编辑替换自定义的内容。

基础信息

数据源:

* 规则名称:

- 进行规则配置

规则配置内容同创建数据质量规则。

规则配置

规则来源: 内置模板 | 规则模板: 汇总值, 1, 7, 30天波动率 | * 负责人: gzw_tenantry_2

* 强/弱规则: 弱规则 | 校验类型: 波动型 | 校验方式: 1,7,30天波动

维度: 准确性

Set Flag: 多条语句之间使用英文逗号(,)分隔, 语句末尾无需添加分号(;

- 数据选择信息

新增: 点击新增按钮您可以选择当前数据源下的数据库、表字段, 并支持批量、单独配置分区表达式。

数据选择

请先添加要监控的字段或者表, 然后再设置分区表达式

[新增](#) [批量设置分区表达式](#)

<input type="checkbox"/>	字段名	所属表	所属数据库	分区表达式	操作
<input type="checkbox"/>	uv	aa_four10	test_database_1	uv=\${yyyyMMdd}	删除 设置分区表达式
<input type="checkbox"/>	yuwen	aa_four10	test_database_1	uv=\${yyyyMMdd-1d}	删除 设置分区表达式

- 配置比较方式和告警配置

比较方式和告警配置信息同创建数据质量规则。

比较方式

* 比较方式: 波动值比较

请选择比较方式:

橙色阈值: 10% 红色阈值: 50%

告警配置

联系人:

机器人:

后续步骤

1. 创建成功后您可以前往**规则管理**查看已创建的规则。
2. 创建成功后您可以调试校验已创建的规则，详细请查看**试运行质量规则**。

4.5.2.5 实例管理

实例管理页面支持查询质检规则的历史执行情况、运行日志详情和告警结果等内容。

任务执行列表

实例ID	数据源	库名	表名	分区表达式	执行方式	执行用户	执行时间	调度时间	执行结果	关联节点	规则数	告警数	操作
3924	test_engine_reconfi_3_Hi VE	default	dim_zzf_test_20240423_001	dt=\${yyyyMMdd}	周期运行	autotestuser	2024-05-22 09:49:48	2024-05-22 09:49:45	成功	dim_zzf_test_20240423_001_yyyyMMdd_autotestuser	1	1	运行日志 质检报告 告警记录 重跑 停止
3923	test_engine_reconfi_3_Hi VE	default	dim_zzf_test_20240314_001	dt=\${yyyyMMdd}	周期运行	autotestuser	2024-05-22 09:49:48	2024-05-22 09:49:45	成功	dim_zzf_test_20240314_001_yyyyMMdd_autotestuser	4	2	运行日志 质检报告 告警记录 重跑 停止
3925	test_engine_reconfi_3_Hi VE	default	dim_zzf_test_20240314_001	dt=\${yyyyMMdd-1d}	周期运行	autotestuser	2024-05-22 09:49:48	2024-05-22 09:49:45	成功	dim_zzf_test_20240314_001_yyyyMMdd_1_autotestuser	1	1	运行日志 质检报告 告警记录 重跑 停止
3922	test_engine_reconfi_3_Hi VE	test_gzw	dim_zzf_test_20240314_007	dt=\${yyyyMMdd-1d}	周期运行	autotestuser	2024-05-22 09:49:47	2024-05-22 09:49:45	成功	dim_zzf_test_20240314_007_yyyyMMdd_1_autotestuser	2	2	运行日志 质检报告 告警记录 重跑 停止
3917	test_engine_reconfi_3_Hi VE	test_db	model_one_041103	z00137=\${yyyyMMdd}	周期运行	zzfl	2024-05-22 09:49:20	2024-05-22 09:49:18	成功	model_one_041103_yyyyMMdd_zzfl	1	1	运行日志 质检报告 告警记录 重跑 停止
3914	test_engine_reconfi_3_Hi VE	test_db	nixiang0423	NOTAPARTITIONTABLE	周期运行	zzfl	2024-05-22 09:49:20	2024-05-22 09:49:18	成功	nixiang0423_zzfl	1	1	运行日志 质检报告 告警记录 重跑 停止
3919	test_engine_reconfi_3_Hi VE	test_db	test_gzw_course	NOTAPARTITIONTABLE	周期运行	zzfl	2024-05-22 09:49:20	2024-05-22 09:49:18	成功	test_gzw_course_zzfl	1	1	运行日志 质检报告 告警记录 重跑 停止

页面说明如下：

功能点	解释
操作-查询	根据搜索和选项框内容查询执行实例，支持重置查询条件。
实例 id	自增任务实例 id。
数据源	当前质检任务实例数据对应的数据源名称。
库名	当前质检任务实例数据对应的数据库名称。

功能点	解释
表名	当前质检任务实例数据对应的数据表名称。
分区表达式	当前质检任务实例数据对应的分区表达式。
执行方式	当前执行方式类型包括：试运行、周期运行、补数、重跑。
执行时间	当前质检任务实例触发时间。
调度时间	当前质检任务实例数据时间（试运行、补数可指定历史调度时间）。
执行结果	当前质检任务实例执行结果（成功/失败）。
关联节点	当前质检任务实例已关联的数据开发节点。
规则数	当前质检任务实例包含的质检规则数量。
告警数	当前质检任务实例触发的告警数量。
操作-运行日志	点击可查看当前质检任务实例的运行日志。
操作-质检报告	点击可查看当前质检任务实例的质检报告。
操作-告警记录	点击可查看当前质检任务实例的告警记录。

功能点	解释
操作-重跑	点击可重新执行当前质检实例。
操作-停止	点击可终止执行当前质检实例。

查看运行日志

实例列表 实例ID:3872日志 x

数据源	test_engine_reconfig_3_HIVE	库名	default
表名	zzf_test_stardand_20240318_001	分区表达式	zhidiaobiaozun005_dt='20240521'
执行时间	2024-05-21 15:25:50	调度时间	2024-05-21 15:25:48

```

[INFO] 2024-05-21 15:25:59.701 +0800 - TenantCode:test_tenant_1 check success
[INFO] 2024-05-21 15:25:59.701 +0800 - ProcessExecDir:/tmp/dolphinscheduler/exec/process/test_tenant_1/12348965832704/13676282766000_0/1062009/89859 check success
[INFO] 2024-05-21 15:25:59.701 +0800 - Resources:{} check success
[INFO] 2024-05-21 15:25:59.701 +0800 - Task plugin: SHELL create success
[INFO] 2024-05-21 15:25:59.702 +0800 - shell task params {"rawScript": "\n /tmp/test_tenant_1/2024-05-21\n /tmp/test_tenant_1/2024-05-21/790d8235-baae-4b06-bb5e-e9277040a791.result\nkeytab=/etc/security/keytabs/hive.keytab\nproxyUser=hive/mozl-metastore-2@datawings.newtest.cn\n #({keytab}) #({proxyUser})\nspark-sql --master yarn --hiveconf hive.cli.print.header=true --executor-memory 2G --num-executors 3 --executor-cores 1 --driver-memory 512M --driver-cores 1 --proxy-user test_ldap_3 --queue default --name test_ldap_3_spark-sql -e "\n ;select count(*) from default.zzf_test_stardand_20240318_001 where zhidiaobiaozun005_dt='20240521' and 1=1;;select count(*) from default.zzf_test_stardand_20240318_001 where zhidiaobiaozun005_dt='20240521' and 1=1;";\n05-21/790d8235-baae-4b06-bb5e-e9277040a791.result";localParams: {}"}
[INFO] 2024-05-21 15:25:59.702 +0800 - Success initialized task plugin instance success
[INFO] 2024-05-21 15:25:59.702 +0800 - Success set taskVarPool: null
[INFO] 2024-05-21 15:25:59.703 +0800 - raw script:
  
```

查看质检报告

实例列表 实例ID:3872质检报告 x

数据源	test_engine_reconfig_3_HIVE	库名	default
表名	zzf_test_stardand_20240318_001	分区表达式	zhidiaobiaozun005_dt='20240521'
执行时间	2024-05-21 15:25:50	调度时间	2024-05-21 15:25:48

规则名称	检测类型	检测方式	检查数	波动值 (橙色预警)	波动值 (红色预警)	比较方式	期望值	采样结果	历史结果	质检结果	是否告警	是否命中抑制规则	执行状态	质检状态
zzf_test_stardand_20240318_001_并行 数: 固定值_20 24/04/26 16:56:15	数值型	与固定值比较	0	-	-	等于	58885.0	0.00	-	0.00	是	否	成功	红色异常

共1条 | 20 | < 1 > 前往 1 页

查看告警记录

实例列表 实例ID:3864告警记录 x

规则名称 zzf_test_standard_20240

基本信息

告警时间	2024-05-21 10:27:36	业务来源	data_quality	
数据来源	zzf_test_standard_20240318_002_表行数_固定值_2024/04/26 16:56:15		告警结果	成功
告警通道	企业机器人站内信	告警等级	警告	

日志查询

告警通道 请选择告警通道 告警对象 请输入告警对象 日志内容 请输入日志内容 查询 重置

仅显示异常

- 告警通知: 通过企业机器人通知了测试环境-数据服务告警 企业机器人
- 告警通知: 通过站内信通知了zzf1 (zzf1) 站内信

4.5.2.6 抑制管理

数据质量模块支持对某业务时间的质检规则进行去噪抑制，在配置的业务时间范围内不触发告警以及不会阻塞下游任务。

抑制策略列表

您可以在抑制管理列表页面查看已创建的抑制策略。

状态说明

上线：已上线的抑制策略才会生效，命中抑制策略的规则在业务时间段内将不会触发告警和阻塞 workflow。

下线：已下线的抑制策略不生效，且支持重新编辑策略。

数据治理 / 数据质量 / 抑制管理

+ 新增 抑制时间段 开始日期 ~ 结束日期 查询 重置

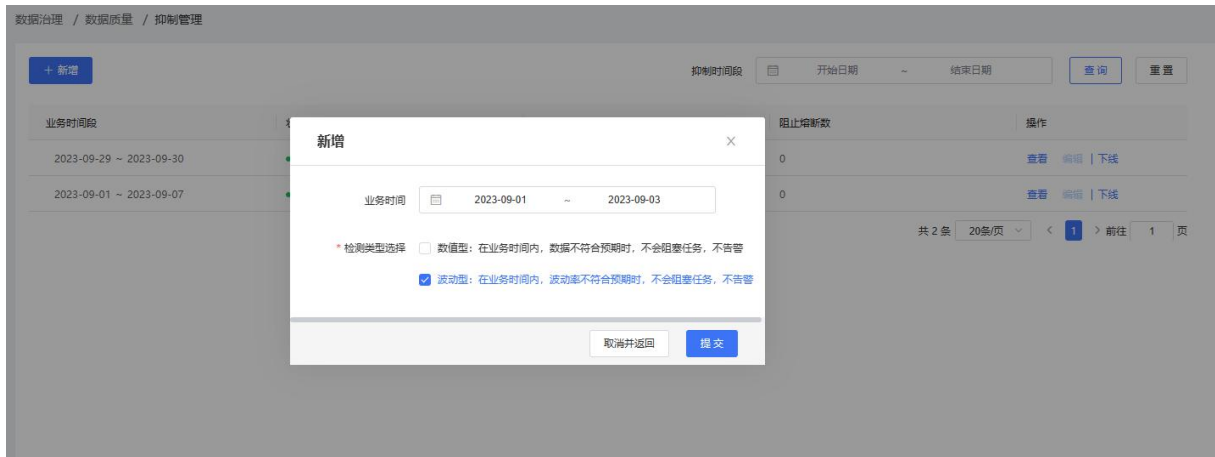
业务时间段	状态	关联任务数	阻止阻断数	操作
2023-09-29 ~ 2023-09-30	● 上线	0	0	查看 编辑 下线
2023-09-01 ~ 2023-09-07	● 上线	0	0	查看 编辑 下线

共 2 条 20条/页 < 1 > 前往 1 页

创建抑制策略

业务时间：所在业务时间的时间段，业务时间在此时间段的质检规则不会触发告警和阻塞 workflow。

检测类型：选择数值型/波动性，此类型的质检规则不会触发告警和阻塞 workflow。



4.5.2.7 质检报告

质检报告管理为您提供质量大屏数据对外触达的能力，根据配置的定时信息、质量图表内容、时间范围等发送质量报告到对应收件人的邮箱。

前提条件

1. 已在【运维中心】配置联系人信息，并配置了联系人的邮箱信息。
2. 已配置且运行质检规则，质量大屏已有可视化数据。

操作步骤

1. 点击数据治理-数据质量进入数据质量模块。
2. 点击质检报告管理-新增质检报告。

创建自定义质检报告

质检报告列表
新增质检报告 ×

基础配置

* 名称 0/30

* 数据范围 天

发送周期

内容配置

* 高线 数据库检测率 数据表检测率 规则维度分布 上线规则趋势 今日调度执行状态分布 调度实例执行趋势 告警数趋势 告警数排行 阻塞数趋势 阻塞数排行

订阅配置

* 联系人

功能	说明
名称	自定义质检报告名称。
数据范围	质量报告图表的数据范围，最大可以访问最近一年的数据。
发送周期	质量报告邮件的发送周期。
内容配置	质量图表的组件范围。
订阅配置	需要邮件触达的收件人列表。

4.5.2.8 数据质量分析

根据数据质量规则调度生成的实例，平台每日会自动统计表的数据质量情况，您可以在数据质量分析中查看表的质量报告。

质量分析列表

质量分析列表按表粒度展示已创建数据质量规则且调度运行的表的最近一次表质量分等信息，

统计周期为 T-1。

点击分析报告，即可查看当前表的详细报告，包括质量分及异常规则实例明细；

点击规则管理，即跳转到规则管理页面查看当前表的规则详情。

数据源	数据库	表名称	责任人	最近一次表质量分	统计日期	操作
test_engine_reconfg_3_HIVE	test_gzw	test01	test_ily_new2	10	2024-05-21	分析报告 规则管理
test_engine_reconfg_3_HIVE	test_gzw	dim_zzf_test_20240314_010	zzf1.userc	0	2024-05-21	分析报告 规则管理
test_engine_reconfg_3_HIVE	test_gzw	dim_zzf_test_20240314_008	zzf1	0	2024-05-21	分析报告 规则管理
test_engine_reconfg_3_HIVE	test_gzw	dim_zzf_test_20240314_007	zzf1	0	2024-05-21	分析报告 规则管理
test_engine_reconfg_3_HIVE	test_db	test_gzvw_student_course	test_ldap_3	20	2024-05-21	分析报告 规则管理
test_engine_reconfg_3_HIVE	test_db	test_gzvw_student	test_ldap_3	0	2024-05-21	分析报告 规则管理
test_engine_reconfg_3_HIVE	test_db	test_gzvw_score	test_ldap_3	20	2024-05-21	分析报告 规则管理
test_engine_reconfg_3_HIVE	test_db	test_gzvw_course	test_ldap_3	0	2024-05-21	分析报告 规则管理
test_engine_reconfg_3_HIVE	test_db	table_0507_test_copy_copy	cqj1.test_tenant_1	0	2024-05-21	分析报告 规则管理
test_engine_reconfg_3_HIVE	test_db	nixiang0423	fr1	0	2024-05-21	分析报告 规则管理
test_engine_reconfg_3_HIVE	test_db	model_one_041103_copy_copy	cqj1	0	2024-05-21	分析报告 规则管理
test_engine_reconfg_3_HIVE	test_db	model_one_041103	cqj1	0	2024-05-21	分析报告 规则管理
test_engine_reconfg_3_HIVE	default	zzf_test_stardand_20240318_004	zzf1	0	2024-05-21	分析报告 规则管理

查看分析报告

分析报告主要包含三部分内容：基本信息、近七天调度异常规则明细、表质量分历史趋势。

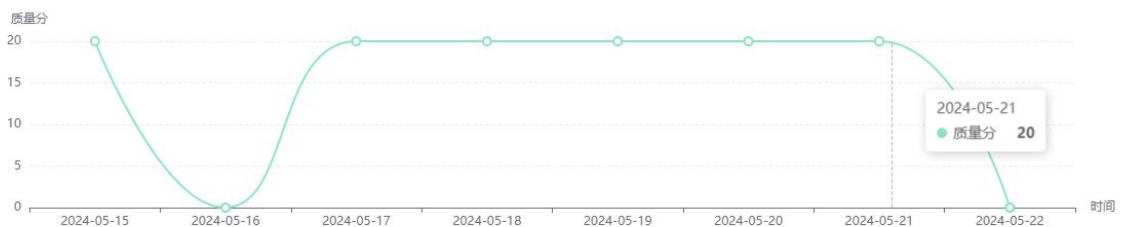
基本信息

数据源	test_engine_reconfig_3_HIVE	数据库	test_db
数据表	test_gzw_student_course	表负责人	test_ldap_3
已上线规则数	1	已上线强规则数	0

近7天调度异常规则明细

分区表达式	规则名称	维度	执行时间	调度时间	执行状态	质检状态	操作
NOTAPARTITIONTABLE	test_gzw_student_cours...	准确性	2024-05-17 09:49:20	2024-05-17 09:49:18	失败	-	查看详情
NOTAPARTITIONTABLE	test_gzw_student_cours...	准确性	2024-05-16 21:49:20	2024-05-16 21:49:18	失败	-	查看详情
NOTAPARTITIONTABLE	test_gzw_student_cours...	准确性	2024-05-15 09:49:19	2024-05-15 09:49:18	失败	-	查看详情

 共3条 < 1 > 前往 页

表质量分历史趋势
 ~


表质量分计算规则

已配置数据质量规则的表会根据维度及维度权重计算该表的数据质量评分，维度权重配置方式如下。

- 计算周期**

1) 表质量分 T-1 计算，按昨日规则实例运行状态统计表质量分。

- 计算规则**

1) 总分数=Σ (当日某维度质量规则运行成功且符合预期的实例数/当日某维度质量规则 * 某维度质量分权重)。

2) 仅统计由数据处理处触发的调度执行实例，试运行质量规则不纳入质量分统计。

维度权重配置

操作路径：点击数据治理-数据质量-配置进入数据质量配置页面，切换质量评分配置。

概览配置	质量评分配置		
修改配置			
维度	维度权重	更新时间	更新人
准确性	20	2024-05-22 08:08:32	autotestuser
完整性	20	2024-05-22 08:08:32	autotestuser
有效性	20	2024-05-22 08:08:32	autotestuser
一致性	5	2024-05-22 08:08:32	autotestuser
唯一性	10	2024-05-22 08:08:32	autotestuser
及时性	25	2024-05-22 08:08:32	autotestuser

4.6.3 数据治理 360

4.6.3.1 数据治理 360 概述

本章节主要介绍数据治理 360 的功能概述及相关概念。

功能概述

数据治理 360 通过对平台中研发、存储、质量等维度的潜在问题进行诊断，基于健康分评估模型与元数据，自动从全局与个人视角进行量化评估，给出健康分与相应治理指南，帮助用户对平台内资源的使用和价值进行有效的管理。

相关概念

健康分：根据对治理项的诊断结果，基于系统预设健康分评估模型，量化得到的平台健康程度，便于用户直观感受当前平台的健康状态与治理效果。

治理项：基于常见系统问题总结出的潜在待治理或优化问题，为健康评估提供依据。

健康指南：基于常见治理项提供的问题说明与解决方案。

白名单：无需处理的治理项问题可添加至白名单，系统将所有白名单集中至白名单页面进行统

一管理。治理项问题被加入白名单后，后续将不会作为治理项问题检测出来。

4.6.3.2 开启使用

首次进入数据治理 360 功能时，通过点击“开启使用”按钮激活功能，开启治理工作。

功能入口

单击左上方的图标，选择全部产品 > 数据治理 > 数据治理 360，进入页面。

操作步骤


新开通数据治理 360 的用户在首次进入页面时，可通过点击下图中的“开启使用”按钮激活功能，开启治理。开启使用后，该租户下的其他子账号无需再次激活，可直接进入页面进行使用。

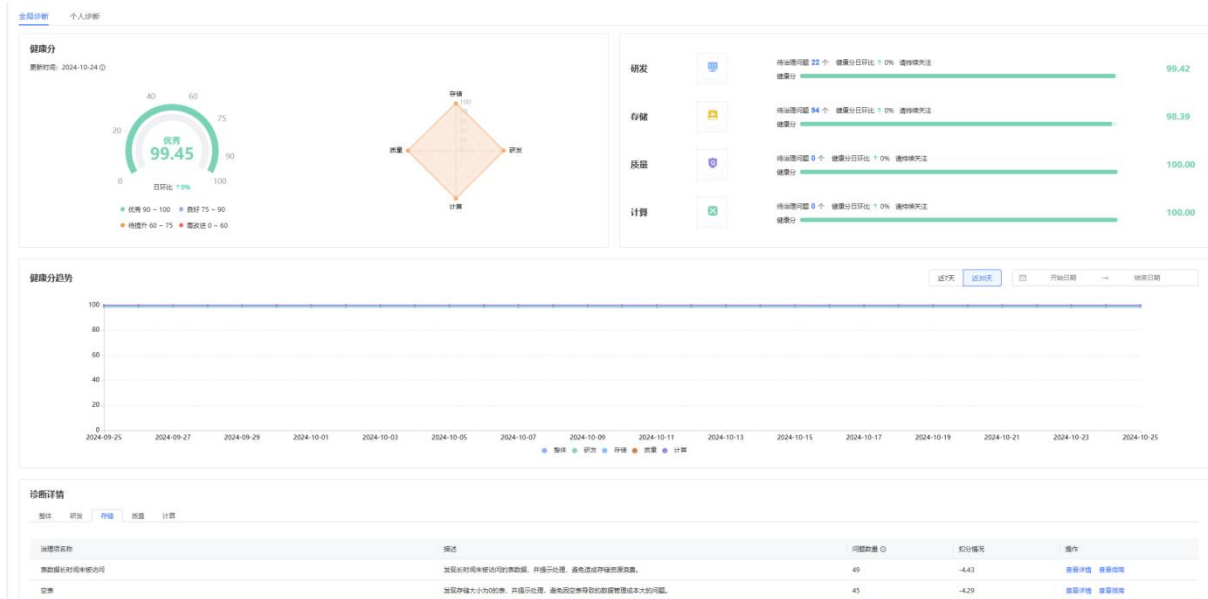


4.6.3.3 健康诊断

从全局与个人视角量化展示平台的健康情况，并引导用户对识别出的健康问题进行治疗。

进入页面

单击左上方图标，选择全部产品 > 数据治理 > 数据治理 360 > 健康诊断，进入页面。



功能说明

1. 健康分：展示前一日总健康分与各维度健康分的诊断结果。

- 健康分计算逻辑可参看如下说明。

说明

- 计算执行时间：T+1计算，根据前一日24点状态为启用的治理项进行诊断统计，按天更新。
- 计算方法：总健康分等于各维度健康分的均值；各维度健康分按具体待治理问题的严重程度进行扣分，每个治理项最多扣10分。
- 日环比计算公式： $(\text{当日健康分} - \text{前日健康分}) / \text{前日健康分} * 100\%$ 。

- 健康分评价。

健康分评价	健康分范围
优秀	90~100
良好	75~89.99
待提升	60~74.99
需改进	0~59.99

2. 查看待治理问题：点击各维度健康分中的待治理问题数量，将页面滑动至诊断详情模块，便于用户查看该维度下的详细待治理问题。

3. 健康分趋势：展示时间周期内总健康分与各维度健康分的变化趋势，当鼠标悬浮于趋势图上，

可查看所选日期的健康分具体数据。

① 说明


- 展示维度：默认展示总健康分与全部维度的健康分数据，用户可通过点击趋势图下方的图例，按需选择想要查看的维度信息，并隐藏不必要信息。
- 时间筛选：默认展示近30天健康分数据，用户可按需选择近7天数据，或在日期栏中进行灵活配置。若实际使用时间少于所选时间范围，将仅展示有数据的时间范围。

4. 诊断详情：展示诊断过程中识别到的待治理问题与扣分情况。用户可点不同页签查看待治理问题，并通过点击“查看指南”前往“健康指南”页面查看所选治理项的详细信息。

4.6.3.4 治理项问题

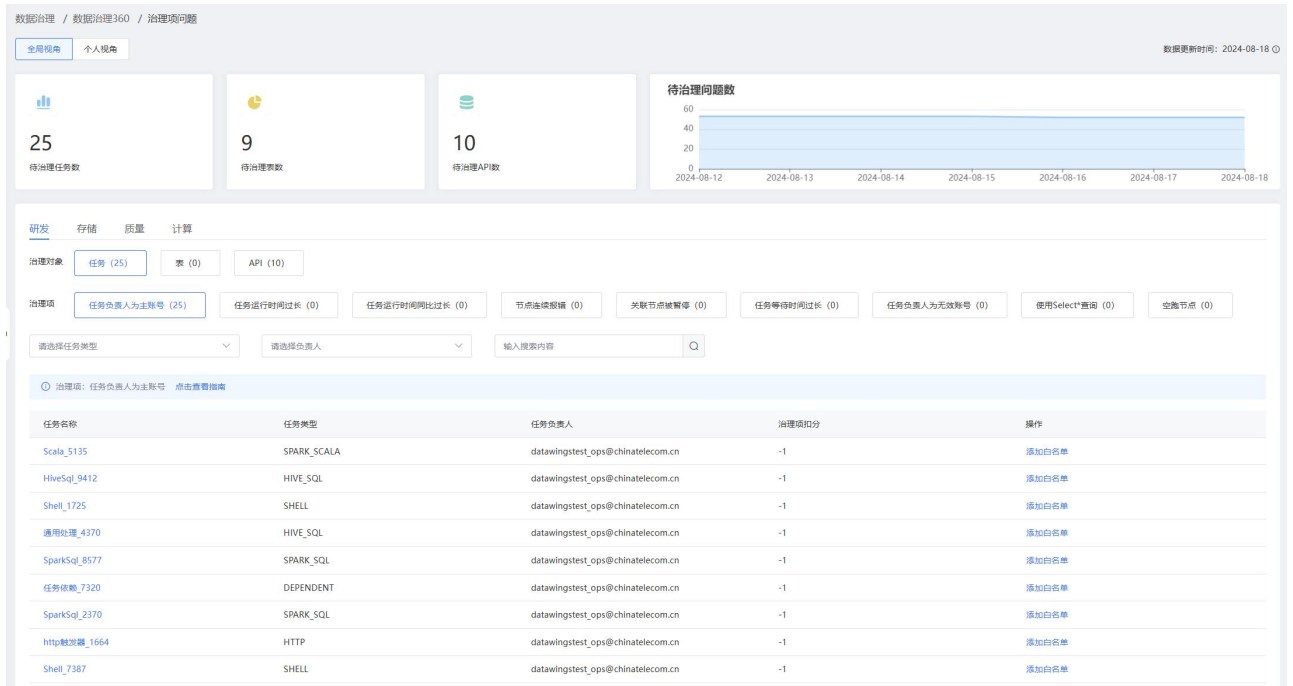
支持按照全局和个人维度，展示基于治理项检测出的问题详情信息，包含统计信息展示（待治理任务数、表数、API数、问题数趋势）和治理项问题详情。数据更新时间根据为前一天的 24:00 进行全局全量检测并形成汇总信息。

进入页面

单击左上方图标，选择全部产品 > 数据治理 > 数据治理 360 > 治理项问题，进入页面。

维度说明

- 1.全局视角：展示租户范围内的全部治理项问题和统计信息。
- 2.个人视角：展示责任人归属于用户自身的治理项问题和统计信息。



统计信息说明

- 待治理任务数：以任务为对象，展示检测出的治理问题数。
- 待治理表数：以表为对象，展示检测出的治理问题数。
- 待治理 API 数：以 API 为对象，展示检测出的治理问题数。
- 待治理问题数：趋势图，展示按照最新的统计时间计算，近 7 天的治理问题数的趋势图。

治理项问题列表说明

1. 筛选说明：

- 维度：按照研发、存储、质量等维度展示，可切换维度查看。
- 治理对象：每个治理维度均有对应的治理对象。
- 治理项：展示治理维度下，治理对象下的治理项。

2. 列表说明：

根据所选的治理对象不同，列表参数、支持的筛选项、可检索内容以及跳转落地页有所不同。

- 任务：展示任务名称、任务类型、任务负责人、治理项扣分信息，支持对任务类型和负责人的筛选，支持对任务名称的检索。点击任务名称将跳转至【数据处理】-【任务调度】-【任务运维】的节点定义页签，查看具体的任务节点信息。
- 表：展示表名、数据库名称、数据源名称、数据源类型、表负责人、治理项扣分信息，支持对数据源类型和负责人的筛选，支持对表名称、数据库名称、数据源名称的检索。点击表名将跳转至【数据地图】-【数据检索】页，查看具体的表详情信息。
- API：展示 API 名称、API 分组、负责人、治理项扣分信息，支持对 API 负责人的筛选，支持对 API 名称的检索。点击表名将跳转至【数据服务】-【服务管理】-【API 管理】页，查看具体的 API 详情信息。

3.操作说明：

添加白名单：单击目标治理项问题操作列的“添加白名单”：可将无需处理的治理项问题添加至白名单，治理项问题被加入白名单后，后续将不会作为治理项问题检测出来。

4.6.3.5 治理项配置

支持管理员对各维度的治理项进行管理与启停控制，便于用户按需选择需要诊断的治理项。当前治理 360 支持从研发、存储、质量等维度对系统进行诊断与治理配置，计算维度正在加紧开发中，敬请期待。

进入页面

选单击左上方图标，择全部产品 > 数据治理 > 数据治理 360 > 治理项配置，进入页面。

数据治理 / 数据治理360 / 治理项配置

研发 存储 质量 计算

请输入治理项名称 请选择状态

[查询](#) [重置](#)

治理项名称	描述	诊断类型	状态	处理指南
任务运行时间比过长	发现任务运行时长较同期有明显上涨的任务实例，并...	研发	<input checked="" type="checkbox"/>	查看指南
节点连续报错	发现连续运行报错的节点，并提示处理，避免阻塞下...	研发	<input checked="" type="checkbox"/>	查看指南
任务运行时间过长	发现运行时间过长的任务，并提示处理，避免因逻辑...	研发	<input checked="" type="checkbox"/>	查看指南
任务等待时间过长	发现等待时间过长的任务，并提示处理，避免因资源...	研发	<input checked="" type="checkbox"/>	查看指南
任务负责人为无效账号	发现已被禁用或删除的任务负责人，并提示处理，避...	研发	<input checked="" type="checkbox"/>	查看指南
任务负责人为主账号	发现任务负责人是主用户的情况，并提示处理，避免...	研发	<input checked="" type="checkbox"/>	查看指南
使用Select查询	发现任务中未指定查询字段，直接使用Select进行查...	研发	<input checked="" type="checkbox"/>	查看指南
空跑节点	发现不会有数据产出的空跑任务节点，并提示用户及...	研发	<input checked="" type="checkbox"/>	查看指南
长时间未调用API	发现长时间未被调用的已发布上线的数据服务API，...	研发	<input checked="" type="checkbox"/>	查看指南
异常API调用	发现已发布上线的数据服务API的异常调用情况，并...	研发	<input checked="" type="checkbox"/>	查看指南
关联节点被暂停	发现处于暂停状态，但被下游所依赖的节点，并提示...	研发	<input checked="" type="checkbox"/>	查看指南
表负责人为无效账号	发现已被禁用或删除的表负责人，并提示处理，避免...	研发	<input checked="" type="checkbox"/>	查看指南

共计12条记录 < 1 > 20条/页 跳至 1 页

操作说明

1.按维度查看治理项：用户可通过点击页面上方的页签查看各维度治理项。

研发 存储 质量 计算

2.状态：点击状态栏中的启停开关，可控制所选治理项启停状态。启用状态下，该治理项将计入健康分诊断，停用后开关置灰，该治理项将不再被诊断。

说明

健康分模型将依据当日24点，状态为启用的治理项进行诊断计算。

研发 存储 质量 计算

请输入治理项名称 请选择状态

查询 重置

治理项名称	描述	诊断类型	状态	处理指南
任务运行时间同比过长	发现任务运行时长较同期有明显上涨的任务实例，并...	研发	<input checked="" type="checkbox"/>	查看指南
节点连续报错	发现连续运行报错的节点，并提示处理，避免阻塞下...	研发	<input type="checkbox"/>	查看指南
任务运行时间过长	发现运行时间过长的任务，并提示处理，避免因逻辑...	研发	<input checked="" type="checkbox"/>	查看指南
任务等待时间过长	发现等待时间过长的任务，并提示处理，避免因资源...	研发	<input checked="" type="checkbox"/>	查看指南
任务负责人为无效账号	发现已被禁用或删除的任务负责人，并提示处理，避...	研发	<input checked="" type="checkbox"/>	查看指南
任务负责人为主账号	发现任务负责人是主用户的情况，并提示处理，避免...	研发	<input checked="" type="checkbox"/>	查看指南
使用Select*查询	发现任务中未指定查询字段，直接使用Select*进行查...	研发	<input checked="" type="checkbox"/>	查看指南
空跑节点	发现不会有数据产出的空跑任务节点，并提示用户及...	研发	<input checked="" type="checkbox"/>	查看指南
长时间未调用API	发现长时间未被调用的已发布上线的数据服务API，...	研发	<input checked="" type="checkbox"/>	查看指南
异常API调用	发现已发布上线的数据服务API的异常调用情况，并...	研发	<input checked="" type="checkbox"/>	查看指南
关联节点被暂停	发现处于暂停状态，但被下游所依赖的节点，并提示...	研发	<input checked="" type="checkbox"/>	查看指南
表负责人为无效账号	发现已被禁用或删除的表负责人，并提示处理，避免...	研发	<input checked="" type="checkbox"/>	查看指南

共计12条记录 < 1 > 20条/页 跳至 1 页

3.查看指南：点击“查看指南”按钮将跳转至“健康指南”页面，并默认展开所选治理项的详情信息。

研发 存储 质量 计算

请输入治理项名称 请选择状态

查询 重置

治理项名称	描述	诊断类型	状态	处理指南
任务运行时间同比过长	发现任务运行时长较同期有明显上涨的任务实例，并...	研发	<input checked="" type="checkbox"/>	查看指南
节点连续报错	发现连续运行报错的节点，并提示处理，避免阻塞下...	研发	<input type="checkbox"/>	查看指南
任务运行时间过长	发现运行时间过长的任务，并提示处理，避免因逻辑...	研发	<input checked="" type="checkbox"/>	查看指南
任务等待时间过长	发现等待时间过长的任务，并提示处理，避免因资源...	研发	<input checked="" type="checkbox"/>	查看指南
任务负责人为无效账号	发现已被禁用或删除的任务负责人，并提示处理，避...	研发	<input checked="" type="checkbox"/>	查看指南
任务负责人为主账号	发现任务负责人是主用户的情况，并提示处理，避免...	研发	<input checked="" type="checkbox"/>	查看指南
使用Select*查询	发现任务中未指定查询字段，直接使用Select*进行查...	研发	<input checked="" type="checkbox"/>	查看指南
空跑节点	发现不会有数据产出的空跑任务节点，并提示用户及...	研发	<input checked="" type="checkbox"/>	查看指南
长时间未调用API	发现长时间未被调用的已发布上线的数据服务API，...	研发	<input checked="" type="checkbox"/>	查看指南
异常API调用	发现已发布上线的数据服务API的异常调用情况，并...	研发	<input checked="" type="checkbox"/>	查看指南
关联节点被暂停	发现处于暂停状态，但被下游所依赖的节点，并提示...	研发	<input checked="" type="checkbox"/>	查看指南
表负责人为无效账号	发现已被禁用或删除的表负责人，并提示处理，避免...	研发	<input checked="" type="checkbox"/>	查看指南

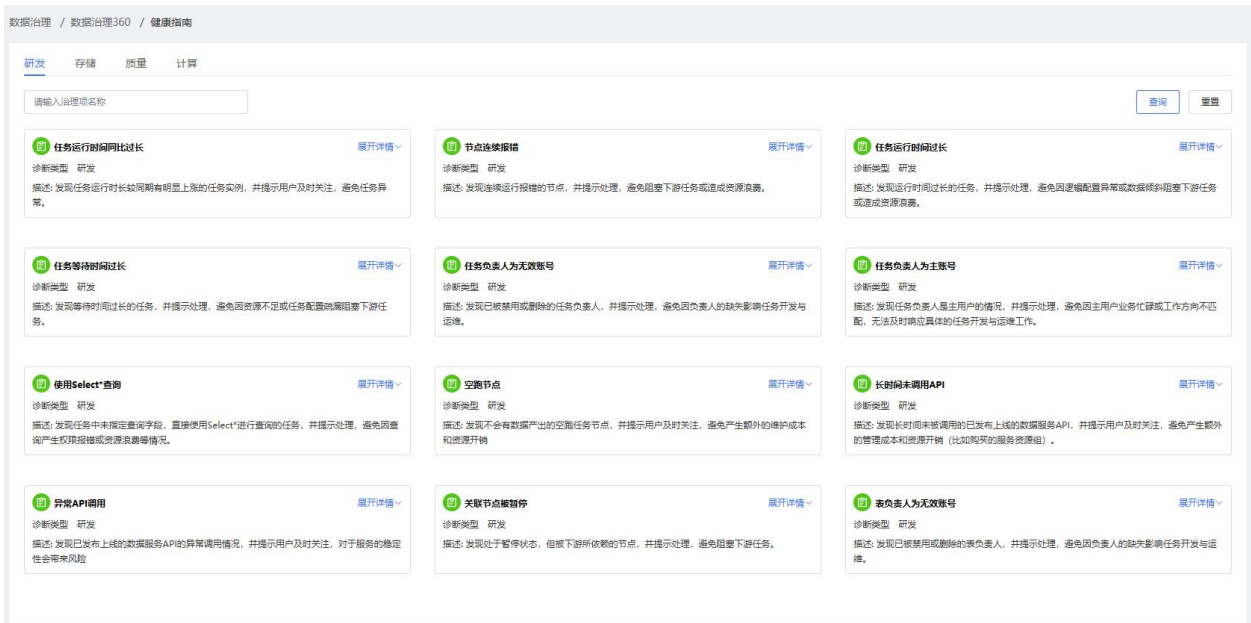
共计12条记录 < 1 > 20条/页 跳至 1 页

4.6.3.6 健康指南

提供各治理项的详情，包含描述、判断规则与操作指南，用户可参考指南进问题治理。

进入页面

单击左上方图标，选择全部产品 > 数据治理 > 数据治理 360 > 健康指南，进入页面。



操作说明

1.按维度查看指南：用户可通过点击页面上方页签查看各维度的指南。



2.展开详情/收起：点击“展开详情”可查看所选治理项的详细信息，包含描述、判断规则与处理指南；点击“收起”将折叠隐藏所选治理项的详细信息。



4.6.3.7 白名单

无需处理的治理项问题可添加至白名单，系统将所有白名单集中至白名单页面进行统一管理。


治理项问题被加入白名单后，后续将不会作为治理项问题检测出来。

前提条件

- 已启用治理项。

- 已将治理项问题添加至白名单。

进入页面





















单击左上方图标，选择全部产品 > 数据治理 > 数据治理 360 > 白名单，进入页面。

查看并管理白名单

在白名单页面，您可以查看白名单列表。同时，也可以将治理对象从白名单移除。

- 搜索：可直接输入资源名称进行搜索相关白名单。
- 筛选：通过治理对象、资源类型、负责人、治理项、申请人等条件进行筛选，查看目标条件下的白名单列表。
- 移除：单击操作列的“移除”，即可移除目标白名单。如果您需要重新检测白名单中的治理项，可以移除该白名单；移除后，治理项问题后续会被重新检测。

数据治理 / 数据治理360 / 白名单

资源名称	治理对象	资源类型	负责人	治理项	申请人	申请时间	操作
任务依赖_2556	任务	DEPENDENT	test_tenant_1			2024-08-14 10:18:55	移除
HiveSql_3274	任务	HIVE_SQL	test_tenant_1			2024-08-14 08:10:08	移除
HiveSql_1172	任务	HIVE_SQL	test_tenant_1			2024-08-13 08:12:22	移除
HiveSql_7861	任务	HIVE_SQL	test_tenant_1			2024-08-12 08:10:19	移除
HiveSql_1667	任务	HIVE_SQL	test_tenant_1			2024-08-11 08:12:27	移除
任务依赖_2916	任务	DEPENDENT	test_tenant_1			2024-08-10 08:11:38	移除
HiveSql_5838	任务	HIVE_SQL	test_tenant_1			2024-08-09 08:11:49	移除
Shell_7782	任务	SHELL	test_tenant_1			2024-08-08 08:12:28	移除
HiveSql_7194	任务	HIVE_SQL	test_tenant_1			2024-08-07 08:12:11	移除
Scala_6429	任务	SPARK_SCALA	test_tenant_1			2024-08-06 08:11:38	移除

共计52条记录 < 1 2 3 4 5 6 > 10条/页 跳至 1 页

4.7 数据地图

4.7.1 数据地图概述

本章节介绍数据地图所包含的功能及作用。

总体说明

数据地图是在元数据基础上提供的企业数据目录管理模块，涵盖数据发现、地图配置、资产首页、数据总览、数据检索、我的数据等功能，同时提供表详情信息，含基础信息、业务信息、技术信息、表结构等表详情，还提供血缘信息、数据预览等功能。数据地图可以帮助您更好地查找、理解和使用数据。

各模块说明

- **数据发现：**提供元数据采集功能，支持将不同系统中的元数据进行统一汇总管理，方便用户在数据地图查看从各数据源汇集而来的元数据信息。
- **地图配置：**提供自定义类目功能，可将集成到数据地图的数据资产按类目进行组织和管理。配置完类目后，可在数据检索按自定义类目进行筛选。
- **资产首页：**提供资产搜索入口，支持按数据建模类目、自定义类目查看统计数据及按指定类目查看资产对象，展示热门搜索、最近浏览及最近收藏的资产对象。
- **数据总览：**提供不同数据源类型的总使用标签数、总表数、总存储量、总数据源数等信息。同时提供这些指标的近 14 天趋势，表存储量 TOP15。方便用户从总体把握数据资产情况。
- **数据检索：**支持按照丰富的筛选条件检索数据以及查看数据表详情。
- **我的数据：**包括我负责的表和我收藏的表。方便快速管理我的数据。

详细介绍文档按照用户使用顺序进行展开。

4.7.2 数据发现

数据发现提供元数据采集功能，方便您将不同系统中的元数据进行统一汇总管理，您可以在数据地图查看从各数据源汇集而来的元数据信息。

前提条件

- 外部数据源：通过数据源管理功能，已成功创建的外部数据源。
- 内置数据源：通过集群注册功能，已成功配置 Hive 组件信息，且成功创建对应引擎；系统自动创建数据源和元数据采集器，按 T+1 更新元数据。

创建元数据采集器

功能入口

- 单击左上方的图标，选择全部产品 > 数据地图 > 数据发现，进入数据发现页面。
- 在数据发现页面，单击  按钮。



数据地图 / 数据发现

新建采集器 数据源管理

输入采集器名称、数据源名称、数据源类型、采集器创建人、采集计划查询

序号	采集器名称	数据源名称	数据源类型	采集器创建人	采集计划	最近执行开始的时间	最近执行结束的时间	最新采集状态	启用状态	操作	
>	1	doris	doris	Doris	yanshi	0 0 0 * * ? *	2024-04-23 00:00:00	2024-04-23 00:00:17	成功	已启用	运行全量获取 手工同步 停止运行 更多
>	2	oracle	oracle	Oracle	yanshi	0 0 0 * * ? *	2024-04-23 00:00:00	2024-04-23 01:10:37	成功	已启用	运行全量获取 手工同步 停止运行 更多
>	3	mysql	mysql	MySQL	yanshi	0 0 0 * * ? *	2024-04-23 00:00:00	2024-04-23 00:00:25	成功	已启用	运行全量获取 手工同步 停止运行 更多
>	4	hive系统内置-poc_engine_HIVE-20240409104728	poc_engine_HIVE	Hive	system	0 0 0 * * ? *	2024-04-23 00:00:00	2024-04-23 00:03:41	成功	已启用	运行全量获取 手工同步 停止运行 更多
>	5	doris系统内置-poc_engine_DORIS-20240409104729	poc_engine_DORIS	Doris	system	0 0 0 * * ? *	2024-04-23 00:00:00	2024-04-23 00:00:20	成功	已启用	运行全量获取 手工同步 停止运行 更多

共5条 10 < 1 > 前往 1 页

操作步骤

1. 选择数据源类型，点击“下一步”。



2. 选择采集对象：按规范填写采集器名称，并选择已经在数据源建立的数据源名称，点击“下一步”。

新建采集器 ×

① 选择数据源类型 ② 选择采集对象 ③ 配置执行计划

采集对象

* 采集器名称

* 数据源名称

*

- Doris数据源
- test1
- test2
- doris复制

3. 配置执行计划。“执行计划”默认为每天的零时零分零秒执行，也可按需配置执行计划。点击“确定”，则按照配置的执行计划进行执行；点击“确定并运行全量获取”，则马上执行一次全量获取元数据。

新建采集器

选择数据源类型 ———— 选择采集对象 ———— 3 配置执行计划

执行计划

每一秒钟

每隔 秒执行从 秒开始

具体秒数(可多选) x

周期从 到 秒

后续步骤

元数据采集成功后，数据地图的其他模块可以启动使用。


4.7.3 地图配置

地图配置提供自定义类目功能，可将集成到数据地图的数据资产按类目进行组织和管理。配置完类目后，可在数据检索按自定义类目进行筛选。

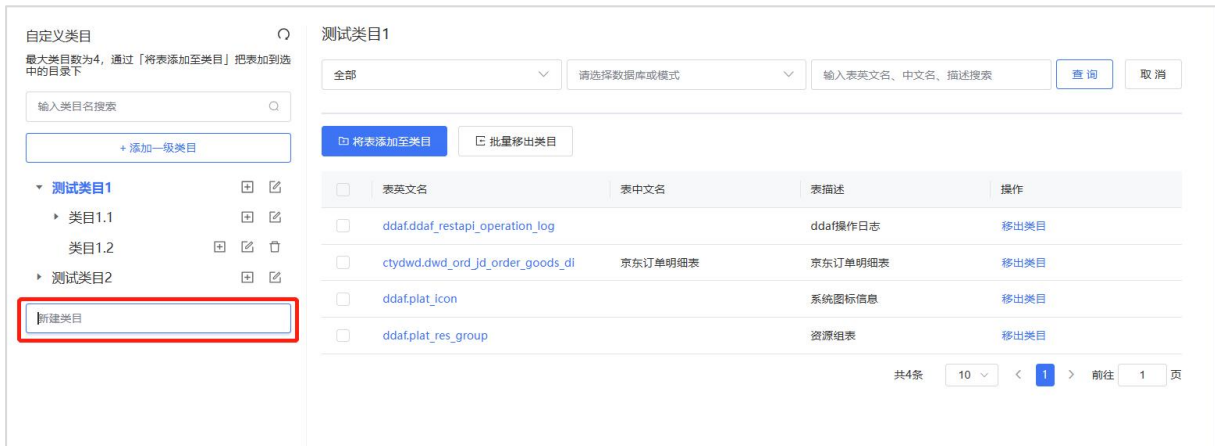
前提条件

已创建元数据采集器，且采集任务状态为成功。

创建及管理类目

1. 单击左上方的，选择全部产品 > 数据地图 > 地图配置。
2. 单击左侧抽屉栏“自定义类目”，进入“自定义类目”页面。

3. 单击 **+ 添加一级类目** 按钮，在输入框中输入类目名称，单击键盘回车键保存名称。



自定义类目

最大类目数为4，通过「将表添加至类目」把表添加到选中的目录下

输入类目名搜索

+ 添加一级类目

测试类目1

- 类目1.1
- 类目1.2
- 测试类目2
- 新建类目

测试类目1

全部 请选择数据库或模式 输入表英文名、中文名、描述搜索 查询 取消

将表添加至类目 批量移出类目

<input type="checkbox"/>	表英文名	表中文名	表描述	操作
<input type="checkbox"/>	ddaf.ddaf_restapi_operation_log		ddaf操作日志	移出类目
<input type="checkbox"/>	ctydw.dwd_ord_jd_order_goods_di	京东订单明细表	京东订单明细表	移出类目
<input type="checkbox"/>	ddaf.plat_icon		系统图标信息	移出类目
<input type="checkbox"/>	ddaf.plat_res_group		资源组表	移出类目

共4条 10 < 1 > 前往 1 页

说明

- 最多支持创建4级。
- 其他操作：点击 **+** 创建子类目；创建完成后可点击 **✎** 修改类目；点击 **✕** 删除类目（没有子类目或者该类目没有挂载具体表，方可删除）。

管理类目下的表

将表添加至类目

1. 选择需要挂载的类目，单击“将表添加至类目”按钮。
2. 在弹窗勾选需添加的表。

将表添加至类目 ×

当前类目: 类目1

仅显示未添加到类目的表

<input type="checkbox"/>	表英文名	表中文名	表描述	操作
<input checked="" type="checkbox"/>	Test_integration_mysql.lalani...			添加
<input checked="" type="checkbox"/>	Test_integration_mysql.test_or...			添加
<input checked="" type="checkbox"/>	Test_integration_mysql.weewrwe			添加
<input type="checkbox"/>	Test_integration_mysql.www			添加
<input type="checkbox"/>	bdmp.DWA_T_NEW			添加
<input type="checkbox"/>	bdmp.GaussDB_newtable			添加
<input type="checkbox"/>	bdmp.ORDDCM_CT_PRED_SET...			添加
<input type="checkbox"/>	bdmp.ORDDCM_INSTALL_DOCS			添加
<input type="checkbox"/>	bdmp.ORDDCM_PROTOCOL_S...			添加
<input type="checkbox"/>	bdmp.ORDDCM_PR_ADMIN_D...			添加

共6344条 < 2 3 4 5 6 ... 635 > 前往 页

3. 点击“批量添加”按钮，将已勾选的表添加到当前类目。

批量移出类目

1. 勾选需移出类目的表。

将表添加至类目 X

当前类目: 类目1

仅显示未添加到类目的表

<input type="checkbox"/>	表英文名	表中文名	表描述	操作
<input checked="" type="checkbox"/>	Test_integration_mysql.lalani...			添加
<input checked="" type="checkbox"/>	Test_integration_mysql.test_or...			添加
<input checked="" type="checkbox"/>	Test_integration_mysql.weewrwe			添加
<input type="checkbox"/>	Test_integration_mysql.www			添加
<input type="checkbox"/>	bdmp.DWA_T_NEW			添加
<input type="checkbox"/>	bdmp.GaussDB_newtable			添加
<input type="checkbox"/>	bdmp.ORDDCM_CT_PRED_SET...			添加
<input type="checkbox"/>	bdmp.ORDDCM_INSTALL_DOCS			添加
<input type="checkbox"/>	bdmp.ORDDCM_PROTOCOL_S...			添加
<input type="checkbox"/>	bdmp.ORDDCM_PR_ADMIN_D...			添加

共6344条 < **1** 2 3 4 5 6 ... 635 > 前往 页

2. 点击“批量移出类目”按钮，则将已勾选表从当前类目移出。

后续步骤


在“数据检索”模块，点击“自定义类目”可按父级或子级类目进行筛选。



4.7.4 资产首页

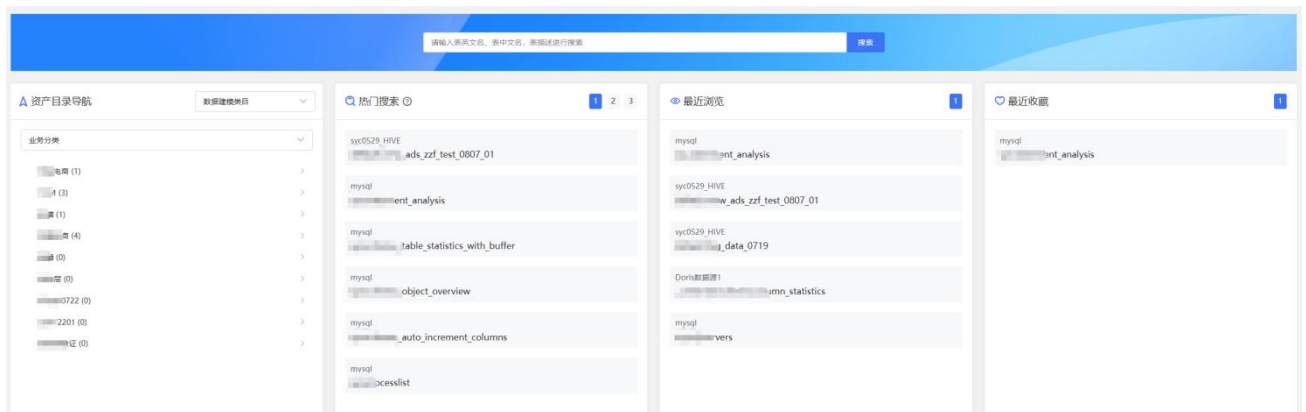
本章节主要介绍资产首页的功能。

功能入口

单击左上方的，选择全部产品 > 数据地图 > 资产首页。

功能描述

提供资产搜索入口，支持按数据建模类目、自定义类目查看统计数据及按指定类目查看资产对象，展示热门搜索、最近浏览及最近收藏的资产对象。



- 资产搜索：支持输入表英文名、表中文名及表描述进行搜索。
- 资产目录导航：支持切换数据建模类目、自定义类目查看统计数据；点击指定类目可查看


资产对象。

- 热门搜索：展示近 15 天，当前租户内所有账号搜索次数 top30 的资产对象。
- 最近浏览：展示当前账号浏览过资产清单详情页的最近 30 条资产对象。
- 最近收藏：展示当前账号最近收藏的资产对象，不超过 30 条。

4.7.5 数据总览

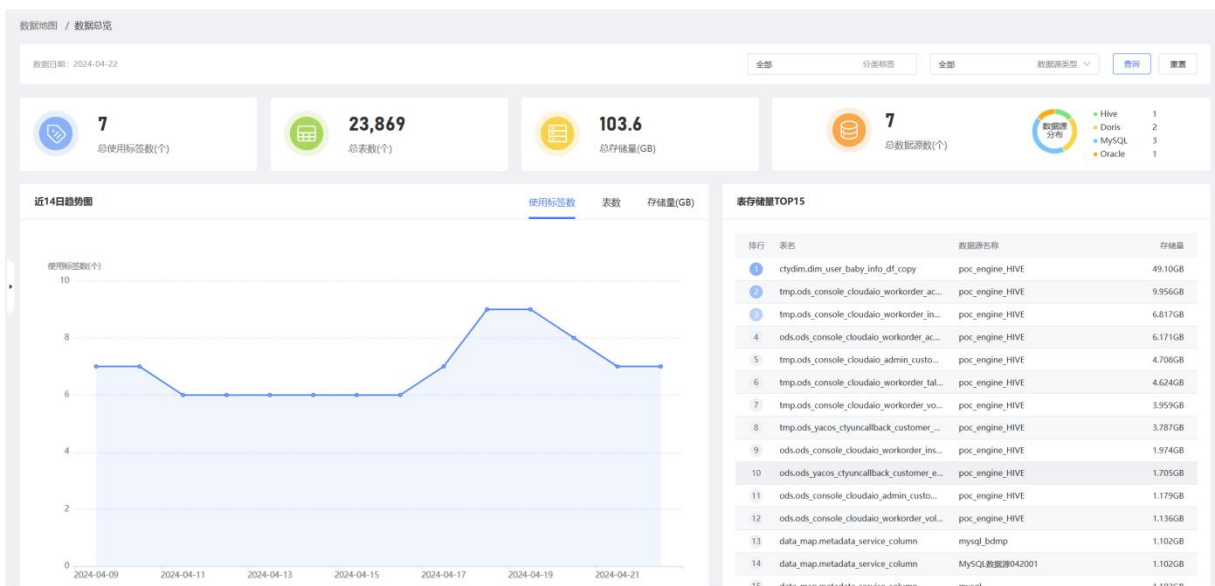
本章节介绍数据总览所展示的内容。

功能入口

单击左上方的，选择全部产品 > 数据地图 > 数据总览。

功能描述


提供不同数据源类型（Hive、Doris、MySQL、Oracle 等）的总使用标签数、总表数、总存储量、总数据源数等信息。同时提供这些指标的近 14 天趋势，表存储量 TOP15。方便用户从总体把握数据资产情况。



4.7.6 数据检索

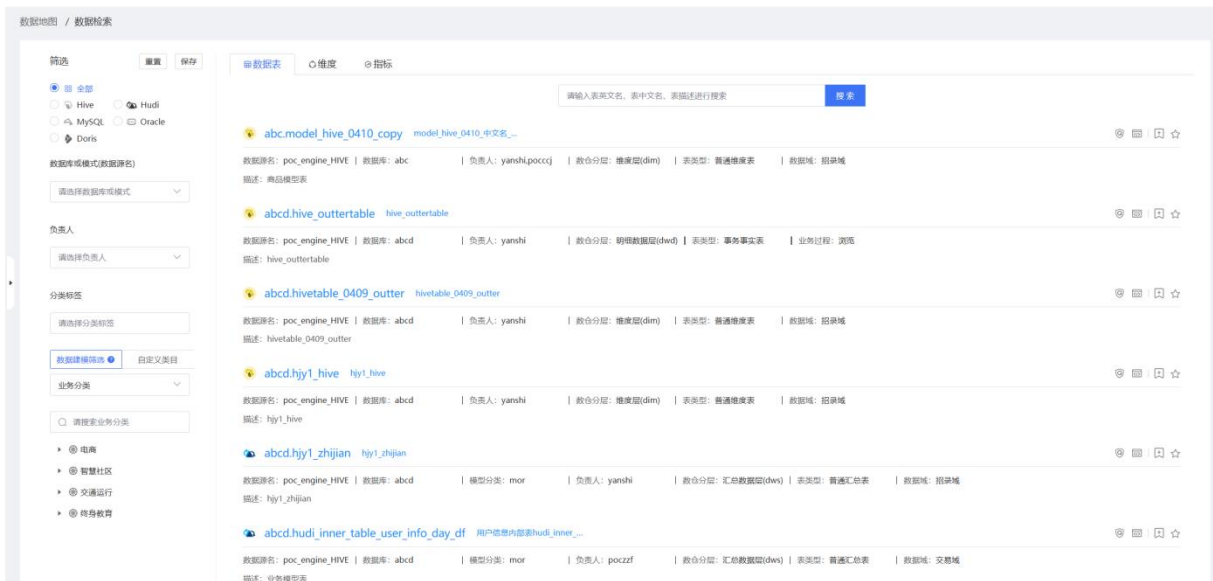
本章节介绍数据检索的功能。

功能入口

单击左上方的，选择全部产品 > 数据地图 > 数据检索。

功能描述

数据检索支持按照丰富的筛选条件检索数据以及查看数据表详情。



数据表检索

检索列表页面支持筛选和搜索两种方式检索数据表：

- 按照数据表类型、数据库或模式(括号内是数据源名称)、负责人、分类标签、数据建模筛选、自定义类目等规则检索数据表。
- 输入表英文名、表中文名、表描述检索数据表。

数据表操作

- 数据质量配置：点击右侧^⑨ 跳转到数据质量按钮，可新建数据质量检测规则、查看该数据表的质检规则等。



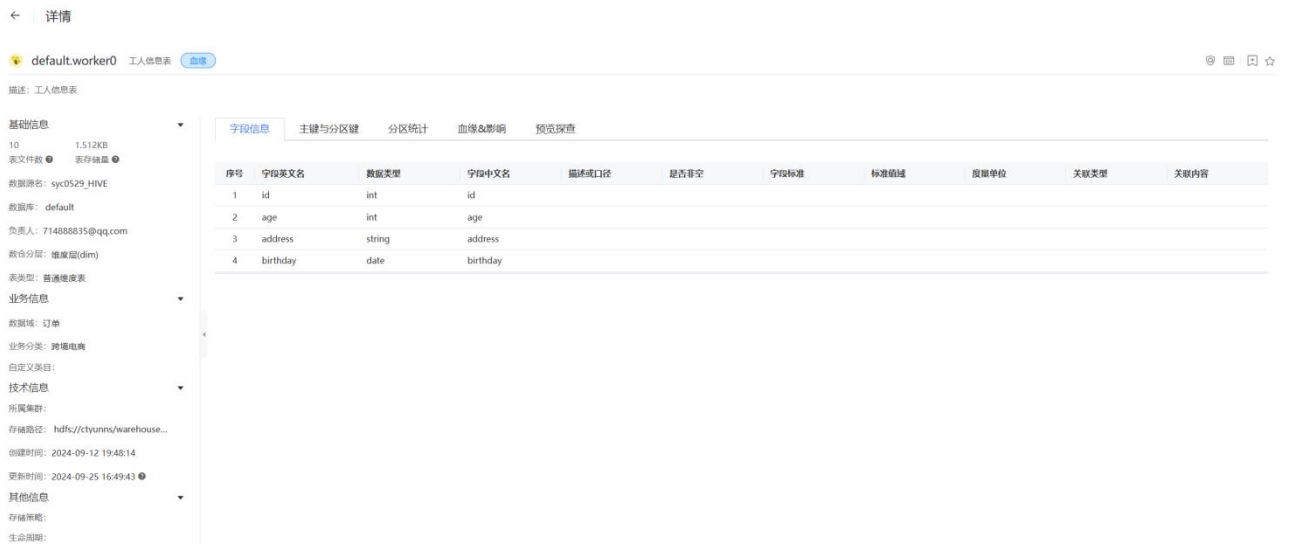
- 跳转到数据开发：点击右侧^⑩ 跳转到数据开发按钮，可直接跳转到数据开发页面。
- 添加标签：点击右侧^⑪ 标签按钮，可快速创建标签、添加标签，后续检索数据表时可按标签检索数据表。



- 收藏：点击右侧☆收藏按钮，可将该数据表加入到我收藏的表，后续可在我的数据-我的收藏的表查看数据表。

数据表详情

点击蓝色英文表名，则可跳转到表详情。如下以 Hive 表举例：



- 表信息：如页面左侧所示：含基础信息、业务信息、技术信息、其他信息。
- 字段信息：主要包括字段英文名、数据类型、字段中文名、描述或口径、是否非空、字段标

准、标准值域、度量单位、关联类型、关联内容等。

- 主键与分区键：表的主键与分区键。
- 分区统计：展示分区名、记录数、存储量等信息。此统计指标仅供参考，数据为离线数据，实际以引擎侧为准。
- 血缘&影响：展示表及字段血缘。

说明
展示基于调度作业（含HiveSQL节点、SparkSQL节点）实例解析得出的表与表、字段与字段之间的血缘关系。离线数据按T+1更新血缘。其他频率的数据，系统半小时采集一次血缘。

字段信息 主键与分区键 分区统计 **血缘&影响** 预览探查

表血缘 字段血缘




```
graph TD;
  W4[default.worker4] --> W0[default.worker0];
  W4 --> W1[default.worker1];
  W4 --> W2[default.worker2];
  W0 --> W1;
  W0 --> W2;
  W3[default.worker3] --> W1;
  W3 --> W2;
```

- 预览探查：支持查看该数据表的 10 条数据，可以方便快捷的了解当前数据表的数据存储信息。

4.7.7 我的数据

本章节介绍我的数据的功能。

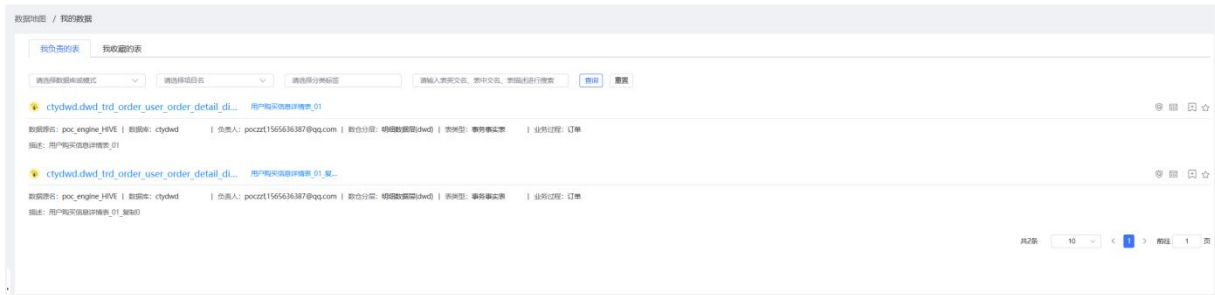
功能入口

单击左上方的，选择全部产品 > 数据地图 > 我的数据。

功能描述

我的数据中，包括我负责的表和我收藏的表，支持多种方式检索数据表。方便快捷管理我的数据。

- 我负责的表：显示当前用户是表负责人的所有表。
- 我收藏的表：显示当前用户收藏的所有表。



4.8 数据服务

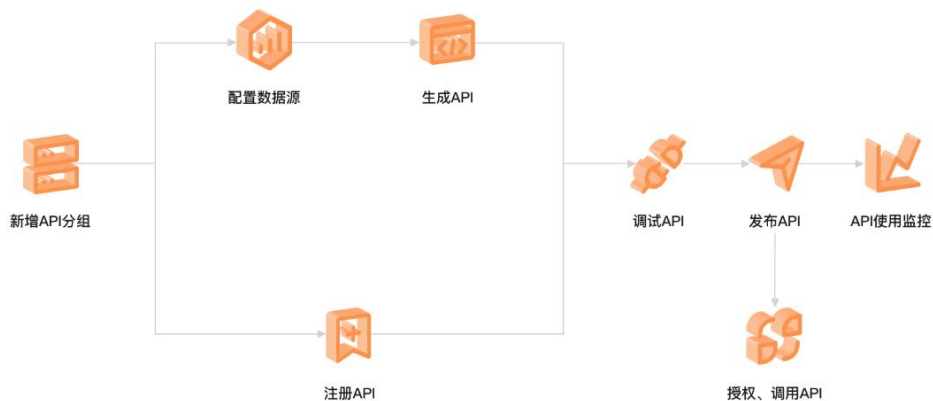
4.8.1 数据服务概述

数据服务旨在为企业提供全面的数据服务及共享能力，帮助企业统一管理面向内外部的 API 服务。您能够将数据表快速生成数据 API，或将已有 API 快速注册至本平台进行统一管理与发布。数据服务工具屏蔽了下层的存储组件的查询方式，可为上层业务提供灵活的数据查询。通过统一的接口，让数据使用者和物理数据源隔离。数据服务提供了标准化、规范化的数据开发手段，实现了 API 开发、API 测试、API 发布、API 管理等全流程功能。数据服务对平台运行所需的统一数据库提供了封装技术能力，支持快速生成 API 以对外提供服务，用户还可以对生成的 API 进行授权管理。

前提条件

- 已创建目标数据源，详情请参见[数据源管理](#)。

操作流程



4.8.2 服务开发

服务开发模块以可视化方式创建数据服务 API 并发布，降低服务开发难度。用户可先行创建 API 分组，随后可使用向导模式或脚本模式配置 API，其中，向导模式以可视化方式完成 API 的配置和发布操作，脚本模式需用户自行编写 SQL 脚本并发布。

向导模式开发 API

服务开发支持以向导模式开发 API，无需编写代码即可完成服务开发和发布。

操作步骤

1. 登录平台，进入数据服务，选择“服务开发”。
2. 在目录区域选中需要生成 API 的分组，点击右侧的添加+图标，选择“生成 API”。在“生成 API”弹窗选择 API 模式为“向导模式”。

配置基本信息

生成API
×

API模式 向导模式 脚本模式

* API名称

* API路径

* 请求方式

* 返回类型

标签 ⓘ

描述

填写 API 名称、API 路径、请求方式、返回类型、标签、描述等信息，上述信息的约束如下：

字段	说明
API 名称	支持中文、英文、数字、下划线 (_) ，且只能以英文或中文开头，4~128 个字符。 API 名称不可重复。
API 路径	表示 API 存放的路径，即相对于服务 host，API 的请求路径。例如/user。支持英文、数字、下划线 (_) 和连字符 (-) ，且只能以 "/" 开头，不得超过 64 个字符。
请求方	可选择 GET 或 POST。

字段	说明
式	
返回类型	可选择 JSON 类型。
标签	最多输入 5 个标签，标签长度不超过 32 字符。
描述	最多输入 255 字符。

配置数据资源

配置前需在数据源管理模块添加数据源的配置信息，相关操作可参见[数据源管理](#)模块。此处需选择数据源类型、数据源名称、数据库、数据表等信息。数据源类型当前支持 MySQL、Oracle、Doris。

数据资源 若还未创建数据源，请前往数据源管理进行创建。 [创建数据源](#)

* 数据源类型

* 数据源名称

* 数据库

* 数据表名称

参数配置

选择数据资源后，所选表的字段信息将显示在参数配置区域，用户可勾选字段设置为请求参数、返回参数和排序字段等，勾选后的字段将显示在对应的参数区域。

参数配置

请输入字段名称

序号	<input type="checkbox"/> 设为请求参数	<input type="checkbox"/> 设为返回参数	字段名	字段类型	字段描述	<input type="checkbox"/> 设为排序字段	<input type="checkbox"/> 过滤
1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	id	INTEGER		<input type="checkbox"/>	<input type="checkbox"/>
2	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	name	VARCHAR		<input type="checkbox"/>	<input type="checkbox"/>
3	<input type="checkbox"/>	<input checked="" type="checkbox"/>	age	INTEGER		<input type="checkbox"/>	<input type="checkbox"/>
4	<input type="checkbox"/>	<input checked="" type="checkbox"/>	address	VARCHAR		<input type="checkbox"/>	<input type="checkbox"/>
5	<input type="checkbox"/>	<input checked="" type="checkbox"/>	version	INTEGER		<input type="checkbox"/>	<input type="checkbox"/>
6	<input type="checkbox"/>	<input type="checkbox"/>	height	FLOAT		<input type="checkbox"/>	<input type="checkbox"/>

配置请求参数

设为请求参数的字段将显示在请求参数区域，可配置参数名称、参数类型、比较方式、是否必填、示例值、默认值、描述等信息。

请求参数

序号	参数名称	绑定字段	参数类型	参数位置	比较方式	是否必填	示例值	默认值	描述	操作
1	<input type="text" value="id"/>	<input type="text" value="id"/>	<input type="text" value="bigint"/>	<input type="text" value="QUE"/>	<input "="" type="text" value="="/>	<input type="checkbox"/>	<input type="text" value="请输入"/>	<input type="text" value="请输入"/>		删除
2	<input type="text" value="name"/>	<input type="text" value="name"/>	<input type="text" value="string"/>	<input type="text" value="QUE"/>	<input "="" type="text" value="="/>	<input type="checkbox"/>	<input type="text" value="请输入"/>	<input type="text" value="请输入"/>		删除

配置返回参数

设为返回参数的字段将显示在返回参数区域，可配置参数名称、示例值、描述、脱敏配置等信息。

返回参数

序号	参数名称	绑定字段	参数类型	示例值	描述	脱敏配置	操作
1	id	id	bigint	请输入		请输入	删除
2	age	age	bigint	请输入		请输入	删除
3	address	address	string	请输入		请输入	删除
4	version	version	bigint	请输入		请输入	删除
5	name	name	string	请输入		请输入	删除

配置排序字段

设为排序字段的字段将显示在排序字段区域，可配置字段的排序方式，并支持对排序字段进行上移、下移、删除等操作。

排序字段

序号	字段名	字段类型	排序方式	操作
1	id	INTEGER	升序	上移 下移 删除
2	name	VARCHAR	升序	上移 下移 删除

过滤配置

设为过滤字段的字段将显示在过滤配置区域，支持按列设置过滤规则，可添加一或多组条件，条件默认为 AND 关系。

序号	字段名	字段类型	过滤规则	过滤参数	操作
1	id	INTEGER	>	3	删除

高级配置

可配置 API 的超时时间和分页逻辑，超时时间默认为 1000ms，用户可手动修改。若用户设置“返回结果分页”为“是”，可设置每页条数，每页条数最大为 2000。当返回结果记录数大于 2000 时

请选择分页，不分页则最多返回 500 条记录。

高级配置

* 超时时间 ms


返回结果分页 是 否 ?

* 每页条数 条

脚本模式开发 API

服务开发支持以脚本模式开发 API，支持手动编辑 API 的查询 SQL 语句，满足个性化发布 API 的需求。

操作步骤

1. 登录平台，进入数据服务，选择“服务开发”。
2. 在目录区域选中需要生成 API 的分组，点击右侧的添加  图标，选择“生成 API”。在“生成 API”弹窗选择 API 模式为“脚本模式”。

配置基本信息

生成API
✕

API模式 向导模式 脚本模式

* API名称

* API路径

* 请求方式

* 返回类型

标签 ⓘ

描述

填写 API 名称、API 路径、请求方式、返回类型、标签、描述等信息，上述信息的约束如下：

字段	说明
API 名称	支持中文、英文、数字、下划线 (_) ，且只能以英文或中文开头，4~128 个字符。 API 名称不可重复。
API 路径	表示 API 存放的路径，即相对于服务 host，API 的请求路径。例如/user。支持英文、数字、下划线 (_) 和连字符 (-) ，且只能以“/”开头，不得超过 64 个字符。

字段	说明
请求方式	可选择 GET 或 POST。
返回类型	可选择 JSON 类型。
标签	最多输入 5 个标签，标签长度不超过 32 字符。
描述	最多输入 255 字符。

配置数据资源

配置前需在数据源管理模块添加数据源的配置信息，相关操作可参见[数据源管理](#)模块。此处需选择数据源类型、数据源名称、数据库、数据表等信息。数据源类型支持 MySQL、Oracle、Doris。

数据资源 若还未创建数据源，请先前往数据源管理进行创建。 [创建数据源](#)

* 数据源类型

* 数据源名称

* 数据库

编写查询 SQL

编写查询SQL



支持mybaits的 where,foreach,if 动态SQL; 比较字符 "<" 请用 "<" 替代; 比较字符 ">" 请用 ">" 替代; 若数据源类型为Oracle, 字段名区分大小写可能导致错误, 建议表名和字段名均加上双引号("")

请填写SQL语句

可在此区域填写 SQL 查询语句。SELECT 查询的字段即 API 的返回参数，WHERE 条件处的参数即 API 的请求参数，可使用#{ }标识请求参数。

输入 SQL 语句时，需要遵循以下规则：

- 1) 支持同一数据源下的单表查询、多表关联查询和嵌套查询。
- 2) 不支持以下语句：
 - 不支持多条 SQL 语句。
 - 不支持写入注释。
 - 不支持 INSERT、UPDATE 和 DELETE 等非 SELECT 语法。
 - 不支持 SELECT *，必须明确指定查询的列。
 - 不支持将#{param}放在引号中。例如' {id}'、'abc {xyz}123'。如果有相关需求可通过 concat('abc', #{xyz}, '123')实现。
 - 不支持设置参数为可选。
- 3) 如果 SELECT 查询列的列名带有表名前缀（例如 t.name），则必须取别名作为返回参数名（例如 t.name as name）。
- 4) 如果使用聚合函数（min、max、sum 和 count 等），必须取别名作为返回参数名。例如 sum(num)

as total_num。

5) SQL 中的#{param}统一作为请求参数进行替换，包含字符串中的#{param}。当#{param}前包含转义符 (\) 时，作为普通字符串处理。

脚本模式部分 SQL 可自动解析请求参数和返回参数，不支持解析请求参数和返回参数的需用户根据编写的脚本手动配置请求参数和返回参数信息。

配置请求参数

可手动添加一行或多行请求参数，配置参数名称、参数类型、比较方式、是否必填、示例值、默认值、描述等信息。点击操作栏的【删除】可删除已添加的请求参数。

请求参数

序号	参数名称	参数类型	参数位置	比较方式	是否必填	示例值	默认值	描述	操作
1	<input type="text" value="请输入"/>	<input type="text" value="请选择"/>	<input type="text" value="QUER"/>	<input type="text" value="请选择"/>	<input type="checkbox"/>	<input type="text" value="请输入"/>	<input type="text" value="请输入"/>	<input type="text"/>	<input type="button" value="删除"/>
2	<input type="text" value="请输入"/>	<input type="text" value="请选择"/>	<input type="text" value="QUER"/>	<input type="text" value="请选择"/>	<input type="checkbox"/>	<input type="text" value="请输入"/>	<input type="text" value="请输入"/>	<input type="text"/>	<input type="button" value="删除"/>

配置返回参数

可手动添加一行或多行返回参数，配置参数名称、示例值、描述、脱敏配置等信息。点击操作栏的【删除】可删除已添加的返回参数。

返回参数

序号	参数名称	参数类型	示例值	描述	脱敏配置	操作
1	<input type="text" value="请输入"/>	<input type="text" value="请选择"/>	<input type="text" value="请输入"/>	<input type="text"/>	<input type="text" value="请输入"/>	<input type="button" value="删除"/>

高级配置

可配置 API 的超时时间和分页逻辑，超时时间默认为 1000ms，用户可手动修改。若用户设置“返回结果分页”为“是”，可设置每页条数，每页条数最大为 2000。当返回结果记录数大于 2000 时

请选择分页，不分页则最多返回 2000 条记录。

高级配置

* 超时时间 ms


返回结果分页 是 否

* 每页条数 条

注册 API

若用户已有 API，也可以通过注册功能，将 API 通过数据服务平台进行统一管理。在 API 列表页面点击注册 API 可进行注册，注册 API 包含三个步骤：填写基础信息、参数配置以及功能测试。

操作步骤

1. 登录平台，进入数据服务，选择“服务开发”。
2. 在目录区域选中需要生成 API 的分组，点击右侧的添加  图标，选择“注册 API”。

配置基本信息



注册API

* API名称

* API路径

* 请求方式

* 返回类型

标签

描述

注册API API使用监控

填写 API 名称、API 路径、请求方式、返回类型、标签、描述等信息，上述信息的约束如下：

字段	说明
API 名称	支持中文、英文、数字、下划线（_），且只能以英文或中文开头，4~128 个字符。API 名称不可重复。
API 路径	API 存放的路径，基于输入的路径可生成 API 的请求路径，例如 api/v1/dataapi/execute/test/yanshi。。例如/user。支持英文、数字、下划线（_）和连字符（-），且只能以“/”开头，不得超过 64 个字符。
请求方式	可选择 GET 或 POST。
返回类型	可选择 JSON 类型。
标签	最多输入 5 个标签，标签长度不超过 32 字符。
描述	最多输入 255 字符。

配置后端服务

需填写所注册 API 的后台服务 Host、后台服务 Path、后台超时时间等信息。

后端服务配置

* 后台服务Host

* 后台服务Path

* 后台超时时间

配置请求参数

可手动添加一行或多行请求参数，配置参数名称、参数类型、是否必填、示例值、默认值、描述等信息。点击操作栏的【删除】可删除已添加的请求参数。

参数配置

请求参数

序号	参数名称	参数类型	参数位置	是否必填	示例值	默认值	描述	操作
1	<input type="text" value="请输入"/>	<input type="text" value="请选择"/>	<input type="text" value="QUERY"/>	<input type="checkbox"/>	<input type="text" value="请输入"/>	<input type="text" value="请输入"/>	<input type="text"/>	删除

+ 新增参数

配置内容返回实例

可在正常返回结果示例区域和异常返回结果示例区域填写 API 的返回结果示例。

返回内容示例

正常返回结果示例

请填写正常返回结果示例

异常返回结果示例

请填写异常返回结果示例

配置错误码

可手动添加一行或多行错误码信息，添加后需配置错误码、错误信息、解决方案等信息。点击操作栏的【删除】可删除已添加的请求参数。

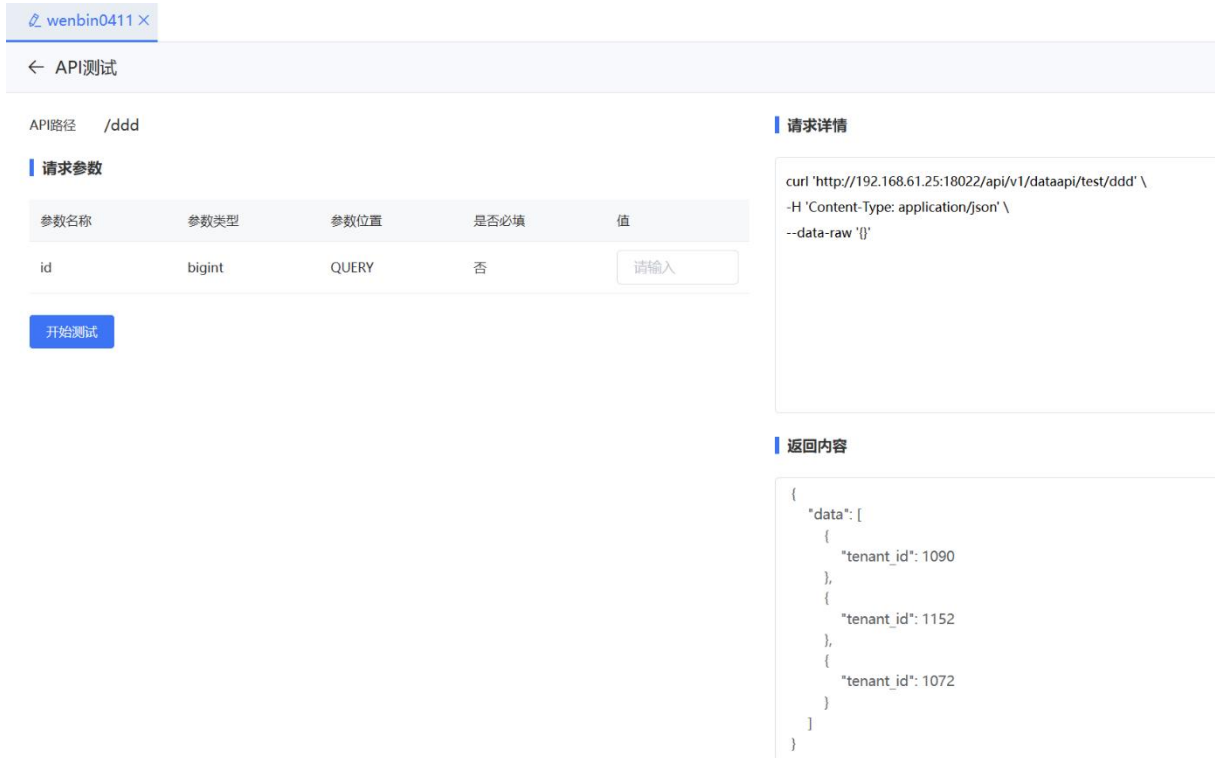
错误码信息

序号	错误码	错误信息	解决方案	操作
1	<input type="text" value="201"/>	<input type="text" value="用户不存在"/>	<input type="text" value="请输入"/>	删除

+ 新增一行

测试 API

API 逻辑配置完成后，点击【保存】，并点击【测试】进行 API 测试，测试通过的 API 才可以进行发布。



The screenshot shows the API configuration interface. At the top, there is a breadcrumb "wenbin0411 X" and a back arrow "API测试". The API path is "/ddd". Under "请求参数" (Request Parameters), there is a table with one parameter:

参数名称	参数类型	参数位置	是否必填	值
id	bigint	QUERY	否	<input type="text" value="请输入"/>

Below the table is a "开始测试" (Start Test) button. To the right, the "请求详情" (Request Details) section shows a curl command:

```
curl 'http://192.168.61.25:18022/api/v1/dataapi/test/ddd' \
-H 'Content-Type: application/json' \
--data-raw '{}'
```

The "返回内容" (Return Content) section shows the following JSON response:

```
{
  "data": [
    {
      "tenant_id": 1090
    },
    {
      "tenant_id": 1152
    },
    {
      "tenant_id": 1072
    }
  ]
}
```

发布 API

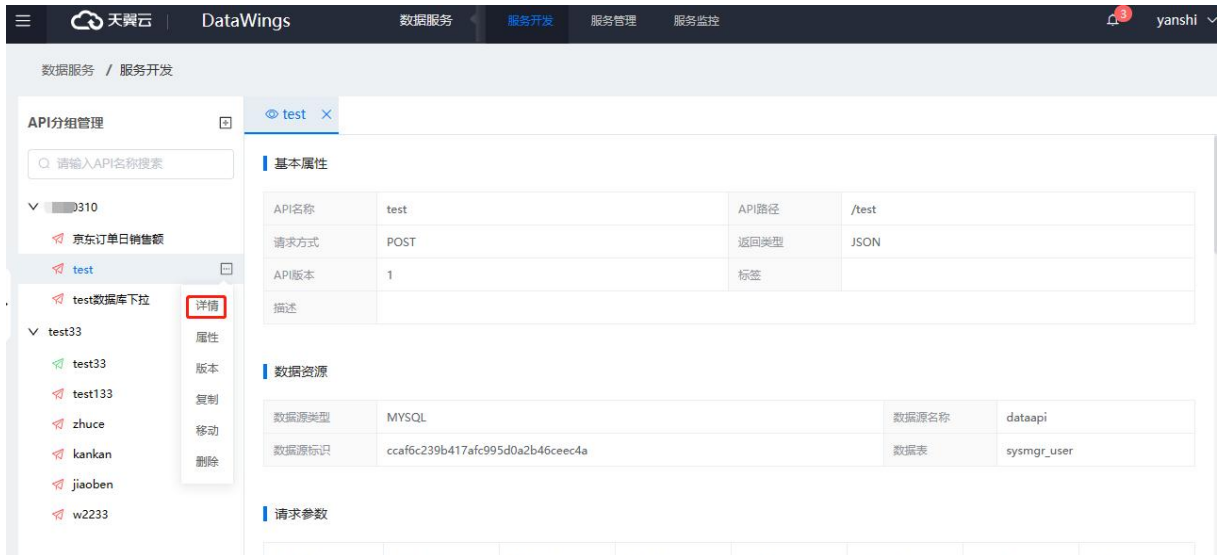
点击【发布】，并填写版本说明，即可将 API 发布上线。



The screenshot shows the "发布API" (Publish API) dialog box. At the top, there are buttons for "保存" (Save), "测试" (Test), and "发布" (Publish). The dialog has a title bar with "发布API" and a close button. The main content area has a label "* 版本描述" (Version Description) and a text input field with the placeholder "请输入版本描述" (Please enter version description). At the bottom right, there are "取消" (Cancel) and "确定" (Confirm) buttons.

[查看 API 详情](#)

1. 登录平台，进入数据服务，选择“服务开发”。
2. 在目录区域选中需要查看详情的 API，点击【详情】操作，可在右侧界面中查看 API 的配置详情。



查看 API 版本

1. 登录平台，进入数据服务，选择“服务开发”。
 2. 在目录区域选中需要查看版本的 API，点击【版本】操作，可进入 API 的版本历史页面。
- 支持查看历史版本详情、版本对比和历史版本回滚。



注意

版本回滚时仅回滚 API 配置信息，并生成新的版本号，仅下线状态的 API 支持回滚版本

复制 API

1. 登录平台，进入数据服务，选择“服务开发”。
2. 在目录区域选中需要复制的 API，点击【复制】操作，可在“复制 API”弹窗中填写新 API 的属性信息，填写完毕后点击【确定】按钮，即可创建和原 API 配置相同的 API。

复制API ×

API模式 向导模式 脚本模式

* 分组名称

* API名称

* API路径

* 请求方式

* 返回类型

标签 ⓘ

描述

4.8.3 服务管理

服务管理可对已发布的 API 服务进行统一监控和管理，主要包括 API 管理、API 测试和 API 调用三个功能，提供对 API 服务的生命周期管理。其中 API 管理可以对已发布的 API 进行管理，支持对 API 进行上/下线切换、测试等操作。API 测试是所有已发布 API 的测试中心，可选择任意 API

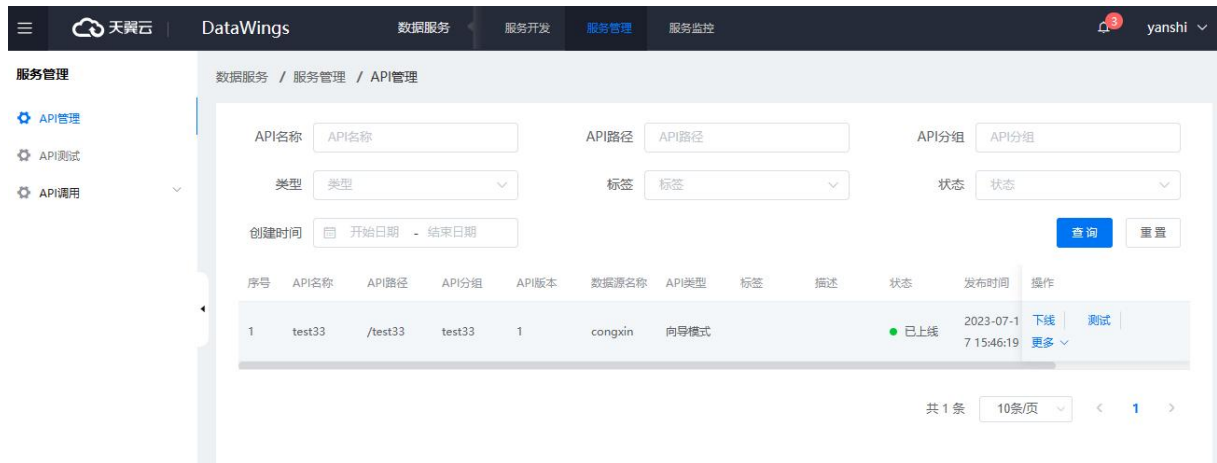
进行单次的调用测试，查看请求详情和返回内容值。API 调用支持基于 API 分组对应用进行授权操作，可选择短期、长期授权，用户可根据授权的 appKey 和 appSecret 对所授权分组下的单个 API 进行调用。

API 管理

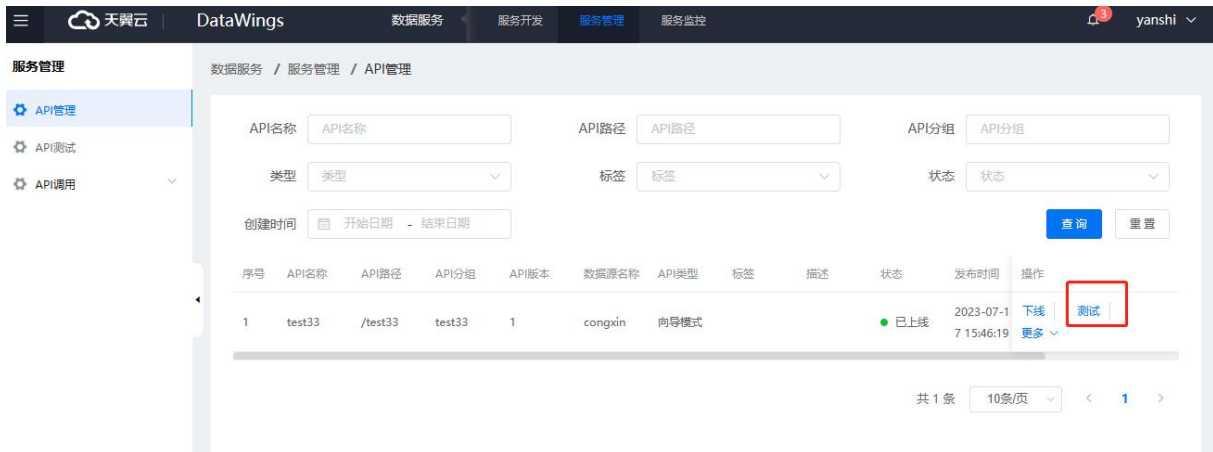
API 管理功能用于管理已发布的 API，可以对 API 执行上线/下线切换、测试、下载文档等操作。

操作说明

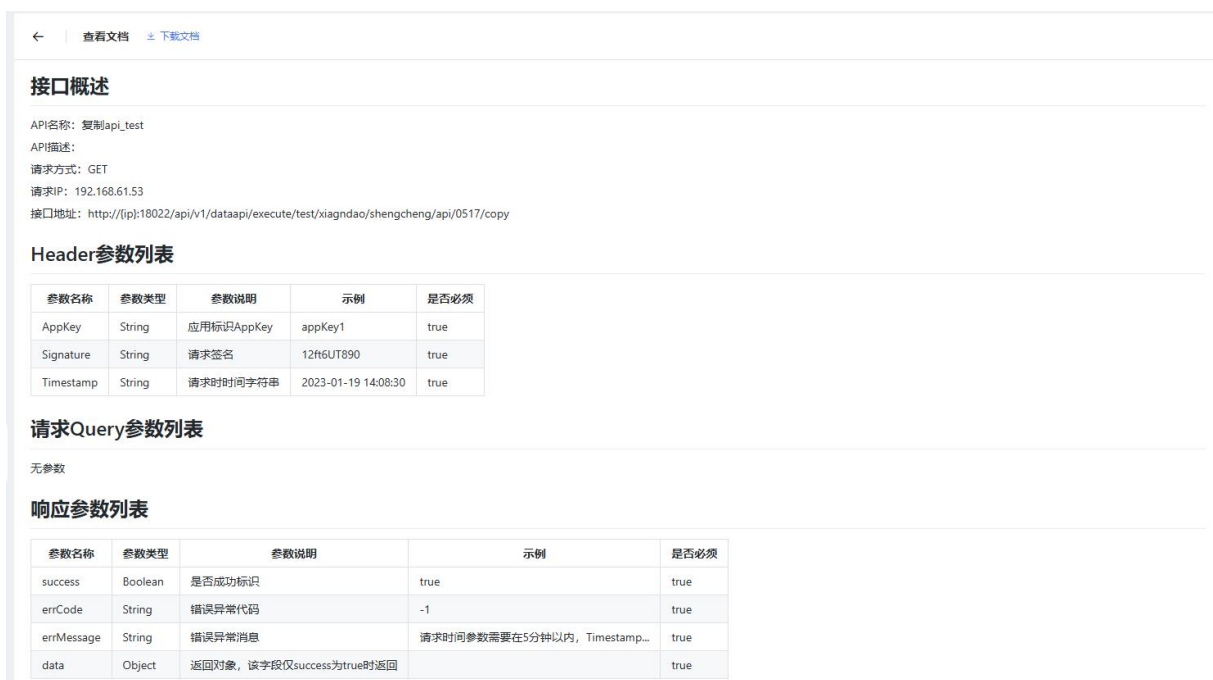
1. 登录平台，进入数据服务，选择“服务管理 > API 管理”，可查看所有已发布 API 的基本信息。



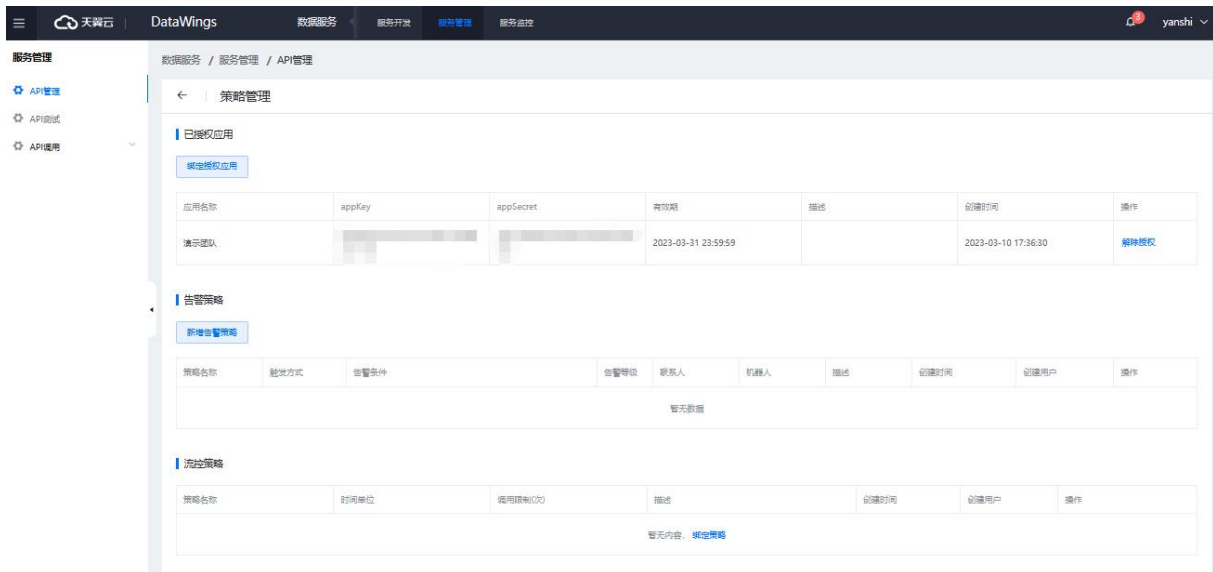
2. 对于选中的 API，点击操作栏的【上线】、【下线】、【测试】、【策略管理】和【查看文档】可执行相应的操作。【上线】、【下线】操作用于切换 API 的状态，已上线的 API 是经过测试、发布的 API，具有权限的应用可以直接调用，已下线的 API 是经过发布但已下线的 API，下线的 API 不可再调用，但可重新上线。若 API 下线后被编辑过，则需测试通过后才可重新发布上线。
3. 点击 API 操作栏的【测试】操作可进入 API 测试模块对 API 进行调用测试。



4. 点击【查看文档】操作，可查看相应 API 的说明文档，并且支持下载 API 文档。



5. 点击【策略管理】操作，可进入 API 的策略管理页面。策略管理页面展示 API 已授权的应用、告警策略和流控策略。可在页面中新增绑定的授权应用或告警策略，以及解绑现有的授权应用和告警策略。一个 API 最多可选择 10 个告警策略，最多可绑定一个流控策略。

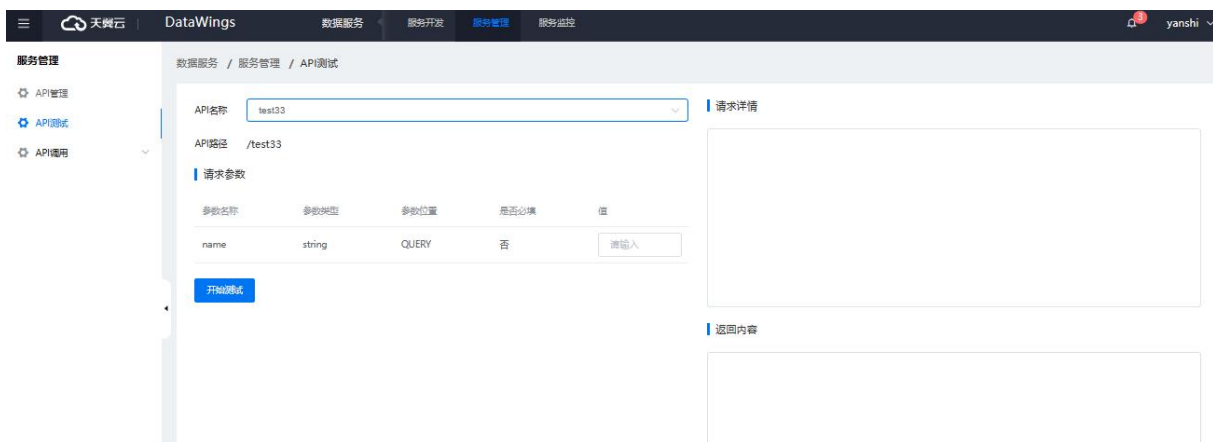


API 测试

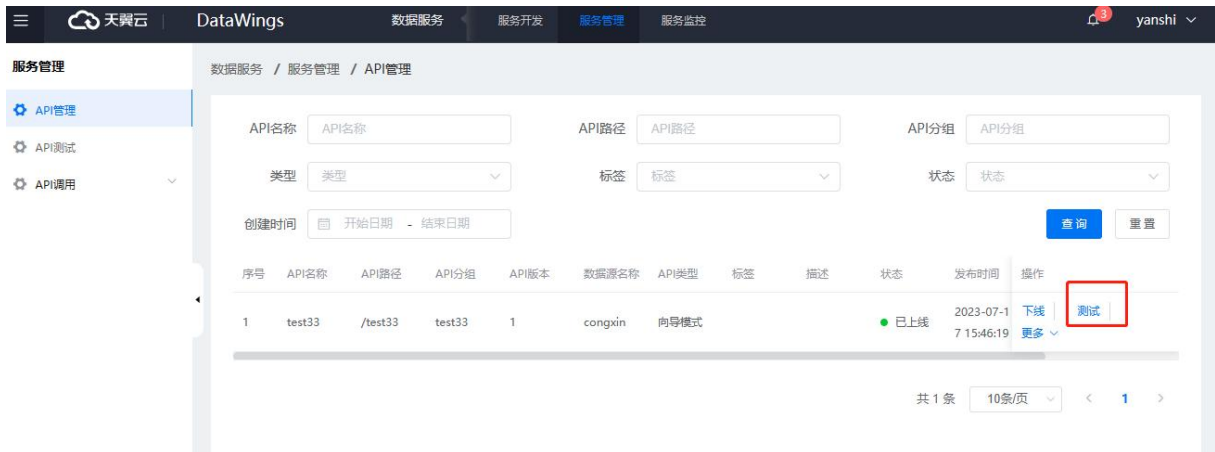
API 测试是所有已发布 API 的测试中心，可选择任意 API 进行单次的调用测试，查看请求详情和返回内容值。

操作说明

1. 登录平台，进入数据服务，选择“服务管理 > API 测试”。
2. 可通过下拉框选择一个 API 进行测试，测试前需填写 API 的参数取值。点击【开始测试】按钮可测试 API，页面右侧将显示请求详情和 API 的返回内容，以便检验数据和功能的有效性。



3. 点击 API 操作栏的【测试】操作可进入 API 测试模块对 API 进行调用测试。



API 调用

应用授权

应用授权功能用于管理可调用 API 的授权应用，支持新增授权应用、管理应用授权和授权有效期等。

操作说明

1. 登录平台，进入数据服务，选择“服务管理 > API 调用 > 应用授权”。可在此页面管理 API 授权应用。
2. 点击【新增应用】按钮，将显示“新增应用”弹窗，填写应用名称、有效期、授权的 API 分组、IP 白名单、描述等信息后，点击【确定】按钮可新增一个授权应用。系统将为用户自动生成密钥（appSecret），生成的应用可用于访问已授权分组里的所有 API。

API授权



* 应用名称 0/128

* 有效期 长期有效

授权的API

IP白名单

描述 0/128

取消

确定

- 3.
4. 支持根据应用名称、创建用户等信息进行应用查询。
5. 点击操作栏的【编辑】按钮，可编辑应用的名称、描述、有效期、授权分组等信息。
6. 点击操作栏的【删除】按钮，可删除相应的应用，删除的应用不可再用于调用 API。仅未绑定 api 的应用支持删除。

流控策略

流控策略功能用于管理作用于 API 的流控策略，页面列表中显示所有已添加的流控策略。支持新增、编辑、删除等操作。

操作说明

1. 点击【新增策略】按钮，可在新增流控策略弹窗中填写流控策略的相关信息。

流控策略



* 策略名称 0/64

* 时间单位

* 调用限制(次)

绑定的API

描述 0/128

取消

确定

字段	说明
策略名称	必填，不可重复，支持中文、英文、数字、下划线（_），且只能以英文或中文开头，1~64 个字符。
时间单位	必选，可选分钟、秒、小时等。
调用限制	必填，可输入单位时间内 API 调用次数的上限。

字段	说明
绑定的 API	非必选，为避免多个流控策略之间含义冲突，一个 API 只能绑定一个流量控制策略，若 API 已经绑定了策略，下拉选项中该 API 不可选。
描述	非必填，流控策略的描述信息。

2. 点击流控策略对应的【编辑】按钮可编辑流控策略，编辑流控策略时，所有内容均可编辑，编辑后的策略自动更新到所绑定的 API。
3. 点击流控策略对应的【删除】按钮可删除流控策略。注意，仅没有绑定 API 的流控策略支持删除。删除时需进行二次确认。

告警策略

告警策略功能用于管理作用于 API 的告警策略，告警策略页面的列表中显示所有已添加的告警策略。支持新增、编辑、删除等操作。

+ 新增策略

序号	策略名称	触发方式	告警条件	告警等级	联系人	机器人	绑定的API	描述	创建时间	创建用户	操作
1	加密验证	满足所有条件	平均响应时间 > 1ms	提示		值班表告...	向导生成...		2024-04-30 17:34:11	hjt1	编辑 删除
2	鉴权告警策略	满足任意条件	平均响应时间 > 1ms	提示	shl1				2024-04-17 20:10:58	shl1	编辑 删除
3	响应时间超时告警	满足任意条件	平均响应时间 > 1ms	严重	shl1	sjfw机器...	向导生成...		2023-12-27 14:37:19	shl1	编辑 删除
4	鉴权告警策略	满足任意条件	调用失败次数 >= 1次	严重	zzf1	测试环境...	api测试...		2024-03-11 16:42:04	zzf1	编辑 删除
5	告警策略	满足所有条件	调用失败次数 > 1次	警告	shl1	sjfw机器...			2023-12-26 10:18:33	shl1	编辑 删除

共 5 条 < 1 >

操作说明

1. 点击【新增策略】按钮，可在新增告警策略弹窗中填写告警策略的相关信息，其中联系

人和机器人需在运维中心的通道管理中配置，目前支持短信、站内信、企业微信机器人、邮箱等告警方式。

告警策略 ×

* 策略名称 0/64

* 触发方式 满足任意条件 满足所有条件

* 告警条件

* 统计频率

* 告警等级

联系人

机器人

绑定的API

描述 0/128

字段	说明
策略名称	必填，不可重复，支持中文、英文、数字、下划线（_），且只能以英文或中文开头，1~64 个字符。

字段	说明
触发方式	必选，支持满足所有条件或满足任意条件时触发告警。
告警条件	告警指标可选调用失败次数、调用失败率、平均响应时间，每项指标只能出现一次。告警条件中，操作符支持>和>=，点击【删除】可删除一行已添加的条件，点击【添加一行】可添加一行告警条件。
统计频率	支持选择每 1 小时、每 24 小时、每 7 天、每 1 个月等。统计频率为整点生效，例如用户在 10:30 添加了告警策略，统计频率为每 1 小时，则在 11:00 时可能触发告警，不会在 11:30 触发告警。
绑定的 API	非必选，一个 API 最多可以绑定 10 个告警策略。若 API 已经绑定了 10 个策略，则下拉选择中该 API 不可选。
告警等级	必选，可选严重、警告或提示。用于标识告警的严重程度。
联系人	可选择告警通知的联系人。
机器人	可选择告警通知的机器人。
描述	非必填，告警策略的描述信息。

2. 点击【编辑】按钮，可编辑修改告警策略的相关信息。
3. 点击【删除】按钮，即可删除当前告警策略，已被使用的告警策略不支持删除。

4.8.4 服务监控

服务监控用于监控已发布的 API，包括监控图表和访问日志两类信息。其中 API 监控功能提供多样化的 API 调用监测图表及统计，包括 API 调用次数、API 调用量排行、API 出错排行、API 调用趋势、高频应用访问等图表，并支持对单个 API 查看请求次数和平均响应时间。访问日志功能提供所有已发布 API 的历史调用记录。

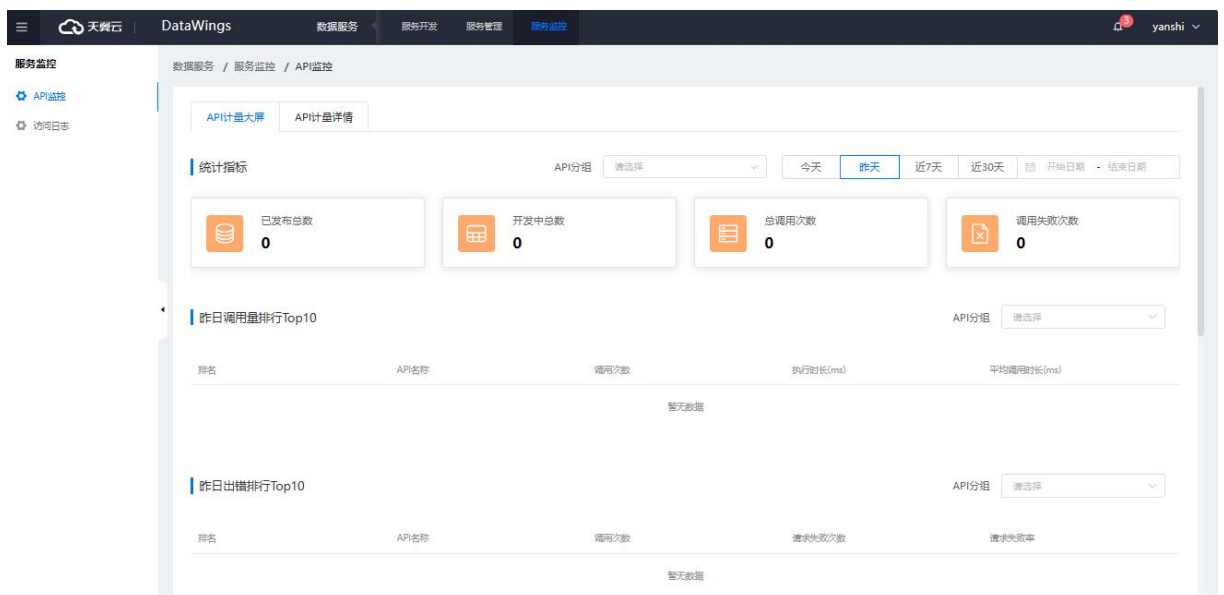
API 监控

API 监控包括“计量大屏”和“计量详情”两个子页面，用于监控 API 的总体情况或单个 API 的运行情况。其中“API 计量大屏”从 API 总数、调用次数、调用量、出错率角度展示 API 的调用信息，“API 计量详情”可查看单个 API 的请求次数和响应时间详情。

API 计量大屏

1. 登录平台，进入数据服务，选择“服务监控 > API 监控”，可查看所有已发布 API 的统计信息。
2. API 计量大屏展示了 API 统计指标、API 调用排行、API 出错排行、API 调用趋势等信息。

3.



1) 统计指标

展示已发布总数、开发中总数、总调用次数、调用失败次数等信息。用户可点击右上角的时间控制栏切换统计的时间段，支持切换查看今天、昨天、近 7 天、近 30 天或任意时间段的统计信息。

2) 昨日调用量排行 Top10

以列表形式展示昨日调用量最多的前 10 个 API 的相关信息，包括 API 名称、调用次数、总执行时长、平均调用时长等。

3) 昨日出错排行 Top10

以列表形式展示昨日调用出错次数最多的前 10 个 API 的相关信息，包括 API 名称、调用次数、失败次数、失败率等。

4) API 调用趋势

以折线图形式展示某个时间段内 API 调用趋势的变化。用户可点击右上角的时间控制栏切换统计的时间段，支持切换查看今天、昨天、近 7 天、近 30 天或任意时间段的统计信息。

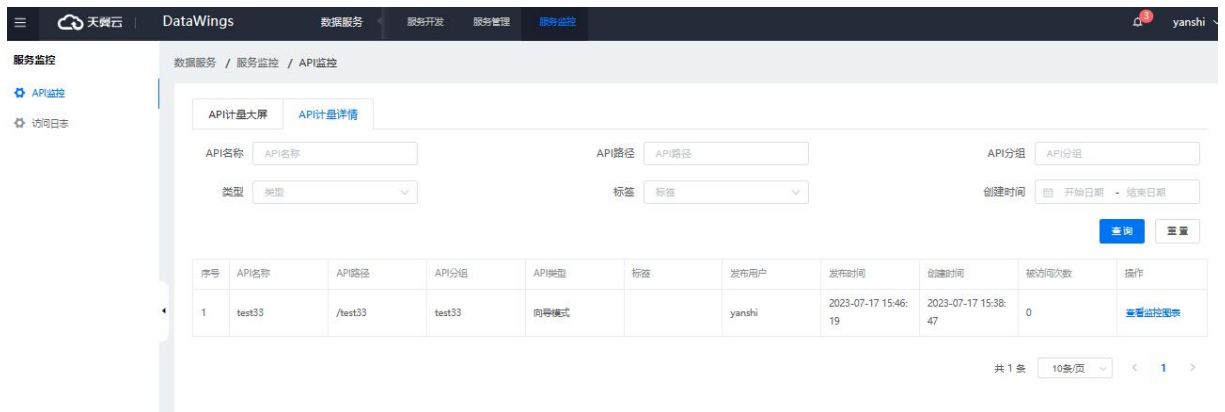
5) 高频访问应用 Top10

以列表形式展示某个时间段内调用次数最多的前 10 个 API 的相关信息。用户可点击右上角的时间控制栏切换统计的时间段，支持切换查看今天、昨天、近 7 天、近 30 天或任意时间段的统计信息。

API 计量详情

1. 登录平台，进入数据服务，选择“服务监控 > API 监控 > API 计量详情”，可查看单个 API 的运行情况。
2. API 计量详情展示了所有已发布 API 的基本信息支持点击【查看监控图表】，查看单个

API 的调用信息。



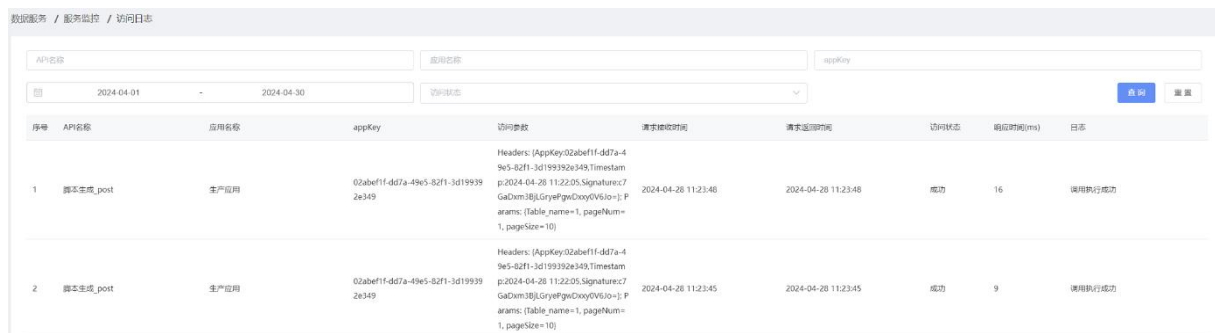
3.

访问日志

访问日志功能用于记录所有已发布 API 的访问记录，按照访问 API、访问应用、访问参数、响应时间、访问结果等维度形成访问日志，提供日常审计的查询入口。

操作说明

1. 登录平台，进入数据服务，选择“服务监控 > 访问日志”，可查看所有已发布 API 的访问日志信息。
2. 访问日志列表支持查看 API 名称、应用名称、appKey、访问参数、请求接收时间、请求返回时间、访问状态、响应时间、日志等信息。支持从 API 名称、应用名称、appKey、请求接收时间等维度进行日志查询。



4.8.5 服务配置

服务配置模块主要用于对数据服务模块进行全局配置。

全局访问控制

全局访问控制功能用于设置数据服务模块全局的 IP 黑名单或 IP 白名单，启用后对整个数据服务模块的所有 API 生效。全局访问控制策略仅可设置黑名单（禁止访问）或白名单（允许访问），不可同时设置两类策略。黑名单表示除黑名单之外的 IP 地址均可调用数据服务 API，白名单表示仅白名单之内的 IP 地址可调用数据服务 API。

注意，以下情况表示任意 IP 均可调用数据服务 API：

- 1) 未添加全局访问控制。
- 2) 全局访问控制策略未添加任何 IP 地址。
- 3) 全局访问控制策略未启用。

若用户首次使用全局访问控制功能，或未添加过全局访问策略，进入此功能时将显示初始空白页面。可点击【立即添加】按钮，点击后进入全局访问控制策略的配置页面。



数据服务 / 服务配置 / 全局访问控制

基本设置 编辑模式 只读模式

* 限制方式 允许访问 禁止访问

* 是否启用 否 是

* IP地址 + 添加IP地址

限制方式：可选允许访问或禁止访问，允许访问表示仅允许指定的 IP 地址可调用 API，禁止访问表示禁止指定的 IP 地址调用 API。

是否启用：使用开关控件表示启用状态，开启表示启用，关闭表示不启用。仅启用状态的全局

访问控制策略会生效。

IP 地址：限制或允许访问的 IP 地址。点击【+添加 IP 地址】按钮可添加一行。点击【删除】按钮可删除所选行。IP 地址输入框内需输入合法的 IP 地址。

4.9 运维中心

4.9.1 查看运维大屏

运维大屏为您展示重点关注的任务运维指标及运行概况，帮助您快速了解大数据管理平台任务的整体概况，及时发现并处理异常任务，提升运维效率。

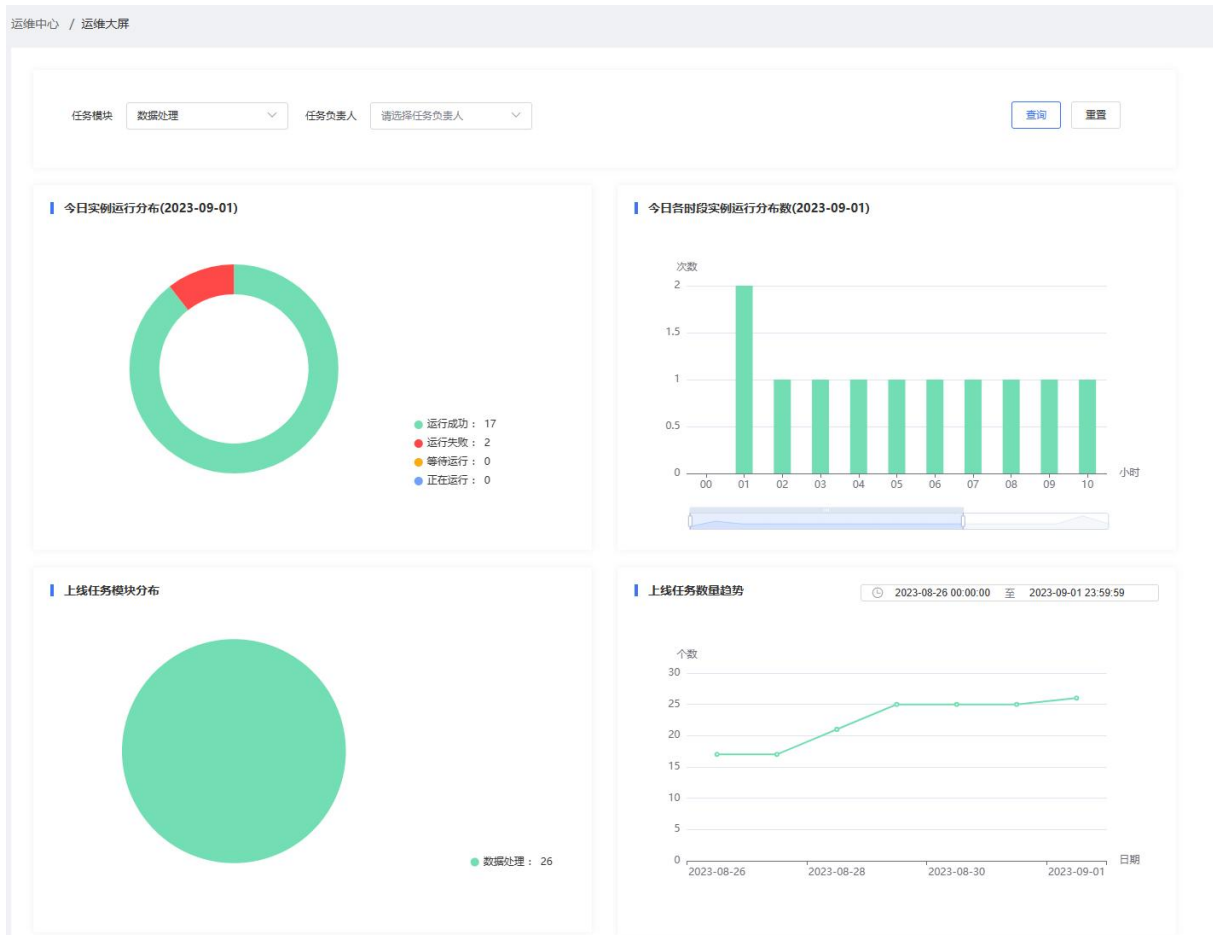
使用说明

- 您可以选择任务所在模块，查看当前模块下实例运行指标和统计数据，目前仅接入了数据处理模块的任务实例。
- 您可以选择任务负责人，查看当前负责人在某一模块下的实例运行指标和统计数据。

查看运维关键指标

任务运维关键指标可详见下表。

图表名称	说明
今日实例运行分布	统计当日 0 点-24 点运行的实例不同运行状态的分布。
今日各时段实例运行分布数	统计当日 0 点-24 点运行的实例在不同小时段的分布。
上线任务模块分布	统计当日已上线任务在不同模块的分布情况。
上线任务数量趋势	统计已上线任务的日趋势。



查看任务耗时及失败排行

任务耗时图表说明详见下表。

图表名称	说明
任务运行耗时排行 top10	统计指定时间内任务的运行最大耗时及平均耗时，并按照平均耗时展示耗时时间最长的 10 个数据，仅统计周期执行的任务实例。
任务运行失败次数排行 top10	统计指定时间内任务的失败次数，并按照失败次数展示失败次数最多的 10 个数据，仅统计周期执行的任务实例。

任务运行耗时排行top10 🕒 2023-08-26 00:00:00 至 2023-09-01 23:59:59

排行	任务/节点id	任务名称	任务模块	任务负责人	运行最大耗时	运行平均耗时
1	10732681817664	data_quality_1436	数据处理	zzf0828	246.00s	229.00s
2	10734606420928	data_quality_2113	数据处理	zzf0828	183.00s	183.00s
3	10724283483840	sql_3133	数据处理	wangktest	183.00s	183.00s
4	10734386072768	data_quality_6611	数据处理	zzf0828	183.00s	183.00s
5	10734606420544	data_quality_4451	数据处理	zzf0828	183.00s	183.00s
6	10734606420160	data_quality_6611	数据处理	zzf0828	183.00s	183.00s
7	10724373008832	hql_644	数据处理	gzw_tenantry_2	124.00s	124.00s
8	10724373009216	sql_1720	数据处理	gzw_tenantry_2	124.00s	124.00s
9	10724506771648	sql_1621	数据处理	zzf0821	123.00s	123.00s

任务运行失败次数排行top10 🕒 2023-08-26 00:00:00 至 2023-09-01 23:59:59

排行	任务/节点id	任务名称	任务模块	任务负责人	运行失败次数
1	10724373008832	hql_644	数据处理	gzw_tenantry_2	1
2	10724373009216	sql_1720	数据处理	gzw_tenantry_2	1
3	10732681817664	data_quality_1436	数据处理	zzf0828	1

4.9.2 告警管理

4.9.2.1 通道及对象管理

告警管理中心主要为您提供通道管理、联系人管理、机器人管理、联系人组管理能力，帮助

您更自由便捷地定义告警配置信息。

通道管理

当前 DataWings 支持的告警通道包括：短信、站内信、邮箱、钉钉机器人、飞书机器人、企微机器人。

说明：

- 1、使用某一通道前你需要先新增一个通道，并上线该通道。
- 2、通道下线后，如果仍有消息发送到此通道，将会发送失败，请谨慎操作通道下线。

运维中心 / 告警管理 / 通道管理 / 通道配置

[+ 新增通道](#)

序号	通道名称	状态	通道说明	操作
1	钉钉机器人	● 上线		编辑 下线
2	邮箱	● 上线		编辑 下线
3	飞书机器人	● 下线		编辑 上线
4	站内信	● 上线		编辑 下线
5	短信	● 上线		编辑 下线
6	企微机器人	● 上线		编辑 下线

新增机器人



类型

* 名称

* 机器人地址

签名密钥

备注 0/100

取消

确定

联系人管理

你可以在联系人管理里面管理告警发送对象，并且支持关联 DataWings 用户、批量导入联系人。

序号	联系人	关联用户名	手机号	邮箱	操作
1	weibin051009		155****9603	f*****@chinaunicom.cn	编辑 删除
2	hjren14				编辑 删除
3	hjren12	hjren12			编辑 删除
4	q8		177****4773	q**@qq.com	编辑 删除
5	tsu1547547				编辑 删除
6	hj2	gnewid27			编辑 删除
7	tsu45715813				编辑 删除
8	tsu4597956				编辑 删除
9	tsu4600722				编辑 删除
10	tsu46019271				编辑 删除

机器人管理

您可以在机器人管理页面管理群机器人配置（目前支持企微机器人、飞书机器人、钉钉机器人），

新增机器人时可配置机器人的 Webhook 信息及签名信息。



序号	机器人名称	类型	备注	操作
1	201201	企业机器人		设置 编辑 删除
2	测试环境-灾备项目-数据服务告警	企业机器人		设置 编辑 删除
3	飞行机器人123	飞行机器人		设置 编辑 删除
4	测试环境123机器人	飞行机器人		设置 编辑 删除
5	飞行机器人_23F	飞行机器人		设置 编辑 删除
6	飞行机器人	飞行机器人		设置 编辑 删除
7	好字文-企业机器人	企业机器人		设置 编辑 删除
8	测试机器人	企业机器人		设置 编辑 删除
9	测试环境告警机器人	企业机器人		设置 编辑 删除
10	企业级项目测试	企业机器人	测试	设置 编辑 删除

联系人组管理

联系人组管理可以帮您分组管理告警对象，配置告警对象时，可以直接指定联系人组，发送告警信息时，可直接对多个联系人生效。



序号	联系人组名称	联系人数量	创建人	操作
1	演示联系人组	4	gzw_tenantry_2	编辑 删除

4.9.2.2 告警实例

您可以在告警实例页面查看所有模块的监控告警信息，包括数据集成、数据质量、数据服务、数据地图、自定义监控规则等产生的告警信息。

进入告警实例页面

1. 点击左上角主菜单按钮，选择【运维中心】，点击【告警管理】。
2. 进入告警管理页面后，点击左侧导航栏【告警实例】。

查看告警实例列表

告警实例列表支持按数据来源、告警通道、告警结果、业务来源、等级、时间段查询检索告警记录。

点击详情您可以查看当前告警实例的详细信息。

序号	告警时间	业务来源	数据来源	等级	结果	告警来源	操作
1	2024-05-22 16:27:40	数据质量	zdf_test_standard_20240319_004_执行数_历史值 Z...	警告	成功	钉钉企业微信机器人	详情
2	2024-05-22 16:27:39	数据质量	zdf_test_standard_20240319_003_执行数_历史值 Z...	警告	成功	钉钉企业微信机器人	详情
3	2024-05-22 16:27:26	数据质量	zdf_test_standard_20240319_001_执行数_历史值 Z...	警告	成功	钉钉企业微信机器人	详情
4	2024-05-22 16:27:25	数据质量	zdf_test_standard_20240319_002_执行数_历史值 Z...	警告	成功	钉钉企业微信机器人	详情
5	2024-05-22 14:48:56	数据质量	空串问题	严重	成功	钉钉企业微信机器人	详情
6	2024-05-22 14:39:44	数据质量	数据异常警告	提示	成功	钉钉企业微信机器人	详情
7	2024-05-22 14:28:35	数据质量	dsm_perf_20240314_001_order_amount_工单量	警告	存在异常	钉钉企业微信机器人	详情
8	2024-05-22 14:28:22	数据质量	特殊字符0423002	警告	存在异常	钉钉企业微信机器人	详情
9	2024-05-22 14:28:21	数据质量	连续检测到0424apocessuser	警告	存在异常	钉钉企业微信机器人	详情
10	2024-05-22 14:27:52	数据质量	连续检测到0423	警告	存在异常	钉钉企业微信机器人	详情

查看告警详情

一个告警实例中可能包含一个或者多个告警信息，您除了可以查看基本信息外，还可以按照告警通道、告警对象、日志内容再做进一步查询，如果告警消息在发送期间终止或者失败，您也可以在告警日志中查看具体的异常信息。

运维中心 / 告警管理 / 告警实例 / 详情

< 详情

基本信息

告警时间	2023-09-04 18:47:18	业务来源	data_quality
数据来源	非default集群_数据延迟行数_固定值	告警结果	成功
告警通道	企业机器人	告警等级	严重

日志查询

告警通道: 告警对象: 日志内容:

仅显示异常

告警通知: 通过企业机器人通知了非default集群告警机器人

暂无日志内容

4.9.2.3 值班表

值班表模块为您提供运维排班能力，您可以通过自定义值班表，在触发监控告警时自动将告警消息发送给值班人员，以便快速响应和解决问题。

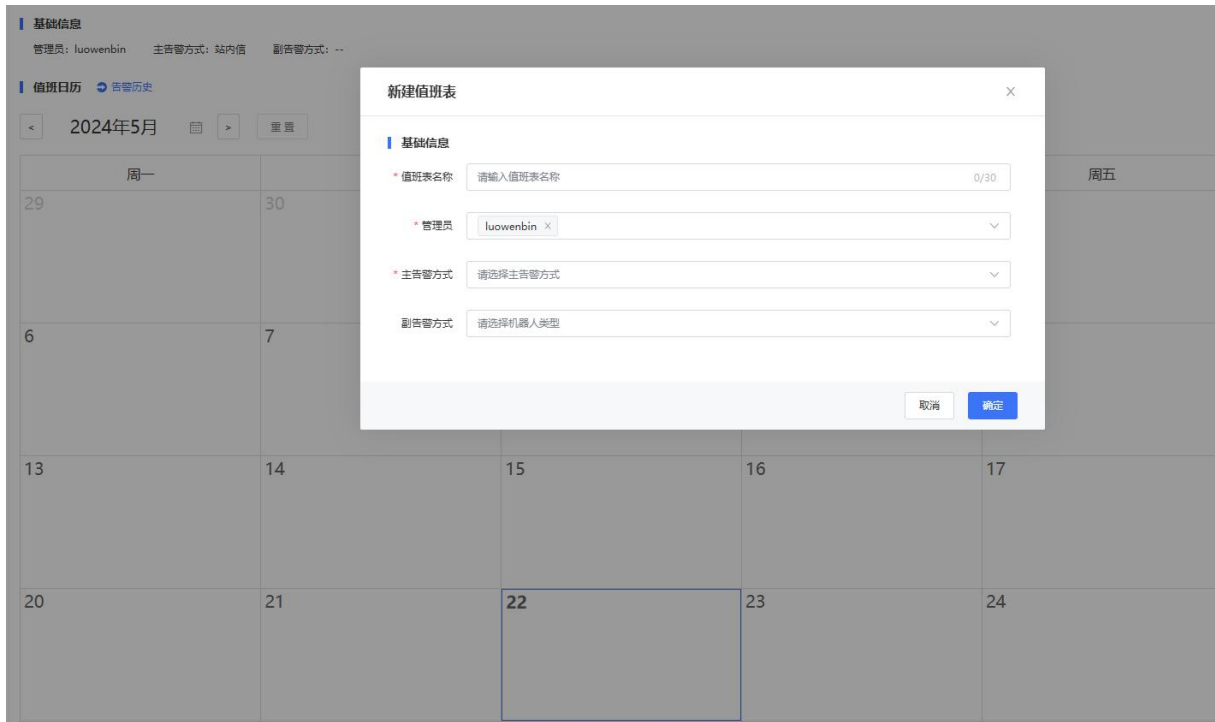
前提条件

1. 已上线邮箱、站内信、短信告警通道中的至少一个。

创建值班表

值班表基础信息说明如下：

配置项	解释说明
值班表名称	输入值班表的名称。
管理员	选择当前值班表的管理员，仅管理员可编辑、删除值班表。
主告警方式	告警优先发送的告警通道，仅支持短信、邮箱、站内信。
副告警方式	主告警方式发送失败后，会将告警消息发送至副告警方式，副告警方式仅支持企微机器人、飞书机器人、钉钉机器人。

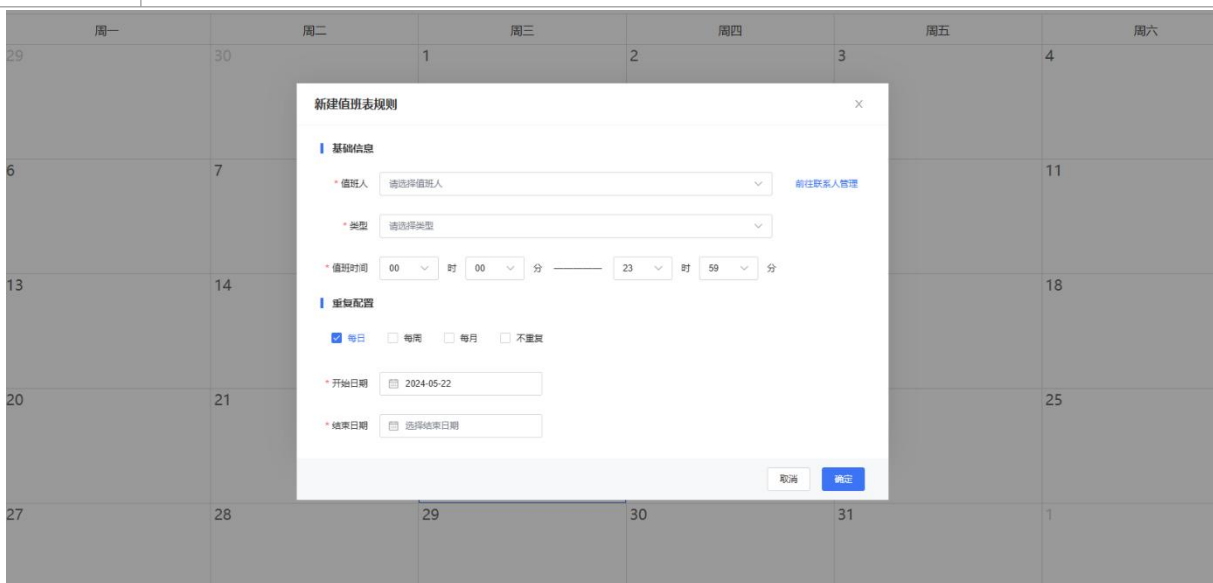


新建值班规则

您可以选中一个值班表，并对当前值班表创建多个值班规则，值班规则说明如下：

配置项	解释说明
值班人	选择当前值班规则的值班人。
类型	主值班：相同时间优先发送告警消息给主值班人。 备值班：同一告警消息发送 2 次后，第三次会将告警消息发给所有主值班人和备值班人。
值班时间	您可以指定当前值班规则的值班时间段。

配置项	解释说明
重复配置	支持每日、每周、每月、不重复四种类型。
开始日期	当前值班规则的开始日期。
结束日期	当前值班规则的结束日期。




4.9.3 资源运维

资源运维模块支持您监控并查看数据管理平台各个资源组的状态及运维使用情况。

前提条件

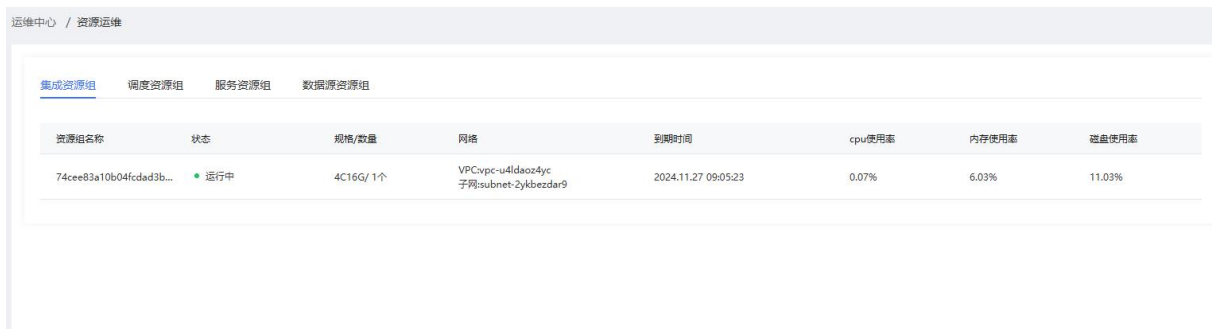
使用前请确保您至少已购买集成资源组、调度资源组、服务资源组中的一种或多种。

页面进入

单击左上方的图标，选择全部产品 > 运维中心 > 资源运维。

查看资源运维信息

- 您可以切换集成资源组、调度资源组、服务资源组、数据源资源组，查看不同类型资源组的信息。
- 资源运维信息包括：资源组名称、状态、规格/数量、网络、到期时间、cpu 使用率、内存使用率、磁盘使用率。
- 您可以点击运维中心 > 监控管理，前往创建资源组使用率的运维监控规则。



截图显示了资源运维的页面，包含资源组列表。列表表头如下：

资源组名称	状态	规格/数量	网络	到期时间	cpu使用率	内存使用率	磁盘使用率
74cee83a10b04fcdad3b...	运行中	4C16G/ 1个	VPCvpc-u4ldaoz4yc 子网:subnet-2ykbzdar9	2024.11.27 09:05:23	0.07%	6.03%	11.03%

注意：

资源组的 cpu 使用率、内存使用率、磁盘使用率为资源组粒度的监控指标，计算方式为资源组下各云主机指标的平均值。

4.9.4 监控规则管理

自定义监控规则为您提供了不同业务指标的监控能力，您可以监控资源组的状态和运维指标、任务基线等，以便及时发现和处理异常问题。

使用限制

- 自定义监控当前仅支持了资源组的使用率指标，其他业务监控敬请期待。
- 定义指标的告警通道和告警对象依赖告警管理的信息配置。

查看自定义监控规则列表

操作说明

功能点	描述
检索	可以按名称、业务来源、状态、等级查询自定义规则。
下线	下线当前规则，下线后当前规则将不再触发告警及运维动作。
上线	上线当前规则，上线后不可编辑。
编辑	重新编辑当前规则，仅下线的规则支持编辑。
删除	删除当前规则，仅下线的规则支持删除。

运维中心 / 监控管理 / 规则管理

规则名称 业务来源 状态 等级

[+ 新建规则](#)

序号	规则脚本名称	告警等级	业务来源	状态	告警通道	备注	操作
1	内存使用率连续10分钟超2...	严重	运维中心	已上线	企业机器人		查看 下线 编辑 删除
2	cpu使用率_host_cpu_util	严重	运维中心	已上线	企业机器人		查看 下线 编辑 删除
3	磁盘使用率_host_disk_util	严重	运维中心	已上线	企业机器人		查看 下线 编辑 删除
4	监控内存使用率_host_me...	警告	运维中心	已上线	企业机器人		查看 下线 编辑 删除
5	测试资源组规则	严重	运维中心	未上线	站内信, 短信		查看 上线 编辑 删除
6	测试自定义规则	严重	运维中心	未上线	站内信		查看 上线 编辑 删除

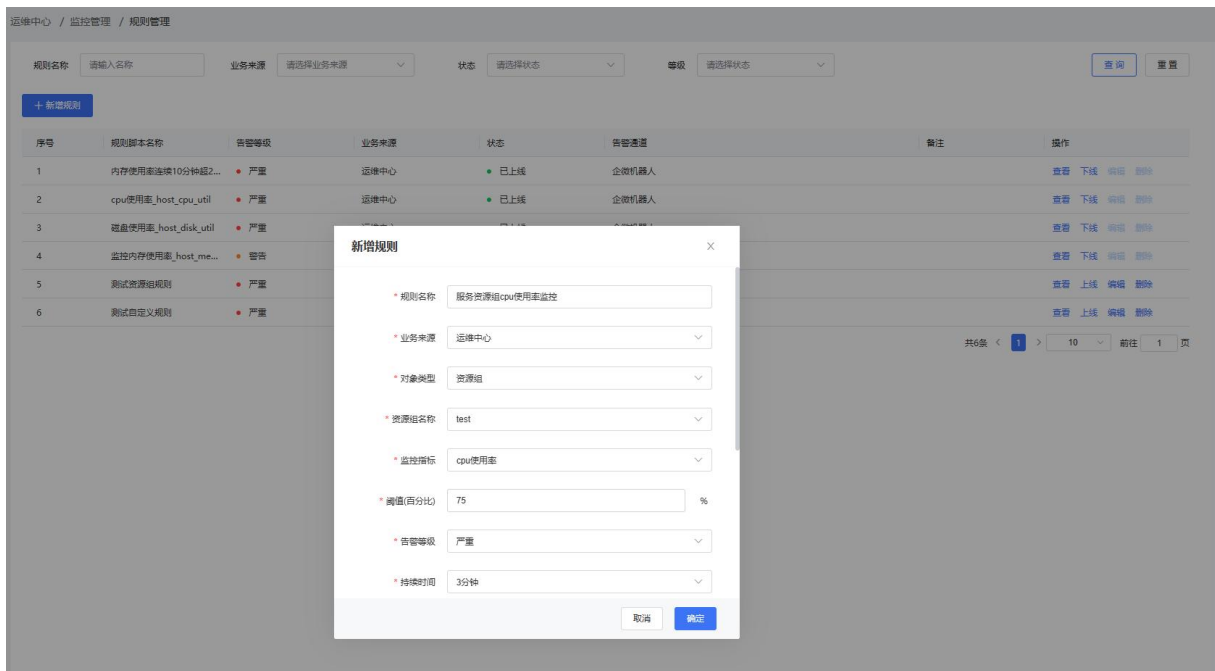
共6条 < 1 > 10 前往 1 页

创建自定义监控规则

规则配置说明

字段	说明

字段	说明
业务来源	当前自定义规则涉及的业务模块。
对象类型	当前自定义规则需要监控的对象类型，当前仅支持资源组。
资源组名称	资源组名称。
监控指标	当对象类型为资源组时，需要填写，可选择：cpu 使用率、内存使用率、磁盘使用率。
阈值（百分比）	当对象类型为资源组时，需要填写，当监控指标大于等于当前阈值且满足持续时间时，触发告警。
告警等级	当前自定义规则的告警级别，可选择：严重、警告、提示。
持续时间	当监控指标大于等于当前阈值且满足持续时间时，触发告警。
最大报警次数	最小告警间隔时间内最大报警次数，超过将不触发告警，用于调节告警发送频率。
最小告警间隔时间	最小告警间隔时间内最大报警次数，用于调节告警发送频率。



4.9.5 操作日志

您可以在操作日志模块查询和检索 DataWings 内进行的关键操作，同时也为您提供了日志总览的能力，查看不同模块的使用情况。

查看操作日志历史记录

功能说明

字段	说明
操作人	日志操作的执行账号。
操作时间	日志操作的记录时间。
操作页面	日志操作的所在页面名称。

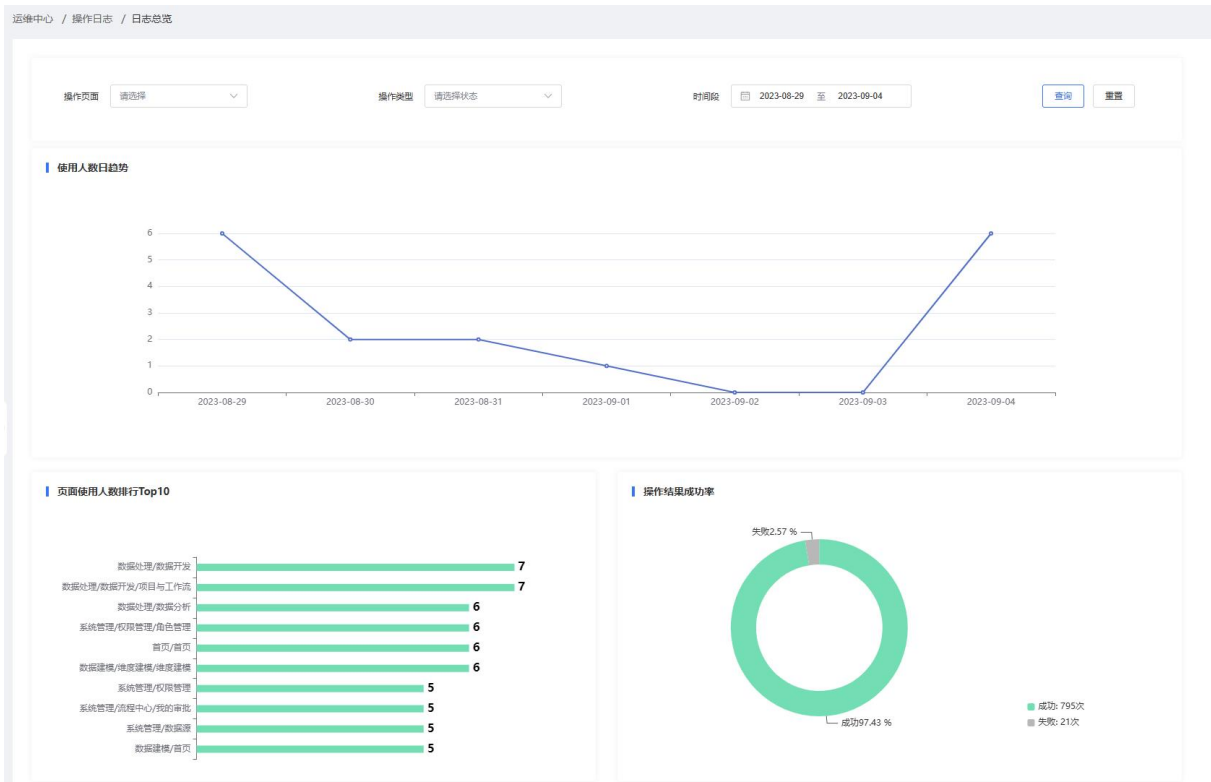
字段	说明
操作类型	日志的操作类型，例如查询、编辑、上线、下线等。
操作对象	日志的操作对象，例如列表、字段标准、数据模型、质量规则等。
操作内容	日志的操作信息。
IP	日志的操作 IP 地址。
地域	日志的操作 IP 地域，基于 IP 地址解析。
操作结果	日志的操作结果，成功/失败。

说明：

- 1、当前日志仅对部分关键操作进行了日志记录，可用于日志审计参考。
- 2、当前日志仅为您保留近半年的日志存储。

查看日志总览

图表名称	说明
使用人数日趋势	可统计展示不同页面、不同类型的日操作人数，按账号 id 进行去重。
页面使用人数排行	可统计当前时段内不同页面的操作人数排行最多的 10 个页面，操作人数按账号 id 进行去重。
操作结果成功率	可统计当前时段所有操作的结果状态分布。



4.10 系统管理

4.10.1 权限管理

4.10.1.1 角色管理

管理员可通过角色管理功能，根据使用权限创建不同角色，并将角色配置给对应子用户，实现对子用户的操作权限管控，加强资源访问的安全性。

背景信息

DataWings 角色管理功能基于 RBAC（Role-based access control）权限模型构建权限管理功能，用户可通过创建角色，对不同角色配置不同权限，并分配角色给所需子用户，以实现菜单、按钮等资源的权限管控。

前提条件

公有云场景下，主用户需先通过天翼云“主子账号及授权中心”对目标子用户授权使用 DataWings，

再使用 DataWings 角色管理功能进行详细权限管控。

名词释义

预设角色：系统内置角色，可进行查看与角色分配，无法进行修改与删除。

自定义角色：用户按需创建角色，可进行查看、角色分配、修改与删除操作。

进入页面

单击左上方的  图标，选择【系统管理】-【权限管理】-【角色管理】进入页面进行角色配置。

操作说明

预设角色

1. 角色说明：当前各版本中均配置了租户管理员、审批管理员与访客角色，功能详情如下：

角色名称	菜单权限	数据权限	默认角色分配	角色说明
租户管理员	全部	全部	主用户	超级管理员
审批管理员	全部	无	-	用于审批菜单 类角色申请
访客	首页、流程中心	无	全部子用户	普通用户

说明

租户管理员可按需将租户管理员角色与审批管理员角色分配给相应子用户。

2. 管理角色：点击列表页操作中的选项，可对角色进行查看与角色分配操作。

操作	描述
查看	点击查看，可在新页面中查看角色详情。
角色分配	点击角色分配，可在弹窗中添加想要分配该权限的用户，或删除某用户以取消其角色权限。

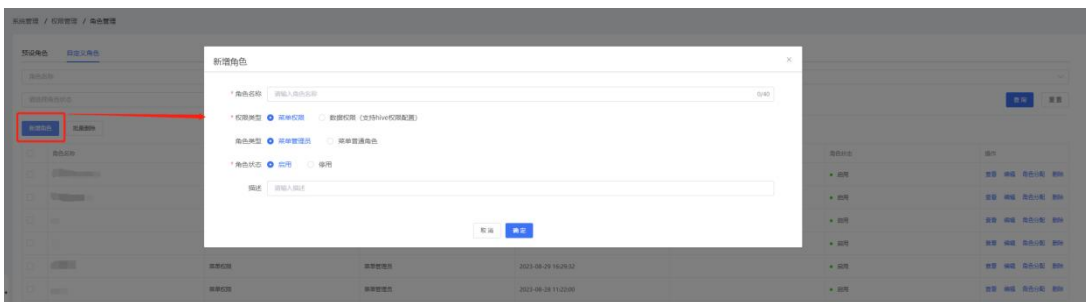
自定义角色

1. 角色说明：管理员可根据业务场景，按需创建角色，并对该角色进行管理。DataWings 支持以下 2 类角色的创建，详情可参考下表：

权限类型	角色类型	使用权限	分配权限
菜单权限	菜单管理员	√	√
	菜单普通角色	√	-

2. 新增角色。

a. 基础信息配置：点击“新增角色”，在弹窗中配置角色的基础信息，包括角色名称、权限类型、角色类型、角色状态与描述，并点击“确定”创建角色。



b. 权限配置。

菜单权限配置：若选择创建“菜单权限”类角色，将进入下图页面进行配置，管理员可按需选

择菜单及权限范围，点击“保存”完成配置。

说明
读写权限包含只读与自定义中全部按钮权限；未选择菜单将默认无权限。

基础信息

角色名称	测试角色	权限类型	菜单权限
角色类型	菜单管理员	角色状态	<input type="radio"/> 停用 <input checked="" type="radio"/> 启用
描述	请输入内容		

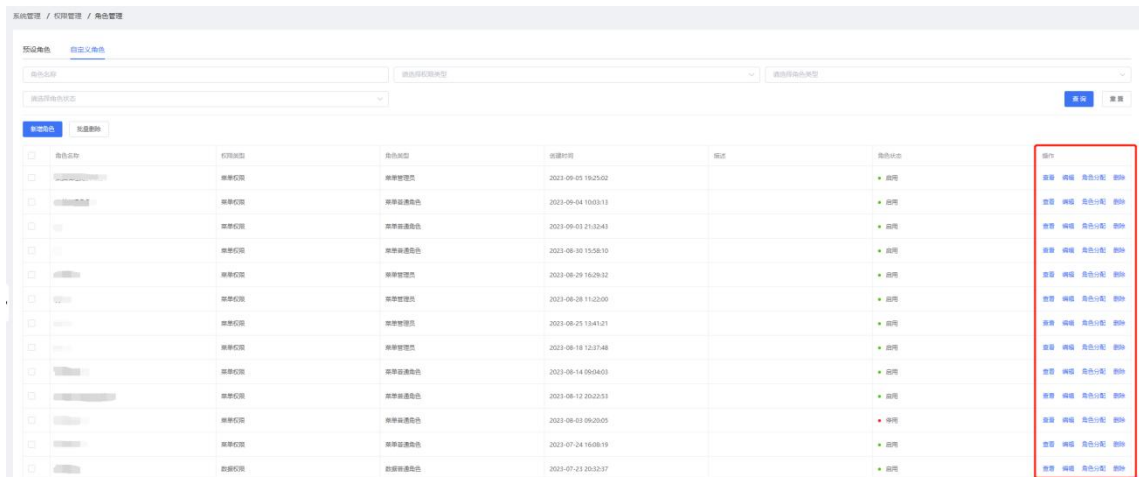
0/200

权限信息

菜单	操作
数据标准	<input type="radio"/> 无权限 <input type="radio"/> 只读 <input type="radio"/> 读写 <input type="radio"/> 自定义
命名词典	<input type="radio"/> 无权限 <input type="radio"/> 只读 <input type="radio"/> 读写 <input type="radio"/> 自定义
字段标准	<input type="radio"/> 无权限 <input type="radio"/> 只读 <input type="radio"/> 读写 <input type="radio"/> 自定义
标签领域	<input type="radio"/> 无权限 <input type="radio"/> 只读 <input type="radio"/> 读写 <input type="radio"/> 自定义
度量单位	<input type="radio"/> 无权限 <input type="radio"/> 只读 <input type="radio"/> 读写 <input type="radio"/> 自定义
规范管理	<input type="radio"/> 无权限 <input type="radio"/> 只读 <input type="radio"/> 读写 <input type="radio"/> 自定义

3. 管理角色：点击列表页操作中的选项，可对角色进行查看、编辑、角色分配、删除操作。此外，可点击页面左上角“批量删除”按钮对角色进行批量删除。

操作	描述
查看	点击查看，可在详情页中查看该角色详细配置信息。
编辑	点击编辑，可在详情页中对角色状态、描述、权限范围进行修改。
角色分配	点击角色分配，可在弹窗中添加想要分配该权限的用户，或删除某用户以取消其角色权限。
删除	点击删除，可删除对应角色。
批量删除	勾选需要删除的角色，并点击批量删除，可批量删除角色。



4.10.2 集群管理

4.10.2.1 集群注册

集群注册页面为您提供集群信息管理的能力，您可以同步和编辑当前主账号已购买的翼 MR 大数据集群。


背景信息

为了能够更好的使用大数据管理平台的能力，若您需要使用数据处理模块，需先购买翼 MR 大数据集群，并绑定大数据引擎。

前提条件

已购买开通翼 MapReduce 产品。

进入页面

单击左上方的图标，选择全部产品 > 系统管理 > 集群管理 > 集群注册。

同步集群信息

- 点击同步集群信息，DataWings 将自动同步当前租户主账号下已购买开通的翼 MapReduce 集群信息（目前仅支持数据湖集群）。
- 当翼 MR 集群信息变更时，您可以再次点击同步按钮更新集群信息，您也可以点击编辑手动更新集群信息。

操作说明

操作	说明
查看	查看当前集群信息。
编辑	编辑当前集群信息。
前往绑定	点击前往绑定，打开引擎管理页面，您可以在引擎管理基于集群创建大数据引擎。
删除	删除当前集群，已绑定引擎的集群不支持删除。

请输入输入集群名称 [查询] [重置]

+ 集群注册 同步集群信息

序号	集群名称	业务场景	创建时间	修改时间	操作
1	datawings0515	数据湖场景	2024-05-15T16:20:26	2024-05-15T16:20:28	查看 编辑 前往绑定 删除

共 1 条 20条/页 < 1 > 前往 1 页

手动注册集群

- 手动注册集群信息需要提前获取翼 MapReduce 集群信息，您可以在已购买的翼 MapRuce 集群的 manager 控制台获取相关信息。
- 集群信息获取可参考 [获取翼 MapReduce 数据湖集群信息最佳实践](#)。

基本信息

* 集群类型

* 集群场景

* 集群名称 0/30

Hadoop

* 配置文件

Yarn

* 版本选择

* 配置文件

Hive

* 版本选择

* 配置文件

Spark

* 版本选择

* 配置文件

Kerberos

是否开启Kerberos

Doris

是否开启Doris

资源组网络配置

hosts

① 请按照格式填写主机配置: 'IP地址 主机名', IP地址只能包含数字和 '.', 主机名只能包含字母、数字、 '-' 或 '.', 每一个配置, 用换行符分隔。

注意:

集群信息涉及底层基础信息, 填写错误信息将大范围影响已创建或执行的大数据任务, 请确保集群信息填写正确。

4.10.2.2 引擎管理

集群信息同步或者注册完成后, 您需要创建翼 MR 计算引擎, 以便通过 DataWings 来执行该引擎类型的大数据计算任务。

前提条件

已购买独享调度资源组, 并已绑定好 VPC 网络, 确保调度资源组与翼 MR 集群间网络连通。


注意:

购买资源组时, 请尽量确保资源组和 DataWings 在同一个 VPC, 否则可能需要额外购买对等连接打通网络。

使用限制

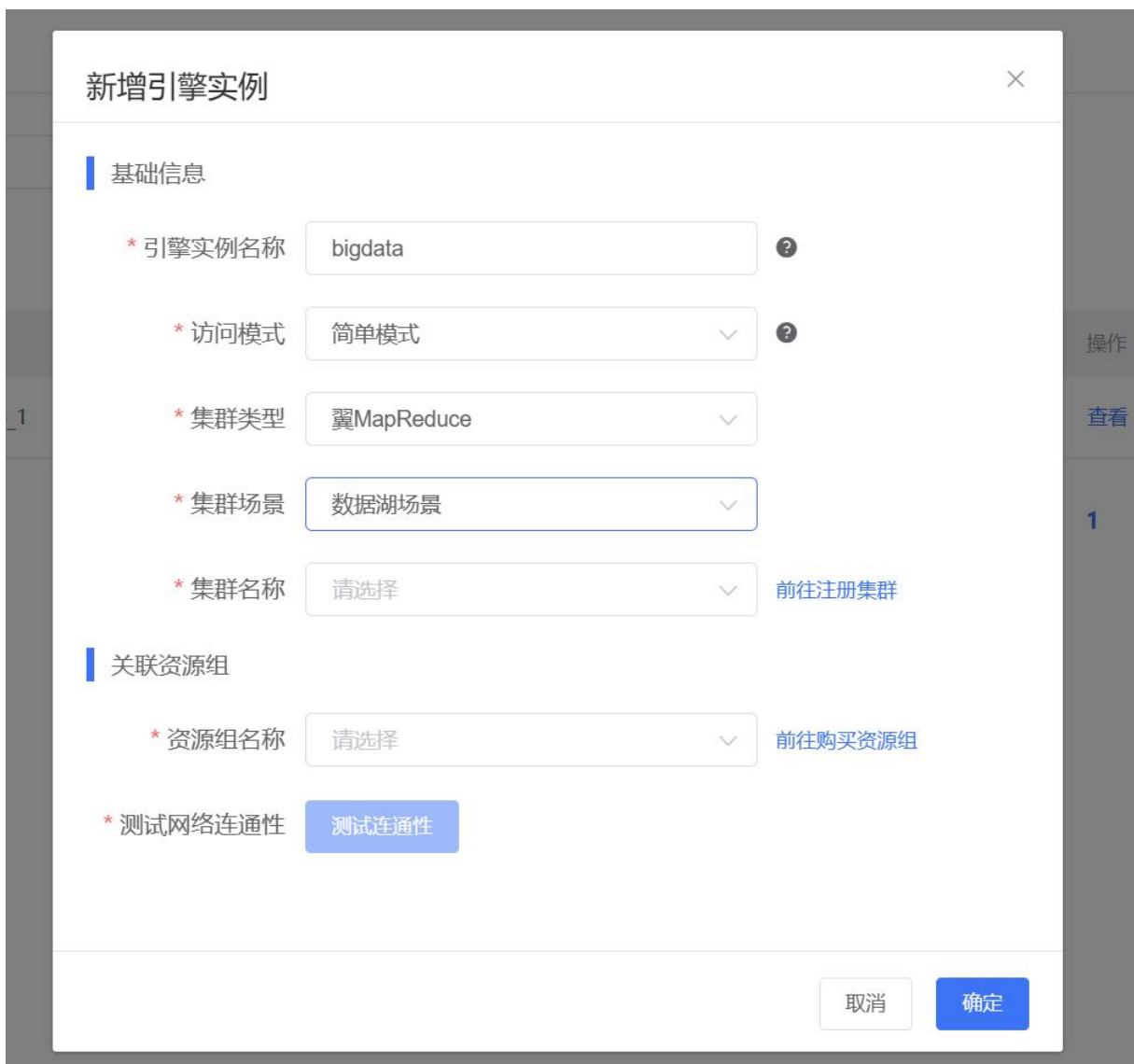
- 仅支持使用独享调度资源组在 DataWings 运行大数据计算任务。

页面进入

单击左上方的图标，选择全部产品 > 系统管理 > 集群管理 > 引擎管理。

功能说明

新增引擎实例



新增引擎实例

基础信息

- * 引擎实例名称: bigdata
- * 访问模式: 简单模式
- * 集群类型: 翼MapReduce
- * 集群场景: 数据湖场景
- * 集群名称: 请选择 [前往注册集群](#)

关联资源组

- * 资源组名称: 请选择 [前往购买资源组](#)
- * 测试网络连通性: [测试连通性](#)

取消 确定

1、配置引擎基本信息

参数	说明
引擎实例名称	当前引擎在 DataWings 的名称，用于标识当前引擎的配置信息。
访问模式	简单模式：DataWings 用户绑定同一个翼 MR 集群用户（LDAP 用户），所有引擎任务使用同一个集群用户身份提交并执行任务。
集群类型	当前仅支持翼 MapReduce 集群。
集群场景	支持数据湖场景和数据分析场景。
集群名称	已注册的集群名称，详情请见 集群注册 。

2、配置资源组信息

资源组名称：已购买的独享调度资源组

测试连通性：您需要确保调度资源组与集群之间网络是互通的，才能成功创建引擎以及执行引擎任务。

配置引擎信息

注意：

引擎初始化完成后，您需要进一步配置引擎信息，才能完成引擎的上线。

参数	说明
----	----

参数	说明
是否启用 hudi	若您购买的翼 MR 集群包含 Spark-Hudi 客户端，您可以选择启用 Hudi，启用后在数据分析模块，您可以创建 Hudi 类型的即席查询任务。
队列	当前引擎可以使用的集群队列列表。
default 队列	当前引擎默认使用的队列。
访问身份配置	集群用户（LDAP 用户）默认配置为 hive 用户，您可以重新输入以修改集群用户。

翼MR
引擎配置 ×

请输入引擎

+ 新增引擎

序号	引擎	状态
1	data	上线

基本信息

引擎实例名称

访问模式

是否启用Hudi ?

队列配置

* 队列

* default队列

访问身份配置

集群用户

配置校验

取消
确定

4、默认数据源及数据地图采集器

数据湖引擎：数据湖引擎上线后，系统将为您自动创建 Hive 数据源及 Doris 数据源（需在集群注册已配置 Doris 信息）、元数据采集器（若您购买的版本包含数据地图功能）。

数据分析引擎：数据分析引擎上线后，系统将为您自动创建 Doris 数据源、元数据采集器（若您购买的版本包含数据地图功能）。

引擎其他操作说明



操作	说明
查看	查看引擎信息。
编辑	编辑引擎信息。
下线	针对已上线状态的引擎，您可以下线当前引擎。 引擎下线后，资源组服务将关闭，您将无法提交任务至当前引擎，已经执行的任务将会执行失败，数据源、元数据采集不受影响。
上线	针对已下线、引擎异常状态的引擎，您可以上线引擎（所有资源组节点或异常节点），

操作	说明
	上线过程中将启动资源组节点服务。
删除	针对已下线的引擎，且您不再使用当前集群，您可以删除此引擎，引擎删除后，将会自动删除此引擎的相关业务信息，包括内置数据源、内置元数据采集器、数据建模、数据质量规则等，请谨慎操作。
日志	针对上线、下线流程，您可以查看引擎相关日志。


4.10.3 流程中心

4.10.3.1 我的申请

用户可通过“我的申请”，查看流程单状态、流转情况和流程单详情。

操作说明

1. 进入页面。

点击左上方的，选择全部产品 > 系统管理 > 流程中心 > 我的申请，进入流程单列表。

2. 查询流程单。

选择状态或输入查询条件，再点击“查询”按钮进行查询。

流程中心

- [我的申请](#)
- [我的审批](#)
- [我的导入](#)
- [权限申请](#)

系统管理 / 流程中心 / 我的申请

我的申请列表
全部 进行中 执行成功或完成

请输入流程单号

请输入标题

请输入申请人

请选择申请日期

请选择申请日期

请输入导入单号

请选择申请模块

请选择流程类型

请选择当前状态

查询
重置

序号	流程单号	标题	导入单号	申请模块	流程类型	提交时间	申请人	当前处理	当前状态
1	MO-20240430105938-206	2024-04-30_新增_维康建模(dwd_test_0240430_1_c...	/	维康建模	新增	2024-04-30 10:59:39	linq4@c...	审批	● 终止
2	D0-20240430103356-184	2024-04-30_新增_派生指标(ywxid1_userid6_hour4)	IO-20240430103356-56	派生指标	新增	2024-04-30 10:33:57	linq4@c...	审批	● 驳回修改
3	D0-20240430103357-185	2024-04-30_新增_派生指标(ywxid2_userid7_hour5)	IO-20240430103356-56	派生指标	新增	2024-04-30 10:33:57	linq4@c...	审批	● 待审批
4	D0-20240430103151-182	2024-04-30_新增_派生指标(ywxid5_userid5_hour8)	IO-20240430103150-55	派生指标	新增	2024-04-30 10:31:52	linq4@c...	审批	● 待审批
5	D0-20240430103151-183	2024-04-30_新增_派生指标(ywxid6_userid6_hour9)	IO-20240430103150-55	派生指标	新增	2024-04-30 10:31:52	linq4@c...	审批	● 待审批
6	D0-20240430103150-178	2024-04-30_新增_派生指标(ywxid1_userid1_hour4)	IO-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq4@c...	审批	● 待审批
7	D0-20240430103150-179	2024-04-30_新增_派生指标(ywxid2_userid2_hour5)	IO-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq4@c...	审批	● 待审批
8	D0-20240430103151-180	2024-04-30_新增_派生指标(ywxid3_userid3_hour6)	IO-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq4@c...	审批	● 待审批
9	D0-20240430103151-181	2024-04-30_新增_派生指标(ywxid4_userid4_hour7)	IO-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq4@c...	审批	● 待审批
10	D0-20240430102247-173	2024-04-30_新增_派生指标(ywxid3_order_id_hour6)	IO-20240430102246-53	派生指标	新增	2024-04-30 10:22:48	linq4@c...	审批	● 待审批

共计14条记录 < 1 2 > 10条/页 跳至 1 页

3. 查看流程单详情。

点击列表的“流程单号”，进入流程单详情。

我的申请列表
D0-2024... x

📁 D0-20240430103357-185 : 2024-04-30_新增_派生指标(ywxid2_userid7_hour5)

基本信息

派生指标:

业务限定:

统计周期:

维度属性:

基本信息

挂载类型:

分析主题:

英文名称:

中文名称:

描述:

取消
终止
撤回申请

4. 操作流程单。

- 撤回流程单。

方式 1: 待审批的流程单，可点击 + 展开，点击“撤回流程单”，则撤回流程单。撤回后，可在对应的列表再进行操作。

我的申请列表 DO-2024... x

全部
 进行中
 执行成功或完结

请输入流程单号
 请输入标题
 请输入申请人
 请选择申请日期 至 请选择申请日期

请输入导入单号
 请选择申请模块
 请选择流程类型
 请选择当前状态

序号	流程单号	标题	导入单号	申请模块	流程类型	提交时间	申请人	当前处理	当前状态
1	M0-20240430105938-206	2024-04-30_新增_维度建模(dw...)	/	维度建模	新增	2024-04-30 10:59:39	linq4@c...	审批	● 终止
2	DO-20240430103356-184	2024-04-30_新增_派生指标(ywxd1_userId6_hour4)	I0-20240430103356-56	派生指标	新增	2024-04-30 10:33:57	linq4@c...	审批	● 撤回修改
3	DO-20240430103357-185	2024-04-30_新增_派生指标(ywxd2_userId7_hour5)	I0-20240430103356-56	派生指标	新增	2024-04-30 10:33:57	linq4@c...	审批	● 待审批

提交申请
 待审批
 撤回申请

提交人: linq4@...
 审批人: lin...

提交时间: 2024-04-30 10:33:57

4	DO-20240430103151-182	2024-04-30_新增_派生指标(ywxd5_userId5_hour8)	I0-20240430103150-55	派生指标	新增	2024-04-30 10:31:52	linq4@c...	审批	● 待审批
5	DO-20240430103151-183	2024-04-30_新增_派生指标(ywxd6_userId6_hour9)	I0-20240430103150-55	派生指标	新增	2024-04-30 10:31:52	linq4@c...	审批	● 待审批
6	DO-20240430103150-178	2024-04-30_新增_派生指标(ywxd1_userId1_hour4)	I0-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq4@c...	审批	● 待审批
7	DO-20240430103150-179	2024-04-30_新增_派生指标(ywxd2_userId2_hour5)	I0-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq4@c...	审批	● 待审批
8	DO-20240430103151-180	2024-04-30_新增_派生指标(ywxd3_userId3_hour6)	I0-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq4@c...	审批	● 待审批
9	DO-20240430103151-181	2024-04-30_新增_派生指标(ywxd4_userId4_hour7)	I0-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq4@c...	审批	● 待审批
10	DO-20240430102247-173	2024-04-30_新增_派生指标(ywxd3_order_id_hour6)	I0-20240430102246-53	派生指标	新增	2024-04-30 10:22:48	linq4@c...	审批	● 待审批

方式 2：点击“流程单号”，进入流程单详情，在下方的操作，点击“撤回申请”，则撤回流程单。撤回后，可在对应的列表再进行操作。

我的申请列表 DO-2024... x

DO-20240430103357-185 : 2024-04-30_新增_派生指标(ywxd2_userId7_hour5)

基本信息

原子指标:
 统计周期:

业务限定:
 维度属性:

基本信息

挂载类型:

分析主题:

英文名称:

中文名称:

描述:

● 终止流程单。

点击“流程单号”，进入流程单详情，在下方的操作，点击“终止”，则终止流程单。终止的流程单，只能查看，不能再次操作。

我的申请列表 D0-2024...

D0-20240430103357-185 : 2024-04-30 新增_派生指标(ywxd2_userid7_hour5)

基本信息

原子指标	userid7	统计周期	hour5
业务限定	ywxd2	维度属性	请选择

基本信息

挂载类型	分析主题
分析主题	订单分析主题
英文名称	ywxd2_userid7_hour5
中文名称	小时5_业务限定2_用户id7
描述	请填写派生指标的描述信息


取消 终止 撤回申请

4.10.3.2 我的审批

用户可通过“我的审批”，查看流程单详情，并进行审批操作。

操作说明

1. 进入页面。

点击左上方的，选择全部产品 > 系统管理 > 流程中心 > 我的审批，进入流程单列表。

2. 查询流程单。

选择状态或输入查询条件，再点击“查询”按钮进行查询。

我的审批列表

全部 待审批 已审批

请输入流程单号 请输入标题 请输入申请人 请选择申请日期 至 请选择申请日期

请输入导入单号 请选择申请模块 请选择流程类型 请选择当前状态

查询 重置

序号	流程单号	标题	导入单号	申请模块	流程类型	提交时间	申请人	当前处理	当前状态
1	M0-20240430105938-206	2024-04-30_新增_维度建模(dw...)	/	维度建模	新增	2024-04-30 10:59:39	linq4@c...	审批	终止
2	D0-20240430103357-185	2024-04-30_新增_派生指标(ywxd2_userid7_hour5)	10-20240430103356-56	派生指标	新增	2024-04-30 10:33:57	linq4@c...	审批	审批
3	D0-20240430103151-182	2024-04-30_新增_派生指标(ywxd5_userid5_hour6)	10-20240430103150-55	派生指标	新增	2024-04-30 10:31:52	linq4@c...	审批	审批
4	D0-20240430103151-183	2024-04-30_新增_派生指标(ywxd5_userid5_hour9)	10-20240430103150-55	派生指标	新增	2024-04-30 10:31:52	linq4@c...	审批	审批
5	D0-20240430103150-178	2024-04-30_新增_派生指标(ywxd1_userid1_hour4)	10-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq4@c...	审批	审批
6	D0-20240430103150-179	2024-04-30_新增_派生指标(ywxd2_userid2_hour5)	10-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq4@c...	审批	审批
7	D0-20240430103151-180	2024-04-30_新增_派生指标(ywxd3_userid3_hour6)	10-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq4@c...	审批	审批
8	D0-20240430103151-181	2024-04-30_新增_派生指标(ywxd4_userid4_hour7)	10-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq4@c...	审批	审批
9	D0-20240430102247-173	2024-04-30_新增_派生指标(ywxd3_order_id_hour6)	10-20240430102246-53	派生指标	新增	2024-04-30 10:22:48	linq4@c...	审批	审批
10	D0-20240430102248-174	2024-04-30_新增_派生指标(ywxd4_order_id_hour7)	10-20240430102246-53	派生指标	新增	2024-04-30 10:22:48	linq4@c...	审批	审批

共计13条记录 < 1 2 > 10条/页 跳至 1 页

3. 查看流程单详情。

点击列表的“流程单号”，进入流程单详情。

我的审批列表 D0-2024...

D0-20240430103357-185 : 2024-04-30_新增_派生指标(ywxd2_userid7_hour5)

基本信息

派生指标: userid7 统计周期: hour5

业务限定: ywxd2 维度属性: 请选择

基本信息

挂载类型: 分析主题

分析主题: 订单分析主题


英文名称: ywxd2_userid7_hour5

中文名称: 小时5_业务限定2_用户id7

描述: 请填写派生指标的描述信息

取消 通过 拒绝 驳回修改

4. 操作流程单。

待审批单，点击列表的“流程单号”；或点击列表右侧的“审批”按钮；或者点击  展开，点击“去审批”按钮，均可进入流程单详情。

我的审批列表 D0-2024... x

全部
 待审批
 已审批

请输入流程单号
 请输入标题
 请输入申请人
 请选择申请日期 至 请选择申请日期

请输入导入单号
 请选择申请模块
 请选择流程类型
 请选择当前状态

序号	流程单号	标题	导入单号	申请模块	流程类型	提交时间	申请人	当前处理	当前状态
1	M0-20240430105938-206	2024-04-30_新增_维度建模(dwtd_test_0240430_1_c...	/	维度建模	新增	2024-04-30 10:59:39	linq14@c...	审批	● 终止
2	D0-20240430103357-185	2024-04-30_新增_派生指标(ywx2_userid7_hour5)	I0-20240430103356-56	派生指标	新增	2024-04-30 10:33:57	linq14@c...	审批	<input type="button" value="审批"/>
<div style="border: 1px solid #ccc; padding: 5px;"> <p> <input checked="" type="radio"/> 提交申请 <input checked="" type="radio"/> 待审批 <input type="button" value="去审批"/> </p> <p> 提交人: linq... 审批人: linq... </p> <p>提交时间: 2024-04-30 10:33:57</p> </div>									
3	D0-20240430103151-182	2024-04-30_新增_派生指标(ywx5_userid5_hour8)	I0-20240430103150-55	派生指标	新增	2024-04-30 10:31:52	linq14@c...	审批	<input type="button" value="审批"/>
4	D0-20240430103151-183	2024-04-30_新增_派生指标(ywx6_userid6_hour9)	I0-20240430103150-55	派生指标	新增	2024-04-30 10:31:52	linq14@c...	审批	<input type="button" value="审批"/>
5	D0-20240430103150-178	2024-04-30_新增_派生指标(ywx1_userid1_hour4)	I0-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq14@c...	审批	<input type="button" value="审批"/>
6	D0-20240430103150-179	2024-04-30_新增_派生指标(ywx2_userid2_hour5)	I0-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq14@c...	审批	<input type="button" value="审批"/>
7	D0-20240430103151-180	2024-04-30_新增_派生指标(ywx3_userid3_hour6)	I0-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq14@c...	审批	<input type="button" value="审批"/>
8	D0-20240430103151-181	2024-04-30_新增_派生指标(ywx4_userid4_hour7)	I0-20240430103150-55	派生指标	新增	2024-04-30 10:31:51	linq14@c...	审批	<input type="button" value="审批"/>
9	D0-20240430102247-173	2024-04-30_新增_派生指标(ywx3_order_id_hour6)	I0-20240430102246-53	派生指标	新增	2024-04-30 10:22:48	linq14@c...	审批	<input type="button" value="审批"/>

针对流程单，可进行如下操作：

我的审批列表 D0-2024... x

D0-20240430103357-185 : 2024-04-30_新增_派生指标(ywx2_userid7_hour5)

基本信息

原子指标
 统计周期

业务限定
 维度属性

基本信息

挂载类型

分析主题

英文名称

中文名称

描述

- 通过：通过该单，进入下一流程。
- 拒绝：拒绝该单，此单流转结束。
- 驳回修改：驳回修改，申请人可进行修改后，再次提交。


- 取消：则不操作此单。

4.10.3.3 我的导入

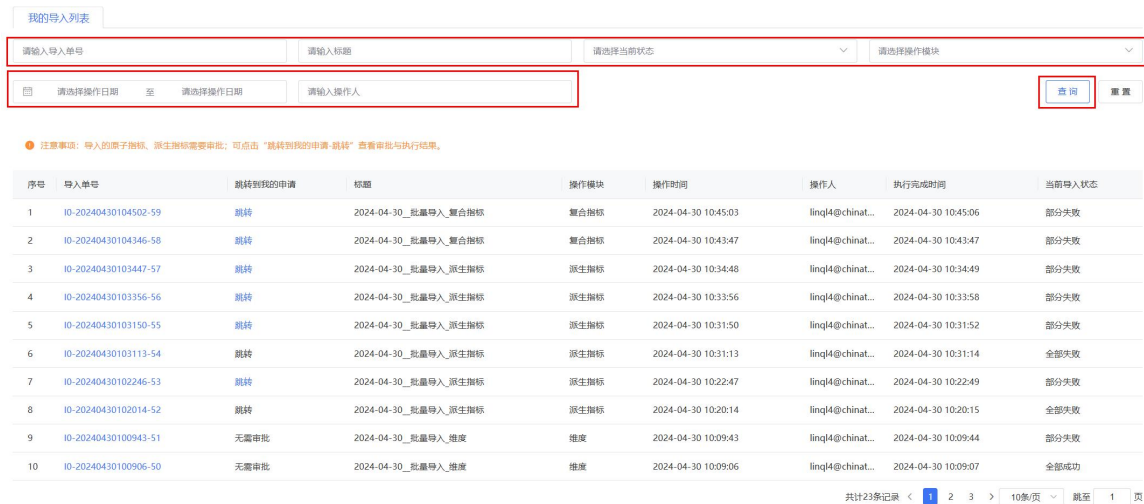
在“我的导入”查看导入状态、导入详情；失败的导入单，可查看导入失败原因。

操作说明

1. 进入页面。

点击左上方的，选择全部产品 > 系统管理 > 流程中心 > 我的导入，进入导入单列表。

2. 查询导入单。



我的导入列表

请输入导入单号 请输入标题 请选择当前状态 请选择操作模块

请选择操作日期 至 请选择操作日期 请输入操作人 查询 重置

● 注意事项：导入的原子指标、派生指标需要审批，可点击“跳转到我的申请-跳转”查看审批与执行结果。

序号	导入单号	跳转到我的申请	标题	操作模块	操作时间	操作人	执行完成时间	当前导入状态
1	10-20240430104502-59	跳转	2024-04-30_批量导入_复合指标	复合指标	2024-04-30 10:45:03	linq14@china...	2024-04-30 10:45:06	部分失败
2	10-20240430104346-58	跳转	2024-04-30_批量导入_复合指标	复合指标	2024-04-30 10:43:47	linq14@china...	2024-04-30 10:43:47	部分失败
3	10-20240430103447-57	跳转	2024-04-30_批量导入_派生指标	派生指标	2024-04-30 10:34:48	linq14@china...	2024-04-30 10:34:49	部分失败
4	10-20240430103356-56	跳转	2024-04-30_批量导入_派生指标	派生指标	2024-04-30 10:33:56	linq14@china...	2024-04-30 10:33:58	部分失败
5	10-20240430103150-55	跳转	2024-04-30_批量导入_派生指标	派生指标	2024-04-30 10:31:50	linq14@china...	2024-04-30 10:31:52	部分失败
6	10-20240430103113-54	跳转	2024-04-30_批量导入_派生指标	派生指标	2024-04-30 10:31:13	linq14@china...	2024-04-30 10:31:14	全部失败
7	10-20240430102246-53	跳转	2024-04-30_批量导入_派生指标	派生指标	2024-04-30 10:22:47	linq14@china...	2024-04-30 10:22:49	部分失败
8	10-20240430102014-52	跳转	2024-04-30_批量导入_派生指标	派生指标	2024-04-30 10:20:14	linq14@china...	2024-04-30 10:20:15	全部失败
9	10-20240430100943-51	无需审批	2024-04-30_批量导入_维度	维度	2024-04-30 10:09:43	linq14@china...	2024-04-30 10:09:44	部分失败
10	10-20240430100906-50	无需审批	2024-04-30_批量导入_维度	维度	2024-04-30 10:09:06	linq14@china...	2024-04-30 10:09:07	全部成功

共计23条记录 < 1 2 3 > 10条/页 跳至 1 页

3. 查看导入单详情。

点击列表的“导入单号”，进入导入单详情。可查看导入单详情。

我的导入列表 I0-2024... x

ID-20240430104502-59-2024-04-30_批量导入_复合指标

导入结果: 总计10条, 导入成功9条, 导入失败1条; 导入失败的记录, 请查看原因, 重新调整后, 再导入。

[跳转到我的申请](#) [刷新查询](#)

*英文名称	*中文名称	挂载类型	挂载类型名	描述	依赖指标	表达式	导入结果	导入失败原因	审批单号
fuhzb010	物流相关指标10	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		CO-20240430104505-200
fuhzb009	物流相关指标9	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		CO-20240430104505-199
fuhzb008	物流相关指标8	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		CO-20240430104505-198
fuhzb007	物流相关指标7	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		CO-20240430104504-197
fuhzb006	物流相关指标6	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		CO-20240430104504-196
fuhzb005	物流相关指标5	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		CO-20240430104504-195
fuhzb004	物流相关指标4	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		CO-20240430104503-194
fuhzb003	物流相关指标3	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		CO-20240430104503-193
fuhzb002	物流相关指标2	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		CO-20240430104503-192
fuhzb001	物流相关指标1	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	失败	跳过, 已存在同名对象	

共计10条记录 10条/页 < 1 > 跳至 1 页

4. 跳转到“我的申请”流程单列表。

方式 1: 导入全部成功或者部分失败的导入单, 且存在需审批的记录, 在列表“跳转到我的申请”点击“跳转”文字, 则可跳转到“我的申请”的流程单列表。一个导入单可有多多个审批单。

我的导入列表 I0-2024... x

请输入导入单号 请输入标题 请选择当前状态 请选择操作模块

请选择操作日期 至 请选择操作日期 请输入操作人 [查询](#) [重置](#)

● 注意事项: 导入的原子指标、派生指标需要审批; 可点击“跳转到我的申请-跳转”查看审批与执行结果。

序号	导入单号	跳转到我的申请	标题	操作模块	操作时间	操作人	执行完成时间	当前导入状态
1	ID-20240430104502-59	跳转	2024-04-30_批量导入_复合指标	复合指标	2024-04-30 10:45:03	linq14@china...	2024-04-30 10:45:06	部分失败
2	ID-20240430104346-58	跳转	2024-04-30_批量导入_复合指标	复合指标	2024-04-30 10:43:47	linq14@china...	2024-04-30 10:43:47	部分失败
3	ID-20240430103447-57	跳转	2024-04-30_批量导入_派生指标	派生指标	2024-04-30 10:34:48	linq14@china...	2024-04-30 10:34:49	部分失败
4	ID-20240430103356-56	跳转	2024-04-30_批量导入_派生指标	派生指标	2024-04-30 10:33:56	linq14@china...	2024-04-30 10:33:58	部分失败
5	ID-20240430103150-55	跳转	2024-04-30_批量导入_派生指标	派生指标	2024-04-30 10:31:50	linq14@china...	2024-04-30 10:31:52	部分失败
6	ID-20240430103113-54	跳转	2024-04-30_批量导入_派生指标	派生指标	2024-04-30 10:31:13	linq14@china...	2024-04-30 10:31:14	全部失败
7	ID-20240430102246-53	跳转	2024-04-30_批量导入_派生指标	派生指标	2024-04-30 10:22:47	linq14@china...	2024-04-30 10:22:49	部分失败
8	ID-20240430102014-52	跳转	2024-04-30_批量导入_派生指标	派生指标	2024-04-30 10:20:14	linq14@china...	2024-04-30 10:20:15	全部失败
9	ID-20240430100943-51	无需审批	2024-04-30_批量导入_维度	维度	2024-04-30 10:09:43	linq14@china...	2024-04-30 10:09:44	部分失败
10	ID-20240430100906-50	无需审批	2024-04-30_批量导入_维度	维度	2024-04-30 10:09:06	linq14@china...	2024-04-30 10:09:07	全部成功

共计23条记录 < 1 2 3 > 10条/页 跳至 1 页

方式 2: 导入全部成功或者部分失败的导入单, 且存在需审批的记录, 在导入单详情点击“跳转到我的申请”, 则可跳转到“我的申请”的流程单列表。一个导入单可有多多个审批单。

我的导入列表 10-2024...

10-20240430104502-59-2024-04-30_批量导入_复合指标

导入结果: 总计10条, 导入成功9条, 导入失败1条; 导入失败的记录, 请查看原因, 重新调整后, 再导入。

[跳转至我的申请](#) [刷新查询](#)

*英文名称	*中文名称	挂载类型	挂载类型名	描述	依赖指标	表达式	导入结果	导入失败原因	审批单号
fuhzb010	物流相关指标10	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104505-200
fuhzb009	物流相关指标9	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104505-199
fuhzb008	物流相关指标8	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104505-198
fuhzb007	物流相关指标7	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104504-197
fuhzb006	物流相关指标6	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104504-196
fuhzb005	物流相关指标5	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104504-195
fuhzb004	物流相关指标4	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104503-194
fuhzb003	物流相关指标3	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104503-193
fuhzb002	物流相关指标2	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104503-192
fuhzb001	物流相关指标1	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	失败	跳过, 已存在同名对象	

共计10条记录 10条/页 < 1 > 跳至 1 页

5. 跳转到“我的申请”流程单详情。

导入成功且需审批的记录, 在流程单详情, 点击“审批单号”, 则可跳转到“我的申请”流程单详情。

我的导入列表 10-2024...

10-20240430104502-59-2024-04-30_批量导入_复合指标

导入结果: 总计10条, 导入成功9条, 导入失败1条; 导入失败的记录, 请查看原因, 重新调整后, 再导入。

[跳转至我的申请](#) [刷新查询](#)


*英文名称	*中文名称	挂载类型	挂载类型名	描述	依赖指标	表达式	导入结果	导入失败原因	审批单号
fuhzb010	物流相关指标10	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104505-200
fuhzb009	物流相关指标9	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104505-199
fuhzb008	物流相关指标8	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104505-198
fuhzb007	物流相关指标7	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104504-197
fuhzb006	物流相关指标6	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104504-196
fuhzb005	物流相关指标5	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104504-195
fuhzb004	物流相关指标4	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104503-194
fuhzb003	物流相关指标3	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104503-193
fuhzb002	物流相关指标2	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	成功		C0-20240430104503-192
fuhzb001	物流相关指标1	分析主题	订单分析主题		小时_订单id,小时4_AP...	小时_订单id+小时4_A...	失败	跳过, 已存在同名对象	

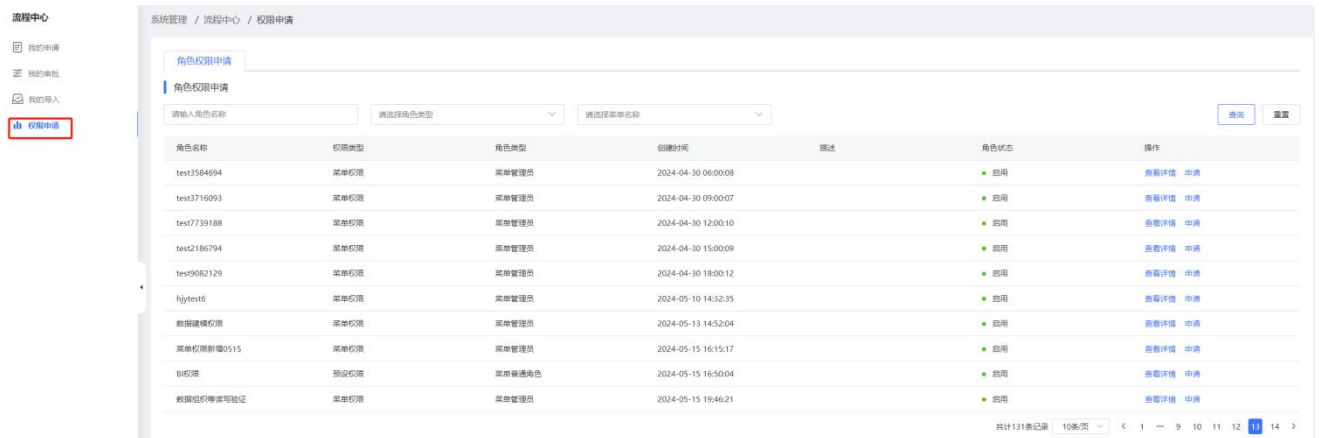
共计10条记录 10条/页 < 1 > 跳至 1 页

4.10.3.4 权限申请

用户可通过权限申请功能, 申请所需菜单权限, 以便进行数据作业。

进入页面

方法 1: 单击左上方图标, 选择全部产品 > 系统管理 > 流程中心 > 权限申请, 进入权限申请页面。



方法 2: 在 DataWings 首页中, 点击 “权限申请” 按钮, 进入申请页面。



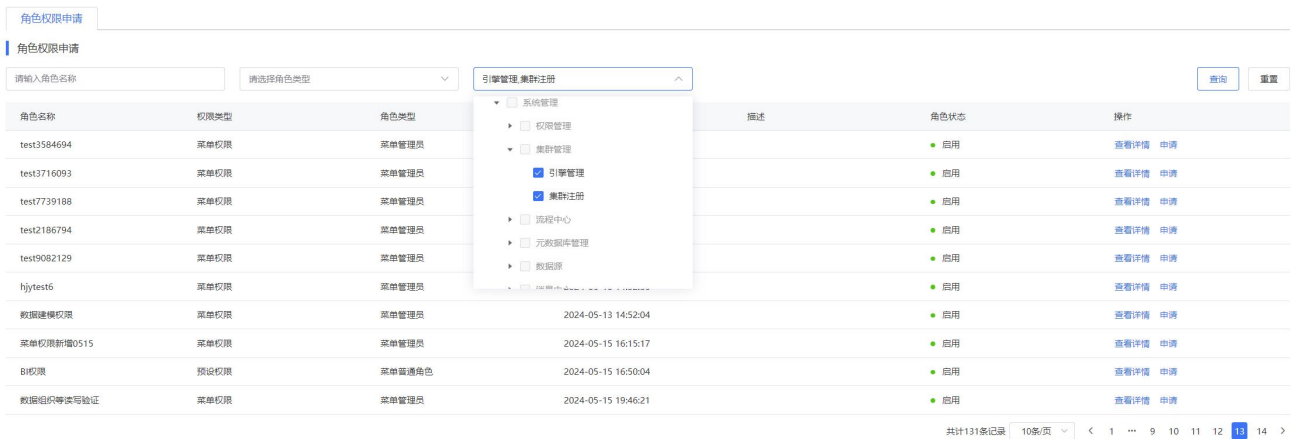
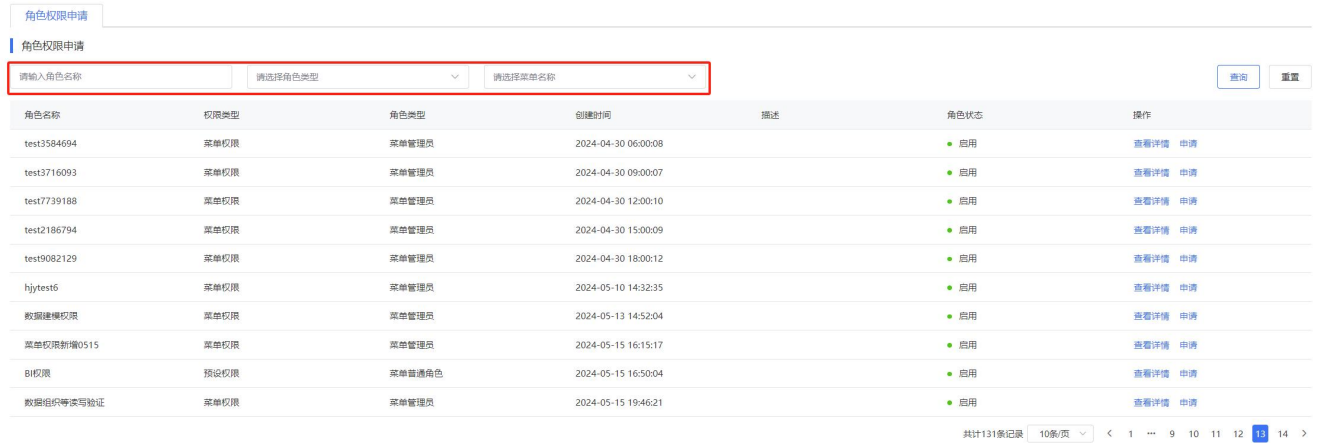
方法 3: 用户进入未被分配权限的菜单页面时, 会提示无当前页面访问权限, 可点击 “权限申请” 按钮, 前往申请页面。

操作说明

角色权限申请

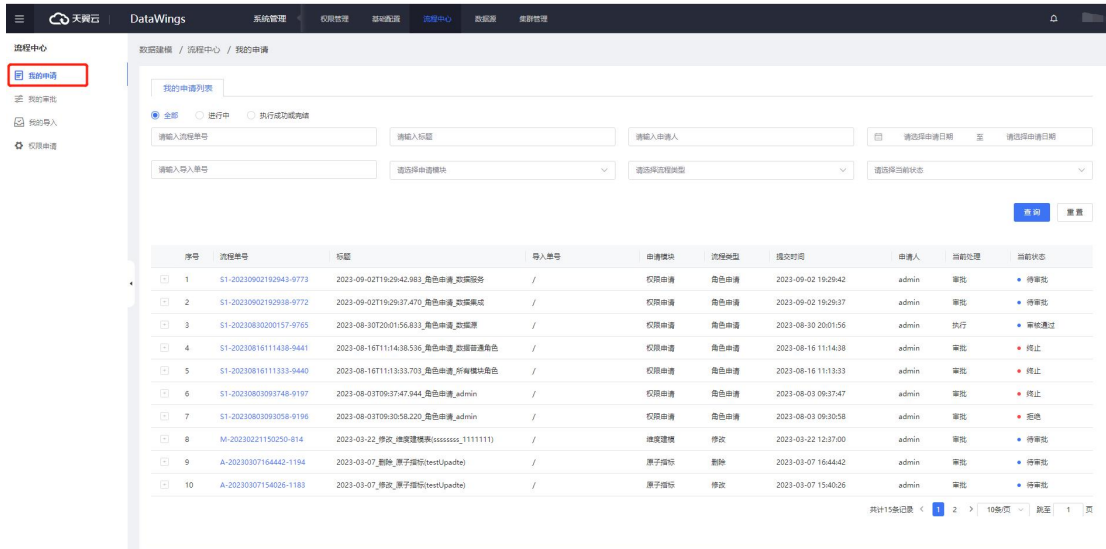
用户可通过 “角色权限申请”, 申请所需菜单对应的角色权限。

- 目标角色检索：用户可通过角色名称、角色类型以及菜单名称检索所需角色，其中，菜单名称检索支持多选，点击“查询”按钮，将返回包含所选全部菜单的目标角色。用户可以通过点击操作栏中的“查看详情”按钮，在详情页中确认该角色权限是否符合需求。



- 申请：点击目标角色右侧操作栏中的“申请”按钮，点击后出现“提交成功”提示，用户可基于申请单号，在“我的申请”功能中查询审批进度与结果。




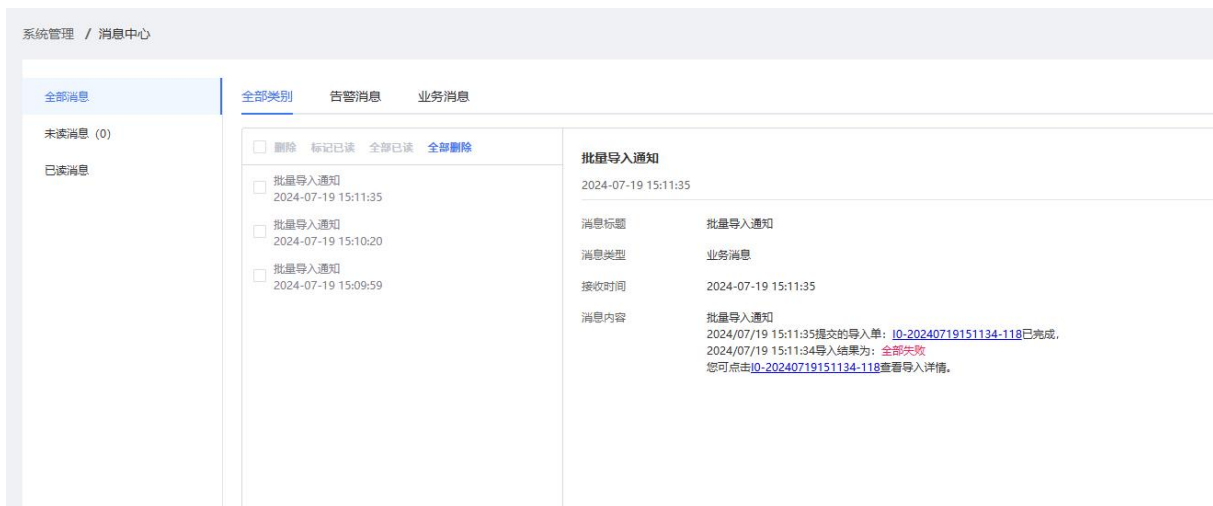


4.10.4 消息中心

用户可在消息中心查看各模块的告警消息和业务消息。

进入页面

点击左上方的 ，选择全部产品> 系统管理 > 消息中心进入消息中心列表。



查看未读消息

点击“未读消息”，用户可查看未读消息列表。点击未读消息标题，即可查看未读消息详情，查看后该消息状态将置为已读，后续用户可在全部消息列表或者已读消息列表查看。

查看已读消息

点击“已读消息”，用户可查看已读消息列表。

查看告警消息

点击“告警消息”，用户可查看不同模块发送给当前用户的告警消息，例如数据开发任务失败告警、数据质量稽核检查告警等。

前提条件

已在业务模块配置告警策略。

告警策略中选择站内信为告警通道。

告警策略中选择当前用户为告警对象。

查看业务消息

点击“业务消息”，用户可查看不同模块的业务消息。

5 最佳实践

5.1 获取翼 MapReduce 数据湖集群信息

本文介绍如何获取翼 MapReduce 数据湖集群的信息。

前提条件

- 已购买翼 MapReduce（以下简称翼 MR）数据湖集群。

操作步骤

1. 进入[翼 MapReduce 控制台](#)。
2. 点击已购买的数据湖集群。

3. 点击“翼 MR Manager”页签，再点击“前往翼 MR Manager”按钮，进入集群管理。



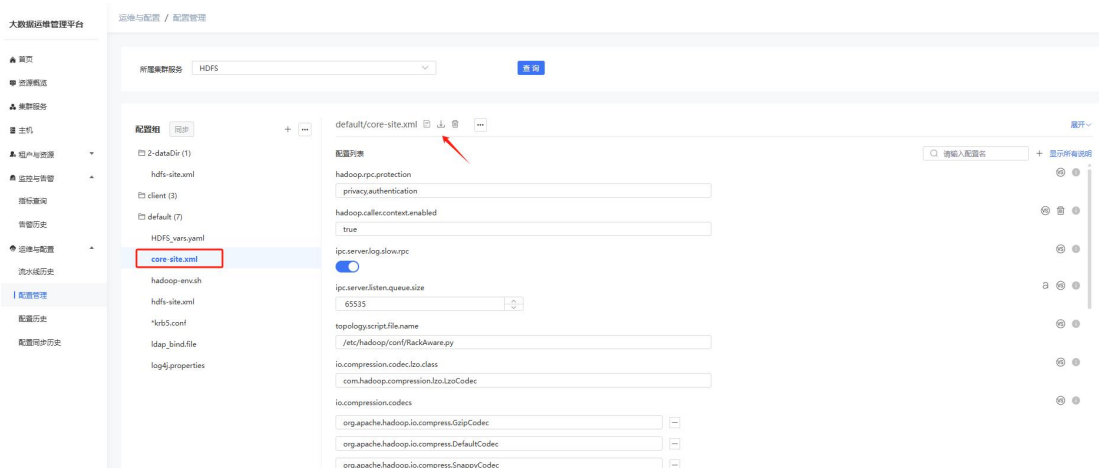
翼 MR 集群信息获取

获取翼 MR 集群配置文件

获取 core-site.xml

操作步骤

- 1) 点击运维与配置-配置管理。
- 2) 所属集群服务选择 HDFS。
- 3) 下载 default 文件夹下 core-site.xml 文件。

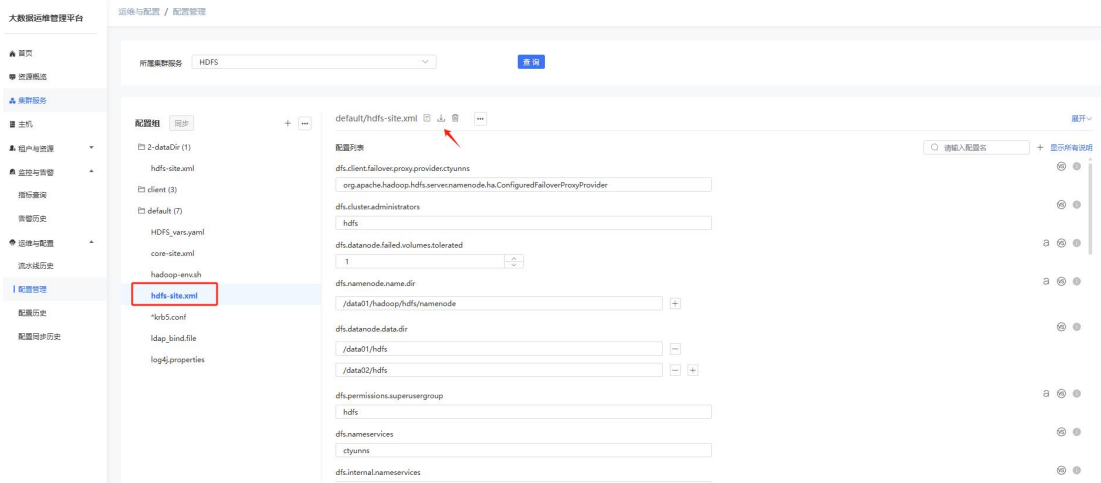


获取 hdfs-site.xml

- 1) 点击运维与配置-配置管理。

2) 所属集群服务选择 HDFS。

3) 下载 default 文件夹下 hdfs-site.xml 文件。

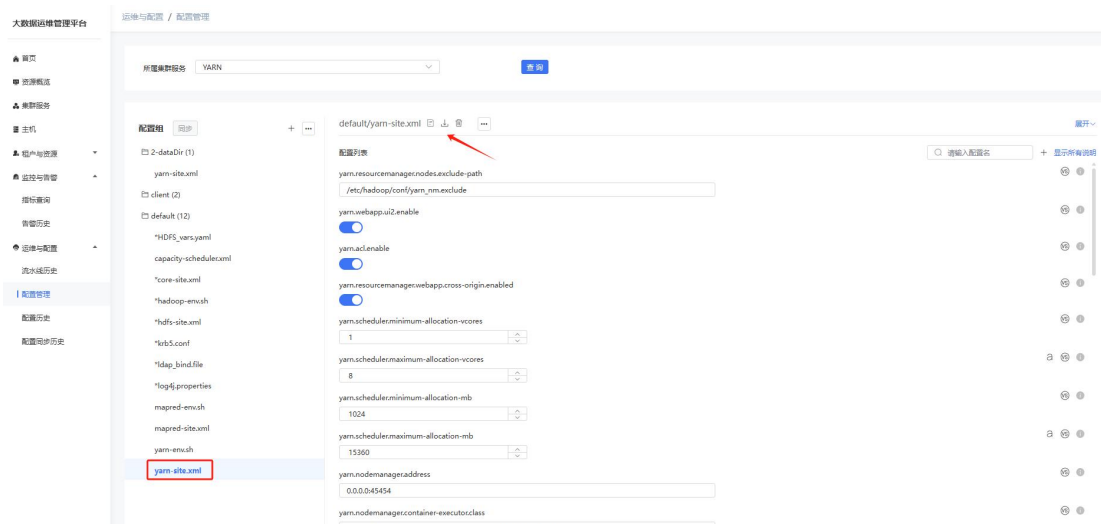


获取 yarn-site.xml

1) 点击运维与配置-配置管理。

2) 所属集群服务选择 YARN。

3) 下载 default 文件夹下 yarn-site.xml 文件。

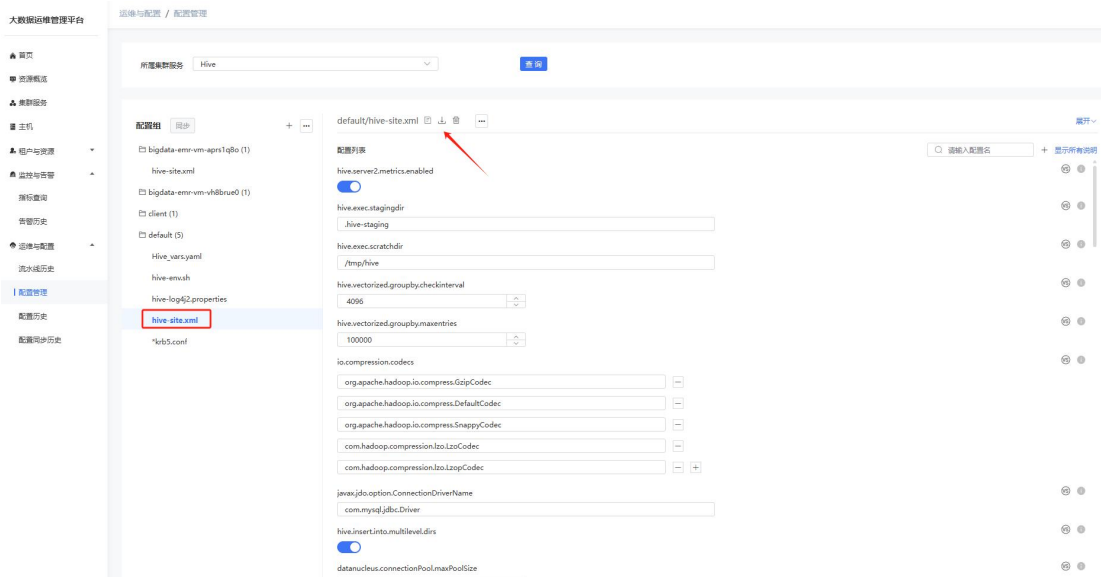


获取 hive-site.xml

1) 点击运维与配置-配置管理。

2) 所属集群服务选择 Hive。

3) 下载 default 文件夹下 hive-site.xml 文件。

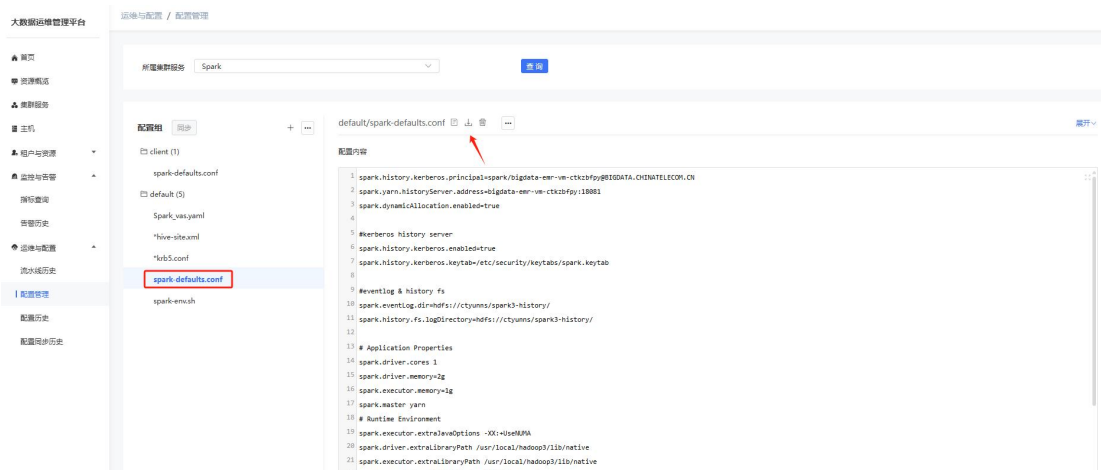


获取 spark-defaults.conf

1) 点击运维与配置-配置管理。

2) 所属集群服务选择 Spark。

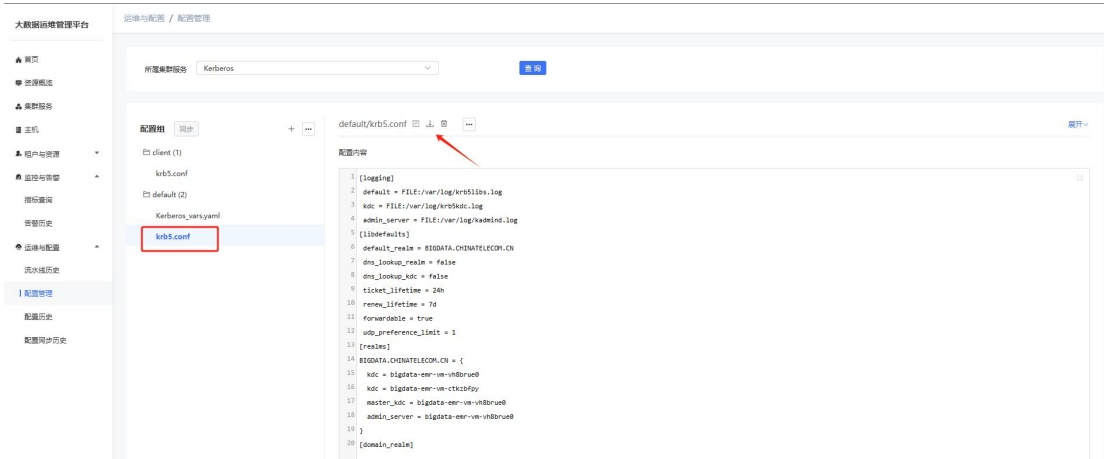
3) 下载 default 文件夹下 spark-defaults.conf 文件。



获取 krb5.conf

1) 点击运维与配置-配置管理。

- 2) 所属集群服务选择 Kerberos。
- 3) 下载 default 文件夹下 krb5.conf 文件。



获取 hive.keytab

为保证数据安全，翼 MapReduce 数据湖集群已强制启用 Kerberos 认证，DataWings 下发大数据计算任务时，将通过代理用户的方式提交，建议您手动创建一个 hive 用户的 keytab。

- 1) 点击租户与资源-Kerberos 安全凭证。
- 2) 点击新建按钮，输入 hive 并点击确定，即可创建 hive 用户的 Principal 和 keytab。



- 3) 点击下载，即可下载 keytab 文件。

用户: 主机名: 重置 查询

操作已选项 删除记录 分发记录 重置

Principal	主机名	创建者	创建时间	操作
had[redacted]@TELECOM.CN		datawingsdev_ops@chinatelecom.cn	2024-09-26 01:17	分发 下载
emr_z[redacted]@TELECOM.CN	bigdata-emr-vm-yooxvrbz	datawingsdev_ops@chinatelecom.cn	2024-07-26 02:01	分发 下载
emr_z[redacted]@TELECOM.CN	bigdata-emr-vm-yooxvrbz	datawingsdev_ops@chinatelecom.cn	2024-06-12 16:47	分发 下载
hive@BIGDATA.CHINATELECOM.CN		datawingsdev_ops@chinatelecom.cn	2024-05-15 12:40	分发 下载

共 4 条 10条/页 1 / 1 页

获取翼 MR 集群配置属性

获取 Kerberos 账号信息

Kerberos 账号可填写 hive@BIGDATA.CHINATELECOM.CN, keytab 使用上节下载的 hive.keytab 即可。

注意

Kerberos 账号需要与 keytab 一一对应，否则后续认证会不通过。

获取翼 MR 集群网络配置信息

进入 Manager 管理平台，点击进入主机页面，即可看到所有的主机信息。

大数据运维管理平台 主机

IP/主机名: 重置 查询

操作已选项 当前页刷新

主机IP	主机名称	初始化状态	主机状态	健康状态	内存(GB)	CPU型号	主机配置	操作系统	描述
192.168.0.20	bigdata-emr-vm-vh88ruu0	成功	使用中	正常	32			ctyunos2.0.1	
192.168.0.3	bigdata-emr-vm-cubatzq6	成功	使用中	正常	32			ctyunos2.0.1	
192.168.0.6	bigdata-emr-vm-yj3hrvc3	成功	使用中	正常	32			ctyunos2.0.1	
192.168.0.7	bigdata-emr-vm-yooxvrbz	成功	使用中	正常	32			ctyunos2.0.1	
192.168.0.8	bigdata-emr-vm-aprs1q8o	成功	使用中	正常	32			ctyunos2.0.1	

100条/页 1 / 1 页 共 6 条记录

手动创建集群时，您需要复制所有主机的主机 IP-主机名称对，配置在集群注册表单中，以确保资源组能成功访问大数据集群，例如：

资源组网络配置

```
hosts
192.168.0.7 bigdata-emr-vm-yxxxxrbz
192.168.0.6 bigdata-emr-vm-wjbrhlz3
192.168.0.9 bigdata-emr-vm-ctkzbfpy
192.168.0.5 bigdata-emr-vm-cubztql6
192.168.0.20 bigdata-emr-vm-vh8brue0
192.168.0.8 bigdata-emr-vm-aprs1q8o
```

① 请按照格式填写主机配置：'IP地址 主机名'，IP地址只能包含数字和 '.'，主机名只能包含字母、数字、 '-' 或 '_'，每行一个配置，用换行符分隔。

5.2 电商行业数据建模实践

本文介绍如何针对电商行业进行数据建模规划设计。

为什么需要数据建模？

<p>如果没有「数据标准」</p> <p>各个数据要素、数据指标、表名等命名五花八门，数据格式各不相同，难以进行共享与流通</p>	<table border="1"> <tr><td>ord</td><td>order</td></tr> <tr><td>trd_ods...</td><td>ods_trd...</td></tr> <tr><td>dws 汇总层</td><td>dws 接入层?</td></tr> </table>	ord	order	trd_ods...	ods_trd...	dws 汇总层	dws 接入层?	<p>如果没有「数据指标」</p> <ul style="list-style-type: none"> 同名不同义 同义不同名 类似指标众多 人工对齐口径费时费力 	<table border="1"> <tr><td>size_id 尺码id</td><td>size_id 商品id</td></tr> <tr><td>goods_amount, goods_amt</td><td>销售额</td></tr> <tr><td>扣满减销售额、扣活动销售额...</td><td></td></tr> </table>	size_id 尺码id	size_id 商品id	goods_amount, goods_amt	销售额	扣满减销售额、扣活动销售额...	
ord	order														
trd_ods...	ods_trd...														
dws 汇总层	dws 接入层?														
size_id 尺码id	size_id 商品id														
goods_amount, goods_amt	销售额														
扣满减销售额、扣活动销售额...															
<p>如果没有「数仓规划」</p> <p>企业数据呈几何倍增长，数据量庞大、复杂，如果没有数仓规划，则数据指标、表模型等往往会出现杂乱无章，数据难以管理和查找</p>	<table border="1"> <tr><td>订单商品明细表</td><td>支付记录表</td></tr> <tr><td>商品信息表</td><td>品牌粒度销售表</td></tr> <tr><td colspan="2">品牌粒度KPI月表</td></tr> <tr><td colspan="2">品牌粒度KPI年表</td></tr> </table>	订单商品明细表	支付记录表	商品信息表	品牌粒度销售表	品牌粒度KPI月表		品牌粒度KPI年表		<p>如果没有「维度建模」</p> <ul style="list-style-type: none"> 通过命令行进行建模，门槛高 缺少必要的业务元数据 缺少数据标准关联 缺少维度、指标的定义与关联 	<table border="1"> <tr><td>create table ...</td></tr> <tr><td>表描述、使用场景、哪个分层、哪个数据域?</td></tr> <tr><td>usr_id标准是user_id, 怎么知道?</td></tr> <tr><td>维度是关联哪个维表? 字段关联哪个指标?</td></tr> </table>	create table ...	表描述、使用场景、哪个分层、哪个数据域?	usr_id标准是user_id, 怎么知道?	维度是关联哪个维表? 字段关联哪个指标?
订单商品明细表	支付记录表														
商品信息表	品牌粒度销售表														
品牌粒度KPI月表															
品牌粒度KPI年表															
create table ...															
表描述、使用场景、哪个分层、哪个数据域?															
usr_id标准是user_id, 怎么知道?															
维度是关联哪个维表? 字段关联哪个指标?															

数据建模解决什么问题？

提供数据建模方法论，自顶向下进行规范构建数仓，包含数据标准、数仓规划、数据指标及维度建模四大模块，帮助政府及企事业单位在搭建数据中台、数据集市建设过程中提升建模能力，并通过数据建模快速构建政府及企事业单位的数据资产。



数据标准

制定数据标准，规范定义，为数仓规划的各个要素、数据指标、表名等提供标准和约束，从源头保障标准化生产。主要包含命名词典、字段标准、标准值域、表名规范、指标规范等。

命名词典

命名词典为您提供管理各个数据要素，如业务名词、物理表、字段、原子指标的词根和词缀及其标准化翻译功能。在各个数据要素命名时，可以参考命名词典进行命名。命名词典模块支持新建、修改、删除命名词典，同时支持批量导入导出命名词典。

字段标准

字段标准是对含义相同但字段名称不同的数据进行统一规范管理的数据准则。可将多个表中含义相同但字段名不同的内容进行关联，并对该字段制定相关的取值范围、度量单位、标准值域等内容。

标准值域

标准值域是字段标准的取值范围，在标准值域中可设置字段标准可选择的数据内容及范围。标准值域与枚举维度相似，但数据治理的作用要强于枚举维度，标准值域可以与数据质量打通，某个字段被关联标准值域后，该字段的取值范围必须在标准值域内，如取值范围不在其内，需要马上治理。

表名规范与指标规范

系统提供规范管理功能，用以规范表模型及指标的命名，制定表模型及指标的命名规范，在模型和指标创建时，系统会根据规范推荐表名和自动生成指标名，使用方能通过表名和指标名快速了解对应的业务信息。不符合表名规范的表，则在“待治理清单”中展示，形成治理闭环。

数仓规划

背景知识

业务分类

所在的集团或公司业务复杂且较为庞大时，可创建业务分类，对数据进行业务区分，便于对业务数据进行更加方便的管理。

数据域

数据域是联系较为紧密的数据主题的集合，通常是根​​据业务类别、数据来源、数据用途等多个维度，对企业的业务数据进行的区域划分，将同类型数据存放在一起。

业务过程

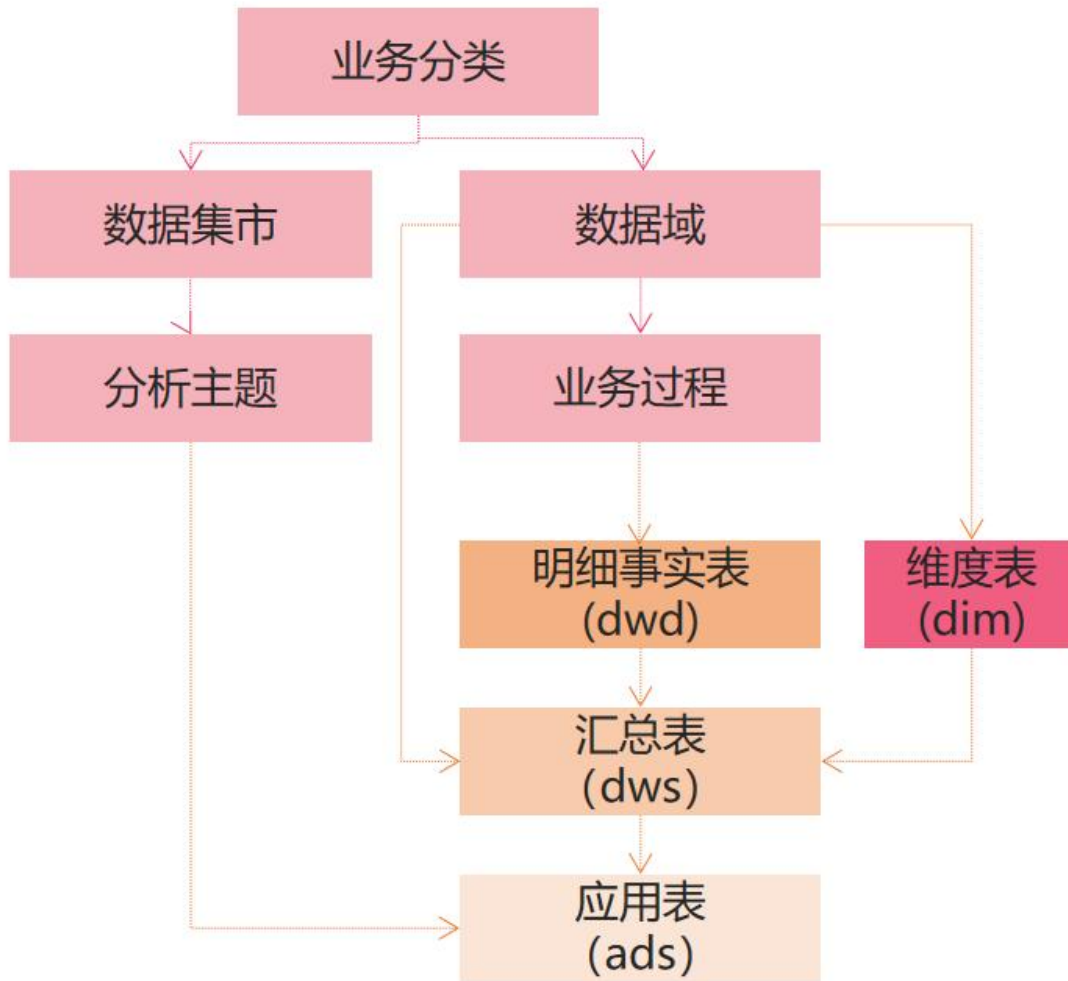
业务过程是企业​​在指定的数据域中所执行的业务活动，是数据建模所需要分析的逻辑主体。例如，交易域中可以有加入购物车、下单、支付等业务过程。

数据集市

数据集市是基于业务分类，面向特定应用场景或者产品的数据组织。通常位于数据应用层，依赖于整合数据。

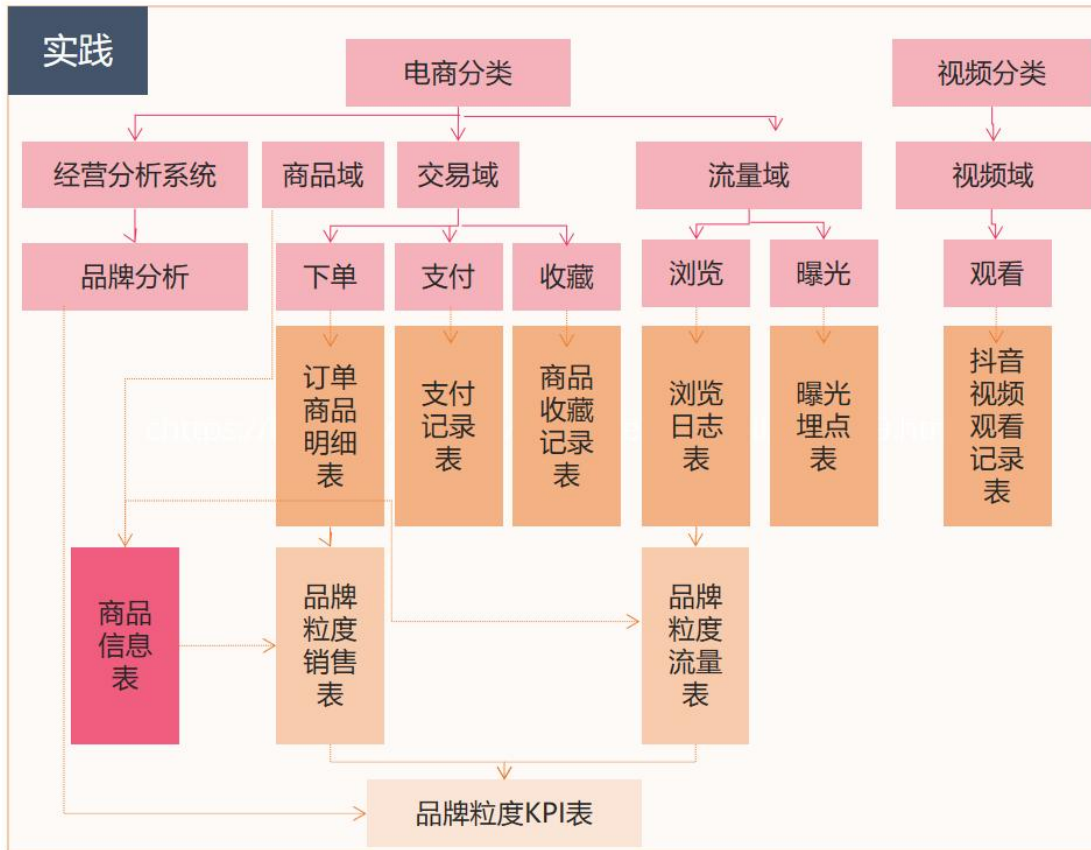
分析主题

分析主题用于将数据集市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。如盘货分析、关键指标分析等。



数仓设计实践

根据规划的数据要素，进行分层划域和表模型设计。



数据指标

原子指标

用于明确业务的统计口径和计算逻辑，是基于用户的业务活动（即业务过程）创建的，用于统计业务活动中某一业务状况的数值。例如，用户的业务活动为购买，则原子指标就可以指定为支付金额。

业务限定

业务限定用于限定统计数据业务范围。

统计周期

统计周期即统计数据的时间范围，例如近 7 天、近 30 天等。用于在统计派生指标时，限定业务统计的时间范围。

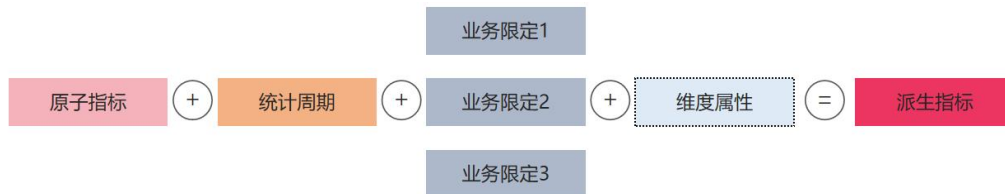
维度及维度属性

维度是人们观察事物的角度，进行统计的对象。例如客户、货品、场景等具体的业务对象；地域、组织关系和产品类目等抽象的业务对象。

维度属性：是用于具象化维度，比如货品的 ID、货品的名称。

派生指标

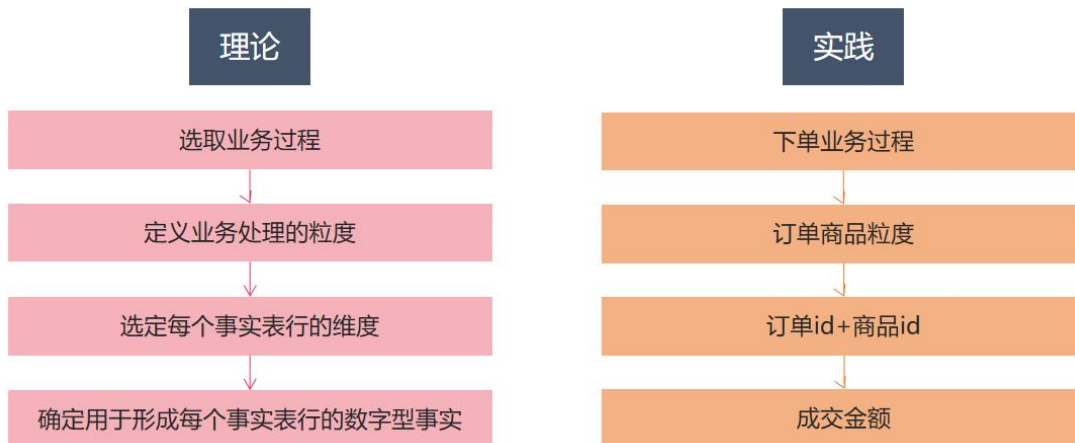
派生指标是由原子指标、业务限定、统计周期、统计粒度（维度属性）组成，用于反映某个粒度某一业务活动在指定时间周期及目标范围中的业务状况。例如，品牌 ID 粒度一周华南地域的销售金额。



统一梳理，统一命名，统一定义和建设

维度建模

维度建模四步曲



1. 选定业务过程为下单。

* 数据源: poc_engine_HIVE

* 数据库名: ctydwd

* 表物理类型: Hive_Table

* 表属性: 内部表 外部表

* 数仓分层: 明细数据层(dwd)

* 表类型: 事务事实表

* 挂载类型: 业务过程

* 业务过程: 1 下单

* 表名主体: order_detail_lq1

表名规范: dwd_数据域缩写_业务过程缩写_表名主体_存储策略 (弱规则)

存储策略: 日增量(d)

生命周期: 请输入大于或等于1的正整数 天

描述: 0/2048

* 表英文名: dwd_trd_order_order_detail_lq_di [使用推荐表名](#)

① 若拼接项有正则表达式, 「使用推荐表名」后, 正则表达式对应内容需手工替换成具体值

表别名: 可由小写英文字母、下划线(_)、数字组成, 并以小写英文字母开头。

* 表中文名: 订单明细表lq1

2. 定义业务处理的粒度：订单商品粒度。
3. 选定每个事实表行的维度：订单 id+商品 id。
4. 确定用于形成每个事实表行的数字型事实。

设置字段属性 共6个字段

序号	字段英文名	字段中文名	数据类型	非空	字段标准	标准值域	度量单位	操作
1	order_id	订单id	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
2	goods_id	商品id	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
3	user_id	用户id	BIGINT	<input type="checkbox"/>	vipUserId	请选择	请选择	
4	xbm	性别码	STRING	<input checked="" type="checkbox"/>	性别码	性别	请选择	
5	buy_count	购买商品数量	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
6	dt	日期	STRING	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	

怎么在维度建模中进行过程治理

1. 设置非空，联动质检。

设置字段属性 共6个字段 批量删除 全局排序

序号	*字段英文名	*字段中文名	*数据类型	非空	字段标准	标准值域	度量单位	操作
1	order_id	订单id	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
2	goods_id	商品id	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
3	user_id	用户id	BIGINT	<input type="checkbox"/>	vip用户id	请选择	请选择	
4	xbm	性别码	STRING	<input checked="" type="checkbox"/>	性别码	性别	请选择	
5	buy_count	购买商品数量	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
6	dt	日期	STRING	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	

2. 关联字段标准，联动质检。

设置字段属性 共6个字段 批量删除 全局排序

序号	*字段英文名	*字段中文名	*数据类型	非空	字段标准	标准值域	度量单位	操作
1	order_id	订单id	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
2	goods_id	商品id	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
3	user_id	用户id	BIGINT	<input type="checkbox"/>	vip用户id	请选择	请选择	
4	xbm	性别码	STRING	<input checked="" type="checkbox"/>	性别码	性别	请选择	
5	buy_count	购买商品数量	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
6	dt	日期	STRING	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	

3. 关联标准值域，联动质检。

设置字段属性 共6个字段 批量删除 全局排序

序号	*字段英文名	*字段中文名	*数据类型	非空	字段标准	标准值域	度量单位	操作
1	order_id	订单id	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
2	goods_id	商品id	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
3	user_id	用户id	BIGINT	<input type="checkbox"/>	vip用户id	请选择	请选择	
4	xbm	性别码	STRING	<input checked="" type="checkbox"/>	性别码	性别	请选择	
5	buy_count	购买商品数量	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
6	dt	日期	STRING	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	

4. 语义化维度/指标。

设置字段属性 共6个字段 批量删除 默认排序

序号	*字段英文名	*字段中文名	*数据类型	非空	字段标准	标准值域	度量单位	操作
1	order_id	订单id	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
2	goods_id	商品id	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
请输入描述口径								
		维度属性	关联类型	poc_engine_HIVE ctydim dim_goods_goods_info_df good_id				
3	user_id	用户id	BIGINT	<input type="checkbox"/>	vip用户id	请选择	请选择	
4	xbm	性别码	STRING	<input checked="" type="checkbox"/>	性别码	性别	请选择	
5	buy_count	购买商品数量	BIGINT	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	
请输入描述口径								
		原子指标	关联类型	购买商品数量 关联内容				
6	dt	日期	STRING	<input checked="" type="checkbox"/>	请选择字段标准	请选择	请选择	

5. 设置主键，联动质检。

设置主键

*字段英文名	*字段中文名	*数据类型	操作
order_id	订单id	BIGINT	
goods_id	商品id	BIGINT	

6 常见问题

6.1 常见产品问题

本章节主要介绍大数据管理平台 DataWings 的常见问题。

大数据管理平台 DataWings 是什么？

天翼云大数据管理平台 DataWings，是一站式、开箱即用的大数据开发与数据资产管理工具，包括数据建模、数据集成、数据处理、数据治理、数据地图、数据服务等能力，致力于解决企业在数字化转型时面临的数据采集、开发、管理、治理等方面的问题，帮助企业高效上云，实现数据驱动业务增长。

如何使用大数据管理平台 DataWings 进行作业开发？

DataWings 提供了开箱即用、高安全保障的标准化产品，作业开发前可参考如下流程进行配置。

1. 开通产品并在控制台查看各资源组运行状态。
2. 注册大数据集群并创建引擎。
3. 开始创建数据作业（按需开展数据建模、数据集成、数据处理等数据作业）。

如何治理我的数据？

1. 数据建模：提供全面的数仓规划、数据标准（含命名词典、字段标准、表名规范等）、数据指标及维度建模能力，为后续的数据使用提供标准规范及治理依据。
2. 数据处理：支持加入数据质量节点，及时对调度任务中的数据质量问题进行治理。
3. 数据质量：支持与数据开发联动，提供强弱规则校验，提供近 40 种以上质检模板规则以及自定义规则。
4. 数据治理 360：从多维度评估数据治理水平，为客户提供全面的治理能力打分，提示用户及时治理系统健康问题。

如何查看我的数据？

用户可以通过“数据地图”功能了解数据资产，不仅提供数据发现、数据总览功能，还支持自定义数据资产编目（自定义数据目录），同时跟数据建模打通，支持按建模分层及类目查看，有

效进行数据资产管理和查看。

如何进行数据处理/分析工作?

用户可以在“数据处理”模块进行数据的处理与分析工作，数据处理功能支持丰富的存算引擎，包含 Spark、Hive、Python、Shell、JDBC 等开发节点，支持画布式自由编排 workflow。

如何分享我的数据?

用户可以通过“数据服务”功能进行数据分享。数据服务功能提供了全面的服务发布及共享能力，帮助企业统一管理面向内外部的 API 服务。通过标准、规范的服务开发过程，实现 API 开发、API 测试、API 发布、API 管理、API 授权等全流程，为上层业务提供灵活的数据查询。

如何进行权限配置与管控?

管理员可通过“权限管理-角色管理”功能，创建所需角色，并将角色分配给相应子用户以实现权限的配置与管控。若子用户有权限需求，也可通过“流程中心-权限申请”功能，主动发起权限申请，经管理员审批通过后，可使用相关权限。

6.2 购买类

本章节主要介绍大数据管理平台 DataWings 的购买类问题。

如何订购大数据管理平台 DataWings?

公测期间暂不支持自行订购，您可在大数据管理平台 DataWings 的官网产品详情页，点击“申请开通”，填写并提交试用信息，等待工作人员联系，若您符合试用条件，将为您开通产品。

如何续订大数据管理平台 DataWings?

您可以进入天翼云官网，点击右上角的“管理中心”，点击“订单管理”-“续订管理”，输入筛选条件，找到需要续订的产品信息，按提示完成续订操作即可。

如何退订大数据管理平台 DataWings?

公测期间暂不支持自行退订，退订请联系客户经理。

资源组的可用规格有哪些？

公测期间，默认开通集成资源组、调度资源组、服务资源组和数据源资源组的基础规格，商用后，您可以选购更多规格。

6.3 计费类

本章节主要介绍大数据管理平台 DataWings 的计费类问题。

大数据管理平台 DataWings 如何收费？

公测期间可免费申请试用大数据管理平台 DataWings。

大数据管理平台 DataWings 计费项有哪些？

公测结束商用后，购买大数据管理平台 DataWings 的费用包含两个部分：

- 大数据管理平台 DataWings 版本服务费用。
- 资源组费用（集成资源组、调度资源组、服务资源组）。