ALIBABA CLOUD

Alibaba Cloud

DataWorks

数据治理

文档版本：20220712

阿里云

法律声明

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

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。

2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。

3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。

4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。

5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。

6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

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

目录

1.安全中心

1.1. 概述

DataWorks的安全中心，帮助您快速构建平台的数据内容、个人隐私等相关的安全能力，满足企业面向高风险场景的各类安全要求（例如，审计），无需您额外配置即可直接使用该功能。

DataWorks的安全中心作为云上大数据体系的安全门户，致力于向您提供面向数据安全生命周期全过程的安全能力，同时在符合安全规范要求的前提下，提供各类安全诊断的最佳实践。其核心功能如下：

- 数据权限管理

  安全中心为您提供精细化的数据权限申请、权限审批、权限审计等功能，实现了权限最小化管控，同时，方便您查看权限审批流程各环节的进展，及时跟进处理流程，详情请参见数据访问控制。

- 数据内容安全管理

  安全中心提供的数据分级分类、敏感数据识别、数据访问审计、数据源可追溯等功能，在处理业务流程的过程中，能够快速及时识别存在安全隐患的数据，保障了数据内容的安全可靠，详情请参见数据保护伞。

- 安全诊断的最佳实践

  安全中心提供的平台安全诊断、数据使用诊断等功能，在符合安全规范要求的前提下，为您提供了诊断各类安全问题的最佳实践。保障您的业务在最佳的安全环境，更有效的执行。详情请参见平台安全诊断及数据使用诊断。

1.2. 数据平台安全

1.2.1. 平台安全诊断

DataWorks的平台安全诊断，为您提供了当前DataWorks工作空间与绑定的引擎在数据传输、存储、运算等过程中，与身份认证、访问权限控制、开发模式等功能相关的安全能力，以及诊断相关安全问题的最佳实践，帮助您及时发现平台的安全隐患，在进行相关工作事务前快速建立基本的安全体系。

背景信息

平台安全诊断根据相关安全问题的最佳实践，为您展示系统诊断出的当前工作空间与绑定的计算引擎，在进行业务交互时存在的风险隐患。您可以根据诊断结果，识别风险类别、风险等级，查看风险详情，及时处理待优化项，保障业务执行过程的安全可靠。平台安全诊断目前支持的安全域说明如下：

- **数据计算与存储安全性诊断**

  用于对数据权限的控制、数据存储加密、数据存储备份等功能进行安全性诊断，及时识别潜在的安全隐患，提升在数据存储与访问过程中的安全性。

- **数据传输安全性诊断**

  用于对数据源访问控制、生产与开发数据源隔离等功能进行安全行诊断，识别数据传输时存在的安全隐患，方便您及时发现并优化，保障数据传输环境的安全可靠。

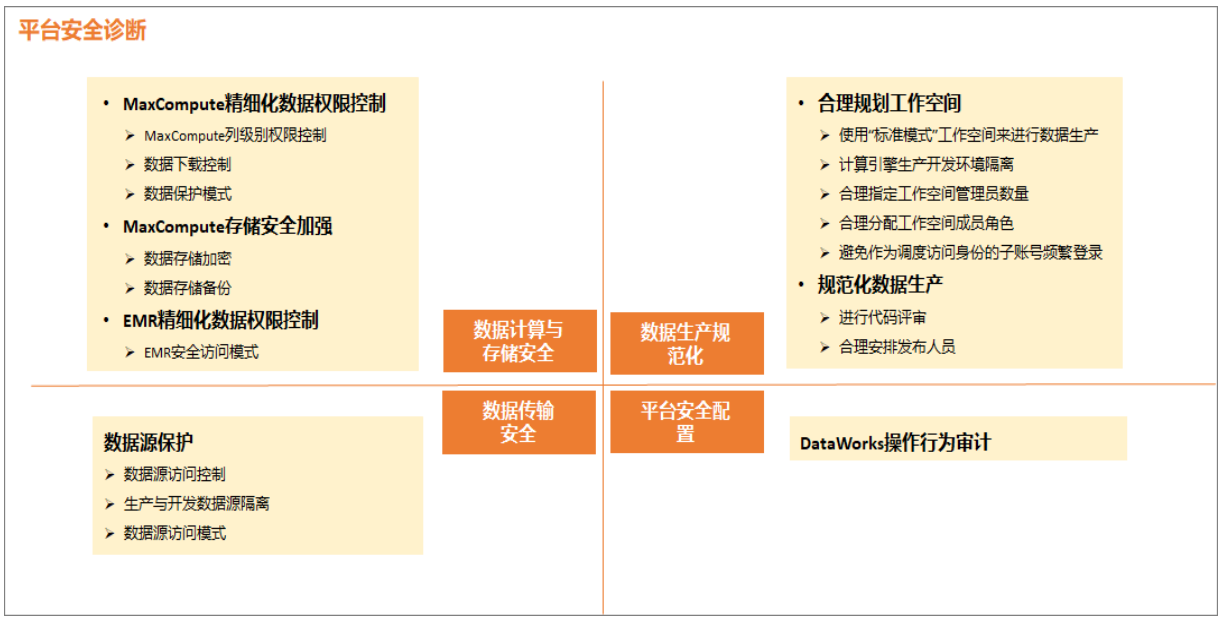
- **数据生产规范化诊断**

  用于对当前工作空间安排的角色、管理员数量、发布人员的合理性等涉及生产环节的安全性进行诊断，及时发现并清理安全隐患，提升数据产出体系的稳定性与安全性。

- **平台安全配置诊断**

  用于对DataWorks操作行为审计等功能进行安全诊断，提升在数据安全通用领域的安全性。

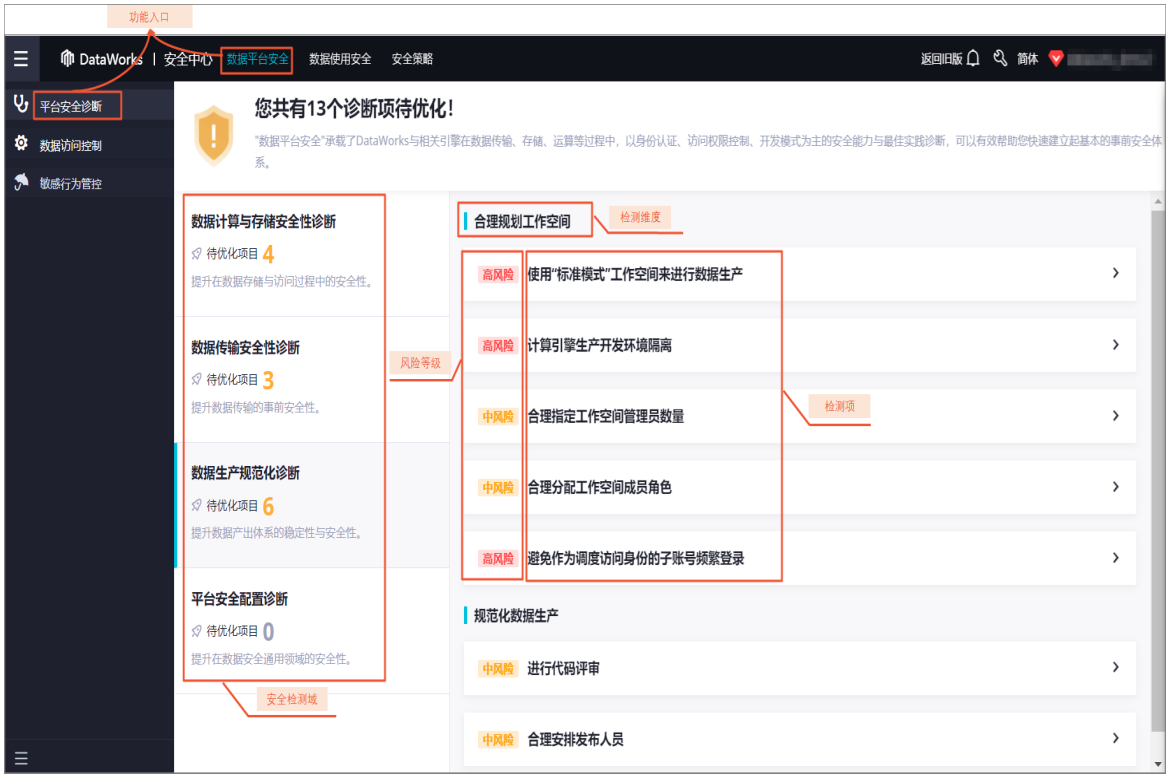
上述安全域各维度检测出的待优化项，通过低、中、高定义待优化项的风险等级。同时在每一项风险问题项中提供相应的诊断结果与改进建议，保障业务执行过程的安全可靠。您可以通过附录：诊断详情列表查看所有安全检测域的各维度诊断规则。



进入平台安全诊断

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上角的图标，选择**全部产品 > 数据治理 > 安全中心**，默认进入**数据访问控制**页面。
5. 单击左侧导航栏的**平台安全诊断**，进入**平台安全诊断**页面。

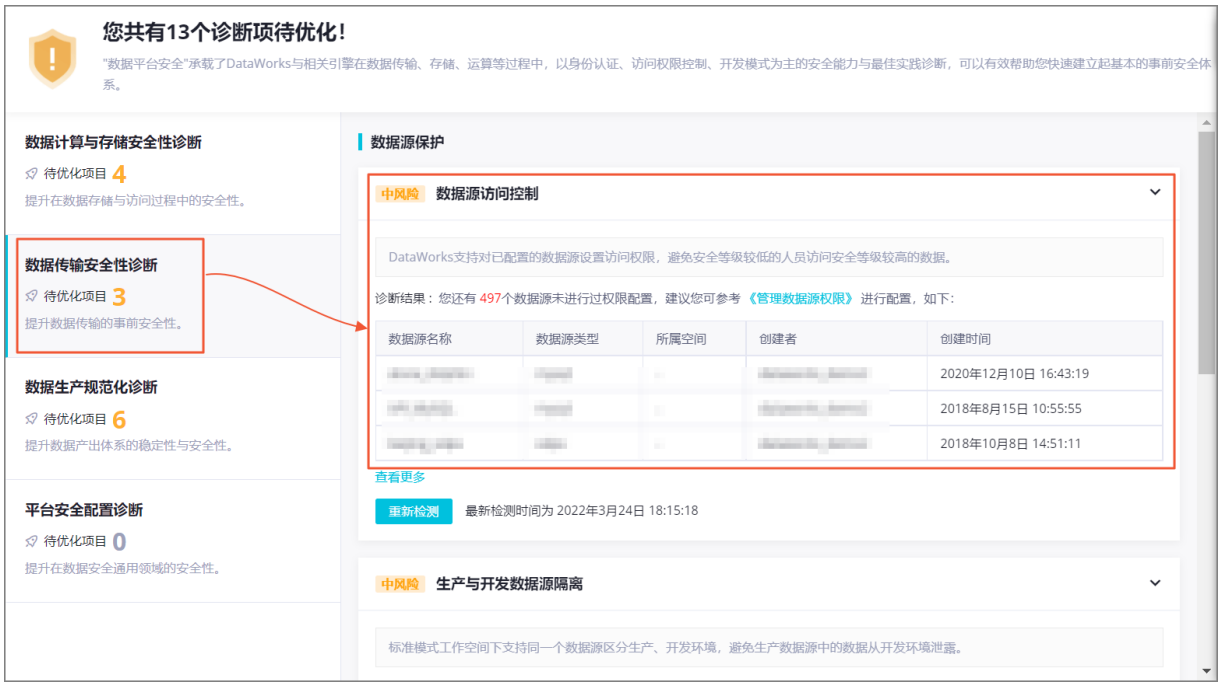
   平台安全诊断界面默认对当前地域的待治理项问题进行检测，量化待优化治理项，并通过低、中、高风险等级定义安全隐患。



查看诊断结果

在**平台安全诊断**界面，对每个安全域下存在的安全隐患、待优化的中、高等级风险问题进行统计，您可以单击对应的目标安全隐患，查看风险详情，并根据诊断建议进行优化。下图以**数据传输安全性诊断**为例查看该安全域下具体待治理项。

查看诊断结果与诊断建议：



- 安全隐患

  数据源未配置相关权限，可能导致安全等级较低的人员访问安全等级较高的数据，造成数据源访问不安全。

- 处理建议

  您可以根据所给的建议进行数据源权限配置，提升数据源访问的安全性。

附录：诊断详情列表

平台安全诊断功能目前支持的安全检测域详情如下。

> **说明** 实际界面展示的诊断项与您工作空间绑定的引擎和现有的待治理项有关。

- **数据计算与存储安全性诊断**

  提升在数据存储与访问过程中的安全性。

安全检测维度	安全检测项	检测对象	检测方式

安全检测维度	安全检测项	检测对象	检测方式
MaxCompute精细化数据权限控制	MaxCompute列级别权限控制 **说明** MaxCompute 2.0安全模型具有更细粒度的数据权限管理能力、更科学的项目分权管控机制、更强大的端识别能力，支持用户实现更加贴合实际场景的安全配置。	MaxCompute项目	列级权限控制依托于MaxCompute2.0权限模型。此安全检测项为您扫描哪些MaxCompute工作空间未开启maxcompute2.0权限模型。
	数据下载控制 **说明** 建议严格控制无关人员直接下载数据（MaxCompute Tunnel方式）到本地，避免非预期的数据泄露事故。	MaxCompute项目	下载权限控制依托于MaxCompute 2.0权限模型与Download权限。此安全检测项为您扫描哪些MaxCompute工作空间未开启maxcompute2.0权限模型。已经开启2.0工作空间中，哪些工作空间未开启下载管控。您可以通过Download权限控制决定是否需要开启权限控制。
	数据保护模式 **说明** 数据保护机制允许用户控制数据流出方法。	MaxCompute项目	此安全检测项为您扫描是否已为部分或所有MaxCompute项目设置项目保护模式。关于MaxCompute项目保护详情请参见：数据保护机制。
MaxCompute存储安全加强	数据存储加密 **说明** MaxCompute支持通过密钥管理服务KMS（Key Management Service）对数据进行加密存储，提供数据静态保护能力，满足企业监管和安全合规需求。详情请参见：《MaxCompute数据存储加密》	MaxCompute项目	此安全检测项为您扫描未开启数据存储加密的工作空间并列出列表。如有需求可通过提交工单对已有工作空间开启存储加密。

安全检测维度	安全检测项	检测对象	检测方式
	数据存储备份 **说明** 系统会自动备份MaxCompute数据的历史版本并保留一定时间，您可以对保留周期内的数据进行快速恢复，避免因误操作丢失数据。详情请参见：《MaxCompute备份与恢复》	MaxCompute项目	MaxCompute工作空间默认拥有该功能，您可以通过《MaxCompute备份与恢复》文档结合实际情况来调整备份天数或恢复数据。
EMR精细化数据权限控制	EMR安全访问模式 **说明** EMR“安全模式”支持不同阿里云主、子账号之间实现数据权限隔离。安全模式详情请参见：安全模式	DataWorks工作空间	此安全检测项为您扫描哪些工作空间下的EMR引擎仍未使用安全模式。

- **数据传输安全性诊断**

  提升数据传输的事前安全性。

安全检测维度	安全检测项	检测对象	检测方式
数据源保护	数据源访问控制 **说明** DataWorks支持对已配置的数据源设置访问权限，避免安全等级较低的人员访问安全等级较高的数据。	DataWorks工作空间数据源	此安全检测项为你扫描哪些工作空间未对数据源进行权限配置。您可以参考管理数据源权限进行配置。
	生产与开发数据源隔离 **说明** 标准模式工作空间下支持同一个数据源区分生产、开发环境，避免生产数据源中的数据从开发环境泄露。您可以根据数据源开发和生产环境隔离进行评估与修改数据源。	DataWorks工作空间数据源	此安全检测项为你扫描哪些标准模式空间下的数据源生产、开发环境配置相同。

安全检测维度	安全检测项	检测对象	检测方式
	数据源访问模式 **说明** DataWorks支持通过角色模式访问OSS数据源，该模式较传统Access Key模式更具安全性，可有效避免Access Key泄露的情况。	DataWorks工作空间数据源	此安全检测项为您扫描哪些工作空间下的OSS数据源仍在使用Access Key模式，您可以根据通过RAM角色授权模式配置数据源对数据源进行改造。

- **数据生产规范化诊断**

  提升数据产出体系的稳定性与安全性。

安全检测维度	安全检测项	检测对象	检测方式
合理规划工作空间	使用“标准模式”工作空间来进行数据生产 **说明** “标准模式”空间比“简单模式”空间具备更强的安全性，详情请参见：简单模式和标准模式的区别。	DataWorks工作空间模式	此安全检测项为您扫描当前地域下哪些工作空间仍为简单模式工作空间。您可以根据实际情况将简单模式空间升级为标准模式空间，升级前请仔细阅读：工作空间模式升级。
	计算引擎生产开发环境隔离 **说明** 标准模式工作空间下支持将计算引擎生产、开发环境进行区分隔离，避免生产环境中的数据从开发环境泄露。	DataWorks工作空间引擎	此安全检测项为您扫描当前地域下哪些工作空间绑定的计算引擎实例存在开发、生产环境配置相同的情况。
	合理指定工作空间管理员数量 **说明** 在单个工作空间内，过多的管理员可能导致管理混乱，建议每个空间设置不超过3个空间管理员。	DataWorks工作空间成员管理	此安全检测项为您扫描当前地域下哪些工作空间设置的空间管理员个数超过3。
	合理分配工作空间成员角色 **说明** 在单个工作空间内，建议一人仅扮演一种角色（专人专职），避免出现一人分饰多种角色而导致的越权情况。	DataWorks工作空间成员管理	此安全检测项为您扫描当前地域哪些工作空间下一人被授予多个角色。建议您参考附录：预设角色权限列表（空间级）深入了解各角色用途后进行适当配置。

安全检测维度	安全检测项	检测对象	检测方式
	避免作为调度访问身份的子账号频繁登录 **说明** 建议禁止登录作为引擎调度访问身份的子账号，以免发生无关成员查看引擎关键数据的情况。	DataWorks工作空间管理	此安全检测项为您扫描当前地域哪些工作空间下，作为调度访问身份的RAM子账号近三个月内登录过阿里云DataWorks。
规范化数据生产	进行代码评审 **说明** DataWorks提供代码评审，标准模式空间开启强制代码评审开关后，开发人员提交的节点必须通过评审人审核后才可以发布。	DataWorks工作空间管理	此安全检测项为您扫描当前地域哪些工作空间未开启或配置代码评审功能与评审范围。您可以参考代码评审对空间进行配置。
	合理安排发布人员 **说明** 在标准模式空间下，实际执行任务发布的人员应与任务开发者进行区分。	DataWorks工作空间管理	此安全检测项为您扫描近30日内是否出现任务仅由同一人开发并发布的情况。

- **平台安全配置诊断**

  提升在数据安全通用领域的安全性。

安全检测维度	安全检测项	检测对象	检测方式
DataWorks操作行为审计	DataWorks操作行为审计 **说明** DataWorks已支持操作行为审计，您可以通过阿里云Action Trail对用户在DataWorks上的操作行为进行审计，延时约为5~10分钟，详情请参见：通过操作审计查询行为事件日志。	DataWorks工作空间管理	DataWorks工作空间默认拥有该能力。开通操作行为审计产品后即可记录DataWorks操作日志。

1.2.2. 数据访问控制

安全中心提供的数据访问控制功能，方便您以可视化的方式申请、审批、审计相关权限，查看审批流程并跟进审批进度，进行权限的管控。

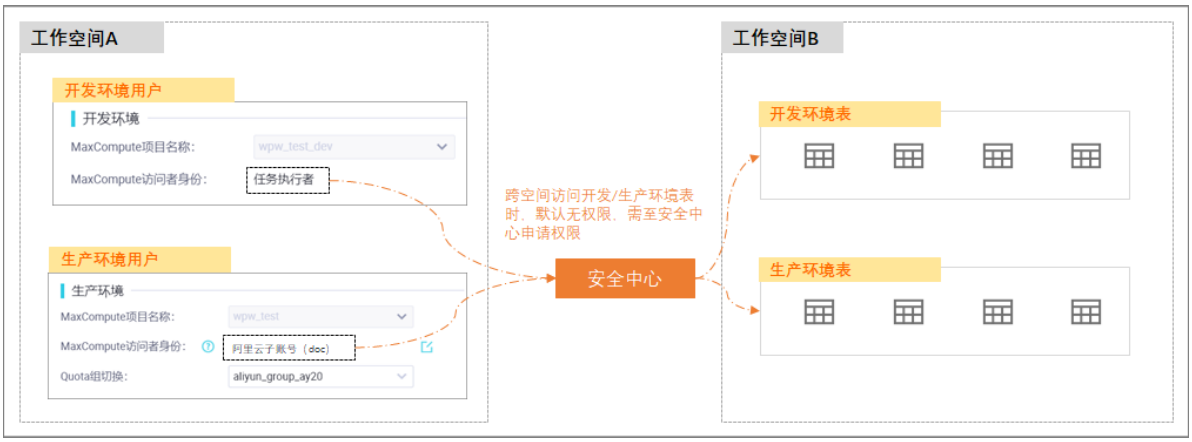
应用场景

为了保障生产数据安全，在标准模式工作空间下，DataWorks对子账号访问MaxCompute表的操作进行了管控。详情请参见：用户、角色与权限概述、简单模式和标准模式的区别。

- 场景一：同工作空间内，开发环境用户访问生产环境表。

当DataWorks的子账号未被添加为生产环境计算引擎访问身份时，默认该账号无法在数据开发界面直接操作本工作空间的生产表，如果子账号需要拥有生产表权限，需要在安全中心发起申请，待审批通过后，便可在数据开发界面对表进行相关操作。

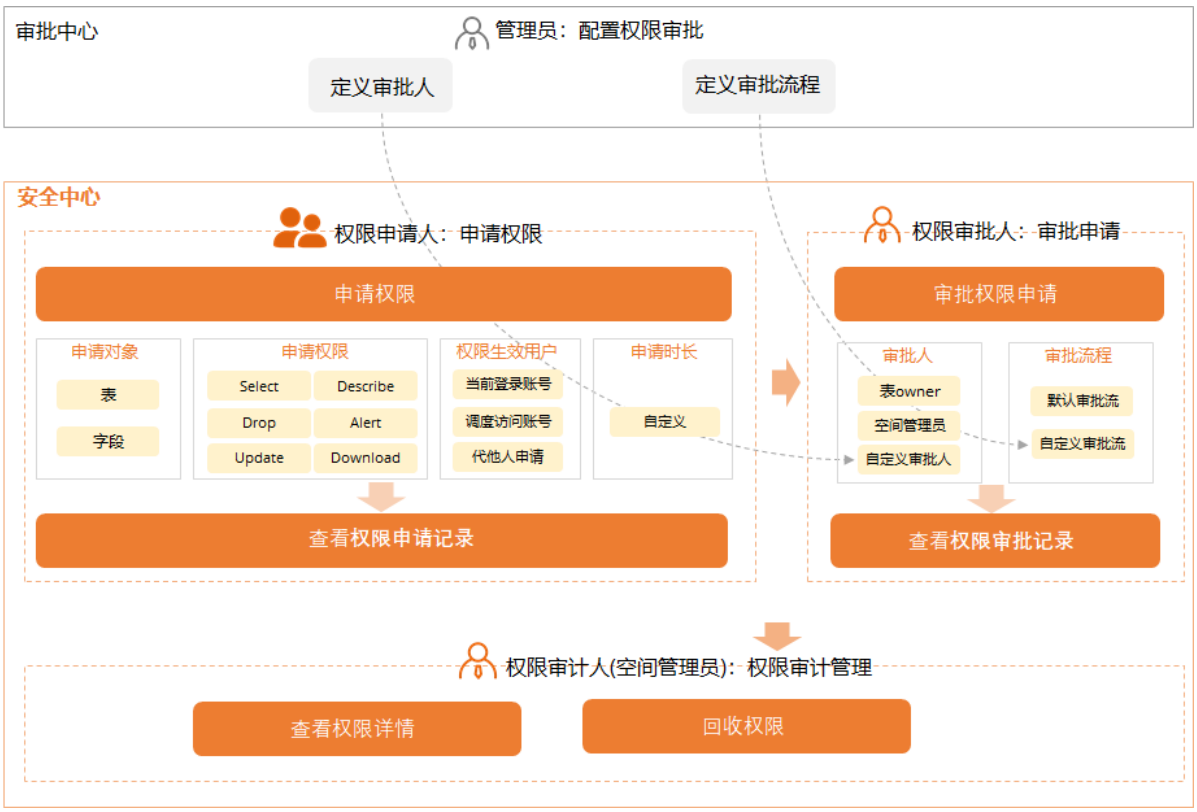
- 场景二：开发或生产环境用户访问跨工作空间的开发或生产环境表。



默认不在工作空间下的子账号无法在数据开发界面跨项目访问开发表或生产表。如果需要跨项目操作开发表或生产表，子账号需要在安全中心发起申请，待审批通过后，便可在数据开发界面对表进行相关操作。

数据访问控制流程

数据访问控制功能支持您进行**权限申请**、**权限审批**、**权限审计**的操作，还支持您查看**权限申请记录**、**权限审批记录**。在子账号开发过程中没有相关表权限的场景下，可以通过**权限申请**界面申请对应权限。待审批人员在**权限审批**页面通过申请后，便可获得权限。



- 权限申请人：可以通过权限申请页面申请MaxCompute表的权限。对于已提交的申请，支持通过权限申请记录页面查看当前登录的阿里云账号提交的申请记录。
- 权限审批人：可以通过权限审批页面查看我作为空间管理员或者表Owner时，需要审批的表权限。对于已审批的申请，支持通过权限审批记录页面查看当前登录的阿里云账号审批通过的表。

还支持阿里云主账号或空间管理员进入**权限审计**页面对工作空间下的成员及其拥有的表权限进行管控，支持对工作空间下某成员所拥有的权限进行回收。详情请参见：权限审计。

> **说明** DataWorks的**安全中心**为您提供默认的权限申请审批流程，同时也支持您在**审批中心**自定义审批流程。当您申请MaxCompute引擎表字段权限时，DataWorks会根据申请的表字段来识别需要进行哪种类型的审批流程。
>
> - 如果表字段在自定义审批流程框定的数据范围内，后续将进行审批中心自定义的审批流程。详情请参见审批中心。
> - 如果表字段不在自定义审批流程框定的数据范围内，则后续将进行安全中心默认的审批流程。默认将审批单发送给表Owner和**工作空间管理员**，只要其中一人审批通过，申请人便可获得表或字段权限。

使用限制

目前仅支持通过**数据访问控制**申请MaxCompute表的权限。

注意事项

数据访问控制页面为您展示的是新版的权限管控平台，如果您需要使用旧版权限管控平台，可以单击页面顶部菜单栏右侧的**返回旧版**进行操作。旧版权限管控平台，详情请参见安全中心。

进入数据访问控制

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上角的图标，选择**全部产品 > 数据治理 > 安全中心**，默认进入**数据访问控制**页面。

权限申请

1. 进入**权限申请**页面。
2. 选择需要申请的表。
   i. 在**申请内容**区域，选择目标**工作空间**及**项目**。

      目前仅支持通过**数据访问控制**申请MaxCompute表的权限。

      因此**申请类型**默认为**表**，**引擎类型**默认为**MaxCopmute**。

   ii. 在**待添加表**区域，勾选需要申请的目标表。

      勾选目标表后，右侧会显示目标表的相关信息。单击**表名称**前 + 图标，显示当前表的所有字段，您可以选择申请目标表的部分或全部字段的权限。默认申请目标表全部字段的权限。

> **说明**
>
> - MaxCompute项目开启Policy权限控制后，该项目下的表才可以定义并在安全中心单独对表中具体字段申请权限。详情请参见：MaxCompute高级配置。关于MaxCompute表中字段的安全等级说明，详情请参见：Label权限控制。
> - 目前支持申请表级别的Select、Describe、Drop、Alter、Update、Download权限。同时支持您针对单个字段单独申请字段权限。

3. 配置**申请信息**。

参数	描述
使用者	○ **当前登录账号**：表示为当前登录DataWorks工作空间的阿里云账号申请目标表权限。 ○ **调度访问账号**：表示为被设置为调度访问身份的云账号申请目标表权限。选择该选项时，需要配置**工作空间**参数。 ○ **代他人申请**：表示当前登录DataWorks工作空间的阿里云账号为其他阿里云账号申请目标表权限。选择该选项时，需要配置**用户名**参数。
工作空间	被设置为调度访问身份的云账号。
用户名	当前登录阿里云账号以外的其他阿里云账号。
申请时长	支持您按需自定义申请表权限的时长，过期权限将自动收回。 **说明** 使用此功能前需要表所在的MaxCompute项目开启Policy授权，详情请参见：MaxCompute高级配置，关于MaxCompute Policy的说明请参见：Policy权限控制。
申请原因	输入申请目标表权限的原因。

4. 单击**申请权限**，提交申请。

   您可以在**权限申请记录**页签，查看当前申请的审批详情及审批记录。

权限审批

1. 查看待审批的申请。

   进入**权限审批**页面，您可以根据**申请账号**、**申请时间**、**工作空间**、**项目名称**、**对象名称**等条件进行筛选，查看目标条件下，当前登录的阿里云账号名下需要审批的申请信息。

> **说明** 同一个申请单提交的多张表权限申请，会按照表Owner的不同自动拆分成多个申请单。

2. 查看审批详情。

   单击目标申请**操作**列的**审批**，您可以在**审批详情**对话框查看目标申请的**申请详情**、**审批记录**等详细信息。

3. 审批申请。

   根据申请的详细内容及当前需求判断是否同意审批该申请，填写**审批意见**，选择**同意**或**拒绝**当前申请。

   您也可以直接在**权限审批**页面，勾选全部申请，单击**批量同意**或**批量拒绝**，填写**审批意见**，批量处理目标申请。

查看权限申请及审批记录

- 进入**权限申请记录**页面，您可以根据**审批状态**、**申请时间**、**工作空间**、**项目名称**、**表名称**等条件进行筛选，查看目标条件下，当前登录的阿里云账号名下涉及的申请记录。

  您可以单击目标申请**操作**列的**查看详情**，查看申请的详细信息。同时，对于**审批状态**为**审批中**的申请，您可以继续后续的审批操作。

- 进入**权限审批记录**页面，您可以根据**申请账号**、**审批结果**、**工作空间**、**项目名称**、**对象名称**、**申请时间**等条件进行筛选，查看目标条件下，当前登录的阿里云账号的审批记录。

  您可以单击目标申请**操作**列的**查看详情**，查看申请的详细信息。

权限审计

您可以在**权限审计**页面，根据**工作空间项目名称**、**对象名称**进行筛选，查看通过安全中心审批的目标**工作空间**、项目或对象所涉及的权限申请。

1.3. 数据使用安全

1.3.1. 数据使用诊断

DataWorks的数据使用诊断，为您提供了对当前DataWorks工作空间的数据内容及数据隐私的安全保护能力，以及诊断相关安全问题的最佳实践及解决方案，帮助您快速建立数据使用时和使用后的基本安全体系。

进入数据使用诊断

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上角的图标，选择**全部产品 > 数据治理 > 安全中心**，默认进入**数据访问控制**页面。
5. 单击顶部菜单栏的**数据使用安全**，进入**数据使用诊断**页面。

查看诊断结果

数据使用诊断根据诊断相关安全问题的最佳实践，为您展示了系统诊断出的当前工作空间，在数据使用过程中存在的风险隐患。您可以根据诊断结果，识别风险类别、风险等级，查看风险详情，及时处理待优化项，保障数据使用过程的安全可靠。

中风险及**高风险**等级的风险问题属于潜在的安全隐患，您可以单击对应级别的目标安全隐患，查看风险详情并及时优化。下图以**敏感数据脱敏**为例查看该检查项目前的风险等级及诊断结果。

- 安全隐患

  DataWorks支持管理员通过智能方式识别敏感数据，前提是您需要对相关数据分级设置数据识别规则。如果未设置数据的识别规则，则系统无法识别敏感数据，可能导致敏感数据流出，造成信息泄露。

- 处理建议

  您可以根据所给的建议，进行敏感数据规则设置，以便全面掌握敏感数据分布情况，保护敏感数据的安全。

诊断详情列表

数据安全防护诊断帮助您提升敏感数据的安全性。

安全检测维度	安全检测项	检测对象	检测方式
敏感数据保护	数据分级分类 说明 数据分级分类是一切敏感数据防护的开始，DataWorks支持管理员对本企业数据进行敏感级别划分。	数据保护伞服务	此安全检测项为您检测是否已在数据保护伞设置分级分类规则。分类分级配置详情请参见：数据分类分级。
	敏感数据识别 说明 从企业纷繁复杂的数据体系中找出重要、敏感的数据是敏感数据防护的关键，DataWorks支持管理员通过智能方式识别敏感数据。	数据保护伞服务	此安全检测项为您检测哪些数据分级目前没有配置数据识别规则。未配置数据识别规则会导致目标数据分级无法匹配到相关数据。您可以参考敏感数据识别进行敏感数据规则设置，以便全面掌握您的敏感数据分布情况。

安全检测维度	安全检测项	检测对象	检测方式
	敏感数据脱敏 说明 在数据使用过程中为避免数据滥用、泄露，实现“数据可用不可见”的目的，DataWorks支持管理员配置脱敏规则，帮助企业在“安全”与“便利”之间获得平衡。	数据保护伞服务	此安全检测项为您检测哪些工作空间已开启脱敏规则，但未开启工作空间级别脱敏开关。您可以参考工作空间配置文档开启空间级别脱敏开关以实现DataStudio页面查询脱敏。

1.3.2. 敏感数据管理（数据保护伞）

数据保护伞是一款数据安全管理产品，为您提供数据发现、数据访问、数据风险、数据审计和规则配置等功能。

进入数据保护伞

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上方的图标，选择**全部产品 > 数据治理 > 数据保护伞**。
5. 单击**立即体验**，进入数据保护伞。

> 说明
> - 如果阿里云主账号已授权，直接进入数据保护伞的首页。
> - 如果阿里云主账号未授权，进入数据保护伞的授权页面。

开通数据保护伞

阿里云主账号在**服务声明**页面，勾选**我已阅读并接受以上协议条款**，单击**立刻开通**。

> 注意 仅阿里云主账号可以进行授权，开通数据保护伞。

数据保护伞的使用请参见概述。

1.4. 安全策略

1.4.1. 实体转交

实体转交支持将目标工作空间下各模块的实体（资源、函数等），通过自动或手动转交触发机制，统一转交给指定实体接收人。转交时除默认规则外，还支持您自定义空间级别转交规则。本文为您介绍如何配置实体转交规则，以及如何查看转交日志等。

背景信息

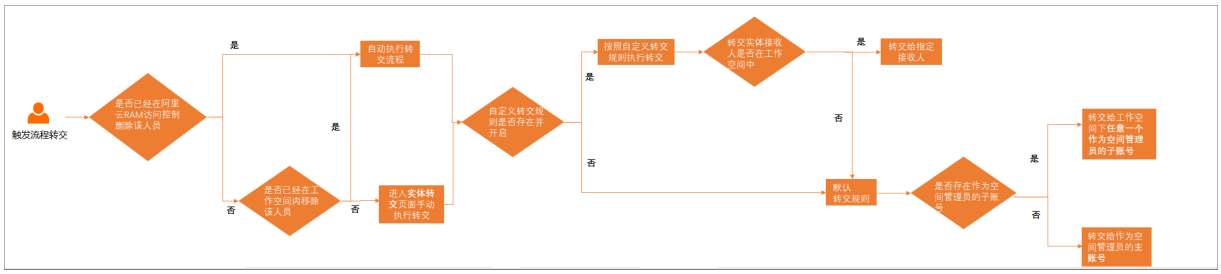
- 在众多实体转交场景中，人员离职场景的实体转交尤为突出，您可以通过实体转交来保障人员离职后DataWorks上的业务安全与稳定，避免因人员离职造成的业务影响。
- 人员离职场景下，其阿里云账号分为账号已删除和账号还未被删除两种情况，针对两种情况，DataWorks为您提供自动触发实体转交和手动触发实体转交两种转交触发机制。同时为实现转交规则可控，除默认转交规则外，DataWorks还支持您通过实体转交页面，自定义空间级别转交规则来对各个模块下的实体指定目标接收人。

使用限制

仅租户安全管理、租户管理员可以进入实体转交页面进行实体转交配置。关于租户权限管控详情请参见：角色及成员管理：全局级。

逻辑说明

自定义转交规则是开启状态时，执行转交后优先转交给自定义规则指定的接收人。如果规则指定的实体接收人不存在（或被移出空间），则实体将按照默认规则进行转交。



- 自动转交触发逻辑：当RAM用户被删除或者已经从工作空间移出时，将会自动触发转交流程，如果对应工作空间没有自定义实体转交接收人，那么人员被移除后将按照默认规则进行转交。默认转交给工作空间中角色为空间管理员的任一个阿里云子账号，当工作空间中没有添加阿里云子账号为空间管理员时，将默认转交给阿里云主账号。如果工作空间已定义了实体转交接收人且接收人未被移出工作空间，则按照实体转交规则转交给指定人员。
- 手动转交触发逻辑：当RAM用户未被删除且还在工作空间成员列表中时，您可以进入**实体转交**页面，手动执行转交。如果对应工作空间没有自定义实体转交规则，那么人员被移除后将按照默认规则进行转交。如果工作空间已定义了实体转交接收人且接收人未被移出工作空间，则按照实体转交规则转交给指定人员。自定义空间级别转交规则详情请参见：配置实体转交规则。

> **说明**
>
> - 如果待转交的实体接收人同时为MaxCompute调度引擎访问身份时，转交规则触发后，将同时修改MaxCompute调度引擎访问身份。关于MaxCompute访问身份详情可参考文档：绑定MaxCompute计算引擎。
> - 自定义转交规则支持按照工作空间维度定义单个工作空间级别的实体转交策略。

进入实体转交

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上角的图标，选择**全部产品 > 数据治理 > 安全中心**，默认进入**数据访问控制**页面。

5. 单击顶部菜单栏的**安全策略**，进入安全策略页面后，在左侧导航栏选择**实体转交**进入实体转交页面。

查看可被转交的实体

在使用说明区域可以查看实体转交功能可被转交的实体、自动转交触发条件及转交注意事项。

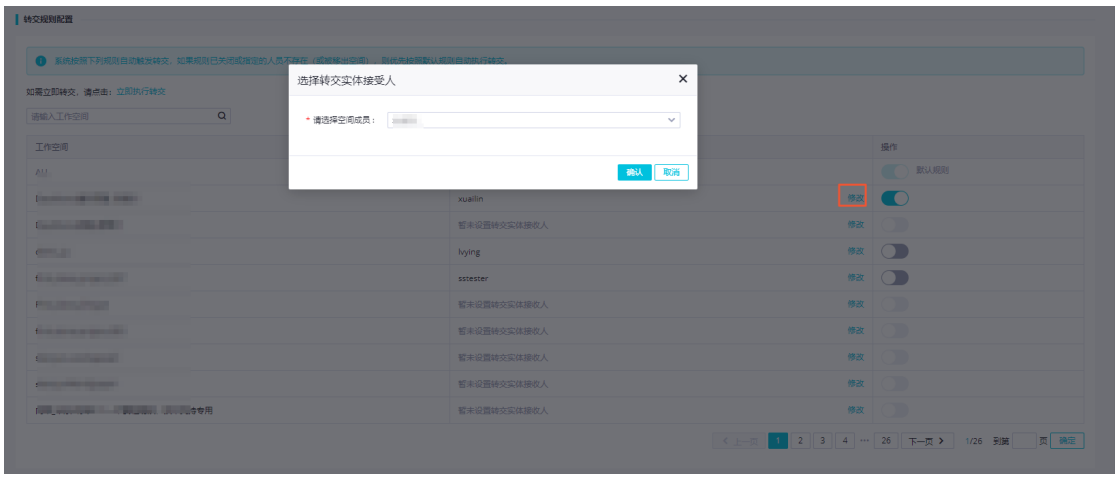
> **说明** 可转交的实体正在逐步丰富中，具体请以产品界面为准。

配置实体转交规则

1. 在转交规则配置区域搜索目标工作空间。

2. 配置实体接收人。

i. 转交规则分为默认转交规则和自定义空间级别转交规则，在**转交规则配置**区域，支持您自定义空间级别实体转交策略。单击对应工作空间右侧**转交实体接收人**列表中的**修改**按钮。在弹出来的**选择转交实体接受人**对话框中选择目标接收人。当触发转交流程时，系统按照下列规则自动触发转交，如果规则已关闭或指定的人员不存在（或被移出空间），则优先按照默认规则自动执行转交。



- 默认转交规则：默认开启状态且不支持关闭，并且对所有未指定空间级别实体转交接收人的工作空间，及指定的实体转交接收人失效的工作空间生效。

  > **说明** 实体接收人失效指的是空间级别实体转交规则指定的接收人在转交时已被移出工作空间。

- 自定义空间级别转交规则：默认关闭，您可以在需要指定实体转交接收人的场景下，通过选择工作空间下成员作为实体接收人，并开启自定义规则的开关。当执行转交规则时，自定义空间级别规则便会生效。

  > **说明** 自定义规则是开启状态时，执行转交后优先转交给自定义规则指定的接收人。如果规则指定的实体接收人不存在（或被移出空间），则实体将按照默认规则进行转交。

ii. 单击操作列的开关按钮开启或关闭转交规则。

- 开启后，实体将按照指定人员进行转交。

  > **说明** 如果规则指定的实体接收人不存在（或被移出工作空间），实体将转交给“默认转交规则的接收人”。

- 关闭后，空间内实体将自动转交给“默认转交规则的接收人”。

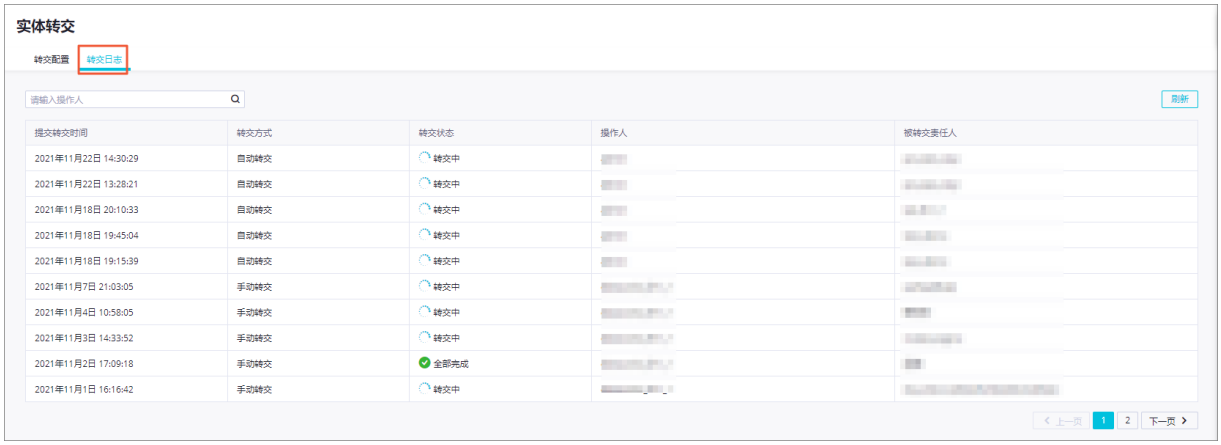
执行实体转交

1. 当RAM用户未被删除且还在工作空间成员列表中时，您可以进入**实体转交**页面，单击**立即执行转交**按钮手动转交指定人员名下的资源。

2. 选择被转交实体的责任人，单击**确认转交**后，当实体责任人在工作空间的成员列表中的时实体将转交给指定责任人。当实体接收责任人不在工作空间的成员列表中时，实体将转交给“默认转交规则的接收人”。

查看转交日志

在**转交日志**页面可以查看转交记录及对应操作人，转交状态、被转交责任人等信息。



2.审批中心

2.1. 审批中心概述

DataWorks审批中心是一个用于管理数据授权、敏感行为管控流程的功能模块，包含审批范围定义、审批流程定义等核心功能，满足不同企业在不同内部合规场景下的审批要求。

功能介绍

在DataWorks上进行数据开发管理的过程中，您可以便捷的对表数据、数据服务API等进行权限管控，进行权限管控时，DataWorks的**安全中心**为您提供默认的权限申请审批流程，同时也支持您在**审批中心**自定义审批流程。

自定义审批流程后，用户在申请某一权限时，DataWorks会自动识别当前用户申请的权限是否需要进过自定义的审批流程，并根据自定义的结果流转审批流程。

DataWorks的**审批中心**当前支持如下功能：

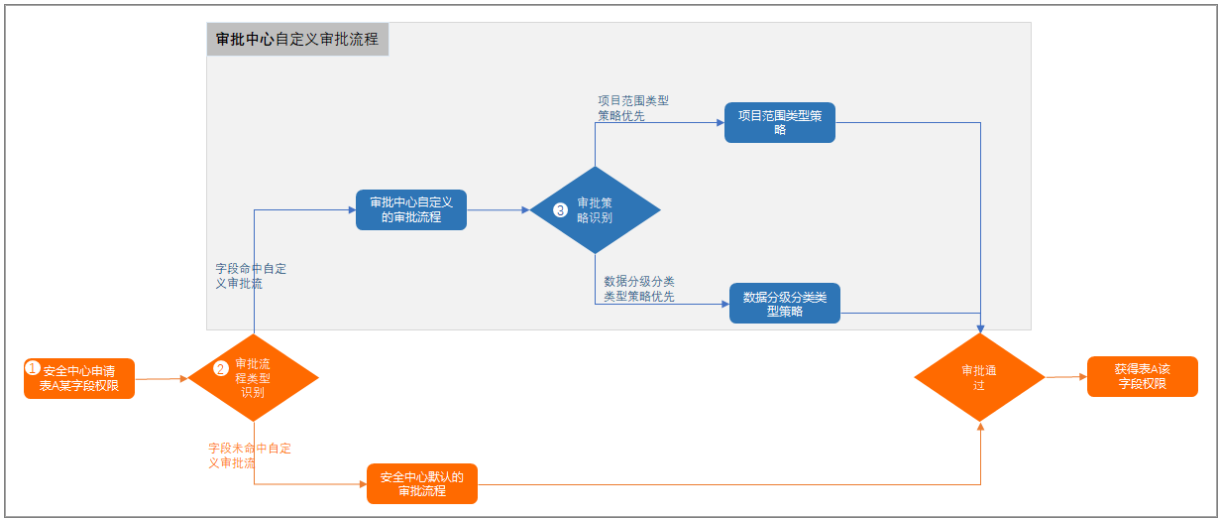
- 定义审批策略：通过圈定审批对象范围、定义审批流程来定义对关键数据资源与敏感行为的管控流程，同时提供短信、邮件、钉钉的方式发送通知。
- 处理审批流程：审批流发起人、审批流执行人可以通过审批中心对流程进行审批。

自定义审批策略的操作详情可参见计算引擎审批策略、数据服务审批策略、数据集成审批策略。

自定义审批策略后，后续进行表权限申请与审批、数据服务权限申请与审批、数据集成任务保存时，流程如表字段权限申请与审批流程、数据服务权限申请与审批流、数据集成任务审批流程所示。

表字段权限申请与审批流程

在**审批中心**自定义审批策略后，当表字段权限申请人在**安全中心**申请权限后，后续的流程如下图所示。



- 在安全中心申请MaxCompute引擎表字段权限时，DataWorks会根据申请的表字段来识别需要进行哪种类型的审批流程。
  - 如果表字段在自定义审批流程框定的数据范围内，则命中自定义审批流，后续将进行审批中心自定义的审批流程。
  - 如果表字段不在自定义审批流程框定的数据范围内，则后续将进行安全中心默认的审批流程。
- 进行自定义审批流程时，DataWorks会根据审批中心设置的审批策略优先级来判断，使用哪种审批策略。

进行自定义审批策略设置时，您可以按照项目范围来框定管控的数据范围并制定审批人、审批通知方式等信息，也可以按照数据分级分类来，并且可以根据需求制定两种方式的优先级，操作详情可参见计算引擎审批策略。

数据服务权限申请与审批流

数据服务审批流程创建完成后，纳入数据服务API发布、函数、服务编排等操作管控的项目在做具体操作时，会触发在审批流程。

当权限申请人在**安全中心**申请权限后，后续的流程如下图所示。

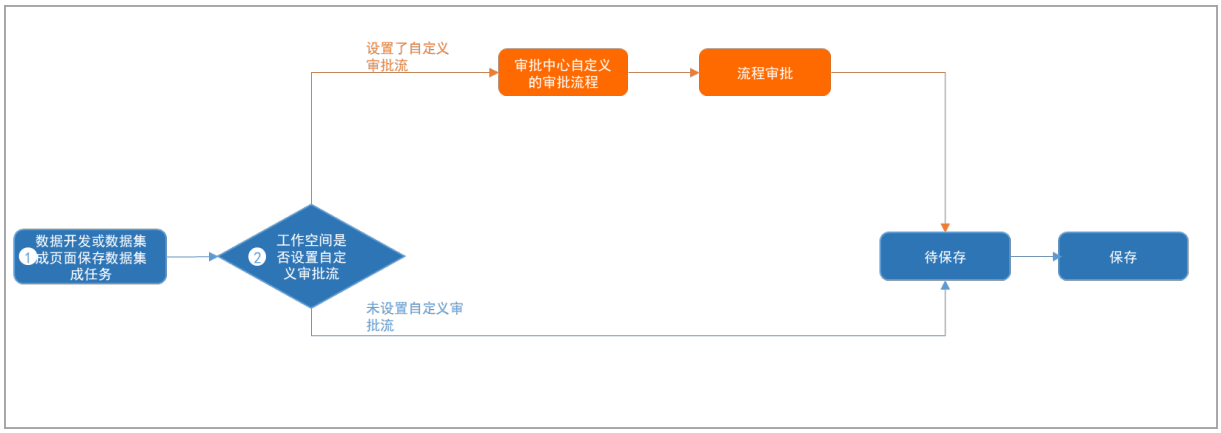
- 在安全中心申请数据服务相关权限时，数据服务根据工作空间是否设置审批流来决定这个工作空间做API，函数，服务编排提交是否走自定义审批流回复完成拒绝。
  - 如果命中自定义审批流，后续将进行审批中心自定义的审批流程。
  - 如果没有命中自定义审批流，则无需进行权限申请即可拥有操作权限。
- 进行自定义审批流程时，DataWorks会根据审批中心设置的审批策略来进行审批流程的流转。

  进行自定义审批策略设置时，您可以按照项目范围来框定管控的数据范围并制定审批人、审批通知方式等信息，操作详情可参见数据服务审批策略。

数据集成任务审批流程

审批中心支持管理员按源端、目的端数据源的组合来定义需要被审批的数据集成任务，包括：在数据集成或数据开发页面保存任务等操作。例如，管理员定义了mysql_1 数据源作为源端、odps_1数据源作为目的端的数据集成任务审批策略，则开发人员在保存相关任务时便会触发审批流程，只有完成权限申请后才能继续执行相关操作。

当权限申请人在**安全中心**申请权限后，后续的流程如下图所示。



- 在数据开发或者数据集成页面保存数据集成任务时，审批中心根据工作空间是否设置任务审批流来决定当前任务保存时是否需要走自定义审批流。
  - 如果命中自定义审批流，后续将进行审批中心自定义的审批流程。
  - 如果没有命中自定义审批流，则无需进行权限申请即可执行保存操作。
- 进行自定义审批流程时，DataWorks会根据审批中心设置的审批策略来进行审批流程的流转。

  进行自定义审批策略设置时，您可以通过指定工作空间，并添加源端、目的端数据源的组合，来定义需要被审批的数据集成任务。同时，还支持您配置审批人、审批通知方式等信息，操作详情可参见数据集成审批策略。

2.2. 审批策略创建与管理

2.2.1. 计算引擎审批策略

您可以自定义MaxCompute引擎的表、资源、函数的审批流程。

背景信息

您可以从**MaxCompute项目**维度或**数据保护伞分级分类**维度定义审批流程适用的数据范围，详情可参见选择配置范围。

使用限制

- 仅空间管理员、被授权AliyunDataWorksFullAccess的子账号有权限进行审批策略的操作。
- 企业版和旗舰版的DataWorks支持使用计算引擎审批策略。

新建审批策略

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上角的图标，选择**全部产品 > 数据治理 > 审批中心**，进入审批中心页面。
5. 进入审批中心页面后，在左侧导航栏选择**审批策略管理 > 计算引擎**，进入计算引擎的审批策略管理页面。

   在计算引擎审批策略的管理页面您可以看到已创建的审批策略列表，也可在此处对其进行编辑、删除等操作。
6. 单击页面右上角的**新建审批策略**，在新建审批策略页面配置审批策略信息。

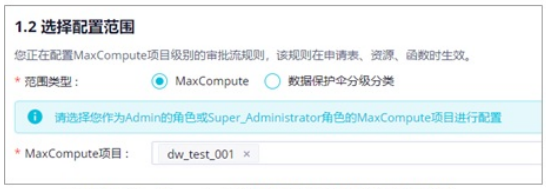
填写基本信息

根据实际审批策略应用场景，填写**审批策略名称**和**审批策略用途**。

选择配置范围

您需要根据实际应用情况，框定本审批策略适用的数据范围。即本审批策略创建成功后，哪些数据的权限审批需要通过本审批策略定义的策略来申请。

对于MaxCompute计算引擎，当前支持通过两种方式来框定审批策略适用的范围：MaxCompute项目方式、数据保护伞分级分类方式。



（a）通过MaxCompute项目空间框定审批策略适用范围

（b）通过数据分类分级框定审批策略适用范围

配置范围时，您需要关注：

- 使用MaxCompute项目方式框定范围时
  - 您需要在**MaxCompute项目空间**配置栏处选择适用的项目空间，后续选中项目空间中的表申请均会使用本策略制定的审批流程。
  - 一个MaxCompute项目只能存在一个MaxCompute审批策略中，否则会提示权限策略冲突。
  - 您可选择当前账号拥有Admin角色或Super_Administrator角色的MaxCompute项目空间，如果您没有在下拉框中找到需要配置的项目空间，可能是因为当前操作账号的权限不对，需要更换为有上述权限的账号进行操作配置。

    > **说明** DataWorks上的管理员角色在底层为role_project_admin角色，非MaxCompute引擎的Admin或者Super_Administrator角色。
    >
    > 如果您不清楚当前账号是什么角色，您可以在DataWorks的数据开发页面执行 `whoami` 命令，获取您的云账号信息，再执行 `show grants for 云账号` ，查看是否有MaxCompute引擎的Admin或者Super_Administrator角色。

- 使用数据保护伞分级分类方式框定范围时
  - 您需要在**选择分级分类**配置栏处选择使用的数据分级分类，后续选中的分级分类下的所有表的权限申请均会使用本策略制定的审批流程。
  - 一个数据分级只能存在于一个数据保护伞分级分类策略中，否则会提示权限策略冲突。
  - 仅支持阿里云主账号、子账号进行配置范围操作，如果使用子账号操作时，子账号需满足如下条件之一：
    - 被授予访问控制RAM策略：AdministratorAccess。
    - 被授予AliyunDataWorksFullAccess和所有MaxCompute项目Project Owner、Super_Administrator角色。

配置通知机制

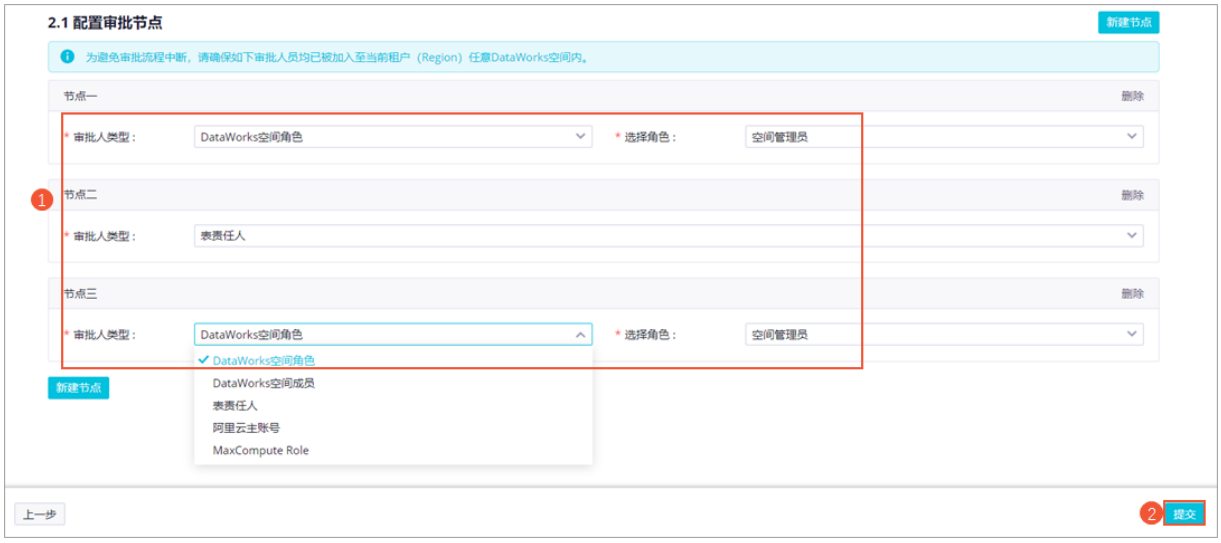
当前支持配置短信、邮件、钉钉机器人这三种通知方式。

选择对应的审批通知方式后，后续有相关的审批流程时，审批任务会通过这里配置的通知方式，通知审批人有待审批的权限申请任务。

> **说明** 此处仅定义审批通知方式，审批人员将在下一步的配置审批节点中定义。
>
> - 为保障审批人能正常通过短信、邮件收到审批任务通知，您还需要将对应角色的用户添加为DataWorks的报警联系人，操作详情可参见查看和设置报警联系人。
> - 为保障审批人能正常通过钉钉收到审批任务通知，您还需要在钉钉机器人管理的配置中，将钉钉机器人的**安全设置**勾选上**自定义关键词**，添加关键词**DataWorks**，并**不要勾选**其他安全设置选项。
>
>   如果您没有添加DataWorks关键词，或者勾选了其他安全设置选项，将无法正常通过钉钉接收到审批通知。

配置审批节点

您可以在**配置审批节点**区域定义每个审批节点的审批人及角色。



配置审批节点时，您需关注：

- 审批流程定义：配置完成后，审批流程会按照已定义的审批节点从上至下进行流转，即上一个节点审批人员审批通过后，下一个节点的审批人员才会收到审批通知并进行审批。
- 审批人员定义：每个节点可以选择不同的审批人类型，审批人类型支持：**DataWorks空间角色**、**DataWorks空间成员**、**表责任人**、**阿里云主账号**、**MaxComputeRole**。

> 说明
>
> - 后续有审批任务时，DataWorks会根据上述步骤配置通知机制中配置的通知方式，给各个审批人发送任务通知，为保障审批人能正常通过短信、邮件收到审批任务通知，您还需要将对应角色的用户添加为DataWorks的报警联系人，操作详情可参见查看和设置报警联系人。
> - 当审批人类型对应的审批角色包含多个人员时，审批通知会发送给所有人员，仅需其中任意一个人员审批完成，审批流程即可往下流转。

设置审批策略优先级

如同时存在MaxCompute项目维度审批策略和数据保护伞分级分类维度审批策略，则可能导致同一数据范围被两个审批流程所覆盖，此时您可以在计算引擎审批策略设置页码选择优先级。

2.2.2. 数据服务审批策略

支持管理员从DataWorks空间级别为每次发布的数据服务API定义发布审批流程。

使用限制

企业版和旗舰版的DataWorks支持使用数据服务审批策略。

新建审批策略

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上角的图标，选择**全部产品 > 数据治理 > 审批中心**，进入审批中心页面。
5. 进入审批中心页面后，在左侧导航栏选择**审批策略管理 > 数据服务**，进入数据服务的审批策略管理页面。

   在数据服务审批策略的管理页面您可以看到已创建的审批策略列表，也可在此处对其进行编辑、删除等操作。
6. 单击页面右上角的**新建审批策略**，在新建审批策略页面配置审批策略信息。

填写基本信息

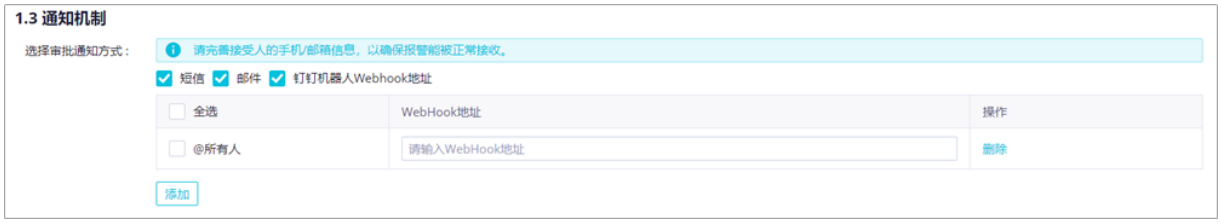
根据实际审批策略应用场景，填写**审批策略名称**和**审批策略用途**。

选择配置范围

您需要根据实际应用情况，框定本审批策略适用的数据范围。即本审批策略创建成功后，哪些数据的权限审批需要通过本审批策略定义的策略来申请。

配置通知机制

当前支持配置短信、邮件、钉钉机器人这三种通知方式。

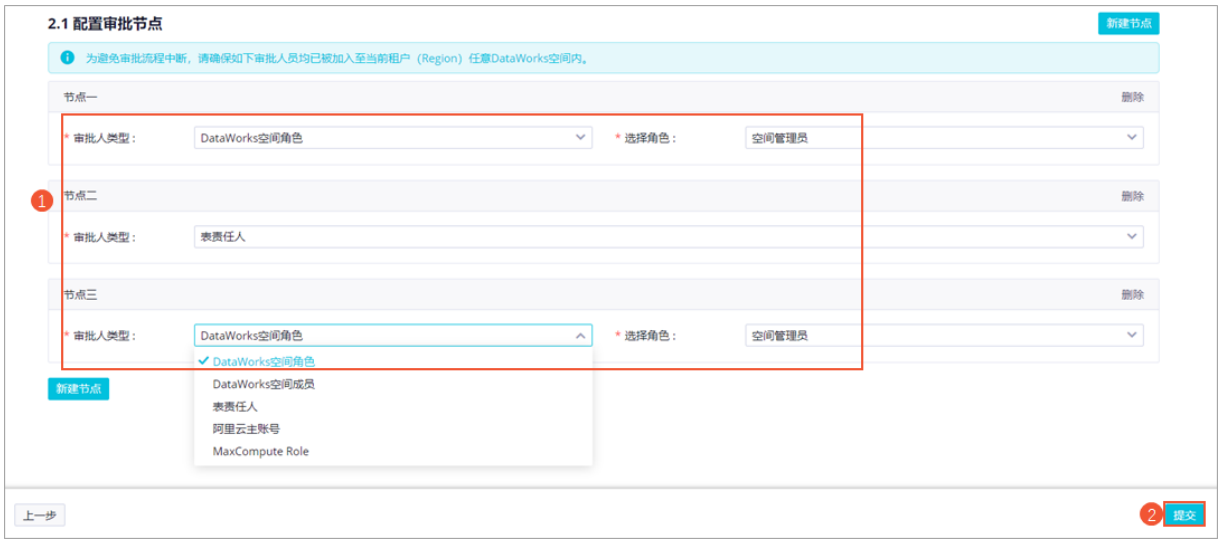


选择对应的审批通知方式后，后续有相关的审批流程时，审批任务会通过这里配置的通知方式，通知审批人有待审批的权限申请任务。

> **说明** 此处仅定义审批通知方式，审批人员将在下一步的配置审批节点中定义。
>
> - 为保障审批人能正常通过短信、邮件收到审批任务通知，您还需要将对应角色的用户添加为DataWorks的报警联系人，操作详情可参见查看和设置报警联系人。
> - 为保障审批人能正常通过钉钉收到审批任务通知，您还需要在钉钉机器人管理的配置中，将钉钉机器人的**安全设置**勾选上**自定义关键词**，添加关键词DataWorks，并**不要勾选**其他安全设置选项。
>
>   如果您没有添加DataWorks关键词，或者勾选了其他安全设置选项，将无法正常通过钉钉接收到审批通知。

配置审批节点

您可以在**配置审批节点**区域定义每个审批节点的审批人及角色。



配置审批节点时，您需关注：

- 审批流程定义：配置完成后，审批流程会按照已定义的审批节点从上至下进行流转，即上一个节点审批人员审批通过后，下一个节点的审批人员才会收到审批通知并进行审批。
- 审批人员定义：每个节点可以选择不同的审批人类型，审批人类型支持：**DataWorks空间角色**、**DataWorks空间成员**、**表责任人**、**阿里云主账号**、**MaxComputeRole**。

> 说明
>
> - 后续有审批任务时，DataWorks会根据上述步骤配置通知机制中配置的通知方式，给各个审批人发送任务通知，为保障审批人能正常通过短信、邮件收到审批任务通知，您还需要将对应角色的用户添加为DataWorks的报警联系人，操作详情可参见查看和设置报警联系人。
> - 当审批人类型对应的审批角色包含多个人员时，审批通知会发送给所有人员，仅需其中任意一个人员审批完成，审批流程即可往下流转。

2.2.3. 数据集成审批策略

数据集成审批策略支持管理员从DataWorks空间级别为数据集成任务的保存操作定义审批策略。本文为您介绍如何创建数据集成审批策略。

背景信息

支持管理员按源端、目的端数据源的组合来定义需要被审批的数据集成任务，包括：在数据集成或数据开发页面保存任务等操作。例如，管理员定义了mysql_1 数据源作为源端、odps_1数据源作为目的端的数据集成任务审批策略，则开发人员在保存相关任务时便会触发审批流程，只有完成权限申请后才能继续执行相关操作。

使用限制

- 仅企业版和旗舰版的DataWorks支持使用数据集成审批策略功能。
- 仅主账号和被授予AliyunDataWorksFullAccess权限的RAM用户可以选择所有工作空间作为管控工作空间。即定义的审批策略会在所有的管控工作空间内生效。
- 工作空间管理员仅能选择自己所在的工作空间作为管控工作空间。即定义的审批策略仅在自己所在的工作空间内生效。

新建审批策略

1. 进入数据集成审批策略管理页面。
   i. 登录DataWorks控制台。
   ii. 在左侧导航栏，单击**工作空间列表**。
   iii. 选择工作空间所在地域后，单击相应工作空间后的**数据开发**。
   iv. 单击左上角的图标，选择**全部产品 > 数据治理 > 审批中心**，进入审批中心页面。
   v. 进入审批中心页面后，在左侧导航栏选择**审批策略管理 > 数据集成**，进入数据集成的审批策略管理页面。
2. 单击页面右上角的**新建审批策略**，在新建审批策略页面配置审批策略信息。

填写基本信息

根据实际审批策略应用场景，填写**审批策略名称**和**审批策略用途**。

选择配置范围

您需要根据实际应用情况，确定本审批策略适用的范围。支持您按源端、目的端数据源的组合来定义需要被审批的数据集成任务。当审批策略配置完成并生效后，开发人员在保存相关任务时便会触发该审批流程，只有完成权限申请后才能继续执行相关操作。

选择配置范围时，您需要关注：

- 支持选择多个工作空间。

  > **说明**
  > - 仅主账号和被授予AliyunDataWorksFullAccess权限的RAM用户可以选择所有工作空间作为管控工作空间。即定义的审批策略会在所有的管控工作空间内生效。
  > - 工作空间管理员仅能选择自己所在的工作空间作为管控工作空间。即定义的审批策略仅在自己所在的工作空间内生效。

- 对每个工作空间，支持您通过**新增配置**添加多个需要审批的源端数据源或目的端数据源。

配置通知机制

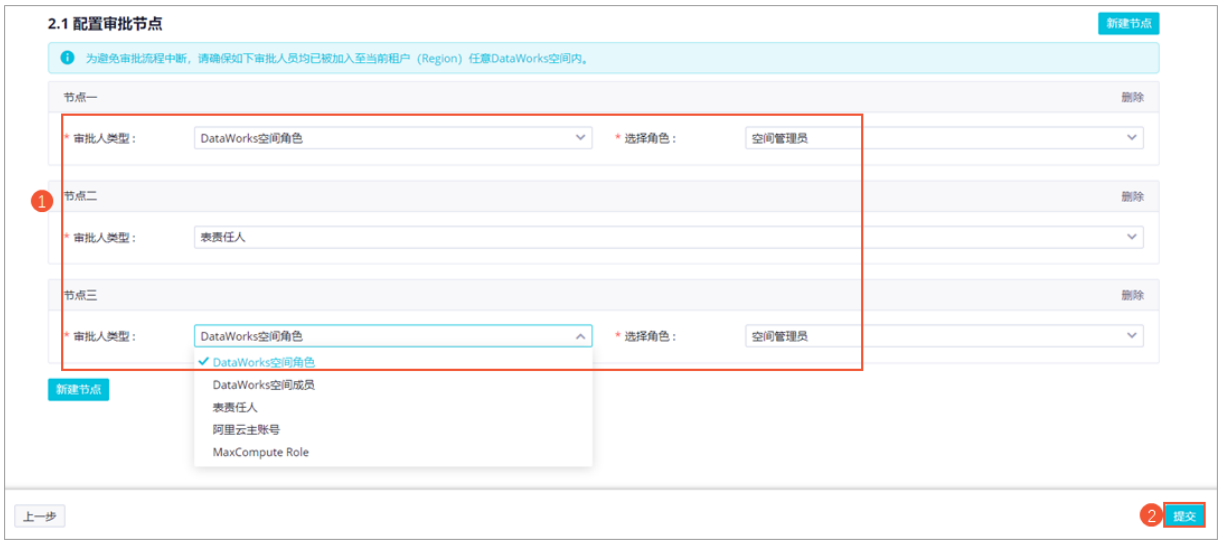
当前支持配置短信、邮件、钉钉机器人这三种通知方式。

选择对应的审批通知方式后，后续有相关的审批流程时，审批任务会通过这里配置的通知方式，通知审批人有待审批的权限申请任务。

> **说明** 此处仅定义审批通知方式，审批人员将在下一步的配置审批节点中定义。
>
> - 为保障审批人能正常通过短信、邮件收到审批任务通知，您还需要将对应角色的用户添加为DataWorks的报警联系人，操作详情可参见查看和设置报警联系人。
> - 为保障审批人能正常通过钉钉收到审批任务通知，您还需要在钉钉机器人管理的配置中，将钉钉机器人的**安全设置**勾选上**自定义关键词**，添加关键词DataWorks，并**不要勾选**其他安全设置选项。
>
>   如果您没有添加DataWorks关键词，或者勾选了其他安全设置选项，将无法正常通过钉钉接收到审批通知。

配置审批节点

您可以在**配置审批节点**区域定义每个审批节点的审批人及角色。



配置审批节点时，您需关注：

- 审批流程定义：配置完成后，审批流程会按照已定义的审批节点从上至下进行流转，即上一个节点审批人员审批通过后，下一个节点的审批人员才会收到审批通知并进行审批。
- 审批人员定义：每个节点可以选择不同的审批人类型，审批人类型支持：**DataWorks空间角色**、**DataWorks空间成员**、**表责任人**、**阿里云主账号**、**MaxComputeRole**。

> 说明
>
> - 后续有审批任务时，DataWorks会根据上述步骤配置通知机制中配置的通知方式，给各个审批人发送任务通知，为保障审批人能正常通过短信、邮件收到审批任务通知，您还需要将对应角色的用户添加为DataWorks的报警联系人，操作详情可参见查看和设置报警联系人。
> - 当审批人类型对应的审批角色包含多个人员时，审批通知会发送给所有人员，仅需其中任意一个人员审批完成，审批流程即可往下流转。

启用或停止审批策略

审批策略创建完成后，请单击对应审批策略右侧的**生效**或**停止**按钮。停用审批策略后，在数据开发、数据集成页面的保存操作将不再需要审批。此外，还支持您查看、编辑和删除审批策略。

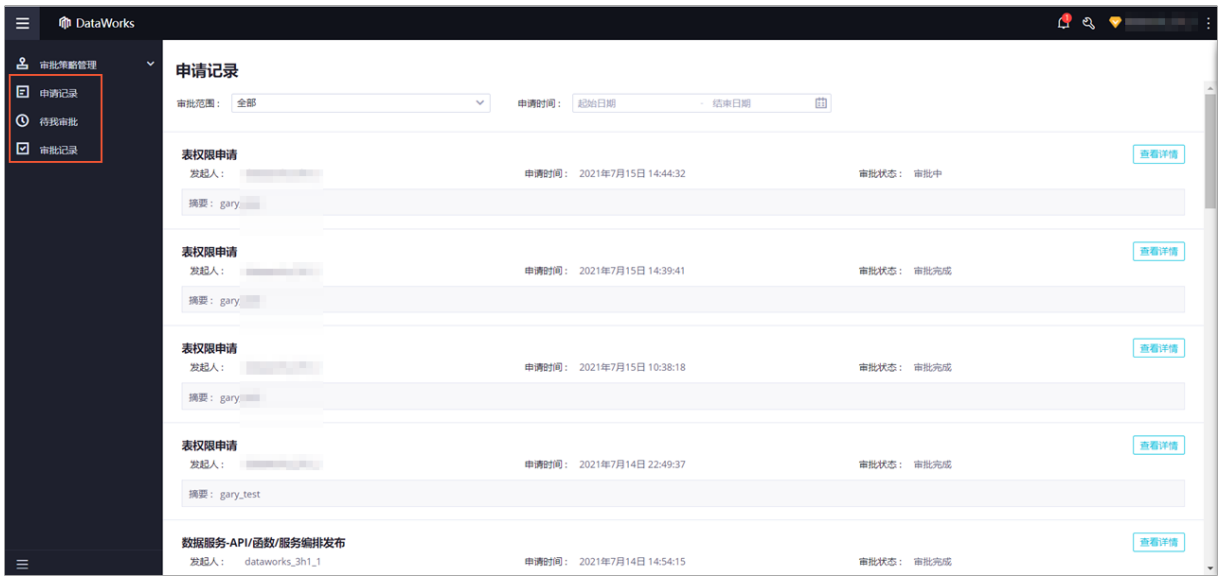
后续步骤

您可以在审批处理与查看查看当前账号待审批的所有申请并执行审批操作。您还可以在此界面查看历史审批记录。

2.3. 审批处理与查看

在审批中心您可以查看当前账号的所有申请记录、审批记录，以及处理待处理的申请记录。

页面概览



- 您可以在**申请记录**页面中查看当前账号下所有的申请记录。
- 您可以在**待我审批**页面中查看所有待审批的申请记录。
- 您可以在**审批记录**页面中查看当前账号下所有已审批完成的申请记录。

3.数据质量

3.1. 数据质量概述

数据质量帮助您第一时间感知源端数据的变更与ETL（Extract Transformation Load）中产生的脏数据，自动拦截问题任务，有效阻断脏数据向下游蔓延。避免任务产出不符合预期的问题数据，影响正常使用和业务决策。同时也能显著降低问题处理的时间成本、避免任务重新运行带来的资源费用浪费。

费用说明

数据质量规则运行产生的费用由两部分组成：

- DataWorks相关收费

  根据数据质量规则实例数进行按量收费，详情请参见：数据质量计费说明。

- 非DataWorks收费

  数据质量规则校验会产生校验SQL并下推到引擎执行，数据质量规则运行将会产生引擎费用，各引擎计费细则请参考各引擎计费文档。例如，假设您使用MaxCompute引擎按量付费模式时，数据质量规则校验将会产生MaxCompute引擎费用，此费用由MaxCompute引擎侧收取，不在DataWorks账单中体现。

功能介绍

数据质量支持对常见大数据存储（MaxCompute、E-MapReduce Hive、Hologres等）和实时数据流（Kafka、DataHub等数据通道）进行质量校验。从完整性、准确性、有效性、一致性、唯一性和及时性等多个维度，配置质量监控规则。并可以将质量监控规则与调度节点进行关联，当任务运行完成后便会触发质量规则校验，帮助您第一时间感知问题数据，按需设置规则的强弱来控制任务是否失败退出，从而避免脏数据影响扩大，有效降低数据恢复处理的时间成本和费用成本。

数据质量各模块功能介绍如下：

名称	描述
概览	数据质量概览页面为您展示数据质量报警与阻塞情况。包括： • 当前登录账号及当前工作空间下离线数据和流式数据的报警和阻塞情况。 • 当前工作空间下各数据源中任务的报警与阻塞趋势图。
我的订阅	我的订阅页面为您展示当前登录账号下通过短信，邮件接收报警的数据质量校验规则。此外，数据质量还支持通过钉钉群机器人、企业微信机器人和飞书群机器人等方式发送报警信息。
规则配置	数据质量支持按表配置或按模板配置质量监控规则，详情请参见：按表配置监控规则、按模板配置监控规则。
任务查询	在任务查询页面您可以通过表或节点搜索表历史校验记录及校验详情。
去噪管理	去噪管理功能支持对当前工作空间某一时间内，数据质量规则校验异常的数据不触发报警，且不阻塞任务运行。
报告模板管理	报告模板管理页面支持您创建报告模板，添加规则配置和规则运行的各项指标，根据设置的统计周期、发送时间和订阅信息，定时生成并发送报告。

名称	描述
规则模板库	数据质量支持自建规则模板库，对通用的自定义监控规则进行统一管理，形成自建的规则模板库，帮助您提升规则配置的效率。

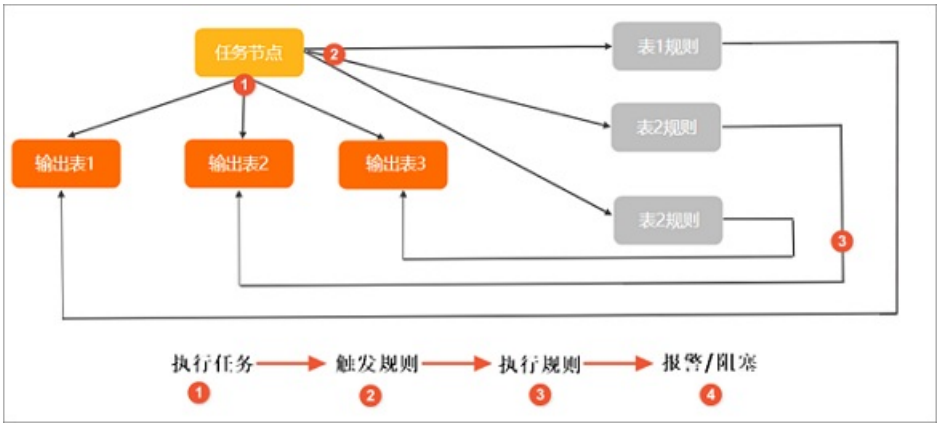
注意事项

- EMR、Hologres、analyticDB for PostgreSQL、CDH在进行数据质量规则配置前，需要先进行元数据采集，详情请参见元数据采集。
- EMR、Hologres、analyticDB for PostgreSQL、CDH配置表数据质量规则后，产出表数据的调度节点需要使用网络已经连通的独享调度资源组执行才可以正常触发数据质量规则校验。
- 一个表可以配置多个数据质量规则。

使用场景

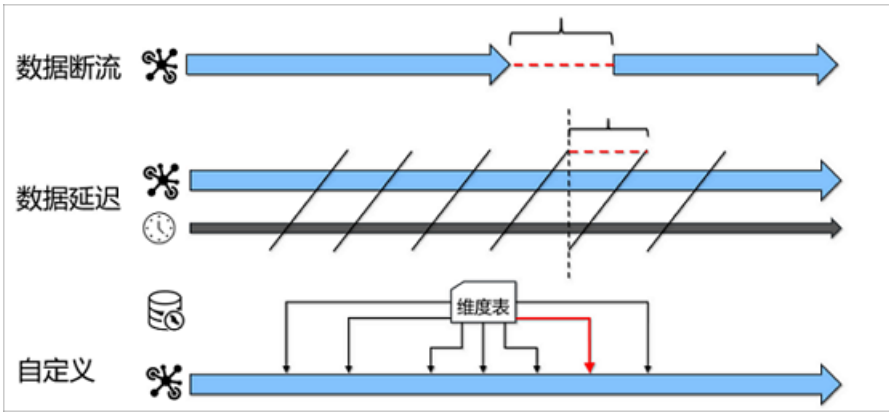
- 离线检验场景

  在离线数据校验场景下，数据质量通过表配置的分区表达式来匹配节点每天产出的表分区，数据质量规则关联产出该表数据的调度节点，当任务运行完成便会触发质量规则校验，您可以设置规则的强弱来控制节点是否失败退出，从而避免脏数据影响扩大，并支持通过报警配置第一时间接收报警信息并处理。



- 流数据校验场景

  在流数据校验场景下，支持对实时数据流（Kafka、DataHub等数据通道）的数据断流、数据延迟等场景进行监控，您可以自定义Flink SQL、维表join、多表join以及窗口函数等，并设置橙色、红色告警等级和告警频次，最大限度地减少冗余报警。



配置规则

- 创建规则：数据质量支持您按表创建数据质量规则，同时，也支持您通过内置规则模板来快速为一批表批量创建数据质量规则。详情请参见：按表配置监控规则、按模板配置监控规则。
- 订阅规则：规则创建完成后，您可以通过规则订阅的方式接收数据质量规则校验报警信息，支持**邮件通知**、**邮件和短信通知**、**钉钉群机器人**和**钉钉群机器人@ALL**、**飞书群机器人**和**企业微信机器人**等方式进行报警。

触发规则校验

在**运维中心**中，当表关联的调度节点运行（执行节点代码逻辑）完成后，将会触发数据质量校验（将会产生一条校验sql在底层执行）。DataWorks平台将会根据数据质量规则强弱和数据质量规则校验结果决定任务是否由于质量规则校验失败退出，并阻塞下游节点执行，防止脏数据影响范围进一步扩大。

查看校验结果

您可以通过运维中心节点运行日志和数据质量任务查询页面查看数据质量校验结果。

- 通过运维中心节点运行日志查看
  i. 查看**实例状态**。当实例状态为质量监控校验失败时，可能是代码运行成功但节点产出的表数据不符合预期，数据质量强规则校验未通过导致任务失败退出并阻塞下游实例运行。

  ii. 打开实例**运行日志**中的**DQC日志**，查看数据质量校验结果。详情请参见查看周期实例。

- 通过数据质量任务查询界面查看。

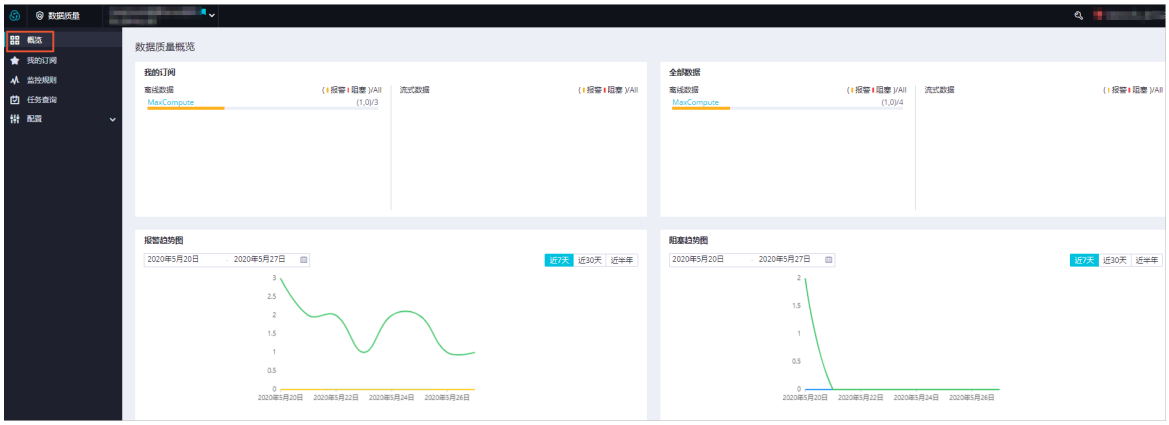
  在任务查询界面通过表或节点搜索校验记录及校验详情。详情请参见：查看监控任务。

3.2. 进入数据质量概览

数据质量概览页面为您展示订阅数据的报警以及任务的阻塞情况。

操作步骤

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上角的图标，选择**全部产品 > 数据质量**，默认进入**概览**页面。



模块	描述
我的订阅	为您展示当前登录用户名下订阅的表，所产生的离线数据和流式数据的报警、阻塞数量，以及订阅的总数。
全部数据	为您展示当前工作空间下离线数据和流式数据的全部数据情况。
报警趋势图	为您展示**近7天**、**近30天**和**近半年**EMR、MaxCompute和DataHub数据源的任务报警趋势图，单位：次。
阻塞趋势图	为您展示**近7天**、**近30天**和**近半年**EMR、MaxCompute和DataHub数据源的任务阻塞情况趋势图，单位：次。

3.3. 查看我的订阅

我的订阅页面为您展示当前账号订阅的EMR（E-MapReduce）、Hologres、AnalyticDB for PostgreSQL、MaxCompute和DataHub数据源的任务。

操作步骤

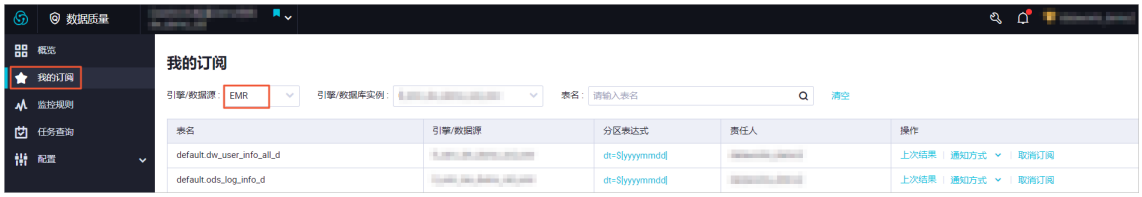
1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上角的图标，选择**全部产品 > 数据治理 > 数据质量**。
5. 在左侧导航栏，单击**我的订阅**。

   数据质量支持EMR、Hologres、AnalyticDB for PostgreSQL、MaxCompute和DataHub等类型的数据源，您可以在**我的订阅**页面选择相应的数据源，查找自己订阅的任务。

   - EMR、Hologres、AnalyticDB for PostgreSQL和MaxCompute数据源

     在**我的订阅**页面左上角的**引擎/数据源**列表中，选择**EMR**，并选择相应的**引擎/数据库实例**，为您显示已订阅的EMR数据表。

您可以通过同样的方式，查看已订阅的Hologres、AnalyticDB for PostgreSQL和MaxCompute数据表。



- 单击相应表名后的**分区表达式**，跳转至**规则配置**页面，详情请参见配置监控规则。
- 单击相应表名后的**上次结果**，跳转至**任务查询**页面，详情请参见查询监控任务。
- 单击相应表名后的**通知方式**，修改规则报警的通知方式。目前支持**邮件通知**，**邮件和短信通知**。

  > **说明** 钉钉群机器人、钉钉群机器人@ALL、企业微信机器人和飞书群机器人的通知方式，以及通过DataWorks配置钉钉群告警的详情请参见自定义规则。

- 单击相应表名后的**取消订阅**，删除该表的订阅信息。

- DataHub数据源

  在**我的订阅**页面左上角的数据源列表中，选择**DataHub**，为您显示已订阅的DataHub数据源。

- 单击相应Topic后的**报警**，进入**报警列表**页面，查看规则报警的具体信息。
- 单击相应Topic后的**通知方式**，修改规则报警的通知方式。
- 单击相应Topic后的**取消订阅**，取消已订阅的Topic。

3.4. 规则配置

3.4.1. 按表配置监控规则

数据质量支持配置EMR（E-MapReduce）、Hologres、AnalyticDB for PostgreSQL、MaxCompute和DataHub数据源的监控规则。本文以配置MaxCompute监控规则为例，为您介绍如何配置监控任务。

前提条件

EMR、Hologres、analyticDB for PostgreSQL、CDH在进行数据质量规则配置前，需要先进行元数据采集，详情请参见采集元数据。

使用限制

- 自动落标规则暂不支持使用。
- EMR、Hologres、analyticDB for PostgreSQL、CDH配置表数据质量规则后，产出表数据的调度节点需要使用网络已连通的独享调度资源组进行调度，才可正常触发数据质量规则校验。

进入监控规则

1. 登录DataWorks控制台。

2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上角的图标，选择**全部产品 > 数据治理 > 数据质量**。
5. 在左侧导航栏，单击**规则配置 > 按表配置**。
6. 选择**引擎/数据源**为**MaxCompute**，并选择相应的**引擎/数据库实例**。

数据质量支持EMR、Hologres、AnalyticDB for PostgreSQL、MaxCompute和DataHub数据源：

- 选择EMR、Hologres、AnalyticDB for PostgreSQL或MaxCompute数据源，显示当前数据源下所有的表。
- 选择DataHub数据源，显示当前数据源下所有的Topic，DataHub数据源配置请参见配置DataHub数据源。

7. 单击相应表后的**配置监控规则**。

   数据质量监控规则包括**模板规则**和**自定义规则**。

> 注意 配置模板规则前，请首先配置分区表达式，详情请参见配置分区表达式。

创建模板规则

1. 单击相应表名后的**配置监控规则**，进入该表的**监控规则**页面。
2. 单击**创建规则**，默认显示**模板规则**对话框。

   您可以通过**添加监控规则**和**快捷添加**两种方式创建**模板规则**：

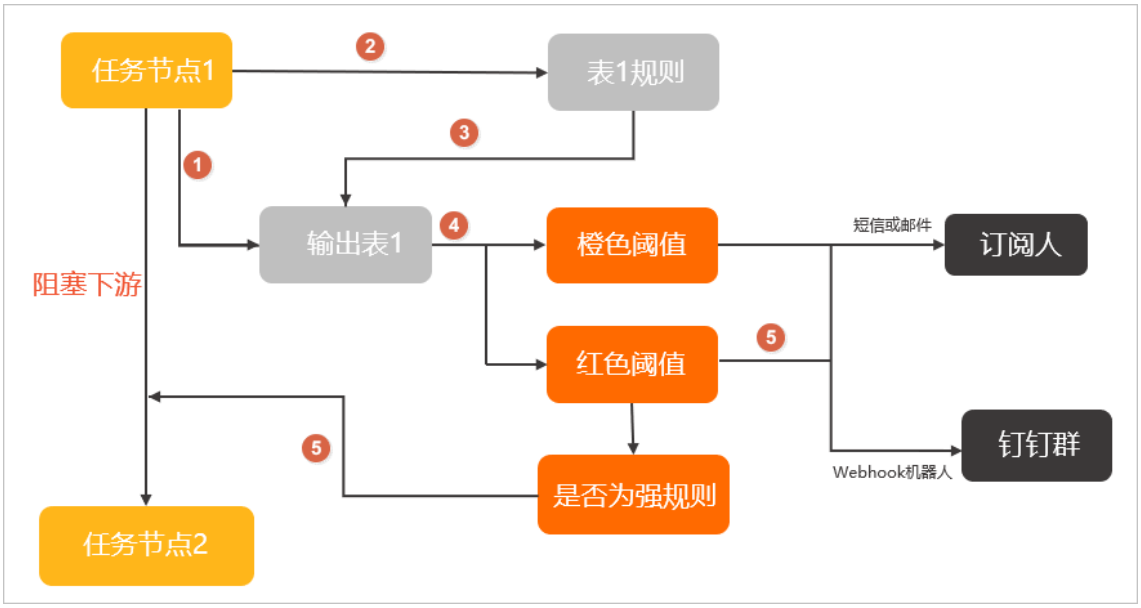
   - **添加监控规则**

     单击**添加监控规则**，下表以**内置模板**为例，为您详细介绍各项参数的配置。

参数	描述
规则名称	请输入规则名称。
强弱	设置强规则或弱规则： ■ 如果设置强规则，红色异常报警并阻塞下游任务节点，橙色异常报警不阻塞。 ■ 如果设置弱规则，红色异常报警不阻塞下游任务节点，橙色异常不报警不阻塞。
动态阈值	根据自身需求，选择是否开启动态阈值。 注意 您需要购买DataWorks企业版及以上版本，才可以使用动态阈值功能。

参数	描述
规则来源	包括**内置模板**和**规则模板库**。 如果您选中**规则模板库**，需要选择相应的规则模板库，详情请参见新建、操作和应用规则模板。 **注意** 您需要购买DataWorks企业版及以上版本，才可以选择**规则模板库**。
规则字段	包括表级规则和字段级规则，字段级规则包括数值型和非数值型。
规则模板	目前共有43种规则，不支持的规则模板将不能被选择。详情请参见内置模板规则。 **说明** 平均值、汇总值、最小值和最大值仅对数值型字段生效。
比较方式	包括**绝对值**、**上升**和**下降**。
波动值比较	■ 计算波动率 您可以根据波动率计算公式（ `波动率=（样本-基准值）/基准值` ）计算结果： ■ 样本 当天采集的具体的样本的值。例如对于SQL任务表行数，1天波动检测，则样本是当天分区的表行数。 ■ 基准值 历史样本的对比值： ■ 如果规则是SQL任务表行数，1天波动检测，则基准值是前一天分区产生的表行数。 ■ 如果规则是SQL任务表行数，7天平均值波动检测，则基准值是前7天的表行数据的平均值。 ■ 计算方差波动 （当次样本-历史N天平均值）/标准差，仅BIGINT和DOUBLE等数值类型可以使用方差。 您可以设置**橙色阈值**和**红色阈值**，对不同严重程度的问题进行监控： ■ 如果校验值的绝对值小于或等于橙色阈值，则返回**正常**。 ■ 如果校验值的绝对值不满足第1种情况，且小于或等于红色阈值，则返回**橙色报警**。 ■ 如果校验值不满足第2种情况，则返回**红色报警**。
描述	对配置的监控规则进行简单描述。

下图为报警与阻塞的实现逻辑。



- **快捷添加**

  单击**快捷添加**，配置各项参数。

参数	描述
规则名称	请输入规则名称。
监控字段	包括表级规则和字段级规则，字段级规则包括数值类型和非数值类型。
快捷规则	包括**表行数大于0**和**表行数动态阈值**。 注意 您需要购买DataWorks企业版及以上版本，才可以选择**表行数动态阈值**。

3. 单击**批量添加**。

创建自定义规则

如果模板规则不能满足您对分区表达式中数据质量的监控需求，您还可以通过创建**自定义规则**来满足个性化的监控需求：

1. 单击相应表名后的**配置监控规则**，进入该表的**监控规则**页面。
2. 单击**创建规则**，默认显示**模板规则**对话框。
3. 单击**自定义规则**。

   您可以通过**添加监控规则**和**快捷添加**两种方式创建**自定义规则**：

   - **添加监控规则**

     添加监控规则时，**规则字段**支持**表级规则**、**自定义SQL**和**字段级规则**：

     - **表级规则**和**字段级规则**

参数	描述
规则名称	请输入规则名称。
强弱	设置强规则或弱规则： ■ 如果设置强规则，红色异常报警并阻塞下游任务节点，橙色异常报警不阻塞。 ■ 如果设置弱规则，红色异常报警不阻塞下游任务节点，橙色异常不报警不阻塞。
规则字段	此处选择**表级规则**。表级自定义规则，支持根据业务属性自定义where过滤条件。
采样方式	支持**count**和**count/table_count**两种方式。 **说明** 这里的count/table_count指的是根据配置的过滤条件过滤后的结果条数与当前分区的表总行数的比值。

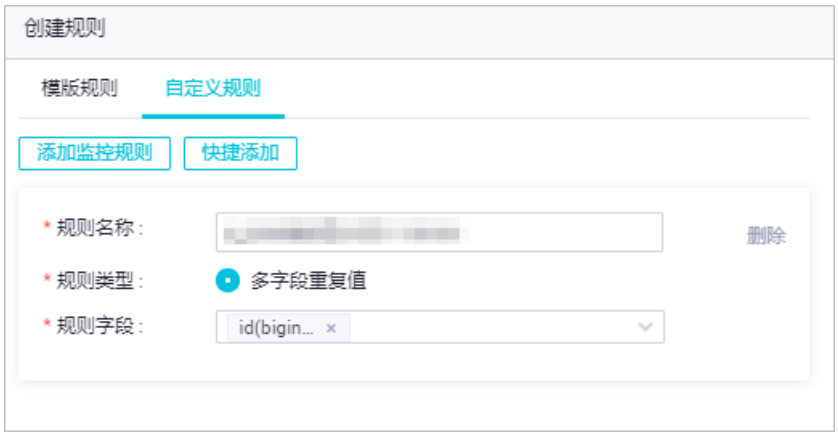
参数	描述
过滤条件	输入过滤条件。例如，您需要查询业务日期下表的分区，可以设置过滤条件为`pt=$[yyyymmdd-1]`。
校检类型	支持**数值型**、**波动率型**和**动态阈值型**。 **说明** 您需要购买DataWorks企业版及以上版本，才可以选择**动态阈值型**。
比较方式	选择的**校检类型**不同，**比较方式**也不同： ■ 如果选择**校检类型**为**数值型**，则比较方式包括**大于**、**大于等于**、**等于**、**不等于**、**小于**和**小于等于**。 ■ 如果选择**校检类型**为**波动率型**，则比较方式包括**绝对值**、**上升**和**下降**。
校检方式	选择的**校检类型**不同，**校检方式**也不同： ■ 如果选择**校检类型**为**数值型**，则校检方式仅支持**与固定值比较**。 ■ 如果选择**校检类型**为**波动率型**，则校检方式包括**7天平均值波动**、**30天平均值波动**、**1天周期比较**、**7天周期比较**、**30天周期比较**、**7天方差波动**、**30天方差波动**、**1,7,30天波动检测**和**上一周期比较**。
期望值	如果选择**校检类型**为**数值型**，需要设置期望值。
波动值比较	如果选择**校检类型**为**波动率型**，则需要设置波动值的橙色阈值和红色阈值。您可以通过拖动进度条来设置，也可以直接输入阈值。
描述	对创建的自定义规则进行描述。

- **自定义SQL**

参数	描述
规则名称	请输入规则名称。
强弱	设置强规则或弱规则： ■ 如果设置强规则，红色异常报警并阻塞下游任务节点，橙色异常报警不阻塞。 ■ 如果设置弱规则，红色异常报警不阻塞下游任务节点，橙色异常不报警不阻塞。
规则字段	此处选择**自定义SQL**，支持自定义SQL逻辑（单行单列输出）。
采样方式	仅支持**自定义SQL**。
Set Flag	输入SQL的前置set语句。

参数	描述
自定义SQL	输入完整的SQL语句，查询结果只能返回一行一列的值。 自定义SQL中，请使用中括号的形式匹配表的分区表达式。示例如下： `select count(*) from table_name where ds=$[yyyymmdd];` **说明** ■ 此处table_name代指当前正在配置监控规则的表名，您需要在实际配置中将其替换为当前实际操作的表名。 ■ 配置分区表达式详情请参见配置分区表达式 ■ 基于自定义SQL创建的数据质量规则校验的表分区由当前SQL条件决定，与上述步骤中的分区表达式配置无关。
校检类型	支持**数值型**和**波动率型**两种类型。
比较方式	选择的**校检类型**不同，**比较方式**也不同： ■ 如果选择**校检类型**为**数值型**，则比较方式包括**大于**、**大于等于**、**等于**、**不等于**、**小于**和**小于等于**。 ■ 如果选择**校检类型**为**波动率型**，则比较方式包括**绝对值**、**上升**和**下降**。
校检方式	选择的**校检类型**不同，**校检方式**也不同： ■ 如果选择**校检类型**为**数值型**，则校检方式仅支持**与固定值比较**。 ■ 如果选择**校检类型**为**波动率型**，则校检方式包括**7天平均值波动**、**30天平均值波动**、**1天周期比较**、**7天周期比较**、**30天周期比较**、**7天方差波动**、**30天方差波动**、**1,7,30天波动检测**和**上一周期比较**。
期望值	如果选择**校检类型**为**数值型**，需要设置期望值。
波动值比较	如果选择**校检类型**为**波动率型**，则需要设置波动值的橙色阈值和红色阈值。您可以通过拖动进度条来设置，也可以直接输入阈值。
描述	对创建的自定义规则进行描述。

- **快捷添加**



参数	描述
规则名称	请输入规则名称。
规则类型	仅支持**多字段重复值**。
规则字段	设置监控字段。

4. 单击**批量添加**。

后续步骤

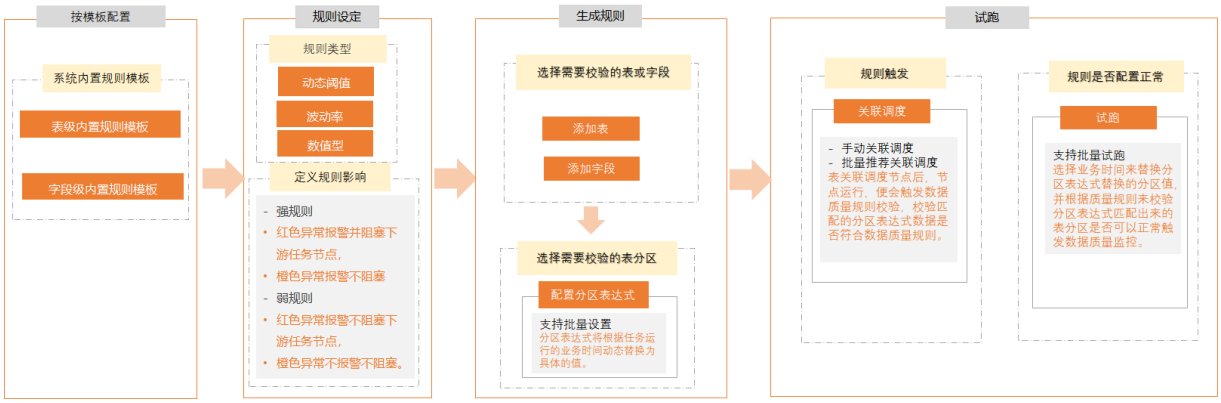
如果您需要在指定业务时间内，不符合质量校验规则的数据不阻塞任务运行，可以用去噪管理功能，详情请参见：去噪管理。

3.4.2. 按模板配置监控规则

数据质量为您提供数十种预设表级别、字段级别的监控模板。本文为您介绍如何按模板配置监控规则。

背景信息

内置模板分为表级和字段级的模板规则，支持您通过内置规则模板来快速为一批表批量创建数据质量规则。



使用限制

按模板配置目前支持配置EMR（E-MapReduce）、Hologres、AnalyticDB for PostgreSQL、MaxCompute数据源的监控规则。

进入按模板配置监控规则页面

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上角的图标，选择**全部产品 > 数据治理 > 数据质量**。
5. 在左侧导航栏选择**规则配置 > 按模板配置**，进入按模板配置页面。

   数据质量提供系统内置的**表级规则模板**和**字段级规则模板**，您可以单击对应模板后的**配置监控规则**来批量对表或字段配置监控规则。

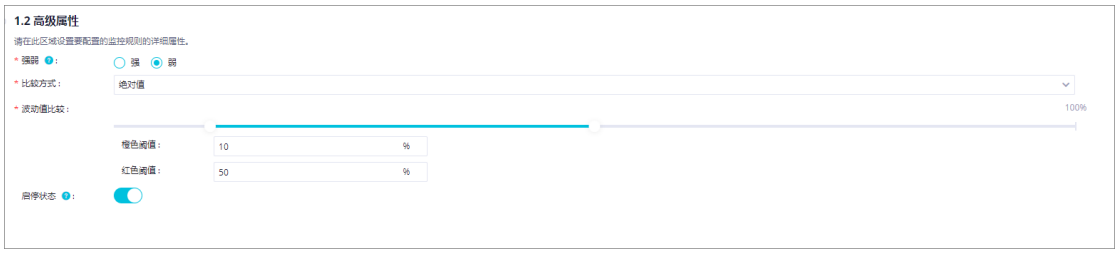
配置监控规则

1. 选择需要进行规则配置的模板，单击操作列的**配置监控规则**，进入该模板的**批量新增监控规则**页面。
2. 配置监控规则的基本属性。

i. 配置监控规则的基本属性。

参数	描述
引擎/数据源	选择后续需要应用此监控规则的表所属的计算引擎或数据源类型。 **说明** 按模板配置目前支持配置EMR（E-MapReduce）、Hologres、AnalyticDB for PostgreSQL、MaxCompute数据源的监控规则。
规则来源	默认来源为**内置模板**。
规则模板	此处展示配置选择的规则模板名称。详情请参见内置模板规则。 **说明** 平均值、汇总值、最小值和最大值仅对数值型字段生效。
规则名称	规则名称系统会自动生成，您可以按需调整名称后缀。
描述	对配置的监控规则进行简单描述。

ii. 配置监控规则的详细属性。



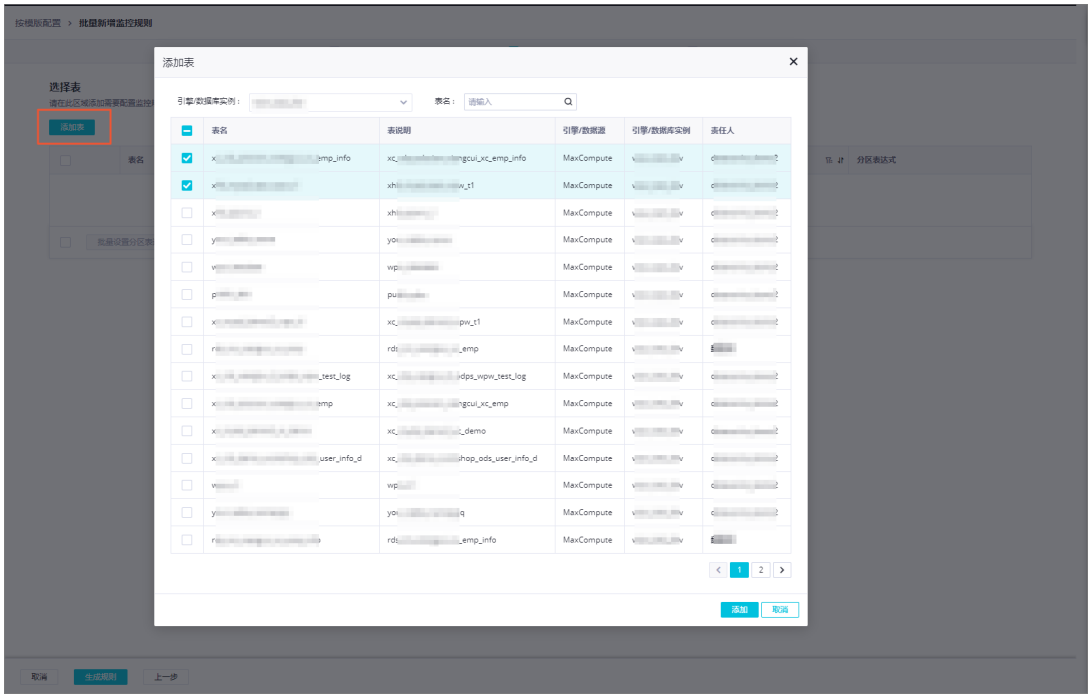
参数	描述
强弱	设置强规则或弱规则： ■ 如果设置强规则，红色异常报警并阻塞下游任务节点，橙色异常报警不阻塞。 ■ 如果设置弱规则，红色异常报警不阻塞下游任务节点，橙色异常不报警不阻塞。
比较方式	■ 当模板的规则类型为数值型时，比较方式包括**大于**、**大于等于**、**等于**、**不等于**、**小于**、**小于等于**。 ■ 当模板的规则类型为波动率型时，比较方式包括**绝对值**、**上升**和**下降**。
期望值	当模板的规则类型为数值型时，您需要填写**期望值**。当触发规则校验时将数据探查结果与期望值进行比较。如果发现数据异常，便会触发报警或阻塞。
波动值比较	当模板的规则类型为波动率型时，您可以设置**橙色阈值**和**红色阈值**，对数据探查结果的波动率与指定时间内数据采样结果的波动率进行比较。支持上升范围、下降范围或波动范围（绝对值）的比较。 例如，假设规则为强规则，并且规则橙色阈值为5%，红色阈值为10%。 ■ 当波动率大于5%且小于等于10%时，将触发橙色报警，任务不会被阻塞，并且发送报警信息。 ■ 当波动率大于10%时，将触发红色报警，任务将被阻塞，并且发送报警信息。
启停状态	单击开关按钮**启用**或**停用**规则，用于控制该规则是否在生产环境中运行。 **注意** 状态设置为**停用**时，规则将无法触发试跑，并且不会被关联的调度任务触发运行。

3. 单击**下一步**，进入生成规则页面。

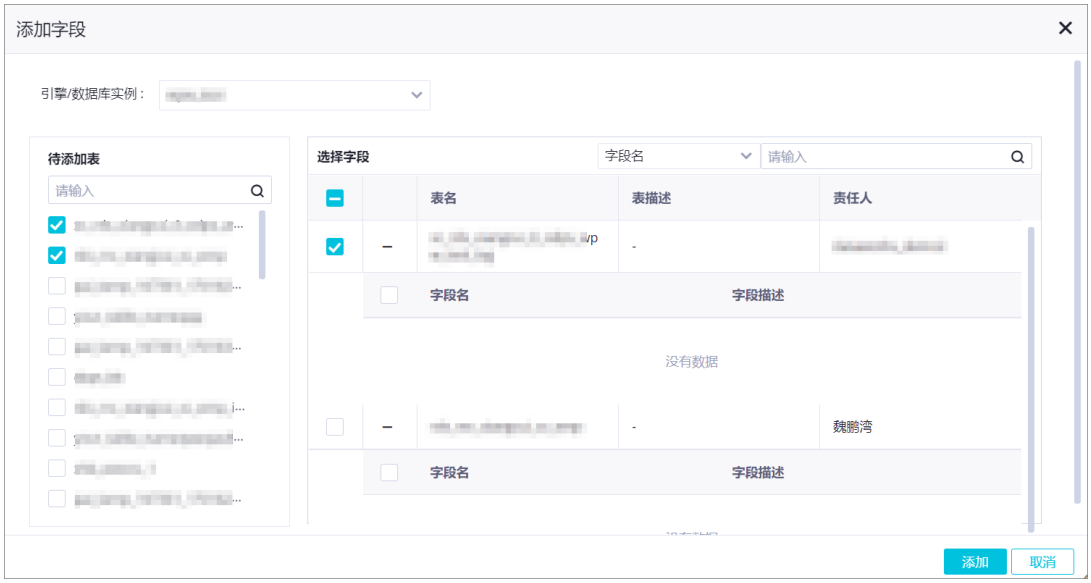
   根据您选择的表级规则模板和字段级规则模板，批量添加需要进行该规则校验的表或字段，添加后，请为分区表配置分区表达式。分区表达式用于确定校验数据的采样范围。对于非分区表，系统会默认配置为NOT APARTITIONTABLE。

i. 添加表或字段。

- 单击**添加表**，在弹出的对话框中，选择目标**引擎/数据库实例**，列表中为您展示当前引擎/数据库中的所有表信息，您还可以输入目标**表名**对结果进行过滤。选中需要配置监控规则的表**添加**至列表中。

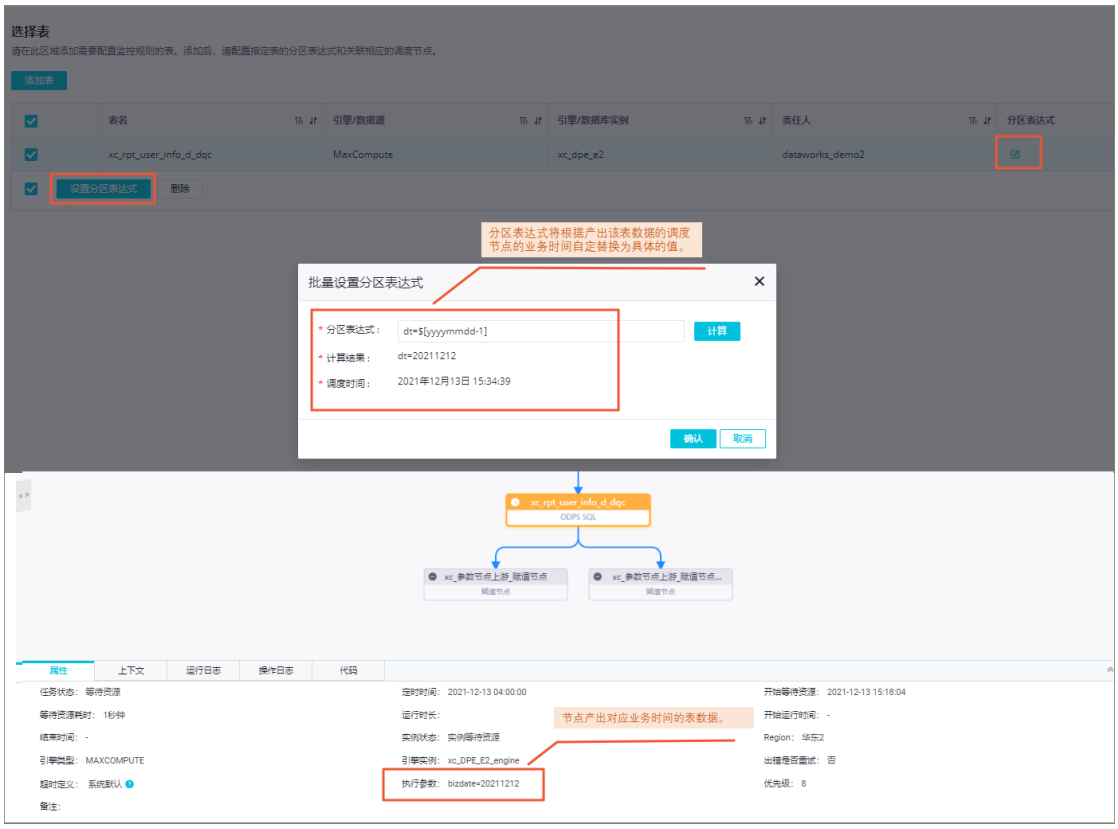


- 单击**添加字段**，在弹出的对话框中，选择目标**引擎/数据库实例**，**待添加表**区域为您展示当前引擎/数据库中的所有表信息，选中要配置监控规则的字段所在的表后，**选择字段**区域为您展示已添加表中的所有字段信息，支持您根据**字段名**和**字段描述**对结果进行过滤。选中需要配置监控规则的字段后**添加**至生成规则页面的列表中。



ii. 配置分区表达式。

单击目标表名右侧的 按钮，在弹出的**批量设置分区表达式**页面输入分区表达式，单击**确认**。数据质量将通过表配置的分区表达式来匹配调度节点每天产出的表分区。如果您需要批量为表配置分区表达式，则可以单击**设置分区表达式**按钮为选中的表批量添加分区表达式。

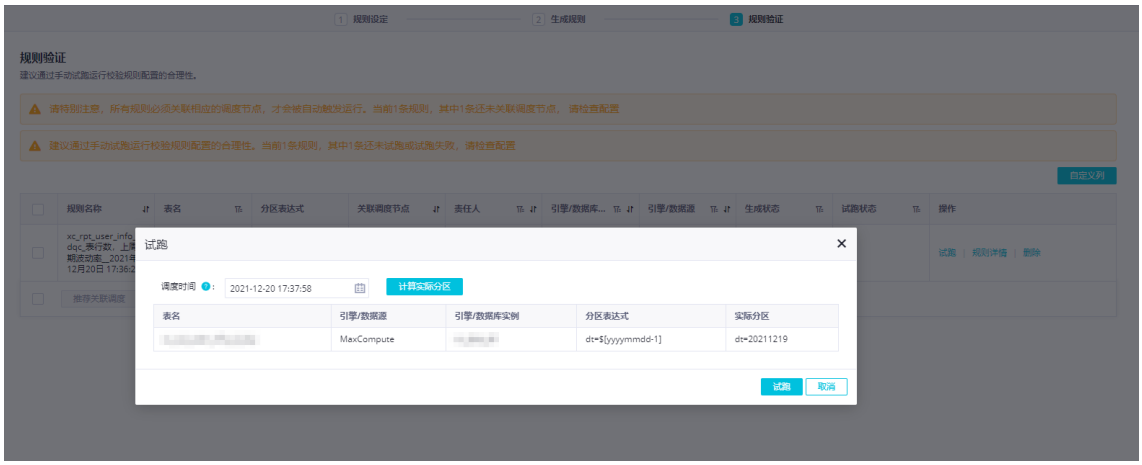


4. 单击**生成规则**，进入**规则验证**页面。

您可以单击自定义列，按需自定义规则详情表中需要显示的内容。在**规则验证**页面您可以进行如下操作：

- 校验规则配置的合理性：试跑

规则创建完成后，您可以选择单个或多个规则进行试跑，在弹出来的**试跑**对话框中选择调度时间（模拟给定校验被触发的时间），系统会根据此时间以及设定的分区表达式，计算要验证的表的具体分区取值。 完成后单击**试跑**，去校验该表该分区下的数据是否符合当前配置的数据质量校验规则。



试跑后您可以单击操作列的**试跑记录**，查看试跑状态详情，并进行相应的处理。

> **说明** 试跑错误的可能原因为：表或表分区不存在、表数据不符合质量校验规则。

- 定义规则触发方式：关联调度

  您可以单击**推荐关联调度**或**手动关联调度**，为单个或多个数据质量规则关联产出表数据的调度节点（运维中心中产出表数据的节点，包括自动调度的周期实例，手动触发的补数据实例，测试实例），当节点任务执行时便会触发该数据质量规则校验，您可以设置规则的强弱来控制节点是否失败退出，从而避免脏数据影响扩大。

  - 推荐关联调度：系统会根据产出该表的节点血缘关系选中的规则自动关联推荐的调度节点。
  - 手动关联调度：您可以为选中的规则手动关联指定的调度节点。

> **注意** 必须关联相应的调度节点，规则才会被自动触发运行。

- 删除规则：您可以选择单个或多个规则进行删除。
- 查看规则详情：您可以单击操作列的**规则详情**，查看规则详情，并对规则进行修改、启停、删除、设置规则强弱、查看日志等操作。

5. 试跑运行成功且关联调度后，单击**保存**。确认是否已完成所有配置，确认无误后单击**确认**完成配置。

后续步骤

- 完成后当您进行按表配置监控规则质量监控规则的时候，即可查看已配置的模板规则详情，并对该规则手动设置**订阅管理**，目前支持通过钉钉群机器人、短信、邮件，报警给指定接收人。
- 如果您需要在指定业务时间内，不符合质量校验规则的数据不阻塞任务运行，可以用去噪管理功能，详情请参见：去噪管理。

3.5. 查看监控任务

数据质量的任务查询模块展示规则的校验结果。规则运行后，您可以在任务查询页面查看运行记录。

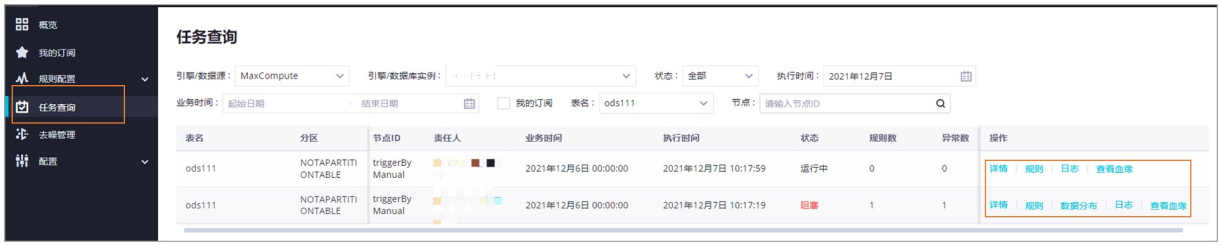
进入任务查询

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。

3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上角的图标，选择**全部产品 > 数据治理 > 数据质量**。
5. 在左侧导航栏，单击**任务查询**。

   您可以在**任务查询**页面，根据**引擎/数据源**、**状态**和**我的订阅**等信息，筛选需要查看的EMR（E-MapReduce）、Hologres、AnalyticDB for PostgreSQL、MaxCompute或DataHub任务。

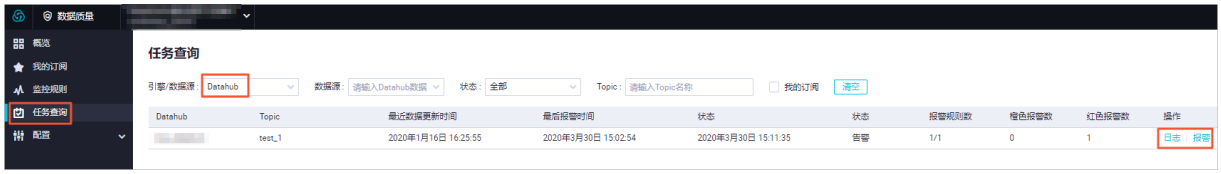
查看EMR、Hologres、AnalyticDB for PostgreSQL和MaxCompute任务



名称	描述
引擎/数据源	根据自身需求，选择**引擎/数据源**为EMR、Hologres、AnalyticDB for PostgreSQL或MaxCompute。
引擎/数据库实例	选择已绑定的EMR、Hologres、AnalyticDB for PostgreSQL或MaxCompute引擎。
状态	支持通过规则运行状态进行过滤，状态包括正常、运行中、报警、阻塞、阻塞和报警。如果是报警或阻塞状态，则需要您多加注意。
业务时间	选择相应的业务时间。
我的订阅	设置是否筛选已订阅的规则。
执行时间	规则的执行时间。
表名	您可以输入表名，快速搜索需要查看的表。
节点	触发规则的节点。
详情	单击相应表名后的**详情**，您可以进行**查看历史结果**、**问题处理**和**处理日志**等操作： • 单击相应规则后的**查看历史记录**，即可查看每次调度后的运行记录。 • 您可以针对当前规则的运行情况填写处理意见。操作如下： i. 单击相应规则后的**问题处理**。 ii. 在**问题处理**对话框中，选择**处理方式**，并输入**处理意见**。 iii. 单击**确认**。 **注意** 您需要购买DataWorks企业版及以上版本，才可以使用问题处理功能。 • 单击相应规则后的**处理日志**，即可查看对该条规则的历史处理记录。
规则	单击相应表名后的**规则**，进入该表的监控规则配置页面。您可以查看之前创建的分区表达式和规则，并进行相应的修改。详情请参见配置MaxCompute监控。

名称	描述
日志	单击相应表名后的**日志**，查看规则的运行日志。
数据分布	单击相应表名后的**数据分布**，查看该规则每次运行的情况，以及表行数、表大小等信息。
查看血缘	点击此处将快速跳转至**数据地图**，查看该表的血缘信息。

查看DataHub任务



名称	描述
引擎/数据源	选择**引擎/数据源**为DataHub。
数据源	选择相应的DataHub数据源。
状态	规则执行结果的状态，如果是报警或阻塞状态，则需要您多加注意。
Topic	您可以输入Topic名称，快速搜索需要查看的Topic。
我的订阅	设置是否筛选已订阅的规则。
清空	单击**清空**，即可清除设置的筛选条件。
日志	单击相应Topic后的**日志**，查看规则的运行日志。
报警	单击相应Topic后的**报警**，查看任务运行的报警详情。 在**报警列表**页面，您可以单击相应报警后的**关闭**，在**注意**对话框中，单击**确认**，关闭当前报警。

3.6. 去噪管理

当任务触发质量规则校验时，您可以使用去噪管理功能，对当前工作空间内，数据质量规则校验异常的数据不触发报警，且不阻塞任务运行（任务不会因为数据质量校验不通过而失败退出）。

前提条件

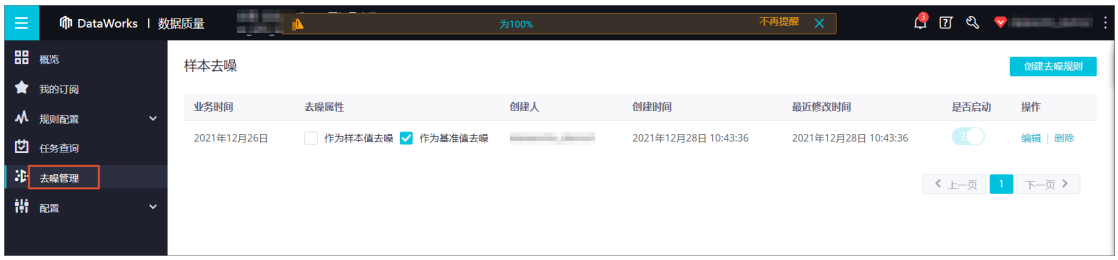
已创建数据质量校验规则，详情请参见按表配置监控规则、按模板配置监控规则。

使用限制

- 仅阿里云主账号支持创建并管理去噪规则。
- 去噪规则仅对当前工作空间内，规则的指定对象（去噪规则中指定的数据质量规则）生效。

1. 进入去噪管理页面。
   i. 进入**数据质量概览**页面，详情请参见进入数据质量概览。

ii. 在左侧导航栏选择**去噪管理**，进入去噪管理页面。



2. 创建去噪规则。

i. 单击右上角的**创建去噪规则**。

ii. 在弹出的**创建去噪规则**页面，单击**添加去噪规则**。

参数	描述

参数	描述
业务日期	选择需要进行去噪的业务日期，开启去噪规则后，该业务日期对应的分区的数据不符合预期时将不阻塞任务运行。
规则类型	选择需要去噪的数据质量校验规则类型，支持选择多种类型的数据质量校验规则。
期望值	您可以将目标数据作为样本值或基准值，为工作空间下指定类型的质量校验规则进行去噪。 ■ 样本值去噪：对指定时间内样本数据进行数据质量校验时，如果校验异常，且需要不阻塞任务运行，则可选择此功能，即上述业务时间数据不符合预期时，不会阻塞任务。 ■ 基准值去噪：当数据作为基准值，在一段时间内该基准值数据不符合预期，且需要不阻塞后续一段时间内的数据质量校验，可选择此功能。即上述业务时间的数据作为基准值校验不符合预期时，不阻塞后续任务运行。 **说明** ■ 样本 当天采集的具体的样本的值。例如对于SQL任务表行数，1天波动检测，则样本是当天分区的表行数。 ■ 基准值 历史样本的对比值： ■ 如果规则是SQL任务表行数，1天波动检测，则基准值是前一天分区产生的表行数。 ■ 如果规则是SQL任务表行数，7天平均值波动检测，则基准值是前7天的表行数据的平均值。
是否启动	选择是否开启该降噪规则，开启后当任务中存在质量规则校验不通过的数据时将不阻塞任务运行。

iii. 单击**保存**，完成去噪规则配置。

管理去噪规则

您可以在**样本去噪**页面，您可以查看已创建的去噪规则详情，并对已创建的去噪规则进行编辑、删除的操作。

3.7. 配置

3.7.1. 新增和操作报告模板

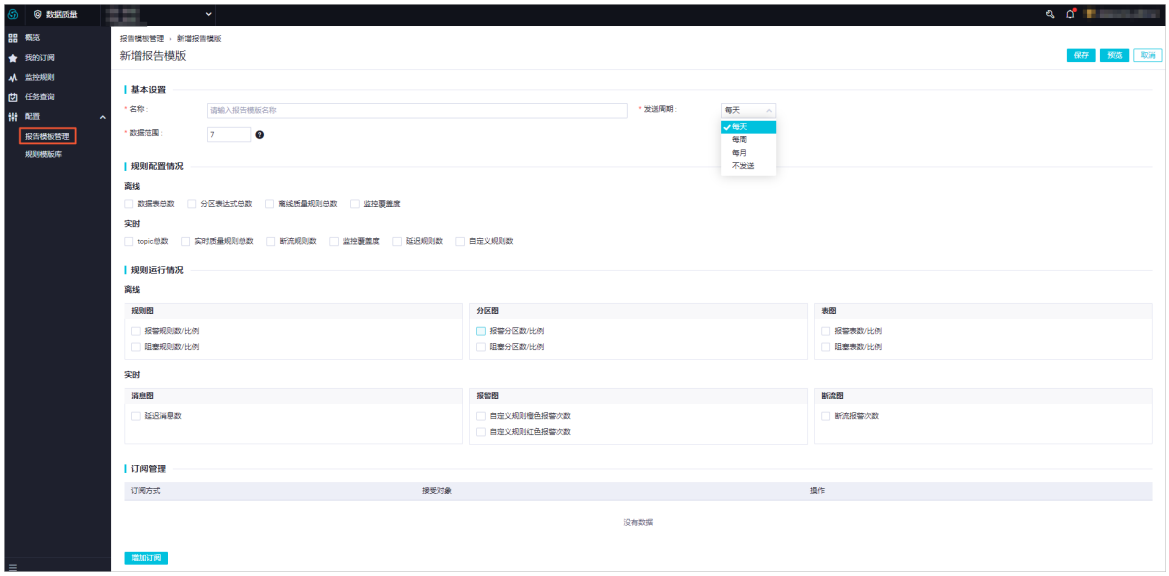
报告模板管理页面支持动态配置报告模板，数据质量根据您配置的报告模板定时生成并发送报告。

前提条件

您需要购买DataWorks企业版及以上版本，才可以使用报告模板管理功能。

新增报告模板

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上角的图标，选择**全部产品 > 数据治理 > 数据质量**。
5. 在左侧导航栏，单击**配置 > 报告模板管理**。
6. 单击**新增报告模板**。
7. 在**新增报告模板**页面，配置各项参数。



分类	参数	描述
基本设置	名称	输入报告模板的名称。
	发送周期	包括每天、每周、每月和不发送。 如果选择每周或每月，则需要继续选择具体的时间。
	数据范围	以当前日期为准，选择当前日期前N天的数据（最大可以选择30天）。

分类	参数	描述
规则配置情况 规则配置情况的指标包括离线规则和实时规则。您可以根据自身需求，勾选需要查看的指标数据。	**离线**	离线监控覆盖度是指该工作空间中配置质量规则的表的比例。包括**数据表总数**、**分区表达式总数**、**离线质量规则总数**和**监控覆盖度**。
	实时	实时监控覆盖度是指该工作空间中配置质量规则的topic的比例。包括**topic总数**、**实时质量规则总数**、**断流规则数**、**监控覆盖度**、**延迟规则数**和**自定义规则数**。
规则运行情况 规则运行情况的指标包括离线规则和实时规则。您可以根据自身需求，勾选需要查看的指标数据。质量报告中，以图表的形式展现被勾选的指标。	**离线**	包括**规则图**、**分区图**和**表图**。
	实时	包括**消息图**、**报警图**和**断流图**。
订阅管理	**订阅方式**	报告的订阅方式以邮件通知为主。
	接收对象	选择接收通知的对象，支持添加多个接收账号。
	操作	您可以**修改**或**删除**已添加的订阅。
	增加订阅	单击**增加订阅**，即可进行设置。

8. 配置完成后，单击页面右上角的**保存**，生成数据质量报告模板。

   您还可以进行以下操作：

   - 单击右上角的**预览**，查看报告模板的展示样式。

     > **说明** 发送给订阅人的邮件报告仅支持表格的展示方式。在数据质量页面中查询报告，包括表格和图表两种展示方式。

   - 单击右上角的**取消**，在**请取消**对话框中，单击**确认**，取消报告模板的新增。

操作报告模板

新增报告模板后，返回**报告模板管理**页面，查看报告模板详情。您还可以进行以下操作：

- 单击相应报告模板后的**编辑**，进入**编辑报告模板**页面进行修改。
- 单击相应报告模板后的**删除**，在**请确认**对话框中，单击**确认**，删除该报告模板。
- 单击相应报告模板后的**查看报告**，输入**查询范围**，查看相应的报告。
- 增加、修改和删除订阅：
  - 增加订阅
    a. 单击相应报告模板后的**订阅管理**。
    b. 在**订阅管理**对话框中，单击**增加订阅**。
    c. 选择**接收对象**，单击**保存**。
    d. 单击**确认**。

- 修改订阅
  a. 单击相应报告模板后的**订阅管理**。
  b. 在**订阅管理**对话框中，单击相应订阅后的**修改**。
  c. 选择**接收对象**，单击**保存**。
  d. 单击**确认**。
- 删除订阅
  a. 单击相应报告模板后的**订阅管理**。
  b. 在**订阅管理**对话框中，单击相应订阅后的**删除**。

3.7.2. 新建、操作和应用规则模板

DataWorks数据质量支持通过统一管理自定义规则，形成自建的规则模板库，帮助您提升规则配置的效率。

前提条件

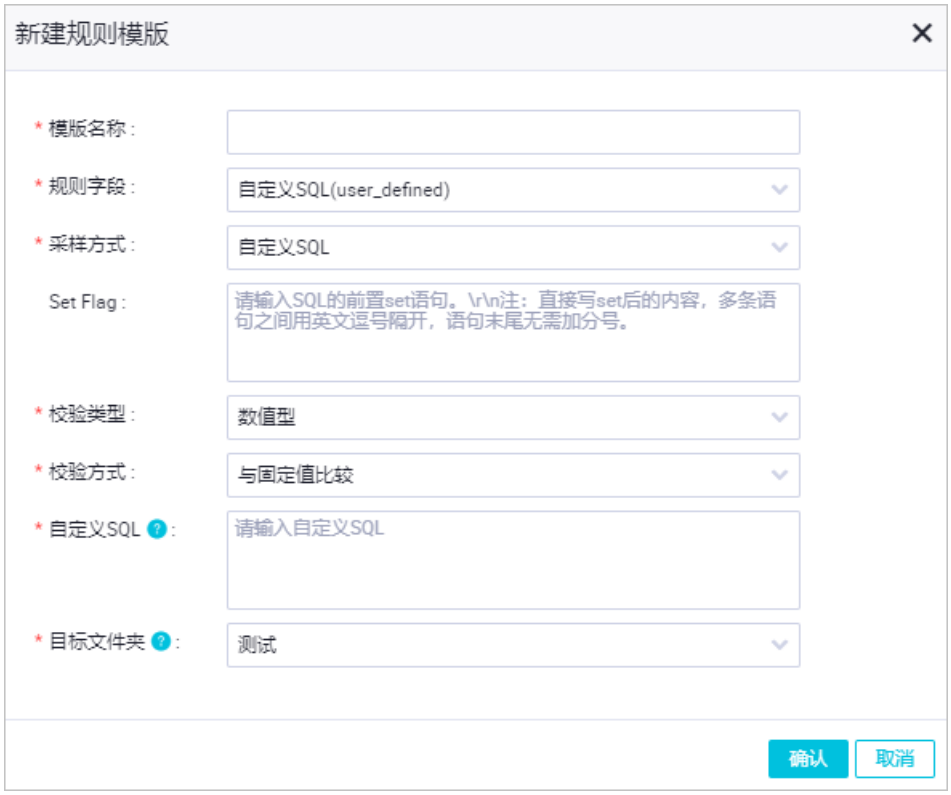
您需要购买DataWorks企业版及以上版本，才可以使用规则模板库功能。

背景信息

您可以通过**规则模板库**和**监控规则**两个页面新建规则模板。新建成功后，即可操作和应用规则模板。

通过规则模板库页面新建规则模板

1. 进入**数据质量**页面。
   i. 登录DataWorks控制台。
   ii. 在左侧导航栏，单击**工作空间列表**。
   iii. 单击相应工作空间后的**进入数据开发**。
   iv. 单击左上角的图标，选择**全部产品 > 数据治理 > 数据质量**。
2. 在左侧导航栏，单击**配置 > 规则模板库**。
3. 单击，选择**新建文件夹**。
4. 在**新建文件夹**对话框中，输入**名称**并选择**目标文件夹**，单击**确认**。
5. 右键单击相应的文件夹名称，选择**新建规则模板**。
   您也可以**重命名**和**删除**相应的文件夹。
6. 在**新建规则模板**对话框，配置各项参数。



参数	描述
模板名称	请输入自定义的模板名称。
规则字段	目前仅支持**自定义SQL**。
采样方式	目前仅支持**自定义SQL**。
Set Flag	请输入SQL的前置`set`语句。 **说明** 多条语句之间使用英文逗号（,）分隔，语句末尾无需添加分号（;）。
校检类型	目前支持**数值型**和**波动率型**。
校检方式	选择不同的校检类型，对应不同的校检方式： ○ 如果选择的**校检类型**为**数值型**，目前校检方式仅支持**与固定值比较**。 仅可以返回经过count、sum等运算后的一个值，与固定值比较。 ○ 如果选择的**校检类型**为**波动率型**，则校检方式包括**7天平均值波动**、**30天平均值波动**、**1天周期比较**、**7天周期比较**、**30天周期比较**、**7天方差波动**、**30天方差波动**、**1,7,30天波动检测**和**上一周期比较**。

参数	描述
自定义SQL	请输入自定义的SQL语句，您可以使用${tableName}表示表名。 说明 自定义SQL的采集结果需要是一个值（一行一列的形式），这样才能跟固定值做比较。
目标文件夹	选择该自定义规则模板需要存放的文件夹名称。

7. 单击**确认**。

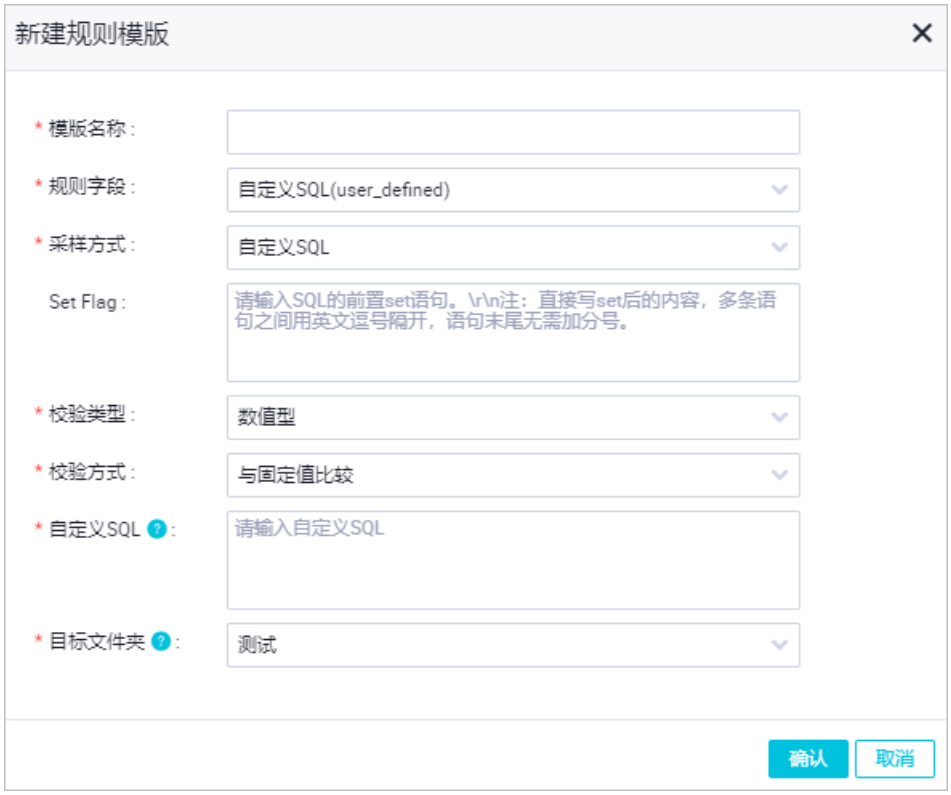
通过监控规则页面新建规则模板

1. 进入**数据质量**页面。
2. 在左侧导航栏，单击**监控规则**。
3. 选择**引擎/数据源**，单击相应**表名**或**Topic名称**后的**配置监控规则**。

   说明 本文以MaxCompute表为例。

4. 选择表的**分区表达式**，单击**自定义规则**。

5. 单击相应自定义规则后的**生成模板**。
6. 在**新建规则模板**对话框，配置各项参数。



参数	描述
模板名称	请输入自定义的模板名称。
规则字段	目前仅支持**自定义SQL**。
采样方式	目前仅支持**自定义SQL**。
Set Flag	请输入SQL的前置 `set` 语句。 **说明** 多条语句之间使用英文逗号（,）分隔，语句末尾无需添加分号（;）。
校检类型	目前支持**数值型**和**波动率型**。
校检方式	选择不同的校检类型，对应不同的校检方式： ○ 如果选择的**校检类型**为**数值型**，目前校检方式仅支持**与固定值比较**。 仅可以返回经过count、sum等运算后的一个值，与固定值比较。 ○ 如果选择的**校检类型**为**波动率型**，则校检方式包括**7天平均值波动**、**30天平均值波动**、**1天周期比较**、**7天周期比较**、**30天周期比较**、**7天方差波动**、**30天方差波动**、**1,7,30天波动检测**和**上一周期比较**。

参数	描述
自定义SQL	请输入自定义的SQL语句，您可以使用${tableName}表示表名。 说明 自定义SQL的采集结果需要是一个值（一行一列的形式），这样才能跟固定值做比较。
目标文件夹	选择该自定义规则模板需要存放的文件夹名称。

7. 单击**确认**。
8. 在左侧导航栏，单击**配置 > 规则模板库**，查看新建的规则模板。

操作规则模板

单击相应规则模板的名称，即可**查看**、**编辑**、**删除**和**复制**该规则模板。

操作	描述
查看	您可以查看相应规则模板的参数配置、**应用列表**和**日志**： • **应用列表**页面为您展示已应用该规则模板的数据质量规则。 • **日志**页面为您展示该规则模板的**操作者**、**操作时间**和**操作内容**。
编辑	单击右上角的**编辑**，在**编辑规则模板**对话框中修改相应的参数，单击**确认**。
删除	单击右上角的**删除**，在**删除模板**对话框中，单击**确认**。
复制	单击右上角的**复制**，在**复制规则模板**对话框中，输入**模板名称**并选择**目标文件夹**，单击**确认**。

应用规则模板

您可以在添加监控规则时，选择和使用创建的自定义规则模板。

1. 进入**数据质量**页面。
2. 在左侧导航栏，单击**监控规则**。
3. 选择**引擎/数据源**，单击相应**表名**或**Topic名称**后的**配置监控规则**。

   说明 本文以MaxCompute表为例。

4. 选择表的分区表达式，单击**创建规则**。

5. 在**创建规则**对话框，单击**模板规则 > 添加监控规则**。
6. 配置新建规则的参数，其中选择**规则来源**为**规则模板库**，并选择相应的**规则模板**，更多参数说明请参见配置监控规则。

7. 单击**批量添加**。

3.8. 使用指南

3.8.1. 配置DataHub监控

监控规则是数据质量（DQC）的核心。数据质量支持EMR（E-MapReduce）、Hologres、AnalyticDB for PostgreSQL、MaxCompute和DataHub监控，本文为您介绍如何配置DataHub监控。

背景信息

DataHub实时数据监控支持以下功能：

- 支持**数据断流**监控模板。
- 自定义Flink SQL、维表JOIN、多流JOIN以及窗口函数等流计算特性。

操作步骤

1. 新增DataHub数据源。
   i. 登录DataWorks控制台。
   ii. 在左侧导航栏，单击**工作空间列表**。
   iii. 选择工作空间所在地域后，单击相应工作空间后的**进入数据集成**。
   iv. 在左侧导航栏，单击**数据源**，进入**数据源管理**页面。
   v. 单击右上方的**新增数据源**添加DataHub数据源，详情请参见配置DataHub数据源。
2. 选择数据源。
   i. 单击左上方的☰图标，选择**全部产品 > 数据治理 > 数据质量**。
   ii. 在左侧导航栏，单击**规则配置 > 按表配置**。

iii. 选择**引擎/数据源**为DataHub，并选择指定的**引擎/数据库实例**，列表中显示当前数据源下所有的Topic。

参数	描述
配置Flink/SLS规则	添加数据源后，Flink或SLS资源会根据数据源拉取相关的信息。
Topic列表	DataHub数据源下所有Topic的名称，您可以在相应Topic后进行下述操作： ■ **配置监控规则**：对当前Topic创建规则，支持创建模板规则和自定义规则。 ■ **订阅管理**：查看当前Topic的订阅人，您可以快捷修改订阅人、报警方式及配置钉钉群报警。此处修改的报警方式对所有订阅人有效。
维度表	对Topic创建自定义规则JOIN时使用。如果采集到的数据有限，则需要对数据流补齐字段。进行数据分析前，将所需的维度信息进行补全，此时需要在数据质量中对这张维度表进行声明。 DataHub支持AliHBase维表、Lindorm维表、RDS维表、OTS（TableStore）维表、TDDL维表、MaxCompute维表。 Flink SQL中没有专门为维表设计的DDL语法，使用标准的create table语法即可。但需要额外增加一行 `period for system_time` 的声明，此行声明定义了维表的周期，即表明该表是一张会变化的表。 **说明** 声明一张维表时，必须指明唯一键。维表JOIN时，on的条件必须包含所有唯一键的等值条件。

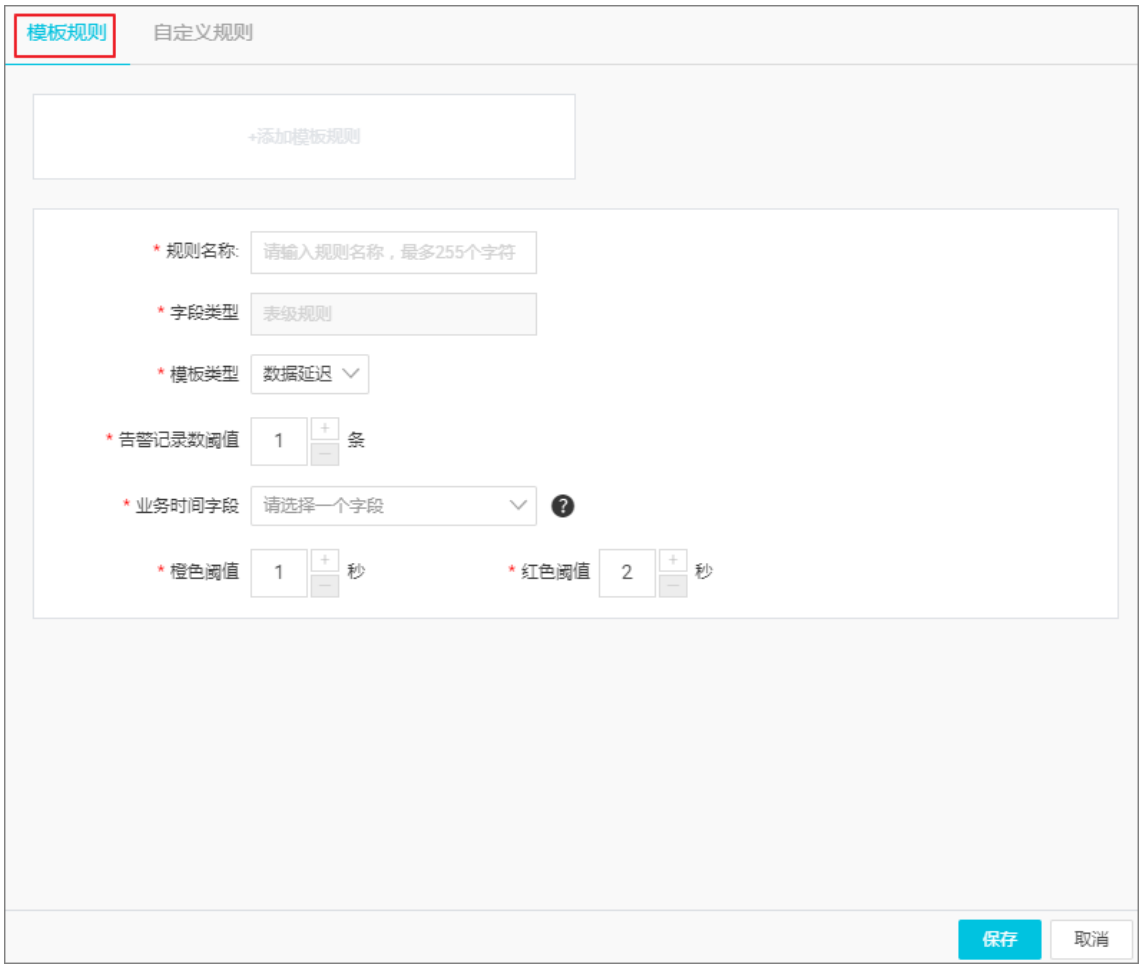
iv. 单击**Topic列表**页签相应Topic后的**配置监控规则**。

3. 在相应Topic的监控规则页面，单击**创建规则**。
4. 配置监控规则。

   数据质量支持**模板规则**和**自定义规则**两种类型：

   ○ 单击**添加模板规则**，模板类型包括**数据延迟**和**数据断流**。

例如选择模板类型为**数据延迟**。



参数	描述
规则名称	输入规则名称，最多255个字符。
字段类型	默认为表级规则。
模板类型	■ **数据延迟**：记录业务时间字段内，数据产生于流入DataHub通道的时间差，超过设定时间立即报警。 **说明** 业务时间字段支持TIMESTAMP和STRING（yyyy-MM-dd HH:mm:ss）两种类型。 ■ **数据断流**：允许在某一时间段内没有数据流入，当超过允许时间，则触发告警。 配置数据断流前，请首先在Flink中购买服务并创建项目。然后单击页面右上方的**配置Flink/SLS规则**，输入**Flink项目名称**，单击**确认**。
告警记录数阈值	允许出现数据延迟的数量上限，超过上限触发数据质量告警。只有模板选择数据延迟时，才需要配置该参数。

参数	描述
业务时间字段	Topic中时间字段的字段名称，支持TIMESTAMP和STRING（yyyy-MM-dd HH:mm:ss）两种类型，只有模板选择为数据延迟时才需配置本参数。
告警频次	告警频次包括10分钟、30分钟、1小时和2小时。
橙色阈值	以秒为单位，仅支持输入整数，且必须小于红色阈值。
红色阈值	以秒为单位，仅支持输入整数，且必须大于橙色阈值。

- 如果对DataHub规则有其它的使用方式，可以单击**添加自定义规则**进行创建。

> **说明**
> - Select的字段必须是一列且能够与橙色阈值和红色阈值进行数值对比。
> - 自定义规则下，From子句必须包含该Topic，且包含该Topic中所有的列。

参数	描述
规则名称	输入规则名称，需要在Topic内唯一，最多支持20个字符。

规则脚本	自定义编写SQL来设定规则，Select的结果字段必须唯一。示例如下：
	■ 简单SQL。
	`select id as a from zmr_tst02;`
	■ 与维表JOIN查询，维表名称（test_dim）。
	`select e.id as eid`
	`from zmr_test02 as e`
	`join test_dim for system_time as of proctime() as w`
	`on e.id=w.id`
	■ 两个Topic进行JOIN查询，另一个Topic名称（dp1test_zmr01）。
	`select count(newtab.biz_date) as aa`
	`from (select o.*`
	`from zmr_test02 as o`
	`join dp1test_zmr01 as p`
	`on o.id=p.id)newtab`
	`group by`
	`id.biz_date,biz_date_str,total_price,'timestamp'`

橙色阈值	以分钟为单位，仅支持输入整数，且必须小于红色阈值。
红色阈值	以分钟为单位，仅支持输入整数，且必须大于橙色阈值。
最小告警间隔	允许告警的最小时间差，以分钟为单位。
附加文本	对当前自定义Topic的描述。

5. 在**订阅管理**对话框中，选择相应的**订阅方式**。

   订阅方式包括**邮件通知**、**邮件和短信通知**、**钉钉群机器人**、**钉钉群机器人@ALL**、**飞书群机器人**和**企业微信机器人**。

   > **说明** 添加钉钉群、飞书群和企业微信机器人获取Webhook地址后，复制Webhook地址至订阅管理中即可。

6. 单击**批量添加**，添加创建的规则至Topic中。
   - 单击**查看日志**，查看该规则的运行日志。
   - 单击**订阅管理**，您可以在该页面查看、修改当前规则的订阅人，也可以修改告警通知方式，对所有订阅人生效。

     您可以将任务报警添加到钉钉群中，支持**邮件通知**、**邮件和短信通知**、**钉钉群机器人**和**钉钉群机器人@ALL**、**飞书群机器人**和**企业微信机器人**。

     > **说明** 添加钉钉群、飞书群和企业微信机器人获取Webhook地址后，复制Webhook地址至订阅管理中即可。

3.8.2. 配置MaxCompute监控

监控规则是数据质量（DQC）的核心。数据质量支持EMR（E-MapReduce）、Hologres、AnalyticDB for PostgreSQL、MaxCompute和DataHub监控，本文为您介绍如何配置MaxCompute监控。

新增MaxCompute数据源

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 单击相应工作空间后的**进入数据集成**。
4. 在左侧导航栏，单击**数据源**，进入**工作空间管理 > 数据源管理**。
5. 单击右上方的**新增数据源**添加MaxCompute数据源，详情请参见配置MaxCompute数据源。

选择数据源

1. 单击当前页面左上方的图标，选择**全部产品 > 数据治理 > 数据质量**。
2. 在左侧导航栏，单击**规则管理 > 按表配置**。
3. 选择**引擎/数据源**为**MaxCompute**，显示当前数据源下所有的表。

   您也可以输入目标表名（支持表名首字母模糊搜索），查找对应的表。
4. 单击相应表后的**配置监控规则**。

配置分区表达式

数据质量通过分区表达式来确定需要配置的规则：

- 如果您的检查对象为非分区表，可以配置分区表达式为NOTAPARTITIONTABLE。
- 如果您的检查对象为分区表，可以配置为业务日期的表达式（例如$[yyyymmdd]），也可以配置为正则表达式。

在数据表的**监控规则**页面，单击**分区表达式**后的**+**，添加分区表达式。

您可以选择新建分区表达式，也可以选择推荐的分区表达式：

- 新建分区的表达式

  在**添加分区**对话框中，您可以根据自身需求编辑符合语法的分区表达式。非分区表可以直接选择推荐的分区表达式中的**NOT APART IT IONT ABLE**。

  - 一级分区的表达式格式：分区名=分区值，分区值可以是固定值，也可以是内置参数表达式。分区表必须配置到最后一级分区。
  - 多级分区表达式格式：1级分区名=分区值/2级分区名=分区值/N级分区名=分区值，分区值可以是固定值，也可以是内置参数表达式。参数必须使用中括号表示，例如$[yyyymmdd-N]。

  分区表达式周期由配置的业务日期决定，例如配置运行时间为前5天，则周期为每5天调度一次。支持的分区表达式如下表所示。

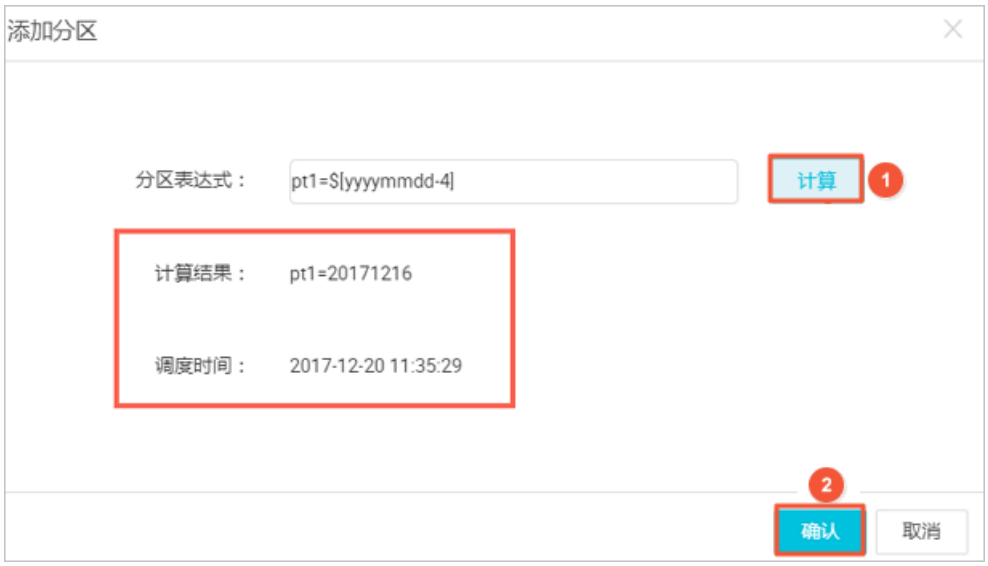
分区表达式	描述
dt=$[yyyymmdd-N]	代表前N天
dt=$[hh24miss-1/24]	代表一个小时前
dt=$[hh24miss-30/24/60]	代表半个小时前
dt=$[add_months(yyyymmdd,-1)]	代表获取上个月今天的日期。精确到天。
$[yyyymmdd]	调度日期
$[yyyymmdd-1]	代表获取业务日期。
$[yyyymmddhh24miss]	格式为 `yyyymmddhh24miss` ，当前实例运行的业务日期： ○ yyyy表示4位数年份 ○ mm表示2位数月份 ○ dd表示2位数天 ○ hh24表示24小时制的时 ○ mi表示2位数分钟 ○ ss表示2位数秒
NOT APART IT IONT ABLE	非分区表可以选择该分区表达式

- 推荐的分区表达式

  下文将以分区名dt为例，为您介绍推荐的分区表达式。动态分区表建议使用包括正则的分区表达式。

i. 在**添加分区**对话框中，单击输入表达式的窗口，会显示数据质量为您推荐的分区表达式：
   - 如果有符合预期的表达式，单击该行，会自动同步至输出窗口。
   - 如果没有满足需求的分区表达式，您可以根据需求自己输入。

ii. 输入分区表达式后，单击**计算**。数据质量会按照当前时间（调度时间）计算出分区表达式的计算结果，以便验证分区表达式的正确性。



iii. 单击**确认**。

如果您有不需要的分区表达式，可以单击相应分区表达式后的**删除**。如果该分区表达式已经配置有规则，删除时会删除该表达式下的所有规则。

配置关联调度

如果您需要在生产链路上监控离线数据质量，需要将数据质量规则与产出表数据的调度节点进行关联：

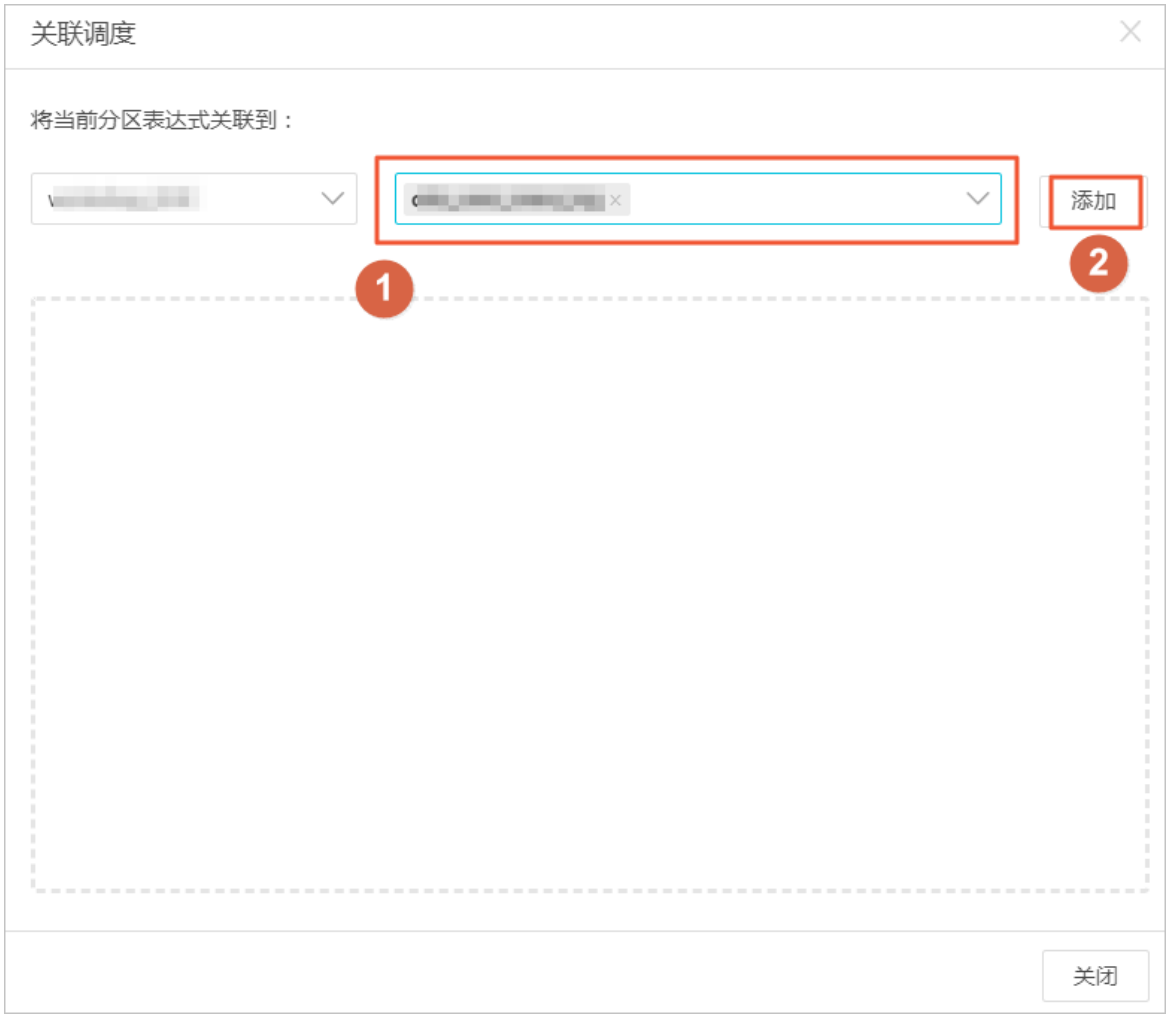
- 关联界面仅能找到已经提交的节点。
- 关联前，请确保您在关联的两个工作空间中，同时拥有**管理员**、**开发**或**运维**中至少一个角色。

数据质量的关联调度可以关联单个或多个节点任务，关联调度完成后，离线数据质量监控任务可以自动运行。

> **说明** 数据质量的关联可以灵活配置，您关联的任务并非一定与您的表有关系。

1. 在相应表的监控规则页面，单击**关联调度**，配置规则与任务的绑定关系。

2. 在**关联调度**对话框中，输入您需要关联的任务节点名称。



3. 单击**添加**。

创建规则

创建规则是数据质量模块的核心内容，您可以根据表的实际需要创建规则。

目前创建规则的方式包括**模板规则**和**自定义规则**，您可以根据自身需求选择相应方式。两种规则又分为**添加监控规则**和**快捷添加**两部分，详情请参见监控规则。

创建完成后单击**批量保存**，即可将创建的所有规则保存到已建好的分区表达式。

添加方式	参数	描述
	规则名称	输入规则名称。
	强弱	配置规则的强弱： • 勾选**强**时，如果触发红色阈值，则报警且任务置为失败状态。如果触发橙色阈值，则报警且任务置为成功状态。 • 勾选**弱**时，如果触发红色阈值，则报警且任务置为成功状态。如果触发橙色阈值，则不报警且任务置为成功状态。

添加方式	参数	描述
添加监控规则	**动态阈值**	您需要购买DataWorks企业版及以上版本，才可以使用动态阈值功能。
	规则来源	包括**内置模板**和**规则模板库**。
	规则字段	包括**表级规则**和**字段级规则**。字段级规则可以针对表中的具体字段配置监控规则。 **说明** 此处选择为表级规则，页面中的其它设置项对应为表级规则配置项。
	规则模板	• 如果您选择**规则来源**为**内置模板**，为您展示系统内置的表级监控规则。 • 如果您选择**规则来源**为**规则模板库**，需要设置**采样方式**、Set Flag等参数，详情请参见新建、操作和应用规则模板。
	比较方式	包括**绝对值**、**上升**和**下降**三种类型。
	波动值比较	设置波动值的橙色阈值和红色阈值。您可以通过拖动进度条来设置，也可以直接输入阈值。
	描述	对配置的规则进行简单描述。
快捷添加	**规则名称**	输入规则名称。
	规则字段	包括表级规则和字段级规则。字段级规则可以针对表中具体字段进行配置监控规则。
	快捷规则	• 选择表级规则，快捷规则支持**表行数大于0**和**表行数动态阈值**。 **注意** 您需要购买DataWorks企业版及以上版本，才可以使用动态阈值功能。 • 选择字段级规则，快捷规则可以选择**字段重复值**、**字段空值**和**唯一值动态阈值**。 **注意** 您需要购买DataWorks企业版及以上版本，才可以使用动态阈值功能。

试跑规则

成功配置规则后，您可以针对某个分区表达式下的所有规则进行试跑，并查看试跑的校验结果。

说明 通过试跑，您可以测试规则配置的正确性、测试订阅发送渠道。试跑是手动运行监控规则的一种方式，您可以根据自身需求选择是否进行试跑。

1. 在相应表的监控规则页面，单击**试跑**。
2. 在**试跑**对话框中，选择**调度日期**。

参数	描述
试跑分区	实际分区会随着业务日期变化而改变。如果为NOT APART IT IONT ABLE，则会自动添加实际分区。
调度时间	选择需要试跑的调度日期，默认为当前时间。

3. 单击**试跑**。
4. 单击**试跑成功，点击查看试跑结果**，进入**任务查询**页面，查看校验结果。详情请参见查询任务。

进行订阅管理

订阅管理默认通知创建者，如果想通知其它用户，您可以手动添加。

1. 在相应表的监控规则页面，单击**订阅管理**。
2. 在**订阅管理**对话框中，选择相应的**订阅方式**。

   订阅方式包括**邮件通知**、**邮件和短信通知**、**钉钉群机器人**、**钉钉群机器人@ALL**、**飞书群机器人**和**企业微信机器人**。

   > **说明** 添加钉钉群、飞书群和企业微信机器人获取Webhook地址后，复制Webhook地址至订阅管理中即可。

3. 单击**保存**。

查看分区操作日志

在相应表的监控规则页面，单击**分区操作日志**。您可以在**操作日志**对话框中查看**操作人**、**操作时间**和**操作内容**。

操作内容显示当前分区表达式设置的所有规则。

查看上一次校检结果

在相应表的监控规则页面，单击**上一次校检结果**，进入**任务查询**页面。您可以查看当前分区表达式下的运行结果情况和历史结果。

复制规则

1. 在相应表的监控规则页面，单击**复制规则**。
2. 在**复制规则**对话框中，选择**目标表达式**。
3. 根据自身需求选中**同步订阅人**或**替换自定义SQL规则中的表名**。
4. 单击**执行复制**。

3.8.3. 内置模板规则

数据质量为您提供内置表级别、字段级别的监控模板。本文为您介绍数据质量的校检逻辑及内置模板规则。

计算说明

计算波动率：您可以根据波动率计算公式（ `波动率=（样本-基准值）/基准值` ）计算结果。

- 样本

  当天采集的具体的样本的值。例如对于SQL任务表行数，1天波动检测，则样本是当天分区的表行数。

- 基准值

  历史样本的对比值：

  - 如果规则是SQL任务表行数，相比7天前的波动率，则基准值是7天前那一天分区产生的表行数。即今天的采样结果与7天前那一天分区的结果比较波动率。
  - 如果规则是SQL任务表行数，7天平均值波动检测，则基准值是前7天的表行数的平均值。即（7天内每天表行数之和）/7。

校检逻辑

数据质量支持与固定值比较、波动值比较和动态阈值三种校检方式。

校检方式	校检逻辑
与固定值比较	1. 根据校验的表达式进行计算，返回布尔值。支持以下比较操作符： `>` 、`<` 、`>=` 、`<=` 和 `!=` 2. 如果上述计算结果为true，返回正常，否则返回红色报警。
波动值比较	• 如果校验值的绝对值小于或等于橙色阈值，则返回**正常**。 • 如果校验值的绝对值不满足第1种情况，且小于或等于红色阈值，则返回**橙色报警**。 • 如果校验值不满足第2种情况，则返回**红色报警**。
动态阈值	您无需手动设置阈值，系统会自动根据算法模型实时检测指标的正确性。如果超出合理的波动范围，便进行报警。 **注意** 您需要购买DataWorks企业版及以上版本，才可以使用动态阈值。

内置模板规则说明

内置模板分为表级和字段级的模板规则，支持您通过内置规则模板来快速为一批表批量创建数据质量规则。详情请参见：按表配置监控规则、按模板配置监控规则。

模板规则		数值型			波动率型								计数
		固定值	上周期差值	1天差值	上周期波动率	1天波动率	7天波动率	30天波动率	1、7、30天波动率	7天平均值波动率	30天平均值波动率	1、7、30天、本月1号波动率	
表级	表行数	是	是	是	是	是	是	是	是	是	是	是	11
	表大小	是	是	是		是	是						5
字段级	**平均值**				否	是			是				2
	汇总值				是	是			是				3
	最大值				是	是			是				3
	最小值				是	是			是				3
	唯一值个数	是							是				2
	唯一值个数/总行数	是											1
	空值个数	是											1
	空值个数/总行数	是											1
	重复值个数	是											1
	重复值个数/总行数	是											1
	离散值（分组个数）	是				是							2
	离散值（状态值）	是											1
	离散值（分组个数及状态值）								是				1
计数		10	2	2	4	7	2	1	7	1	1	1	38

针对数值型字段

表级规则

模板名称	描述
表行数，固定值。	表行数。
表行数，1，7，30天波动率。	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据比对，计算波动率，再与阈值进行比较，只要其中有一个波动率超过阈值就会报警。
表行数，7天平均值波动率。	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。
表行数，30天平均值波动率。	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。
表行数，1天波动率。	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。
表行数，7天波动率。	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。
表行数，30天波动率。	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。
表行数，1，7，30天，本月1号，波动率。	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。
表行数，上周期波动率。	基准值为上一周期产生的分区的表行数，比较当天采集的表行数，对比波动率。
表行数，1天差值。	表的行数，相比1天前的差额。 **说明** 基准值为昨天分区的表行数，比较当天采集的表行数，对比差值。
表行数，上周期差值。	基准值为上一周期产生的分区的表行数，比较当天采集的表行数，对比差值。
表大小，固定值。	表的空间大小（字节）。
表大小，1天波动率。	该模板用于监控表大小的波动情况，样本对比值是与昨天的额度样本进行比较，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。 例如，填写橙色阈值为5%，红色阈值为10%。当波动率大于5%且小于等于10%时，会进行橙色报警。当波动率大于10%时，会进行红色报警。

模板名称	描述
表大小，7天波动率。	该模板用于监控表大小的波动情况，样本对比值是与7天前的额度样本进行比较，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。 例如，填写橙色阈值为5%，红色阈值为10%。当波动率大于5%且小于等于10%时，会进行橙色报警。当波动率大于10%时，会进行红色报警。
表大小，上周期差值。	相比上一周期表大小的差值（字节）。
表大小，相比1天前的差值（字节）。	表的空间大小，相比1天前的差值（字节）。

字段级规则

模板名称	描述
平均值，1、7、30天波动率。	取该字段的平均值，与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。
汇总值，1、7、30天波动率。	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。
最小值，1、7、30天波动率。	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。
最大值，1、7、30天波动率。	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。
唯一值个数，固定值。	去重后的count数与一个期望数字进行比较，即固定值校检。
唯一值个数，1、7、30天波动率。	去重后的count数与1天、1周和1个月前的样本（字段为一值个数）比较进行比较，即固定值校检。
空值个数，固定值。	取该字段的空值数与固定值进行比较。 **说明** 是否为空值，是通过转换为SQL的is null进行判断。

模板名称	描述
空值个数/总行数，固定值。	空值的个数与行总数的比率与一个固定值进行比较。 **说明** 该固定值是一个小数。
重复值个数/总行数，固定值。	重复值个数与总行数的比率与一个固定值进行比较。
重复值个数，固定值。	总行数减去重后的个数，即字段重复值的个数。重复值个数与固定值进行比较。
唯一值个数/总行数。	唯一值个数与总行数的比率与一个固定值进行比较。
离散值（状态值），固定值。	group by之后的分组，每组count数，与固定值进行比较。
离散值（分组个数及状态值），1、7、30天波动率。	group by之后的分组数和分组后每组count数，与1天前的样本、7天前的样本、30天前的样本（离散值）进行比较，计算波动率。
离散值（分组个数），固定值	group by之后的分组数，与固定值进行比较。
离散值（分组个数），1天波动率	group by之后的分组数，与1天前样本进行比较，计算波动率。
平均值，1天波动率	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。
汇总值，1天波动率	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。
最小值，1天波动率	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。
最大值，1天波动率	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。
汇总值，上周期的波动率。	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。
最小值，上周期的波动率。	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。
最大值，上周期的波动率。	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。

4.数据保护伞

4.1. 概述

数据保护伞是一款数据安全管理产品，为您提供数据发现、数据脱敏、数据水印、访问控制、风险识别、数据审计、数据溯源等功能。本文为您介绍如何开通、登录数据保护伞。

使用限制

数据保护伞的数据识别和动态脱敏功能，目前仅支持对EMR、Maxcompute、CDH、Hologres引擎中的数据进行识别和脱敏。

进入数据保护伞

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上方的图标，选择**全部产品 > 数据治理 > 数据保护伞**。
5. 单击**立即体验**，进入数据保护伞。

> **说明**
> - 如果阿里云主账号已授权，直接进入数据保护伞的首页。
> - 如果阿里云主账号未授权，进入数据保护伞的授权页面。

开通数据保护伞

阿里云主账号在**服务声明**页面，勾选**我已阅读并接受以上协议条款**，单击**立刻开通**。

> **注意** 仅阿里云主账号可以进行授权，开通数据保护伞。

登录数据保护伞

开通数据保护伞后，即可登录数据保护伞。

序号	名称	描述
①	功能菜单栏	当前用户有权可见的功能模块，包括**数据开发**、**数据集成**、**运维中心**和**数据保护伞**等。
②	用户信息	当前登录的用户，您可以查看并编辑用户信息，包括邮箱、手机、AccessKey ID和AccessKey Secret。
③	导航栏	对应功能菜单的导航栏。数据保护伞为您提供了如下功能模块。详情请参见数据发现、数据脱敏、数据访问、数据风险（旧版）、数据审计、数据溯源和敏感数据识别。
④	数据保护伞首页	• **数据发现**：为您展示近一周通过数据识别规则命中的、按照项目细分的字段数，以及总量。 • **数据访问**：为您展示通过数据识别规则命中的字段，在统计时间维度内的每日访问量。 **说明** 支持统计**近一周**和**近一月**的每日访问量。 单击**查看详情**，进入**数据访问**页面。 • **数据风险**：为您展示近一周新增风险的数量，以及未处理的风险量。 • **数据审计**：为您展示**近一周**和**近一月**，数据风险的发现量和完成量。 单击**查看详情**，进入**数据审计**页面。 • **数据溯源**：支持上传存在泄露数据的文件，通过系统提取和对比泄露文件中数据的水印信息，帮助您定位到可能会泄露目标数据的责任人。
⑤	引导页切换	单击右上方**引导页**，进入引导页面，查看产品的信息。

4.2. 配置数据规则

4.2.1. 数据分类分级

数据分类分级支持您按照数据的价值、内容敏感程度、影响和分发范围不同对数据进行敏感级别划分。不同敏感级别的数据有不同的管控原则和数据开发要求。本文为您介绍如何编辑和查看分类分级模板。

背景信息

数据分类分级支持您对数据进行敏感级别划分，不同的敏感级别的数据有着不同的管控原则和数据开发要求，定义敏感级别后您可以进入**数据识别规则**按照数据的用途、来源等不同分类定义敏感字段类型并配置数据的识别规则，有效的识别组织内的敏感数据。详情请参见敏感数据识别。

如果您是首次使用数据保护伞的新用户，进入数据分类分级页面后会展示内置分类分级模板，默认有4级分类，您可以根据需求新增或删除分级，如果您是已使用过数据保护伞的老用户，进入数据分类分级页面后，会展示旧版本分级信息模板。

进入数据分类分级

1. 登录**DataWorks控制台**后，进入**数据保护伞**页面，操作详情请参见概述。
2. 单击**开始体验**，默认进入数据保护伞的**首页**。
3. 单击左侧导航栏中的**规则配置 > 数据分类分级**，您可以编辑和查看**分类分级模板**。

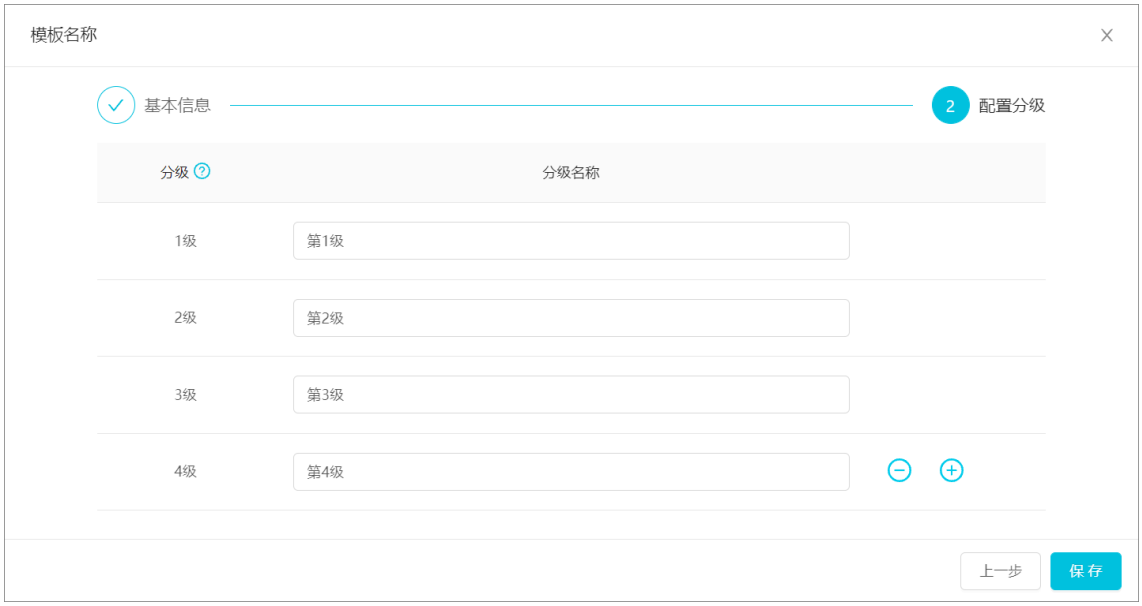
编辑分类分级

数据分类分级页面为您展示分类分级模板的分类分级情况及模板中的敏感字段类型数量，您还可以编辑和查看分类分级模板。本文以首次使用数据保护伞为例进行介绍。

1. 配置基本信息
   i. 在**分类分级模板**的**操作**列，单击按钮。
   ii. 在**模板名称**对话框中，输入**模板名称**、**描述信息**。

> **说明** 分类分级模板名称仅支持中英文、数字和“-”，长度限制1~30个字符。

   iii. 单击**下一步**。
2. 配置分级
   - 在**配置分级**页签支持您自定义分级名称，名称变更后，敏感字段类型所绑定的分级同步更新为新的名称。



说明 分级名称必须唯一，分级数字越大，数据敏感程度越高，最多支持10级。

- 新增分级：单击相应分级后的图标可以新增分级。
- 删除：单击相应分级后的图标可以删除不需要的分级。删除时请确认该分级下是否有敏感字段类型（不区分是否已发布），若有，则需删除全部敏感字段类型后才可以删除分级。详情请参见敏感数据识别。

查看分类分级模板

在**分类分级模板**的**操作**列，单击按钮，跳转至敏感数据识别页面查看分类分级模板中的分类及敏感字段类型的详细信息。

4.2.2. 自生成数据识别模型

DataWorks支持通过您提供的样本字段，进行模型训练，帮助您寻找目标字段的内容特征，生成相应的规则模型。该功能通常用于发现您的数据资产中与该特征内容相似的数据。本文为您介绍如何生成自定义的数据识别模型。

使用限制

- DataWorks不支持对数据量小于10条，并且数据长度小于4大于40的样本字段进行模型训练。
- DataWorks不支持对包含中文字符（包括中文标点符号）的样本字段进行模型训练。

创建模型

1. 进入**数据保护伞**。
   i. 登录DataWorks控制台。
   ii. 在左侧导航栏，单击**工作空间列表**。
   iii. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
   iv. 单击左上方的图标，选择**全部产品 > 数据治理 > 数据保护伞**。

v. 单击**立即体验**，进入数据保护伞。

2. 在左侧导航栏，单击**规则配置 > 数据识别规则**，进入**数据识别规则**页面。
3. 单击**自生成数据识别模型**，进入**自生成数据识别模型**页面。
4. 新建模型并进行模型训练。
   i. 单击**新建模型**。
   ii. 配置**模型名称**，并选择训练样本。

- 选择样本：您可以从当前工作空间下，选择需要训练的样本字段，DataWorks将帮助您找到这些字段的内容特征，生成相应的规则模型。后续您可以使用该规则模型发现您数据资产中与该模型的特征内容类似的数据。

  > **说明**
  > - DataWorks不支持对数据量小于10条，并且数据长度小于4大于40的样本字段进行模型训练。
  > - DataWorks不支持对包含中文字符（包括中文标点符号）的样本字段进行模型训练。

- 过滤字段：如果某些字段容易与样本字段混淆，则您也可以在该规则模型中将其排除，排除后，使用该规则模型识别数据时，排除的字段将不会命中。同时，排除的字段将作为负向样本加入模型训练，以达到不命中混淆数据，提高识别准确率的效果。

iii. 单击**下一步**。

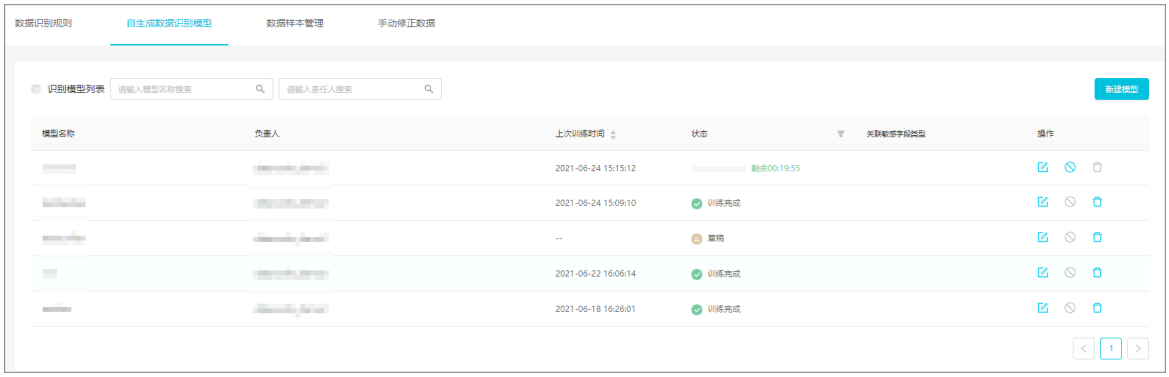
iv. 单击**开始训练**，启动模型训练。

本次模型训练将从您选的样本字段中各随机抽取不超过100条数据进行训练，并根据您的样本字段数量估算耗时。

> **说明** 模型训练时间较长，请您等待。等待过程中，您也可以关闭训练弹窗，操作其他功能，模型将在后台自动运行训练。

5. 查看模型训练结果。

在**自生成数据识别模型**页面，您可以查看目标模型的训练状态及训练结果，并根据训练结果判断该模型是否符合上线使用标准，用于识别数据。



- 查看训练状态。
  - **剩余hh:mm:ss**：表示当前模型正在训练中。
  - **训练完成**：表示当前模型已完成训练，您可以根据训练结果，判断该模型后续是否可用于识别数据。
  - **草稿**：表示该模型已创建，但未进行训练，不能投入识别数据。
- 查看训练结果。

单击完成训练的模型**操作**列的图标，即可查看通过该模型提取的样本特征对样例数据识别的准确率。建议当准确率为100%时，再投入上线使用该模型。

> **说明** 如果模型训练的评估结果准确率达不到100%，则投入上线使用识别的数据可能会有较大误差。建议您增加样本数据，重新训练模型，直至准确率达到100%后再投入上线使用。

6. 单击**确定创建**，完成当前规则模型的创建。

后续步骤

成功创建规则模型后，您可以进入**数据识别规则**页面，上线使用当前模型来识别数据。在**数据识别规则**中使用自定义的模型识别数据，详情请参见敏感数据识别。

4.2.3. 创建并管理样本库

DataWorks支持将您提供的样本文件生成样本库，后续可以将样本库配置为数据识别规则用来识别数据。当需要识别的目标数据包含样本库中的数据时，则会命中该识别规则。该功能通常用于识别可以使用枚举值罗列的数据，例如，员工姓名、用户地址等。本文为您介绍如何创建并管理样本库。

使用限制

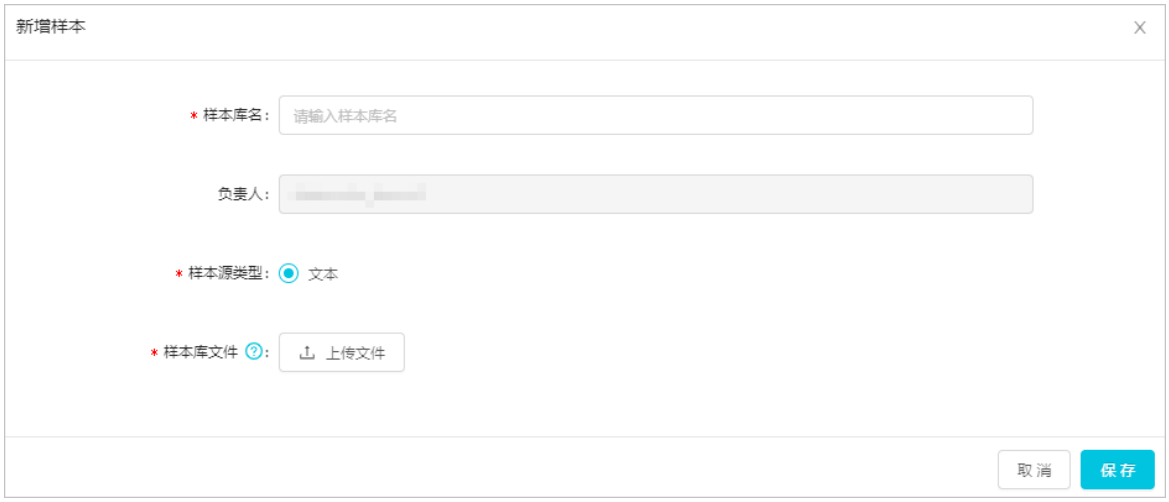
DataWorks仅支持上传大小不超过500KB，*UTF-8*格式的*TXT*文本文件做为样本库文件，并且样本文件中的每个数据占用一行。

> **说明** 一个**数据识别规则**仅支持识别一种类型的数据，因此，建议您的每个样本库中存放同类型的数据。如果您需要使用样本库方式识别多个类型的数据，则需要配置多个样本库。例如，您需要识别员工姓名、家庭住址，则需要配置姓名样本库及家庭住址样本库。

创建样本库

1. 进入**数据保护伞**。
   i. 登录DataWorks控制台。
   ii. 在左侧导航栏，单击**工作空间列表**。
   iii. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
   iv. 单击左上方的☰图标，选择**全部产品 > 数据治理 > 数据保护伞**。
   v. 单击**立即体验**，进入数据保护伞。
2. 在左侧导航栏，单击**规则配置 > 数据识别规则**，进入**数据识别规则**页面。
3. 单击**数据样本管理**，进入**数据样本管理**页面。
4. 单击**新建样本**，配置样本库名称并上传样本文件。

   > **说明**
   > - DataWorks仅支持上传大小不超过500KB，*UTF-8*格式的*TXT*文本文件做为样本库文件，并且样本文件中的每个数据占用一行。
   >
   >   **说明** 一个**数据识别规则**仅支持识别一种类型的数据，因此，建议您的每个样本库中存放同类型的数据。如果您需要使用样本库方式识别多个类型的数据，则需要配置多个样本库。例如，您需要识别员工姓名、家庭住址，则需要配置姓名样本库及家庭住址样本库。
   > - DataWorks支持在一个样本库中上传多个样本文件。

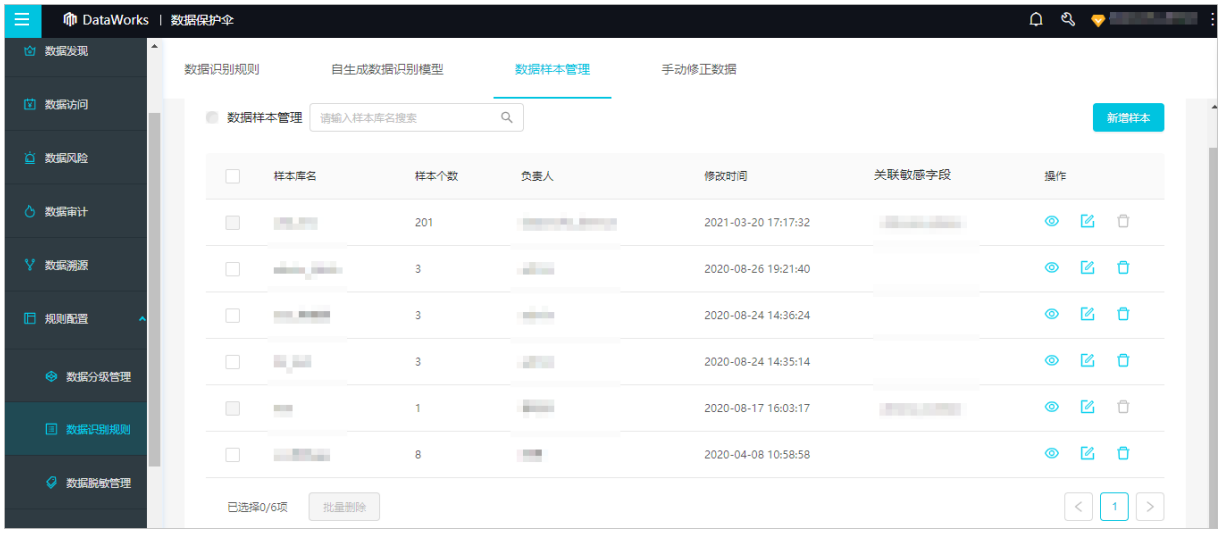


5. 单击**保存**，完成样本库创建。

成功创建样本库后，您可以将该样本库配置为数据识别规则，当需要识别的目标数据包含样本库中的数据时，则命中该识别规则。在**数据识别规则**中使用样本库，详情请参见敏感数据识别。

管理样本库

在**数据样本管理**页面，您还可以对已创建的样本库执行如下管理操作：



- 查看样本库列表。

  您可以在数据样本管理页面查看所有已创建样本库包含的样本个数及关联的数据识别规则。单击目标样本库**操作**列的图标，即可查看该样本库的数据详情。

- 修改样本库文件。

  单击目标样本库**操作**列的图标，您可以为样本库上传新的样本文件，或更换已有的样本文件。

- 删除样本库。

  单击目标样本库**操作**列的图标，即可删除当前样本库。

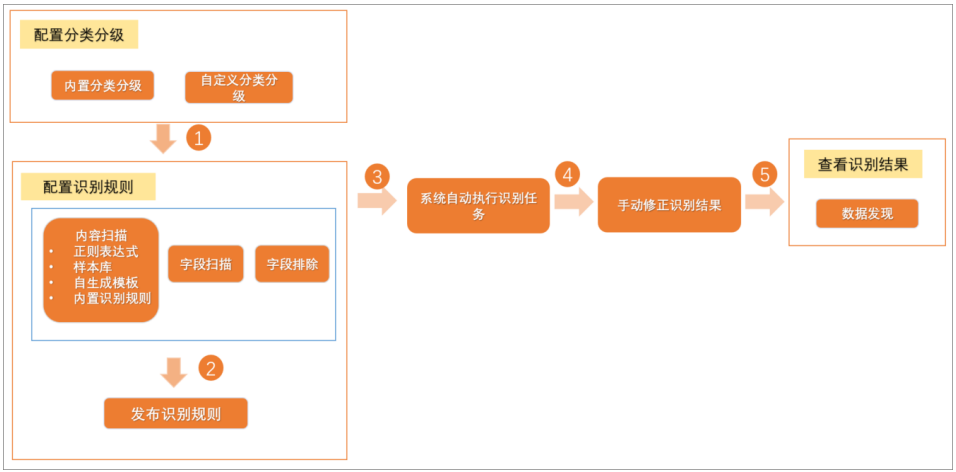
> **说明** 如果目标样本库已被数据识别规则引用，您可以在样本库列表查看该样本库关联的数据识别规则，并在数据识别规则的配置页面取消引用该样本库，取消引用后该样本库才能被删除。配置**数据识别规则**，详情请参见敏感数据识别。

4.2.4. 敏感数据识别

DataWorks支持通过内置敏感字段类型和自定义敏感字段类型，有效识别组织内的敏感数据。本文将为您介绍如何新建、配置数据识别规则。

背景信息

DataWorks支持您按照数据的敏感级别和所属分类定义数据识别规则，帮助您识别组织内的敏感数据，对于识别结果不准确的数据，您可以手动修正数据，并在数据发现模块为您展示最近的通过数据识别规则命中的、按照项目细分的全部敏感字段分布情况，数据识别规则的使用逻辑如下图所示。



> **说明** 对CDH引擎中数据进行识别和脱敏时，您需要通过DataWorks的数据抽样采集器功能，从CDH Hive表中随机抽取表的部分数据用于数据保护伞的敏感数据识别，抽样采集的数据不会存储至DataWorks中，没有数据泄漏风险。详情请参见：CDH Hive数据抽样采集器。

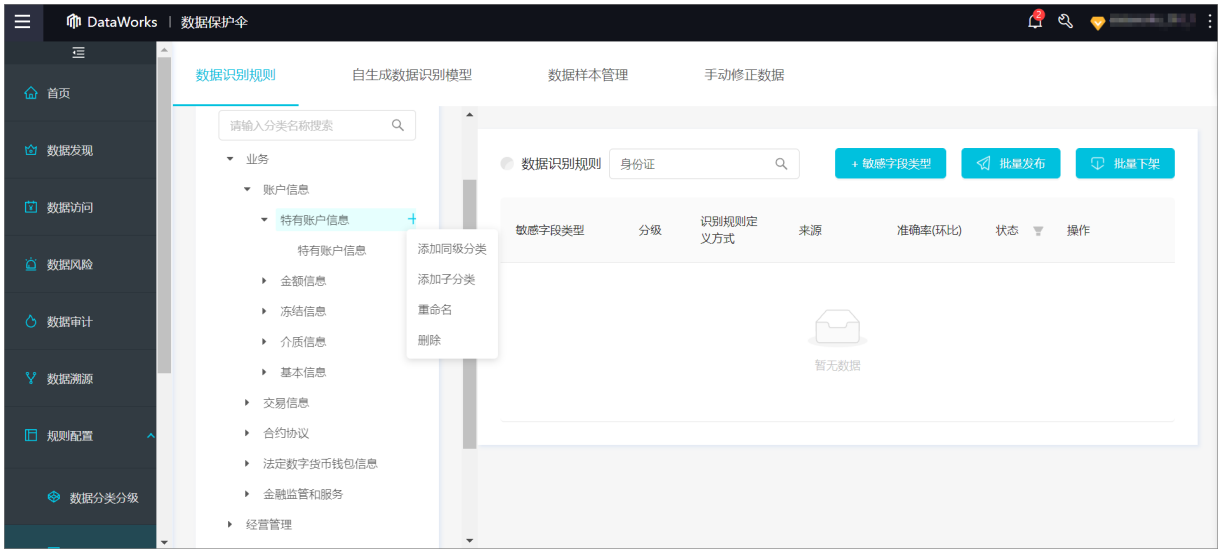
进入数据识别规则

1. 登录**DataWorks控制台**后，进入**数据保护伞**页面，操作详情请参见概述。
2. 单击**开始体验**，默认进入数据保护伞的**首页**。
3. 单击左侧导航栏中的**规则配置 > 数据识别规则**，您可以新增敏感字段类型并配置识别规则。

配置敏感字段所属分类

- 如果您是首次使用数据保护伞的新用户，进入数据识别规则页面后会在左侧区域展示**数据分类分级模板**的默认分类，您可以输入分类名称进行搜索。还支持您单击分类名称后的+图标**添加同层分类**、**添加子分类**、**重命名**和**删除**分类。
- 如果您是已使用过数据保护伞的老用户，进入数据识别规则页面后需要您根据需求在左侧区域创建数据分类。单击**未分类**后的+图标添加分类。

> **说明**
>
> - 分类名称必须唯一，仅支持中英文、数字，长度限制1~30个字符。
> - 删除时请先确认该分类下是否有已发布的敏感字段类型。如果有，请将该分类下全部敏感字段类型下架后方可删除。详情请参见批量下架。



添加敏感字段类型

1. 选择敏感字段所在的数据分类。

   在左侧的**内置分类分级模板**区域选择新增敏感字段所在的数据分类。

2. 新增敏感字段类型并配置识别规则。

   单击右上角的**+敏感字段类型**。

i. 在**基本信息**页签中配置敏感字段类型信息，单击**下一步**。

配置	说明
敏感字段类型	自定义敏感字段类型的名称，例如：姓名、身份证号、手机号等。 **说明** 定义敏感字段类型时，名称必须唯一，当存在重名时系统会提示敏感字段类型重复。
所属分类	下拉列表展示步骤1选中的数据分类，如果您需要修改分类可以在下拉列表进行选择。
所属分级	选择敏感字段类型所属级别，对配置的数据进行等级划分。如果现有的分级不满足需求，请进入**数据分类分级**页面进行设置，详情请参见数据分类分级。
描述信息	对当前敏感字段进行简单描述，长度0~100字符，不包含特殊字符。

ii. 在**规则配置**页签中，配置识别规则命中条件、敏感字段识别规则并测试规则准确性。识别规则配置完成并发布后，即可在识别任务中进行识别。

> **说明** 规则修改后，历史规则命中的字段识别结果将被清理。

配置	说明
识别规则命中条件	您可以在右侧下拉列表中选择识别规则命中条件： ■ **满足以下任一条件即命中规则**：满足数据内容识别或字段名称识别规则其中任何一个条件，即可命中识别规则。 ■ **同时满足以下条件即命中规则**：同时满足数据内容识别和字段名称识别规则时才可以命中识别规则。 > **说明** 识别规则命中条件仅对数据内容识别和字段名称识别规则生效。

配置	说明
数据内容识别	根据规则类型定义敏感数据识别规则的内容，用于匹配敏感数据的文本。 **说明** 数据内容识别的信息为字段的数据内容，例如，字段name，包含张三、李四等数据。则识别的内容为张三、李四等具体的数据内容。 ■ 规则类型选择**正则表达式**时：在**正则表达式**文本框中手动输入该类型的正则表达式，并在**测试数据**输入框中输入样本数据测试识别规则准确性。 ■ 规则类型选择**内置识别规则**时：单击**请选择内置识别规则**下拉框，选择内置识别规则，并在**测试数据**输入框中输入样本数据测试识别规则准确性。 **说明** 仅企业版及以上版本可以选择**内置识别规则**。 ■ 规则类型选择**样本库**时：单击**请选择样本库**下拉框，选择已配置的样本，并在**测试数据**输入框中输入样本数据测试识别规则准确性。样本配置请参见创建并管理样本库。 ■ 规则类型选择**自生成模型**时：单击**请选择自生成模型**下拉框，选择自生成模型，并在**测试数据**输入框中输入样本数据测试识别规则准确性。自生成模型配置请参见自生成数据识别模型。 **说明** 仅MaxCompute引擎支持选择**自生成模型**规则。仅DataWorks企业版及以上才可使用**自生成模型**。 **说明** 仅DataWorks专业版及以上版本，才可以使用内容扫描功能。如果您使用的是低版本的DataWorks，则需要升级至专业版及以上版本，才可以使用该功能。升级DataWorks版本，详情请参见版本服务计费说明。
字段名称识别	在输入框中输入需要识别为敏感数据的字段，支持多个字段匹配，各字段间为或关系。输入格式为：project.table.column，其中任一段可以使用*作为通配符，例如。 ■ abcd.efg.*：abcd的project下efg表中所有字段都会被识别为敏感数据。 ■ ab*.*.salary：ab开头的project下，所有表中的salary字段都会被识别为敏感数据。 ■ *cd.ef*.sa*ry：cd结尾的project下，ef开头的表中，所有以sa开头、ry结尾的字段都会被识别为敏感数据。 **说明** 字段名称识别的信息为字段的名称，例如，字段name，包含张三、李四等数据，则识别的内容为字段名称name。
字段注释识别	识别的信息为字段注释，如敏感字段类型为手机号时，对应字段注释为：手机号、联系方式，则可配置包含手机号、联系方式时，识别为手机号类型。在输入框中输入字段注释，字符长度0-100，字符不限，可添加多个输入框，最多10个。

配置	说明
字段排除	在输入框中输入需要排除的字段，符合字段排除规则的字段将不会被该识别规则命中。输入格式为：project.table.column，其中任一段可以使用*作为通配符，例如。 ■ abcd.efg.*：abcd的project下efg表中所有字段都会被排除，不会识别为该类敏感数据。 ■ ab*.*.salary：ab开头的project下，所有表中的salary字段都会被排除，不会识别为该类敏感数据。 ■ *cd.ef*.sa*ry：cd结尾的project下，ef开头的表中，所有以sa开头、ry结尾的字段都会被排除，不会识别为该类敏感数据。
命中率配置	支持您自定义识别规则命中率，当一列数据中的非空数据，超过命中阈值的数据符合数据内容识别条件时，则认为命中该识别规则。命中率默认配置为50%，命中率计算公式为：`100%*`**该列中命中识别规则的数据条数/该列数据的总条数**。 **说明** 命中率仅对数据内容识别规则生效。

iii. 确认配置无误后，您可以单击**保存草稿**将新增的敏感字段类型状态置为草稿，您还可以单击**发布使用**，发布后，敏感字段类型状态置为**已发布**，并触发新的识别任务。

说明 某列数据可能会命中不同敏感字段类型的识别规则命中条件。当这些敏感字段类型的命中条件个数相同时，识别顺序是**字段名称识别 > 数据内容识别 > 字段注释识别**。当命中条件的个数和类型都相同时，优先命中分级等级高的敏感字段类型的识别规则。

完成敏感字段类型的配置后，可在**数据发现**、**数据访问**和**数据风险**等模块通过筛选已配置的敏感字段类型及级别进行查看。

管理敏感字段类型

- 复制敏感字段类型

  单击相应敏感字段类型后的图标，即可生成一个完全一致的规则。复制后的名称加后缀-副本，复制的规则默认状态为**草稿**，您可以根据需求进行配置。

- 编辑敏感字段类型

  单击相应敏感字段类型后的图标，可以修改敏感字段的**规则配置**。内置敏感字段类型不可修改敏感字段类型名称、所属分类、所属分级信息，自定义敏感字段类型支持修改敏感字段类型信息。

- 删除敏感字段类型

  单击相应敏感字段类型后的图标，在对话框中单击**删除**即可。

> **注意** 删除敏感字段类型影响较大，请仔细阅读以下影响后再确认是否删除。
>
> - 识别结果中该敏感字段类型的记录将会删除。详情请参见手动修正数据。
> - 数据发现中的敏感数据分布信息将不统计该敏感字段类型。详情请参见数据发现。
> - 已配置的风险识别规则中有对应配置项的将会取消该敏感字段类型。详情请参见风险识别管理（旧版）

批量发布

发布对应的敏感字段类型后，系统开始进行敏感数据识别，识别结果请参见数据发现。

1. 单击**批量发布**按钮，勾选需要发布的**敏感字段类型**。

   > **说明** 状态为已发布的敏感字段不可勾选。

2. 单击**发布**，对应敏感字段类型的状态置为已发布。
3. 点击**取消**，恢复原始状态。

敏感数据识别任务

每天早上9点会开始运行敏感数据识别自动任务。您也可以在批量发布任务后，手动触发敏感数据识别任务。

1. 在页面顶端单击**开启任务**按钮开发敏感数据识别任务。
2. 在**开启敏感数据识别任务**面板里，设置扫描范围为**全量**或**自定义范围**。

配置	说明
全量	扫描当前租户授权账号下全部可获取的数据。

配置	说明
自定义范围	○ 项目空间范围默认全部数据引擎和全部项目空间。数据引擎下拉列表目前只能选ODPS，项目空间下拉列表是所选数据引擎下获取到元数据的所有项目空间。 ○ 表名总体长度0-100，字符不限，不填写代表全部。支持 .*通配符，如 .*name表示以name为后缀，private.*表示以private为前缀，多个表名或字段名请用英文逗号分隔。

您可以单击**添加自定义范围**添加多个自定义扫描范围，最多添加10个自定义范围，最终扫描范围取多个自定义范围的并集。

3. 设置完扫描范围后，单击**开启**按钮开启扫描任务，**任务状态**从**无状态**更新为任务进度条，进度计算方式为=（本次任务中已识别的表数量/本次任务中全部要识别的表数量）*100%。如果要结束任务，您可以单击**终止任务**按钮，然后在弹框中单击**确定**按钮。

> **说明** 识别规则修改后，新规则将在下一次自动任务（非实时）中启用，若需要实时触发新任务，您需要手动启动。

4. 单击**查看日志**按钮可以查看最新的50条执行日志记录。
5. 扫描任务结束后，**任务状态**更新为**无任务**。

批量下架

下架对应敏感字段类型后系统将不再进行该类敏感数据的识别，数据发现、手动修正数据等模块中的该类敏感字段类型的记录将会删除。在进行下架操作前，请确认该敏感字段类型是否被**数据脱敏规则**及**风险识别规则**引用，如果有需要先将**数据脱敏规则**置为失效，并取消**风险识别规则**中的引用。详情请参见数据脱敏管理和风险识别管理（旧版）。

1. 单击**批量下架**按钮，勾选需要下架的**敏感字段类型**。
2. 单击**下架**，单击对应敏感字段类型的状态置为草稿。
3. 点击**取消**，恢复原始状态。

任务执行记录

任务执行记录保留近1周已完成任务的记录，不包含当前正在进行中的记录，包括**开始时间**，**结束时间**，**耗时**，**任务类型**，**责任人**和**数据范围**。

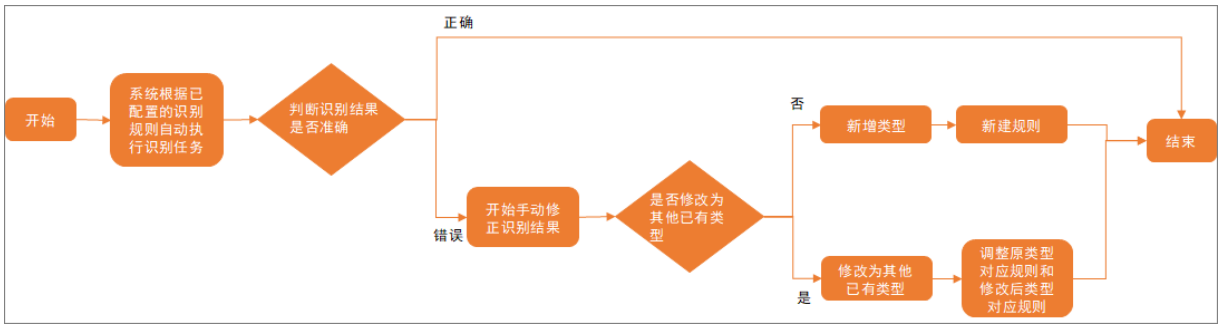
4.2.5. 手动修正数据

本文将为您介绍如何在手动修正数据页面，对规则识别不准确的数据进行手动修正。

> **说明** 手动修正的数据结果，在第2天才会生效展示。

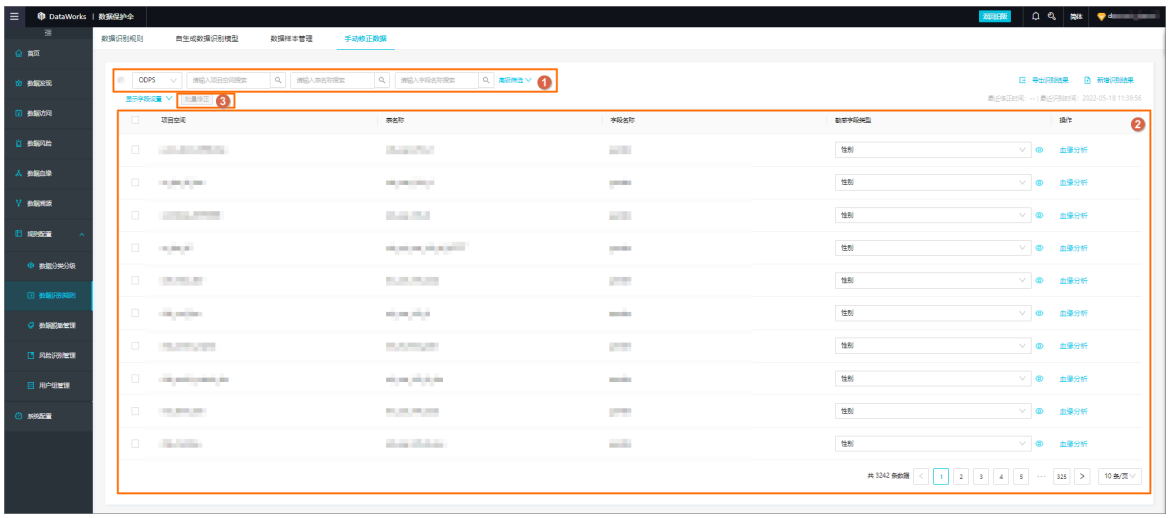
背景信息

DataWorks支持您对数据识别规则识别不准确的敏感数据进行手动修正，手动修正数据的使用逻辑如下图所示。



手动修正数据

1. 进入数据识别规则。详情请参见：进入数据识别规则。
2. 单击**手动修正数据**页签，进入手动修正数据页面。
3. 手动修正识别结果不准确的数据。



操作	描述

操作	描述
筛选	在上图的区域①，您可以通过筛选条件过滤需要查询的识别结果。 你可以根据**引擎类型**、**项目空间名称**、**表名**、**字段名**等条件进行筛选，还支持您单击**高级筛选**，展开更多筛选条件，您可以进一步通过**分类**、**分级**、**敏感状态**等条件进行筛选。 ○ **分类**：当前租户默认分类分级模板中的分类信息。详情请参见：数据分类分级 ○ **分级**：当前租户默认分类分级模板中的分级信息。 ○ **敏感状态**：包括敏感字段和非敏感字段。非敏感字段为您过滤已手动修改为非敏感字段的数据。 **说明** 目前支持对ODPS、EMR、CDH、HOLO引擎中的敏感字段识别结果进行修正。
修正单个数据	在上图的区域②为您展示识别结果列表，您可以单击**显示字段设置**勾选您需要关注的字段信息，刷新识别结果列表详情。列表默认为您展示**项目空间**、**表名称**、**字段名称**、**敏感字段类型**，同时，您还可以单击**操作**列的**血缘关系**进入数据血缘（公测）模块查看字段级别的数据血缘关系。 对于**敏感字段类型**识别结果有误的字段，单击右侧**敏感字段类型**列的下拉框，列表中为您展示当前租户下默认分类分级模板中**已发布**的敏感字段类型。您可以查看已有的敏感字段类型是否满足需求： ○ 满足需求：则选择其他已有敏感字段类型，并单击右侧的图标进入**数据识别规则**页面修改原敏感字段类型对应的识别规则和修改后的敏感字段类型对应的识别规则，以保证后续识别的准确性。 ○ 不满足需求：您可以单击右侧的图标进入**数据识别规则**页面，或滑动至下拉框底部，单击**管理敏感字段类型**，默认跳转至**数据识别规则**页面并打开**新建敏感字段类型**弹窗，新增敏感字段类型，并配置识别规则。详情请参见：敏感数据识别。
批量修正数据	选中需要批量修正的字段，单击上图区域③的**批量修正**按钮，弹出**批量修正识别结果**对话框，**敏感字段类型**下拉框列表中为您展示当前租户下默认分类分级模板中**已发布**的敏感字段类型，你可以选择正确的敏感字段类型，单击**保存**，完成批量修正识别结果的操作。

管理识别结果

对于系统未识别到的数据，您可以单击右上角的**新增识别结果**手动添加识别结果，同时支持您单击**导出识别结果**导出筛选条件下的识别结果至本地。

- **新增识别结果**：在弹出的对话框中选择要新增的字段所在的引擎，并输入格式为**project.table.column**的字段GUID后，选择该字段对应的敏感字段类型（当前租户默认分类分级模板中**已发布**的敏感字段类型），单击**确定**，完成识别结果的导入。

- **导出识别结果**：单击**导出识别结果**自动为您导出当前筛选条件下的识别结果。

  > **说明** 最多支持导出10万条数据。

4.2.6. 数据脱敏管理

DataWorks目前支持动态脱敏和静态脱敏，本文为您介绍如何创建脱敏规则，并在DataWorks中进行脱敏查询。

前提条件

- 您需要购买DataWorks专业版及以上版本，才可以使用数据脱敏管理功能。详情请参见：DataWorks各版本详解。
- 您需要先开启DataWorks项目空间的查询脱敏功能，详情请参见工作空间配置。

背景信息

DataWorks目前支持动态脱敏和静态脱敏。

分类	概念	脱敏场景
动态脱敏	用户在查询敏感数据时在页面展示脱敏后的数据。	当前DataWorks为您内置了全局配置、展示脱敏、数据分析脱敏、底层脱敏等脱敏场景，子场景为动态脱敏的典型应用场景。
静态脱敏	将数据脱敏后存储到指定的数据库位置。	当前DataWorks为您内置了数据集成脱敏场景，子场景为静态脱敏的典型应用场景。

选择脱敏场景

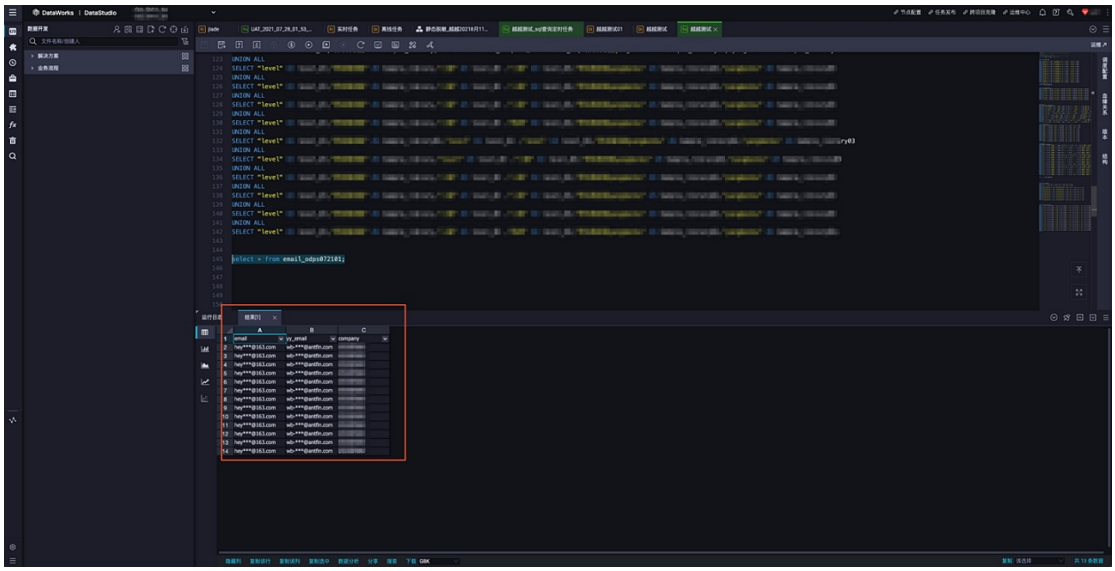
1. 进入数据保护伞。详情请参见：概述。
2. 在左侧导航栏，单击**规则配置 > 数据脱敏管理**。

   在**数据脱敏管理**页面根据需求选择**脱敏场景**。DataWorks为您提供了多种脱敏场景，还支持您自己创建脱敏场景。

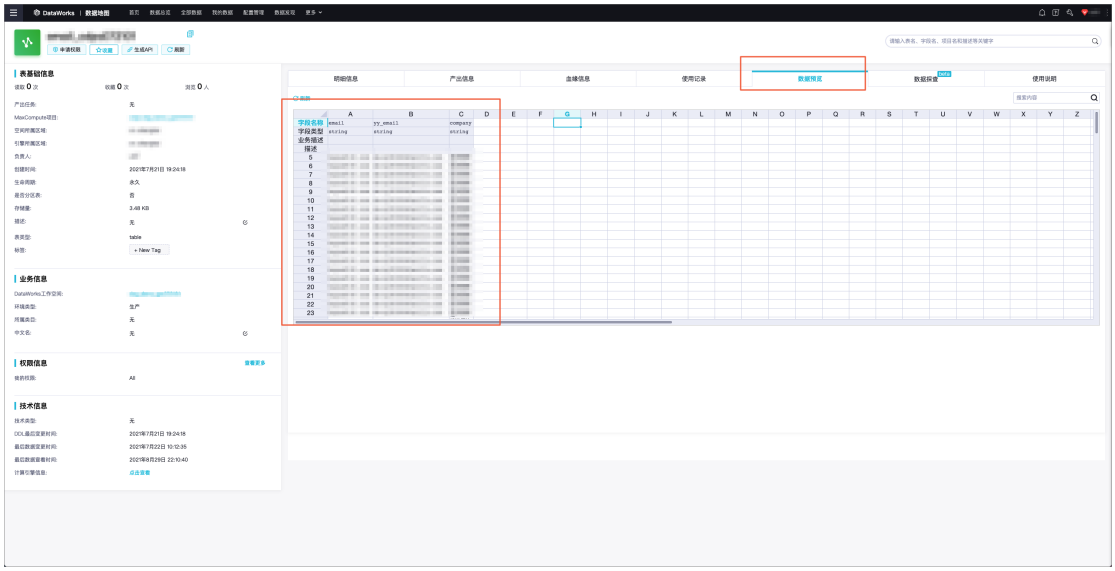
   - **全局配置**：全局配置的脱敏规则和白名单规则会在展示脱敏、数据分析脱敏、底层脱敏等场景中使

用。

- **DataWorks展示脱敏**：
  - 在**数据开发**页面查询数据时，查询的敏感数据（已配置脱敏规则）会经过脱敏。



  - 在**数据地图**的数据预览查询数据时，查询的敏感数据（已配置脱敏规则）会经过脱敏。



- **DataWorks数据分析脱敏**：在**数据分析**进行SQL查询和SQL Notes时，查询的敏感数据（配置了脱敏规则的）会经过脱敏。

- **Hologres展示脱敏**：在**数据开发**或**holostudio**查询holo的数据时按照配置进行脱敏。仅杭州、北京地域可以配置该脱敏规则。默认不开启该功能，如果您需要使用该功能，请提交工单进行开通。

  > **说明** Hologres暂不支持假名脱敏，您配置的全局假名脱敏规则，在Hologres场景下，将被脱敏为“***”。

- **MaxCompute底层脱敏**：从MaxCompute各个访问入口查询数据时，均会被脱敏。仅上海地域可以配置该脱敏规则。
- 新增自定义脱敏场景：在脱敏场景的下拉菜单底部单击**脱敏场景**，弹出**新增脱敏场景**，您可以自定义脱敏场景的**场景名称**（仅包含中文、英文、数字、下划线、中划线）和**场景码**（仅包含数字和英文）。

3. 创建脱敏规则。

   选择场景后，可以在对应场景下新建脱敏规则，便于后续应用的使用。不同场景下创建脱敏规则请参见：

   - 动态脱敏（全局配置、展示脱敏等场景）请参见：创建脱敏规则：全局配置场景。
   - 静态脱敏（数据集成场景）请参见：创建脱敏规则：数据集成场景。

创建脱敏规则：全局配置场景

下面以**全局配置（_default_scene_code）**为例，介绍数据脱敏配置的流程，**Hologres展示脱敏**、**DataWorks展示脱敏**和**DataWorks数据分析脱敏**、**MaxCompute底层脱敏**场景的操作步骤参考**全局配置**场景。

1. 在**数据脱敏管理**页面，选择**脱敏场景**为**全局配置（_default_scene_code）**。
2. （可选）选择脱敏对象并授权。

   > **说明** 仅**Hologres展示脱敏**和**MaxCompute底层脱敏**需要配置。

   单击**选择脱敏project**或**选择脱敏database**，根据界面提示选择需要脱敏的project或database，并勾选同意授权数据保护伞对该project或database脱敏选项。

3. 新建数据脱敏规则。

   i. 在**数据脱敏配置**页面，单击右上方的**新建脱敏规则**。

   ii. 在**新建脱敏规则**对话框中，配置**脱敏规则**和**脱敏方式**。

a. 配置基础信息。

参数	描述
敏感数据类型	你可以根据需要选择已创建的敏感数据类型（系统自动过滤掉已被当前脱敏场景使用过的敏感字段类型）。详情请参见：敏感数据识别。
脱敏规则名称	该文本框会自动代入用户填写的**敏感数据类型**（输入字符限制1~30字，包括：中文、英文、数字），您也可以在该敏感数据类型下新增脱敏规则名称，当与该租户下所有用户创建的脱敏规则出现重名时，提示**规则名称重复**。

b. 配置脱敏方式：DataWorks支持的脱敏方式包括：**保留格式加密**、**掩盖**、**HASH加密**、**字符替换**区间变换、**取整**、**置空**。

- **保留格式加密**（原假名脱敏算法）

  保留格式加密脱敏会将一个值替换成一个具有相同特征的脱敏信息。脱敏后数据和脱敏前数据的格式保持一致。

  - **数据水印**：数据水印可提供数据溯源能力，发生数据泄漏后，可以帮您定位到可能的泄漏源。您可以根据需求选择是否开启**数据水印**。

    > **说明** 仅DataWork企业版及以上版本支持使用数据水印功能。

  - **脱敏特征值**：默认选中5，可选范围0~9，不同脱敏特征值的脱敏策略规则不一致，即相同的待脱敏数据在不同的脱敏特征值脱敏出来的结果信息不一致。例如，原始数据为a123，脱敏特征值设置为0时，脱敏成b124，脱敏特征值设置为1时，脱敏成c234。原始数据相同时，如果脱敏特征值相同则脱敏后的数据也是相同的。
  - 当选择的**敏感数据类型**的识别规则为非内置时，用户需要配置**替换字符集**。

    替换字符集：非内置敏感数据类型需要配置该参数。遇到字符集中的字符，即会被替换为其他相同类型的字符，不支持中文，若需要脱敏的数据不符合字符集范围则不脱敏（可输入大写字母、小写字母、和数字，多个字符请用英文逗号隔开），例如，敏感数据脱敏前是0~3的数字和a~d的字母组成，那么脱敏后也会脱敏成在这个范围内的数字和字母。

- **掩盖**

  掩盖脱敏是对部分信息进行掩盖，将对应位置上的字符用“*”替换，达到脱敏的效果。

  - **推荐方式**：下拉框可选择**只展示前1位和最后1位**、**只展示前3位和最后2位**、**只展示前3位和最后4位**。
  - **自定义**：自定义提供了更加灵活的设置方式，按从左至右顺序配置分段是否脱敏，以及需要脱敏（或者不脱敏）的字符长度。最多可添加10个分段，必须要有且仅有1个分段是**剩余位数**。

图标	描述
①	可选择**位数**、**剩余位数**。
②	输入范围为【1，100】。
③	可选择**脱敏**、**不脱敏**。

例如，脱敏前3位，剩余位数不脱敏。

- **HASH**
  - **数据水印**：数据水印可提供数据溯源能力，发生数据泄漏后，可以帮您定位到可能的泄漏源。您可以根据需求选择是否开启**数据水印**。

    > **说明** 仅DataWork企业版及以上版本支持使用数据水印功能。

  - **加密算法**：包括MD5、SHA256、SHA512、SM3。
  - **加盐值**：设置各加密算法的盐值。默认选中5，可填值为0~9。

    > **说明** 在密码学中，通过在密码任意固定位置插入特定的字符串，让散列后的结果和使用原始密码的散列结果不相符，这种过程称之为加盐。盐值即插入的特定字符串。

- **字符替换**：将指定位置的字符按照您选择的替换方式进行替换。
  - **替换位置**：下拉框可选择**替换全部**、**替换前3位**、**替换后4位**，同时支持您自定义替换位置。

    替换位置选择自定义时，用户可以自定义分段，并配置每个分段如何替换字符，最多可添加10个分段，必须要有且仅有1个分段是**剩余位数**。

图标	描述
①	可选择**位数**、**剩余位数**。
②	输入范围为【1，100】。
③	可选择**替换方式**，包括**随机替换**、**样本值替换**、**固定值替换**。

  - **替换方式**：包括**随机替换**、**样本值替换**、**固定值替换**。
    - **随机替换**：随机替换对应位置上的字符，替换前后字符位数不变。
    - **样本值替换**：您需要选择指定样本库，选择后用样本库中的值替换对应位置上的字符。
    - **固定值替换**：您需要在**替换值**文本框中输入字符（字符不限，长度1~100，不可包含空字符），输入后用该替换值替换对应位置上的字符。
- **区间变换**：仅适用对数值类型的数据进行脱敏。可将指定数值范围内的数据脱敏为固定的值，可添加多个区间范围，至少1个，至多10个。
  - **原始数值范围 [m,n)**：脱敏前数据的数值范围，有效值为大于等于0的数值，最多支持小数点后2位。
  - **脱敏后数值**：脱敏之后的值，有效值为大于等于0的数值，最多支持小数点后2位；
- **取整**
  - **原始数据类型**：仅支持选择数值类型。
  - **保留小数点位数**：有效值范围为0~5，剩余部分四舍五入。例如，原始数值3.1415，保留小数点位数2位，脱敏后为3.14。
- **置空**：脱敏时，对应的敏感字段置为空字符串。

c. 验证脱敏配置结果：您可以在**样本数据**文本框中输入脱敏前样本数据（输入字符限制0~100字符），单击**脱敏验证**，在**脱敏效果**中会返回脱敏后的数据。

iii. 单击**保存**。

iv. 在**数据脱敏配置**页面，设置脱敏策略的状态为**生效**或**失效**。

设置成功后，单击相应脱敏规则后的操作列的图标，可以执行**删除**脱敏规则、**修改**脱敏规则和**查询详情**的操作。

> 说明
>
> - 生效的规则不允许执行**删除**和**修改**的操作。您需要先将规则**失效**，失效时判断是否有相关任务使用到该规则，请联系安全管理员二次确认；
> - **失效**状态下您可以修改脱敏方式，但是**敏感数据类型**和**脱敏规则名称**不可修改。
> - 修改完成后开启**生效**，配置该脱敏规则的任务可继续脱敏。

4. 新增白名单。

i. 单击菜单栏中的**白名单配置管理**。

ii. 在**白名单配置管理**页面，单击右上方的**新增白名单**。

iii. 在**新增白名单**对话框中，配置相关信息。

> 说明
>
> - Hologres展示脱敏场景不需要配置白名单。
> - 设置白名单生效时间后，对于符合白名单条件的敏感数据，将在指定有效期内不进行脱敏处理。
> - 白名单条件不可以全部设置为**全部**。


新增白名单
基础信息
* 白名单名称
请输入
* 分级
全部
* 分类
全部
* 用户组
全部
* 生效时间
短期
永久
30天
2022-05-25
2022-06-23
高级设置
收起
敏感字段类型
全部
项目空间范围
全部引擎
全部项目
数据表范围
字段范围
取消
保存

a. 配置基础信息。

参数	描述
白名单名称	您可以输入白名单名称（输入字符限制1~30字，不可输入特殊字符）。
分级	您可以单击右侧下拉框选择内置或所有用户创建的分级。配置分类分级详情请参见：数据分类分级。
分类	您可以单击右侧下拉框选择内置或所有用户创建的分类。
用户组	您可以选择在**管理用户组**中已配置的用户组，最多可选50个用户组。添加用户组至白名单后，用户组内的账号获取到的数据为脱敏前的原始数据。用户组管理配置详情请参见：创建并管理用户组。
生效时间	您可以设置白名单生效时间。设置后，如果不在白名单脱敏时间的区间内，该用户在查询该敏感信息时将会继续脱敏。 **说明** 设置为短期后，表示从当前时间开始到指定天数内的数据将不进行脱敏。

b. 高级设置。

参数	描述
敏感字段类型	您可以右侧下拉框选择已创建的敏感数据类型（包括内置和所有用户创建的敏感数据类型）。
项目空间范围	您可以选择当前地域支持的引擎及引擎中的项目空间。不选择代表全部。 **说明** 项目空间仅支持选择当前账号有权限的项目空间。
数据表范围	您可以填写数据范围。不填写代表全部。 **说明** 支持 .*通配符，例如，.*name表示以name为后缀，private.*表示以private为前缀，多个数据之间请用英文逗号分隔，字符总长度不超过100。
字段范围	您可以填写字段范围，不填写代表全部。 **说明** 支持 .*通配符，，例如，.*name表示以name为后缀，private.*表示以private为前缀，多个字段之间请用英文逗号分隔，字符总长度不超过100。

iv. 单击**保存**完成白名单配置。

5. 后续步骤：创建完成脱敏规则后，您在数据开发、数据预览、数据分析等页面查询数据时，查询的敏感数据（已配置脱敏规则）会经过脱敏。详情请参见脱敏场景。

创建脱敏规则：数据集成场景

1. 在**数据脱敏管理**页面，选择**脱敏场景**为DataWorks数据集成脱敏（dataworks_data_integration_desense_code）。
2. 新建数据脱敏规则。
   i. 在**数据脱敏配置**页面，单击右上方的**新建脱敏规则**。
   ii. 在**脱敏规则**对话框中，选择需要设置的**敏感数据类型**、**脱敏规则名称**、**脱敏方式**、**安全域**和**替换字符集**。

a. 配置基础信息

参数	描述
敏感数据类型	■ 默认为**选择已有**，右侧下拉框选择已创建的敏感数据类型（包括内置和所有用户创建的敏感数据类型）。你可以根据需要选择已创建的敏感数据类型。 ■ 可切换**新增类型**，右侧输入框可输入敏感数据类型名称（输入字符限制1~30字，包括：中文、英文、数字）。 用户输入**新增敏感数据类型**，系统会判断文字与已有敏感数据类型名称是否相同（包括：内置和该租户下所有用户配置的敏感数据类型），如果名称相同则提示**敏感字段类型重复**。 **说明** 内置敏感数据类型：手机号、身份证号、银行卡号、邮箱_内置、IP、车牌号、 邮政编码、座机号、MAC地址、地址、姓名、公司名、民族、星座、性别、国籍。
脱敏规则名称	该文本框会自动代入用户填写的**敏感数据类型**（输入字符限制1~30字，包括：中文、英文、数字），您也可以在该敏感数据类型下新增脱敏规则名称，当与该租户下所有用户创建的脱敏规则出现重名时，提示**规则名称重复**。

b. 配置脱敏方式与规则：DataWorks支持的**脱敏方式**包括**假名**、**HASH**和**掩盖**三种方式。

■ **假名**

假名脱敏会将一个值替换成一个具有相同特征的脱敏信息。脱敏后数据和脱敏前数据的格式保持一致。

■ 当选择的**敏感数据类型**为内置敏感数据类型（手机号、身份证号、银行卡号、邮箱_内置、IP、车牌号、 邮政编码、座机号、MAC地址、地址、姓名、公司名）时，用户需要配置**安全域**。

安全域：可选范围0~9，不同安全域的脱敏策略规则不一致，即相同的待脱敏数据在不同的安全域脱敏出来的结果信息不一致。例如，原始数据为a123，安全域设置为0时，脱敏成b124，安全域设置为1时，脱敏成c234。原始数据相同时，如果安全域相同则脱敏后的数据也是相同的。

■ 当选择的**敏感数据类型**为非内置时，用户需要配置**替换字符集**。

替换字符集：遇到字符集中的字符，即会被替换为其他相同类型的字符，不支持中文，若需要脱敏的数据不符合字符集范围则不脱敏（可输入大写字母、小写字母、和数字，多个字符请用英文逗号隔开），例如，敏感数据脱敏前是0~3的数字和a~d的字母组成，那么脱敏后也会脱敏成在这个范围内的数字和字母。

■ **哈希**

可将原始数据加密成固定长度的数据。HASH脱敏方式需要选择**安全域**。

安全域：可选范围0~9，不同安全域的脱敏策略规则不一致，即相同的待脱敏数据在不同的安全域脱敏出来的结果信息不一致。例如，原始数据为a123，安全域设置为0时，脱敏成b124，安全域设置为1时，脱敏成c234。原始数据相同时，如果安全域相同则脱敏后的数据也是相同的。

- **掩盖**

  掩盖脱敏是对部分信息进行掩盖，将对应位置上的字符用“*”替换，达到脱敏的效果。

  - **推荐方式**：下拉框可选择**只展示前1位和最后1位**、**只展示前3位和最后2位**、**只展示前3位和最后4位**。
  - **自定义**：自定义提供了更加灵活的设置方式，按从左至右顺序配置分段是否脱敏，以及需要脱敏（或者不脱敏）的字符长度。最多可添加10个分段，至少要有1个分段是**剩余位数**。

图标	描述
①	可选择**位数**、**剩余位数**。
②	输入范围为【1，100】。
③	可选择**脱敏**、**不脱敏**。

例如，脱敏前3位，剩余位数不脱敏。

例如，脱敏后3位，剩余位数不脱敏。

c. 验证脱敏配置结果：您可以在**样本数据**文本框中输入脱敏前样本数据（输入字符限制0~100字符），单击**脱敏验证**，在**脱敏效果**中会返回脱敏后的数据。

iii. 单击**确定**。

iv. 在**数据脱敏配置**页面会新增一条脱敏规则，设置脱敏策略的状态为**生效**或**失效**。

设置成功后，单击相应脱敏规则后的操作列的图标，可以执行**删除**脱敏规则、**修改**脱敏规则和**查询详情**的操作。

> 说明
>
> - 生效的规则不允许执行**删除**和**修改**的操作。您需要先将规则**失效**，失效时判断是否有相关任务使用到该规则，请联系安全管理员二次确认；
> - **失效**状态下您可以修改脱敏方式，但是**敏感数据类型**和**脱敏规则名称**不可修改。
> - 修改完成后开启**生效**，配置该脱敏规则的任务可继续脱敏。

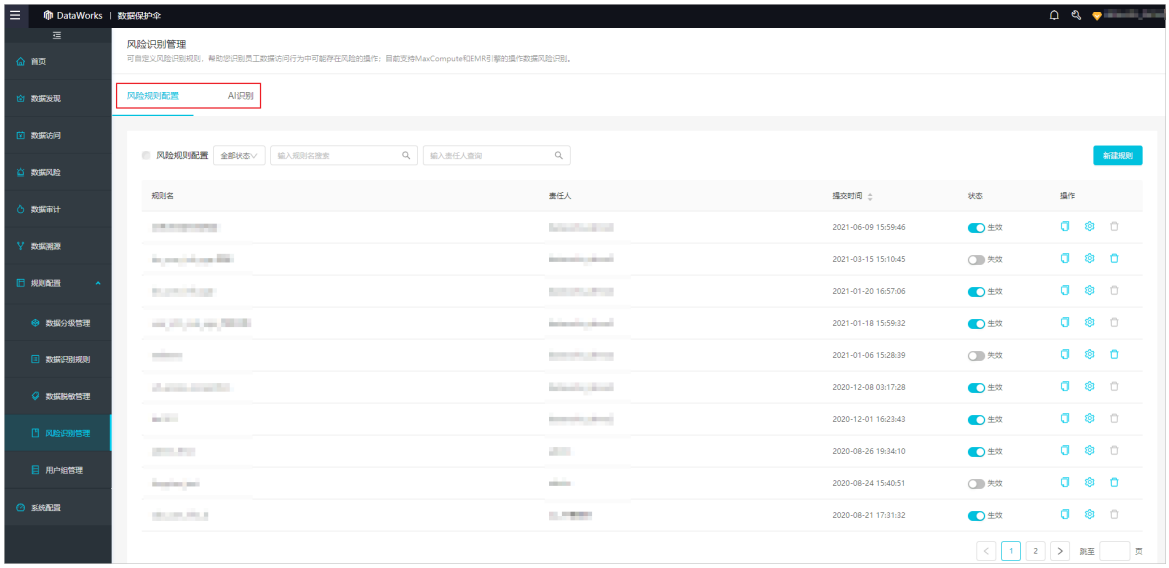
3. 后续步骤：创建数据集成脱敏规则后，您可以在创建实时同步单表数据任务的时候使用该脱敏规则。详情请参见配置数据脱敏。

4.2.7. 风险识别管理（旧版）

风险识别管理页面提供风险数据的规则配置，可以帮助您识别日常访问中的风险。同时，您可以通过启动数据风险的AI识别功能自动识别数据风险。

识别后的风险数据统一在**数据风险**页面进行展示和审计操作，同时会在**数据访问**页面相应的数据后打上识别标志。

1. 登录DataWorks控制台，单击相应工作空间后的**进入数据开发**。
2. 单击左上方的图标，选择**全部产品 > 数据治理 > 数据保护伞**。
3. 单击**立即体验**，默认进入数据保护伞的**首页**。
4. 单击左侧导航栏中的**规则配置 > 风险识别管理**。您可以在该页面新建、复制、配置和删除风险规则，并可以设置AI识别。



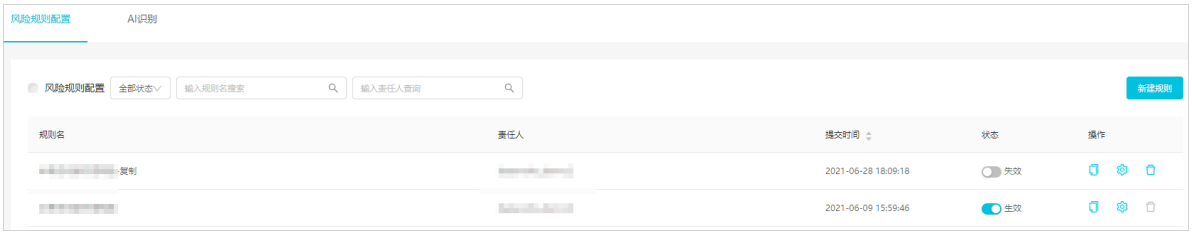
风险规则配置

- **新建规则**

  单击右上角的**新建规则**，在**新建规则**对话框中，输入**规则名**、**负责任人**和**备注**，单击**确定**。

- 复制规则

单击相应规则后的图标，即可生成1个完全一致的规则。



复制的规则默认状态为**失效**，您可以根据自身需求进行配置。

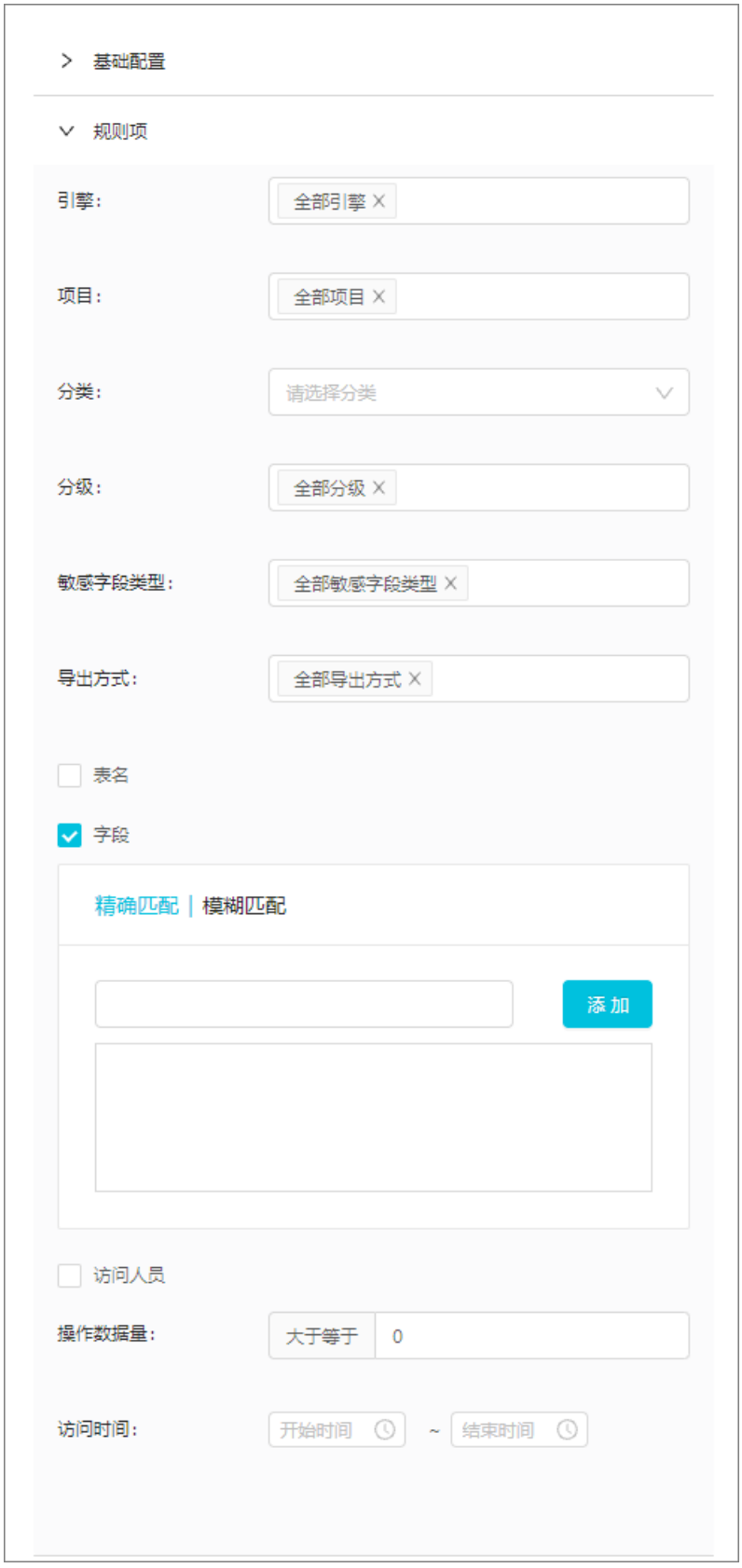
- 配置规则

  如果需要修改已有的规则，操作如下：

  i. 设置相应规则的状态为**失效**。

  ii. 单击相应规则后的图标。

  iii. 在右侧的**规则配置**对话框中，修改**基础配置**和**规则项**。



iv. 修改完成后，单击**保存**。

v. 确认规则无误后，更改状态为**生效**。

- 删除规则

如果您需要删除规则，单击相应规则后的图标，在对话框中单击**删除**即可。

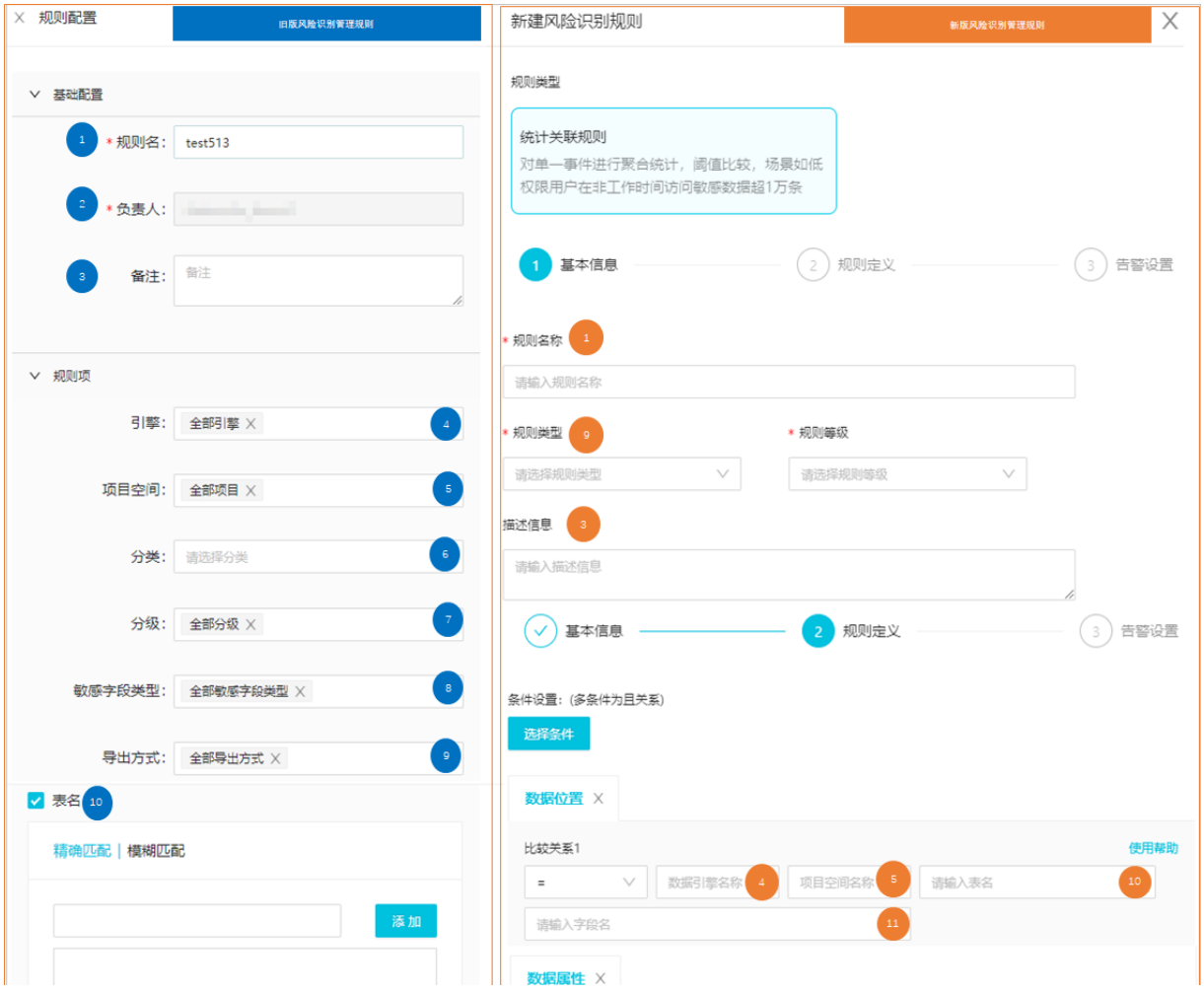
AI识别

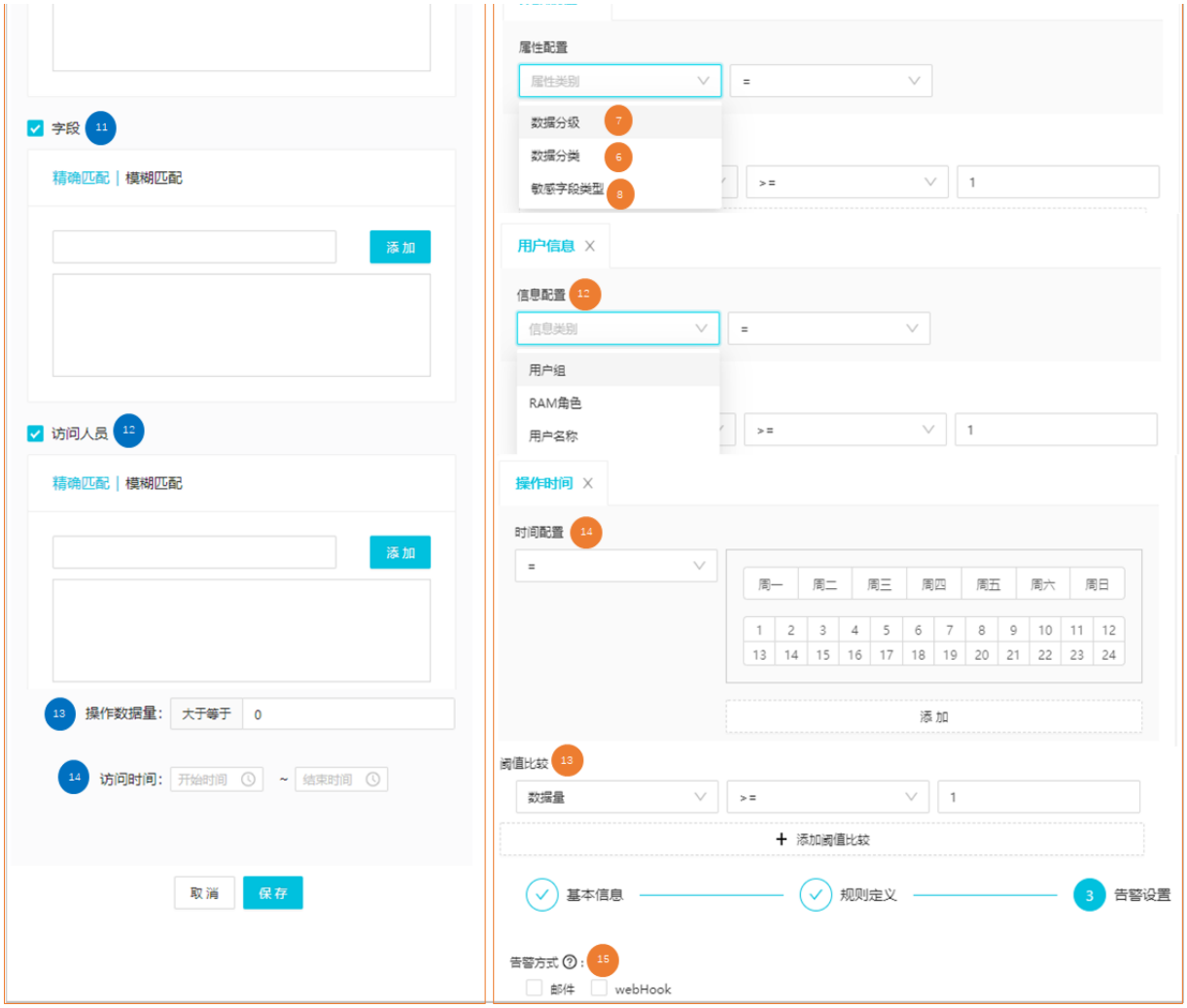
单击**风险识别管理 > AI识别**，该页面仅支持相似SQL查询。

设置相应名称后的**状态**为**生效**，即可开启AI识别。

> **说明**
> - 生效当前规则后，第2天在风险数据中展示命中该条规则的SQL查询数据。
> - 启动后可以更改状态为**失效**，不会删除之前识别的数据。

新旧版风险识别规则配置项位置对比说明





新旧版本风险识别规则配置项的配置位置存在差异，具体如下表。

> **说明** 新版风险识别规则的配置详情，请参见新建风险识别规则，旧版风险识别规则的配置详情，请参见风险规则配置。

序号	风险识别规则配置项	旧版配置位置	新版配置位置
1	规则名称	**基础配置 > 规则名**	**基本信息 > 规则名称**
2	规则责任人	**基础配置 > 负责人**。 默认规则的负责人为当前登录账号。	无该配置项，DataWorks自动记录规则责任人。
3	规则的描述信息	**基础配置 > 备注**	**基本信息 > 描述信息**
4	规则生效的引擎	**规则项 > 引擎**	**规则定义 > 条件设置**区域，**选择条件**选择**数据位置**时，所配置的**数据引擎名称**。
5	规则生效的项目空间	**规则项 > 项目空间**	**规则定义 > 条件设置**区域，**选择条件**选择**数据位置**时，所配置的**项目空间名称**。

序号	风险识别规则配置项	旧版配置位置	新版配置位置
6	规则所属的分类	**规则项 > 分类**	**规则定义 > 条件设置**区域，**选择条件**选择**数据属性**时，属性类别选择**数据分类**。
7	规则所属的分级	**规则项 > 分级**	**规则定义 > 条件设置**区域，**选择条件**选择**数据属性**时，属性类别选择**数据分级**。
8	命中规则的敏感字段类型	**规则项 > 敏感字段类型**	**规则定义 > 条件设置**区域，**选择条件**选择**数据属性**时，属性类别选择**敏感字段类型**。
9	规则所属的操作类型	**规则项 > 导出方式** 取值包括： • **全部导出方式** • **Tunnel下载** • **表访问**	**基本信息 > 规则类型** 取值包括： • **数据访问风险** • **数据导出风险** • **数据操作风险** • **其他**
10	规则生效的表	**规则项 > 表名**	**规则定义 > 条件设置**区域，**选择条件**选择**数据位置**时，所配置的**表名**。
11	规则生效的字段	**规则项 > 字段**	**规则定义 > 条件设置**区域，**选择条件**选择**数据位置**时，所配置的**字段名**。
12	用于配置哪些用户访问时会触发该规则。	**规则项 > 访问人员**	**规则定义 > 条件设置**区域，**选择条件**选择**用户信息**时，所选择的**信息类别**。
13	规则限制的操作数据量，当超出限制的数据量范围时命中规则。	**规则项 > 操作数据量**	在**规则定义 > 条件设置**区域，**阈值比较**下配置**数据量**范围。
14	规则限制的访问时间范围，当访问操作与该时间范围匹配时命中规则。	**规则项 > 访问时间**间	**规则定义 > 条件设置**区域，**选择条件**选择**操作时间**时，所设置的**时间配置**。
15	触发规则后的告警方式	不支持	**告警设置 > 告警方式**

4.2.8. 风险识别管理（新版）

风险识别管理提供了多维度的关联分析及算法，智能化的分析技术帮助您通过风险识别规则，主动发现风险操作并预警，使用可视化方式进行一站式审计。DataWorks内置了多种场景的风险识别规则，您可以直接使用，也可以根据业务场景自定义规则。本文为您介绍如何创建并管理风险识别规则。

背景信息

数据输入DataWorks后会经过数据保护伞进行过滤处理，旧版风险识别管理的风险识别功能仅当涉及敏感数据时才会被识别为风险，不支持操作审计相关场景及事件统计聚合场景的识别。因此，DataWorks为解决该问题，为您提供了功能更加全面的新版风险识别管理功能。具体如下：

- 易用性好

  包含**数据访问风险**、**数据导出风险**、**数据操作风险**、**其他风险类型**等4类风险类型，并支持**访问时间**、**敏感类型**、**访问量**等多种维度组合识别各类风险。

- 精准度高

  增加事件聚合统计比较，通过比较时间窗口内事件发生次数的阈值，更精准识别风险，减少大量误报。例如，在10分钟内发生相同事件3次以上才会命中风险。

- 精细化管理

  支持配置**高**、**中**、**低**的风险级别，根据风险级别精细化管理风险。

- 规则灵活

  内置了多种场景的常用规则，可以直接使用；同时，您也可以基于业务需求，自定义风险识别规则。详情请参见系统内置风险识别规则及新建风险识别规则。

新旧版本风险识别规则配置项的配置位置存在差异，详情请参见新旧版风险识别规则配置项位置对比说明。

使用限制

- 版本限制
  - 仅DataWorks专业版及以上版本支持使用新版风险识别管理功能。
  - 仅DataWorks企业版及以上版本支持内置风险识别规则。
- 新旧版本切换
  - 旧版风险识别管理运行的时间将保留至2022年06月30日（请以界面实际显示的保留时间为准），保留时间到期后，已创建的风险识别规则及相关风险数据将自动清除，后续则只能使用新版风险识别管理功能。请您及时将需要使用的规则及风险数据导出备份，导出备份操作，详情请参见风险识别管理（旧版）。
  - 新版风险识别管理在旧版保留时间段内会同时运行，您可以切换至新版使用。切换至新版后，旧版中已创建的风险识别规则及风险数据不会同步至新版，您需要在新版中重新创建相关规则。
- 告警方式

  仅支持邮件和WebHook告警方式。

  > **说明** DataWorks支持使用钉钉群、企业微信和飞书的WebHook地址。其中，仅企业版及以上版本支持推送报警信息至企业微信或飞书。

进入风险识别管理

1. 进入**数据保护伞**。
   i. 登录DataWorks控制台。
   ii. 在左侧导航栏，单击**工作空间列表**。
   iii. 选择工作空间所在地域后，单击相应工作空间后的**数据开发**。
   iv. 单击左上方的☰图标，选择**全部产品 > 数据治理 > 数据保护伞**。
   v. 单击**立即体验**，进入数据保护伞。

2. 进入**风险识别管理**。

   在**数据保护伞**页面的左侧导航栏，单击**规则配置 > 风险识别管理**，默认进入旧版风险识别管理页面。您需要单击**体验新版本**切换至新版，使用新版风险识别管理功能创建并管理风险识别规则。

   新版风险识别管理内置了多种场景的常用规则，可以直接使用；同时，您也可以基于业务需求，自定义风险识别规则。详情请参见系统内置风险识别规则及新建风险识别规则。

系统内置风险识别规则

新版风险识别管理功能支持的内置规则如下表。

规则名称	规则类型	规则等级	规则配置
非工作时间查询大数据量敏感数据	**数据访问风险**	**低**	如下时间段查询数据量大于10000时命中该规则。 • 周一至周五：`22:00~24:00`。 • 周六至周日：`00:00~24:00`。
相似SQL查询	**数据访问风险**	**低**	十分钟内查询相似SQL大于等于10次时，命中该规则。
批量查询大量敏感数据	**数据访问风险**	**中**	单次查询数据量大于10000时命中该规则。
批量导出大量敏感数据	**数据导出风险**	**高**	单次导出数据量大于10000时命中该规则。
非工作时间导出大数据量敏感数据	**数据导出风险**	**高**	如下时间段导出数据量大于10000时命中该规则。 • 周一至周五：`22:00~24:00`。 • 周六至周日：`00:00~24:00`。

新建风险识别规则

1. 新建规则的规划和准备工作。

   您可以基于实际场景，选择从**数据位置**、**数据属性**、**用户信息**、**操作时间**等维度识别风险数据，配置更精细的风险识别条件。当使用**数据属性**及**用户信息**的不同细分类别配置风险识别条件时，需要进行如下准备工作。

风险识别维度	细分类别	描述
数据属性	**数据分级**	识别指定级别的风险数据，您需要提前定义数据的分级，详情请参见数据分类分级。
	数据分类	识别指定类别的风险数据，您需要提前定义数据的类别，详情请参见敏感数据识别。
	敏感字段类型	识别指定敏感字段的风险数据，您需要提前定义敏感字段类型，详情请参见敏感数据识别。
用户信息	**用户组**	识别当前登录账号下指定用户组的风险数据，您需要提前配置用户组，详情请参见创建并管理用户组。
	RAM角色	识别当前登录账号下RAM用户的风险数据，您需要提前在当前阿里云账号下添加RAM用户，详情请参见创建RAM用户。

2. 在**风险识别管理**页面右上角，单击**+风险识别规则**。
3. 在**新建风险识别规则**对话框配置规则信息。

> **说明** 当前仅支持新建**统计关联规则**。**统计关联规则**用于对单一事件进行聚合统计，阈值比较，当事件数量超过设置的阈值时，则命中规则。例如，设置的规则为低权限用户在非工作时间访问敏感数据超过1万条。

i. 配置规则基本信息。

参数	描述
规则名称	新建的风险识别规则名称。名称长度为1~30字符，并且不能包含特殊字符。
规则类型	风险识别规则的类型，取值如下： ■ **数据访问风险**：访问数据时可能存在风险。 ■ **数据导出风险**：导出数据时可能存在风险。 ■ **数据操作风险**：创建、修改、删除数据时可能存在风险。 ■ **其他**：其他类型。
规则等级	风险识别规则的级别。包含**低**、**中**、**高**等级。您可以根据实际需求，将识别重要数据信息的规则设置为高等级。
描述信息	风险识别规则的描述信息。长度为1~100字符。

ii. 单击**下一步**。

iii. 配置风险识别条件及阈值。

- 配置风险识别的条件。

DataWorks支持设置从**数据位置**、**数据属性**、**用户信息**、**操作时间**等维度识别风险数据，帮助您基于实际场景配置更精细的风险识别条件。

> **说明** 当前最多支持添加10个条件。单击所选维度中的**+添加比较关系**即可添加多个识别条件，并且添加的多个条件逻辑关系为**且**。

- **数据位置**

  用于设置识别风险数据的位置范围，可以精确到字段。

参数	描述	是否必配
是否过滤所选位置	选择是否过滤所选位置的风险数据。取值如下： ■ ≠：表示过滤目标位置，即配置的风险识别规则不会识别所选位置中的风险数据。 ■ =：表示仅识别目标位置，即配置的风险识别规则仅识别所选位置中的风险数据。	是
数据引擎名称	选择识别规则限定的引擎范围。 **说明** ■ 当前仅支持识别MaxCompute引擎中的风险数据。 ■ 每条比较关系只能选择一个引擎，如果您需要限定多个引擎范围，则请单击**+添加比较关系**配置多条风险识别条件。	是

参数	描述	是否必配
项目空间名称	选择识别规则限定的目标项目空间。**项目空间名称**需要配置为所选引擎中的项目空间，您可以在下拉列表选择，也可以输入项目空间名称进行搜索。 **说明** ■ 下拉列表最多显示100个项目空间名称。 ■ 名称搜索支持模糊匹配，即输入关键词，即可搜索到名称包含关键词的项目。 ■ 每条比较关系只能选择一个项目空间，如果您需要限定多个项目空间范围，则请单击**+添加比较关系**配置多条风险识别条件。	是
表名	输入识别规则限定的目标表。您可以输入单个或多个表名称，多个表名称之间使用英文逗号（,）分隔。输入表名称的说明如下： ■ 单个表名不超过30个字符，输入的表名整体不超过100字符。 ■ 支持使用 `*` 通配符。例如，`*name` 表示识别所有名称为 `name` 后缀的表数据。	否。不配置默认识别所选项目空间中所有表的风险数据。
字段名	输入识别规则限定的目标字段。您可以输入单个或多个字段名称，多个字段名称之间使用英文逗号（,）分隔。输入字段名称的说明如下： ■ 单个字段名不超过30个字符，输入的字段名整体不超过100字符。 ■ 支持使用 `*` 通配符。例如，`*name` 表示识别所有名称为 `name` 后缀的字段数据。	否。不配置默认识别所有表字段的风险数据。

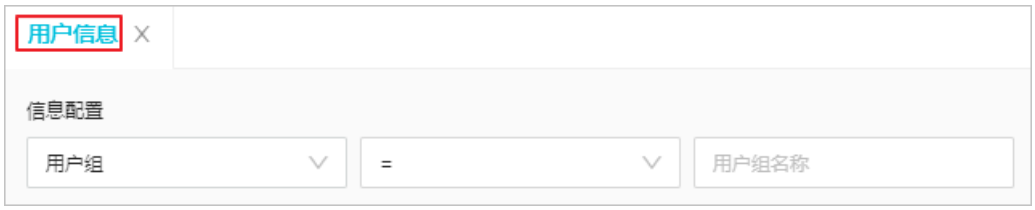
- **数据属性**

  用于筛选识别风险数据的属性范围。

参数	描述
属性	您可以根据业务需求选择识别风险数据的属性类别。当前支持的属性类别如下： ■ **数据分级**：用于指定识别哪个级别的风险数据。您需要提前定义数据的分级，详情请参见数据分类分级。 ■ **数据分类**：用于指定识别哪个类别的风险数据。您需要提前定义数据的类别，详情请参见敏感数据识别。 ■ **敏感字段类型**：用于指定识别哪类敏感字段的风险数据。您需要提前定义敏感字段类型，详情请参见敏感数据识别。
是否过滤所选属性	选择是否过滤所选属性的风险数据。取值如下： ■ **≠**：表示过滤目标属性，即配置的风险识别规则不会识别所选属性中的风险数据。 ■ **=**：表示仅识别目标属性，即配置的风险识别规则仅识别所选属性中的风险数据。

- **用户信息**

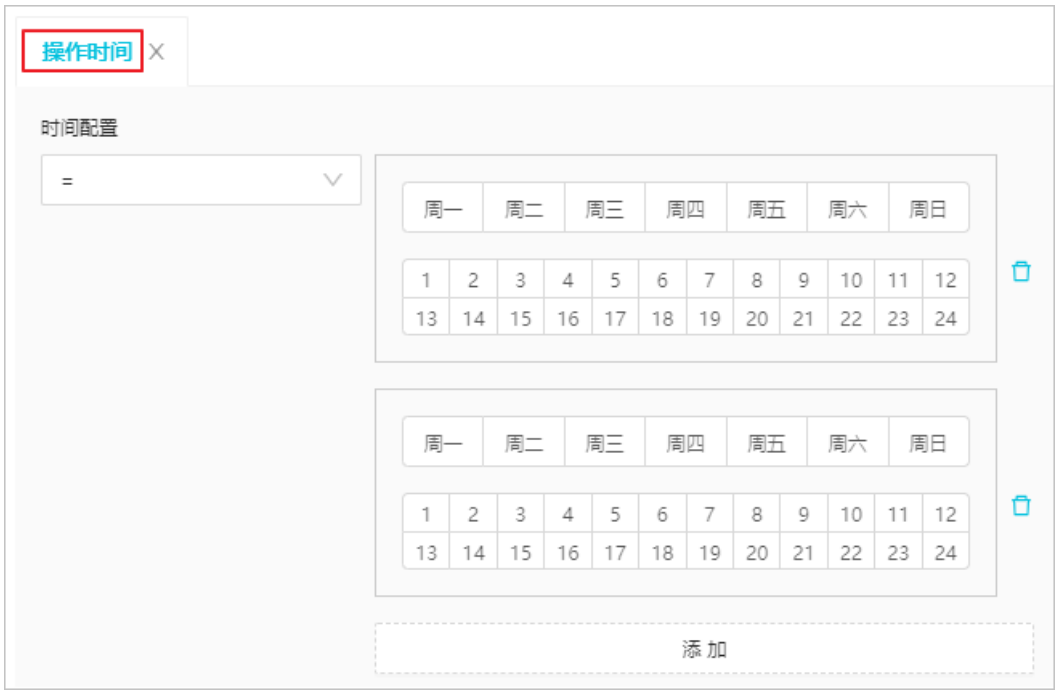
  用于筛选识别风险数据的用户信息范围。



参数	描述
信息类别	选择识别风险数据的用户信息类别，取值如下： ■ **用户组**：当前登录账号下的用户组名称。您需要提前配置用户组，详情请参见创建并管理用户组。 ■ **RAM角色**：当前登录账号下的RAM用户。您需要提前在当前阿里云账号下添加RAM用户，详情请参见创建RAM用户。 ■ **用户名称**：当前登录用户。
是否过滤所选用户信息	■ **≠**：表示过滤目标用户信息，即配置的风险识别规则不会识别所选用户信息中的风险数据。 ■ **=**：表示仅识别目标用户信息，即配置的风险识别规则仅识别所选用户信息中的风险数据。

- **操作时间**

  用于筛选识别风险数据的操作时间范围。



参数	描述
选择时间范围	单击目标星期及小时，即可选择所需的时间范围。您可以选择周一至周日的任意时间，精确到小时。
是否过滤所选时间	■ ≠：表示过滤目标操作时间，即配置的风险识别规则不会识别所选操作时间中的风险数据。 ■ =：表示仅识别目标操作时间，即配置的风险识别规则仅识别所选操作时间中的风险数据。

- 配置阈值。

  DataWorks支持事件聚合统计，您也可以选择通过比较时间窗口内事件发生次数的阈值识别风险数据。单击**+添加阈值比较** 即可配置多个风险识别阈值条件。

参数	描述
阈值类别	■ **数据量**：通过操作的数据量大小识别风险数据。当操作的数据量超过设置的阈值时，则此次操作命中风险。数据量取值为1至1000万的整数，单位为条。默认取值为1。 ■ **发生次数**：通过指定时间范围单一事件发生的次数识别风险数据。当在指定时间范围内，某单一事件的发生次数超过设置的阈值时，则命中风险。发生次数取值为1至10000的整数，单位为次。默认取值为10。 **说明** DataWorks自动帮您归类识别单一事件。
时间窗口	事件发生次数限定的时间范围，默认为10分钟。取值如下： ■ **分钟**：取值范围为 `1~59` 。 ■ **小时**：取值范围为 `1~23` 。 ■ **天**：取值范围为 `1~7` 。 **说明** 仅当**阈值类别**取值为**发生次数**，时需要配置该参数。

iv. 单击**下一步**。

v. 配置告警方式。

在识别到数据风险后，您可以根据配置的告警方式，及时接收到告警信息，以便对风险进行相应的处理。支持您选择**邮件**和**webHook**的告警方式。

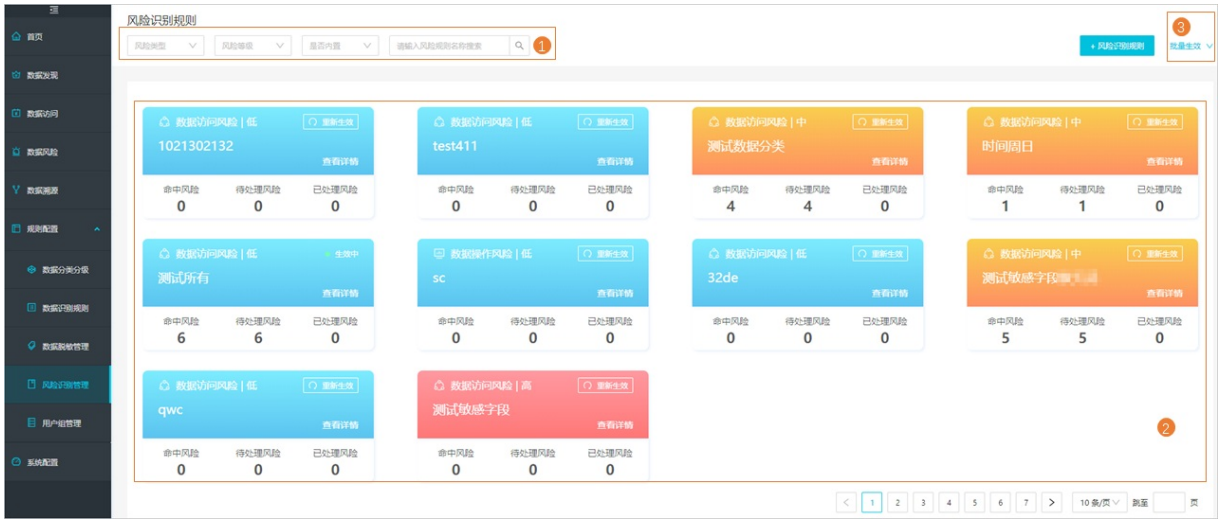
> **说明** 选择告警方式前请确保已在系统配置中完成邮箱及webHook的相关配置。

vi. 单击**保存**，规则创建完成。

创建的自定义规则默认不生效，您需要在**风险识别规则**页面单击目标规则的**重新生效**，手动生效规则。

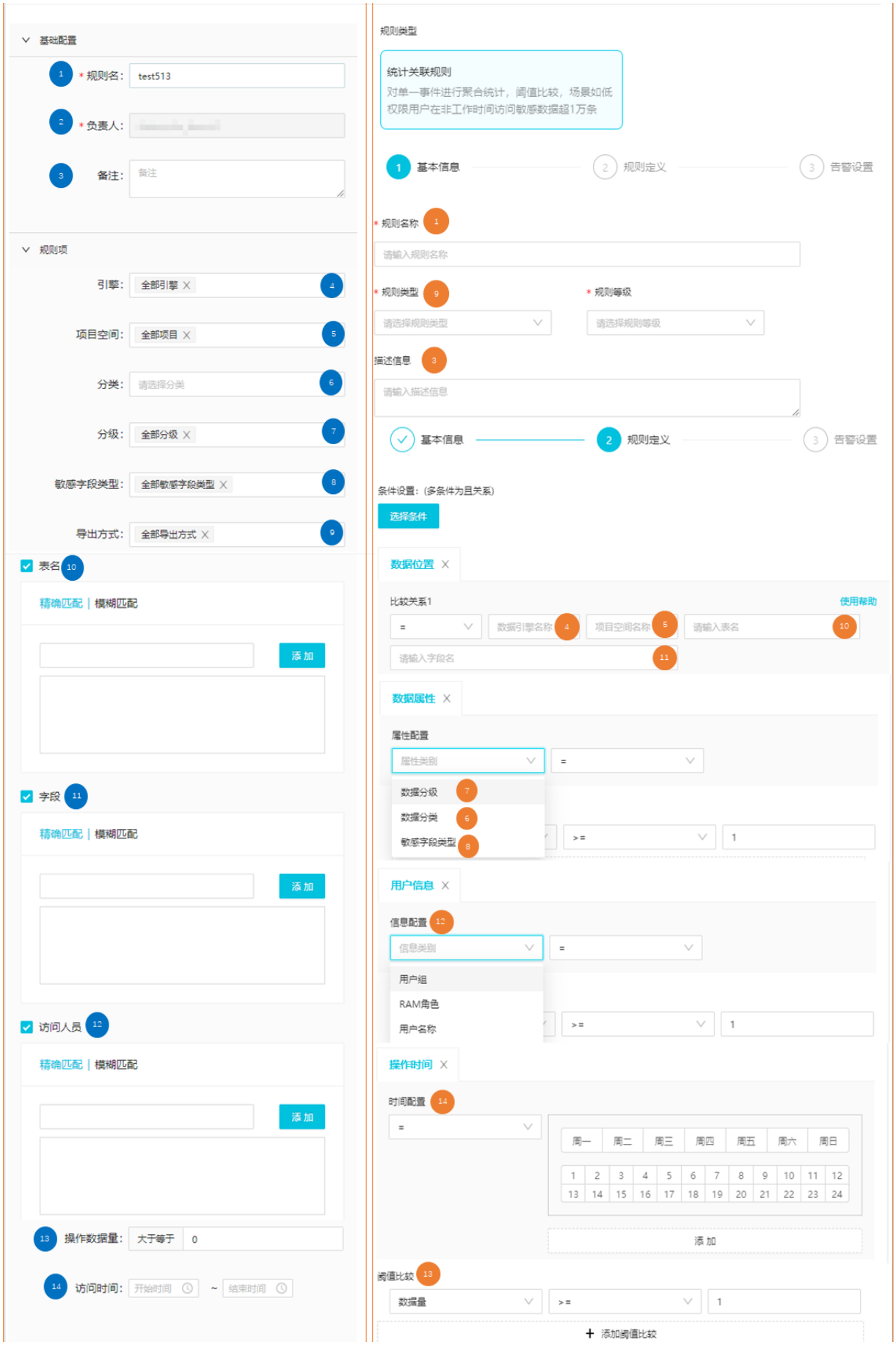
管理风险识别规则

在**风险识别管理**页面，您可以查看已创建的规则列表及规则详细信息。同时，也可以编辑修改目标规则，或批量操作多个规则。



区域	描述
1	在该区域，您可以通过**风险类型**、**风险等级**、**风险类型**、**是否内置**、**风险规则名称**等条件进行筛选，查看目标条件的规则列表。 **说明** 通过名称搜索识别规则时支持模糊匹配，输入关键词即可搜索名称中包含关键词的风险识别规则。
2	在该区域，您可以执行如下操作： • **查看规则基本信息**：查看已创建规则的风险类型、风险等级、生效状态等基本信息，以及该规则识别到的风险数据。您可以基于**命中风险**、**待处理风险**、**已处理风险**，了解当前租户存在的风险及风险处理情况。 • **查看规则详情并编辑规则**：单击**查看详情**，即可查看规则的详细配置信息，您也可以根据实际需求修改规则。 • **重新生效规则**：单击重新生效图标，即可使失效的规则重新生效。 **说明** DataWorks仅支持失效状态的规则执行该操作。
3	在该区域，您可以批量操作目标规则。当前支持执行**批量生效**、**批量失效**、**批量删除**等批量操作，单击图标，即可切换批量操作类别。 **说明** DataWorks不支持删除系统内置的风险识别规则，仅支持删除**失效**状态的自定义规则。

新旧版风险识别规则配置项位置对比说明


基础配置
规则名：test513
负责人：
备注：
规则项
引擎：全部引擎
项目空间：全部项目
分类：请选择分类
分级：全部分级
敏感字段类型：全部敏感字段类型
导出方式：全部导出方式
表名
精确匹配 | 模糊匹配
添加
字段
精确匹配 | 模糊匹配
添加
访问人员
精确匹配 | 模糊匹配
添加
操作数据量：大于等于 0
访问时间：开始时间 ~ 结束时间
规则类型
统计关联规则
对单一事件进行聚合统计，阈值比较，场景如低权限用户在非工作时间访问敏感数据超1万条
1 基本信息
2 规则定义
3 告警设置
规则名称
请输入规则名称
规则类型
请选择规则类型
规则等级
请选择规则等级
描述信息
请输入描述信息
基本信息
2 规则定义
3 告警设置
条件设置：(多条件为且关系)
选择条件
数据位置
比较关系1
使用帮助
数据引擎名称
项目空间名称
请输入表名
请输入字段名
数据属性
属性配置
属性类别
数据分级
数据分类
敏感字段类型
用户信息
信息配置
信息类别
用户组
RAM角色
用户名称
操作时间
时间配置
周一 周二 周三 周四 周五 周六 周日
添加
阈值比较
数据量
添加阈值比较

新旧版本风险识别规则配置项的配置位置存在差异，具体如下表。

> **说明** 新版风险识别规则的配置详情，请参见新建风险识别规则，旧版风险识别规则的配置详情，请参见风险规则配置。

序号	风险识别规则配置项	旧版配置位置	新版配置位置
1	规则名称	**基础配置** > **规则名**	**基本信息** > **规则名称**
2	规则责任人	**基础配置** > **负责人**。 默认规则的负责人为当前登录账号。	无该配置项，DataWorks自动记录规则责任人。
3	规则的描述信息	**基础配置** > **备注**	**基本信息** > **描述信息**
4	规则生效的引擎	**规则项** > **引擎**	**规则定义** > **条件设置**区域，**选择条件**选择**数据位置**时，所配置的**数据引擎名称**。
5	规则生效的项目空间	**规则项** > **项目空间**	**规则定义** > **条件设置**区域，**选择条件**选择**数据位置**时，所配置的**项目空间名称**。
6	规则所属的分类	**规则项** > **分类**	**规则定义** > **条件设置**区域，**选择条件**选择**数据属性**时，属性类别选择**数据分类**。
7	规则所属的分级	**规则项** > **分级**	**规则定义** > **条件设置**区域，**选择条件**选择**数据属性**时，属性类别选择**数据分级**。
8	命中规则的敏感字段类型	**规则项** > **敏感字段类型**	**规则定义** > **条件设置**区域，**选择条件**选择**数据属性**时，属性类别选择**敏感字段类型**。
9	规则所属的操作类型	**规则项** > **导出方式** 取值包括： • **全部导出方式** • **Tunnel下载** • **表访问**	**基本信息** > **规则类型** 取值包括： • **数据访问风险** • **数据导出风险** • **数据操作风险** • **其他**
10	规则生效的表	**规则项** > **表名**	**规则定义** > **条件设置**区域，**选择条件**选择**数据位置**时，所配置的**表名**。
11	规则生效的字段	**规则项** > **字段**	**规则定义** > **条件设置**区域，**选择条件**选择**数据位置**时，所配置的**字段名**。

序号	风险识别规则配置项	旧版配置位置	新版配置位置
12	用于配置哪些用户访问时会触发该规则。	**规则项** > **访问人员**	**规则定义** > **条件设置**区域，**选择条件**选择**用户信息**时，所选择的**信息类别**。
13	规则限制的操作数据量，当超出限制的数据量范围时命中规则。	**规则项** > **操作数据量**	在**规则定义** > **条件设置**区域，**阈值比较**下配置**数据量**范围。
14	规则限制的访问时间范围，当访问操作与该时间范围匹配时命中规则。	**规则项** > **访问时间**间	**规则定义** > **条件设置**区域，**选择条件**选择**操作时间**时，所设置的**时间配置**。
15	触发规则后的告警方式	不支持	**告警设置** > **告警方式**

后续步骤

风险识别规则创建并生效后，您可以进入**数据风险**页面，查看规则命中的风险详情，及时处理存在风险。详情请参见数据风险（新版）。

4.2.9. 创建并管理用户组

DataWorks的用户组管理功能，帮助您根据业务需求，快速将具有相同数据访问权限的目标账号批量添加至一个用户组，后续在配置数据脱敏时，可以直接将该用户组添加至白名单，使用户组内的账号获取到的数据为脱敏前的原始数据。本文为您介绍如何创建并管理用户组。

创建并使用用户组

1. 进入**数据保护伞**。
   i. 登录DataWorks控制台。
   ii. 在左侧导航栏，单击**工作空间列表**。
   iii. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
   iv. 单击左上方的☰图标，选择**全部产品** > **数据治理** > **数据保护伞**。
   v. 单击**立即体验**，进入数据保护伞。
2. 在左侧导航栏，单击**规则配置** > **用户组管理**，进入**用户组管理**页面。
3. 新建用户组。

i. 单击**新建用户组**，在**新建用户组**对话框中配置**用户组名称**，并添加目标账号至用户组。

您可以通过如下两种方式添加目标账号至用户组。

方式1：上传账号信息文本

方式2：选择账号

- 方式一：通过**文本**添加。此种方式通过上传账号信息文本批量添加账号，适用于需要添加大量账号的场景。
- 方式二：**选择已有账号**添加。此种方式直接选择需要添加的阿里云账号或RAM用户，适用于待添加账号较少的场景。

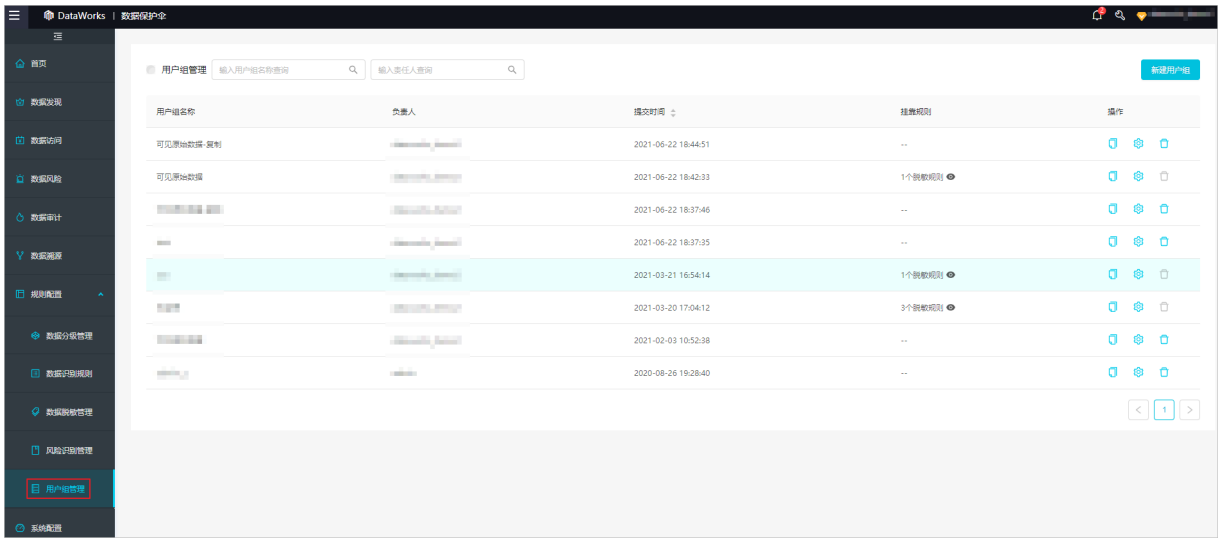
> **说明** DataWorks仅支持上传*UTF-8*格式的*TXT*文本文件，并且每个账号信息占用一行，最多支持配置1000行，即上传1000个账号。

ii. 单击**确定**。

成功创建用户组后，您可以使用**数据脱敏管理**功能，配置目标脱敏规则的白名单，将该用户组添加至白名单中，则使用目标脱敏规则进行脱敏的数据，对该用户组中的用户仍然显示为脱敏前的原始数据。配置脱敏规则的白名单，详情请参见数据脱敏管理。

管理用户组

在**用户组管理**页面，您还可以对已创建的用户组进行如下管理操作：



- 查看用户组列表。

  您可以查看所有已创建的用户组的基本信息，包括**用户组名称**、**负责人**、**提交时间**，以及使用该用户组的脱敏规则数目和名称。

> **说明**
> - 您可以通过**用户组名称**或**负责人**搜索目标用户组，并且使用名称或责任人搜索支持模糊匹配，输入关键词后，即可显示包含关键词的所有用户组。
> - 您可以按照**提交时间**的升降序，根据创建时间对所有用户组进行排序，方便通过时间维度查找目标用户组。
> - 如果目标用户组已经被相关脱敏规则使用，则单击用户组名称**挂靠规则**列的图标，即可查看使用该用户组的脱敏规则名称。

- 操作目标用户组。
  - 复制用户组：单击目标用户组**操作**列的图标，即可快速复制当前用户组。如果您需要创建与当前用户组配置相同的用户组，则可以使用该功能。
  - 编辑用户组：单击目标用户组**操作**列的图标，即可修改已创建用户组的配置信息。您可以快速为目标账号添加白名单权限，或撤销已添加白名单的用户权限。
  - 删除用户组：单击目标用户组**操作**列的图标，即可删除当前用户组。

4.3. 数据发现

您在配置规则后，数据发现页面可以帮助您有效识别工作空间内的敏感数据。

前提条件

阿里云主账号已授权阿里云主账号开通数据保护伞，详情请参见概述。

背景信息

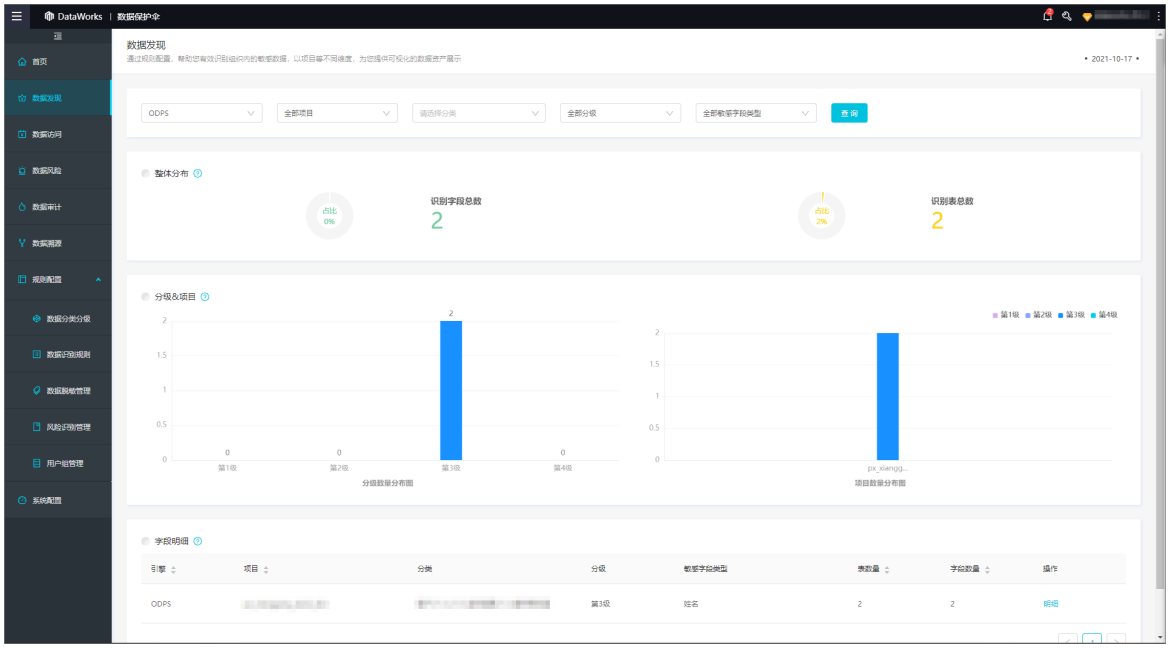
您可以在完成敏感数据规则配置的第二天，查看数据分布情况。

> **说明** 安全管理员可以通过在**系统配置**页面配置**权限控制模式**，来指定可以查看该页面数据的成员。

操作步骤

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上方的图标，选择**全部产品 > 数据治理 > 数据保护伞**。
5. 单击**立即体验**，进入数据保护伞**首页**。
6. 单击左侧导航栏的**数据发现**，进入数据发现页面。

   数据发现页面从工作空间、分级等不同的维度，为您提供可视化的数据资产展示。您可以在该页面查看命中识别规则的字段总数、表总数及对应占比，命中规则的字段各分级、项目数量分布和清单。



4.4. 数据访问

数据访问页面为您展示基于配置规则识别出的敏感数据的访问量、访问趋势、导出量和导出明细等，帮助您掌控每一次访问敏感数据的情况。该页面E-MapReduce计算引擎的操作数据展示暂只支持上海region。

前提条件

阿里云主账号已授权开通数据保护伞，详情请参见概述。

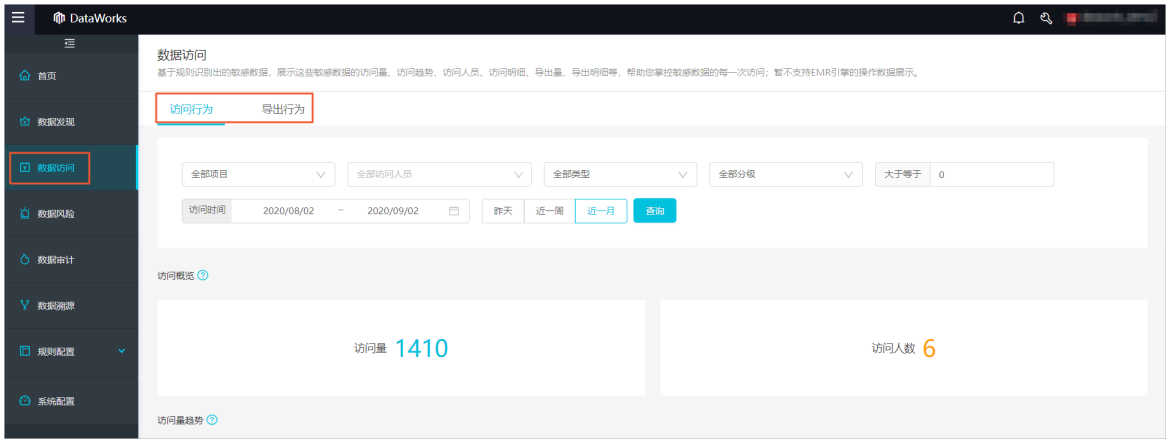
背景信息

您可以在完成敏感数据规则配置的第二天，查看数据的访问和导出情况。

> **说明** 安全管理员可以通过在**系统配置**页面配置**权限控制模式**，指定可以查看该页面数据的成员。

操作步骤

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上方的图标，选择**全部产品 > 数据治理 > 数据保护伞**。
5. 单击**立即体验**，进入数据保护伞。
6. 在左侧导航栏，单击**数据访问**。



数据访问页面包括**访问行为**和**导出行为**：

- **访问行为**：包括Create、Insert操作，但不包括访问失败的行为。

  您可以在完成敏感数据规则配置的第二天，进入**访问行为**页签，查看数据的使用情况。包括**访问概览**、**访问量趋势**和**访问记录**。

- **导出行为**：数据从MaxCompute tunnel通道导出的行为。

  您可以在完成敏感数据规则配置的第二天，进入**导出行为**页签，查看访问人员从MaxCompute中导出数据至外部的情况。包括查询时间段内的数据导出的总量、每天导出的数据量和数据导出总量的前五名。

4.5. 数据风险（旧版）

数据风险页面通过手工打标、风险规则、AI算法等多种方式，识别数据的潜在风险，并支持备注人工审计的结果。该页面E-MapReduce计算引擎的操作数据展示暂只支持上海region。

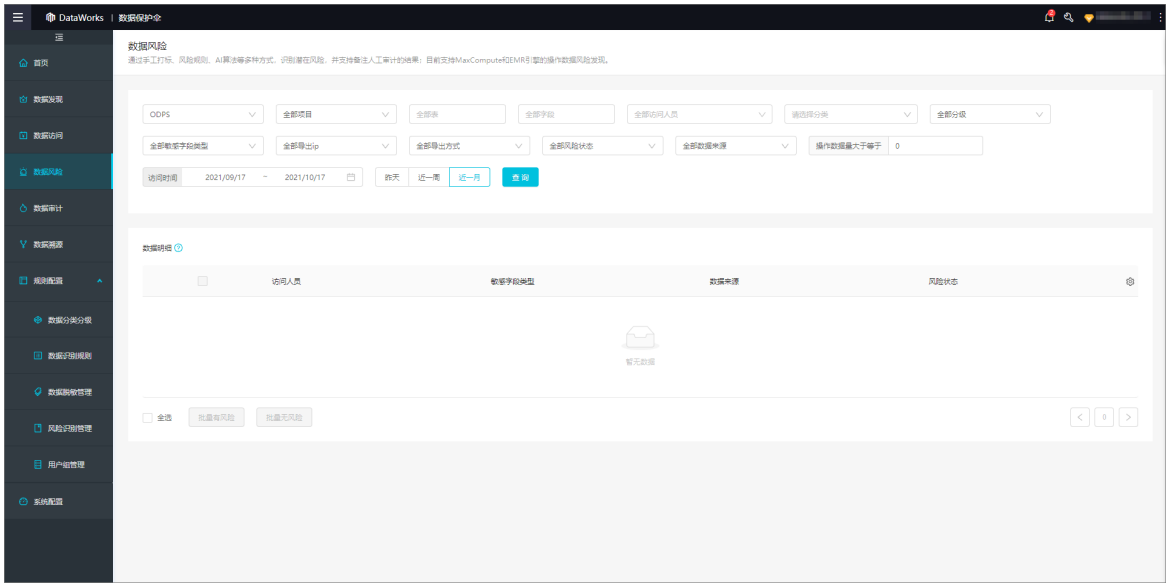
前提条件

租户管理员已授权开通数据保护伞，详情请参见概述。

操作步骤

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上方的☰图标，选择**全部产品 > 数据治理 > 数据保护伞**。
5. 单击**立即体验**，进入数据保护伞。
6. 在左侧导航栏，单击**数据风险**。

   您可以在该页面查看风险数据的明细，并进行**批量有风险**和**批量无风险**的标注。



如果您需要更新**数据明细**列表项的显示，请进行以下操作：

i. 在**数据明细**区域的右侧，单击图标。

ii. 在**列表选项**对话框中，根据业务需求选中或取消相应的选项。

iii. 单击**确定**。

4.6. 数据风险（新版）

数据风险从多维度呈现了通过配置的风险识别规则命中的风险数据，方便您了解不同维度的风险分布、指定时间的风险趋势及风险项目空间排名，获取风险高发的时间及项目空间，也可以查看产生风险的用户、时间、操作等详情，及时定位并处理风险。本文为您介绍数据风险的相关内容。

前提条件

已启用新版风险识别管理功能。新版数据风险页面与新版风险识别管理功能配合使用，您需要通过新版风险识别管理配置相关风险识别规则，命中规则的风险数据才会在呈现在数据风险页面。配置风险识别管理，详情请参见风险识别管理（新版）。

进入数据风险

1. 进入**数据保护伞**。
   i. 登录DataWorks控制台。
   ii. 在左侧导航栏，单击**工作空间列表**。
   iii. 选择工作空间所在地域后，单击相应工作空间后的**数据开发**。
   iv. 单击左上方的图标，选择**全部产品 > 数据治理 > 数据保护伞**。
   v. 单击**立即体验**，进入数据保护伞。
2. 在**数据保护伞**页面的左侧导航栏，单击**数据风险**，进入**数据风险**。

在该页面，您可以查看当前租户下的风险概况及风险详细信息，详情请参见查看风险概览及查看风险明细。

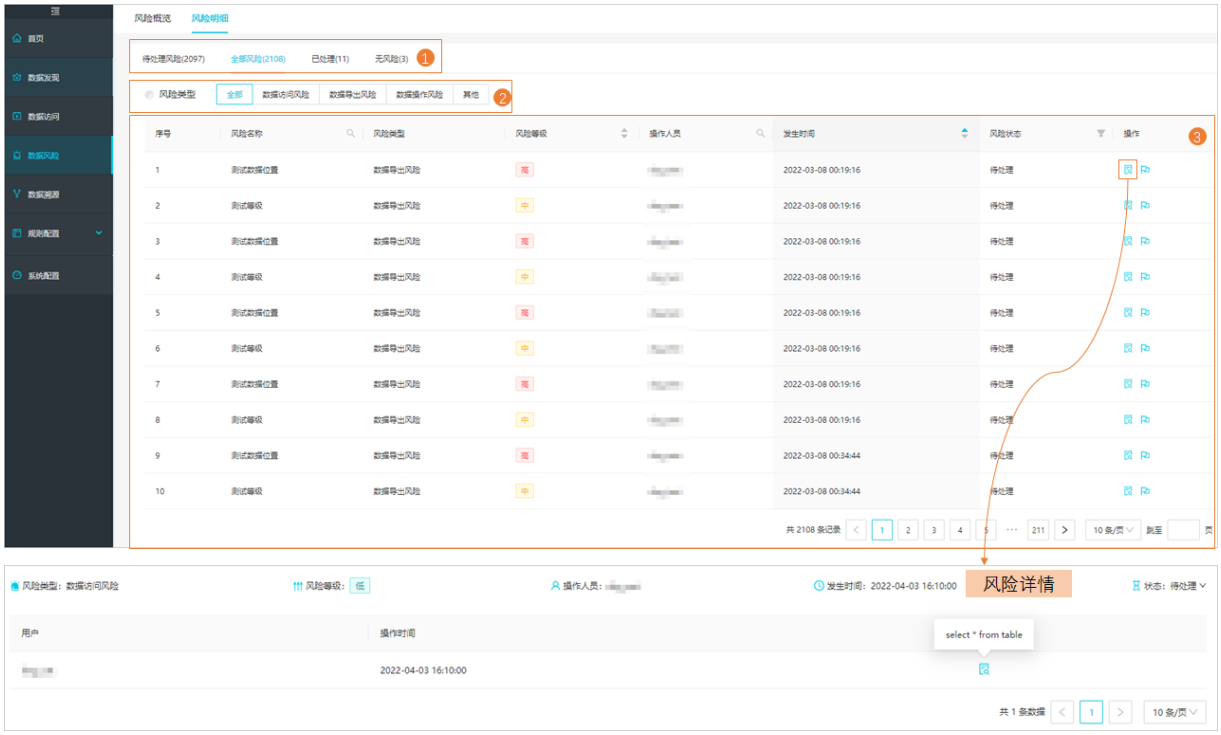
查看风险概览

风险概览页面从操作类别、风险等级、风险趋势等不同维度为您展示了当前租户下的风险数据，帮助您整体了解并把控风险情况。

区域	描述
1	在该区域，您可以查看当前租户下**待处理风险**、**风险总数**、**高风险数**、**中风险数**、**低风险数**等类别的风险数量。
2	在该区域，您可以通过饼状图直观的看到当前租户下**高**、**中**、**低**不同等级风险的分布状况。单击目标等级色块，即可看到对应等级的风险数量。
3	在该区域，您可以通过饼状图直观的看到当前租户下**数据访问**、**数据操作**、**数据导出**等不同操作类型的风险分布状况。单击目标等级色块，即可看到对应操作类别的风险数量。
4	在该区域，您可以选择时间区间，查看指定时间段的风险数量趋势。鼠标悬停至折线图中的数据点，即可显示对应日期当天产时生的风险数量。
5	在该区域，您可以查看当前租户下风险数量较多的项目空间，快速识别风险项目，并集中治理相关风险问题。

查看风险明细

风险明细页面为您展示了风险的名称、类型、操作人员、发生时间、风险状态等详细信息，您可以根据详细信息追溯风险的产生原因并及时定位处理。本文以**全部风险**界面示例，为您展示风险明细页的相关内容。



区域	描述
1	在该区域，您可以查看当前租户下**待处理风险**、**全部风险**、**已处理风险**、**无风险**等类别的风险数量。
2	在该区域，您可以筛选风险类型，查看指定类型的风险数量。默认查看所有类型的风险。

区域	描述
3	在该区域，您可以查看所选条件的风险列表。风险列表为您展示了风险的**风险名称**、**风险类型**、**风险等级**、**操作人员**、**发生时间**、**风险状态**等信息。您可以在该区域执行如下操作： • **精确查找风险** 在**风险名称**、**操作人员**单击图标，即可通过风险名称及操作人员搜索目标风险。 **说明** 搜索操作支持模糊匹配，输入关键词即可查找所有名称包含关键词的风险。 • **风险排序** 单击**风险等级**、**发生时间**后的图标，即可按照风险等级或风险发生时间的升降序对风险排序。 • **风险状态筛选** 单击**风险状态**后的图标，即可选择显示**已处理**、**待处理**状态的风险。  **说明** 仅当区域1选择**全部风险**时，会显示**风险状态**列。 • **查看风险详情** 单击目标风险**操作**列的图标，即可进入风险详情页查看风险详细信息。您可以根据详细信息追溯风险的产生原因，精准定位并快速处理该风险。 • **标记风险** 单击目标风险**操作**列的图标，即可根据实际情况修改该风险的状态为**已处理**或**无风险**。

4.7. 数据审计

数据审计页面多维度展示您的风险处理结果和风险分布情况，该页面E-MapReduce计算引擎的审计情况暂只支持上海Region。

操作步骤

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 单击左上方的图标，选择**全部产品 > 数据治理 > 数据保护伞**。

5. 单击**立即体验**，进入数据保护伞。
6. 在左侧导航栏，单击**数据审计**。

   您可以在该页面查看**昨天**、**近一周**和**近一月**的**风险概述**、**风险量趋势**和**风险事件维度分析**：

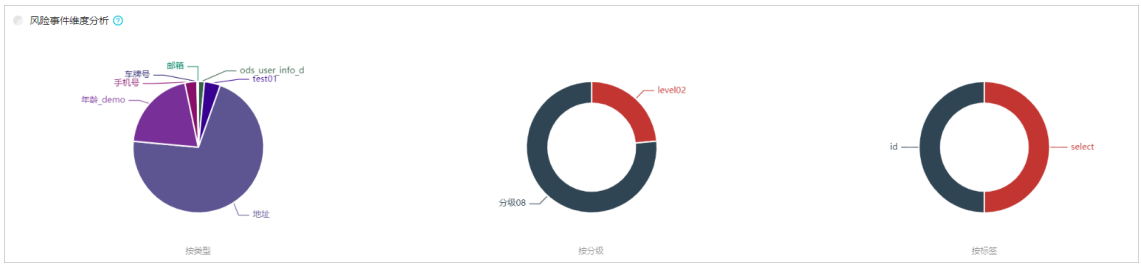
   - **风险概述**：为您展示指定时间段内的风险总量、已处理风险数和未处理风险数。

   - **风险量趋势**：为您展示指定时间段内的风险总量、已处理风险数和未处理风险数的变化趋势。

   - **风险事件维度分析**：根据类型、分级和标签三个维度，为您展示各类风险数的占比。

4.8. 数据溯源

DataWorks的数据溯源功能，支持通过提取数据泄露文件中数据的水印信息，帮助您定位到可能会泄露目标数据的责任人。本文为您介绍如何创建溯源任务，并通过该任务查找可能会泄露数据的责任人。

前提条件

1. 已创建**数据识别规则**，详情请参见敏感数据识别。
2. 为目标**数据识别规则**开启**数据水印**功能，详情请参见数据脱敏管理。

背景信息

通过DataWorks的**数据保护伞**的**数据脱敏管理**，开启目标**数据识别规则**的**数据水印**功能后，则在DataWorks中，对命中该规则的数据所执行的所有操作（例如查询、下载等）均会自动生成水印信息。水印信息用于记录用户的访问行为，并且唯一标识此次访问。后续如果该数据被泄露，您可以通过数据溯源功能，提取泄露数据的数据水印，定位出可能会泄露数据的责任人。

使用限制

- DataWorks仅支持对小于200MB的CSV格式文件进行数据溯源。

- DataWorks仅支持对开启**数据水印**功能之后所执行的数据访问操作进行溯源。

  > **说明** 例如，您查询表A之前未开启**数据水印**功能，此时，即使您开启了**数据水印**功能并启动对该数据文件的溯源任务，仍然无法通过**数据溯源**功能溯源到此次查询操作。

创建并执行数据溯源任务

1. 进入**数据溯源**页面。
   i.
   ii.
   iii.
   iv.
   v.
   vi.
   vii. 在左侧导航栏，单击**数据溯源**，进入**数据溯源**页面。
2. 创建溯源任务。
   i. 单击**新建数据溯源任务**。
   ii. 在**溯源任务**对话框，单击**上传文件**，上传需要溯源的目标文件。

   > **说明**
   > - DataWorks仅支持对小于200MB的CSV格式文件进行数据溯源。
   > - 您可以将DataWorks中的数据文件导出或下载至本地，再上传至溯源任务中进行溯源，也可以将外部系统的数据保存至CSV文件，再上传至溯源任务中进行溯源。

   目标文件上传成功后，您还可以选择**替换**或**下载**该文件。

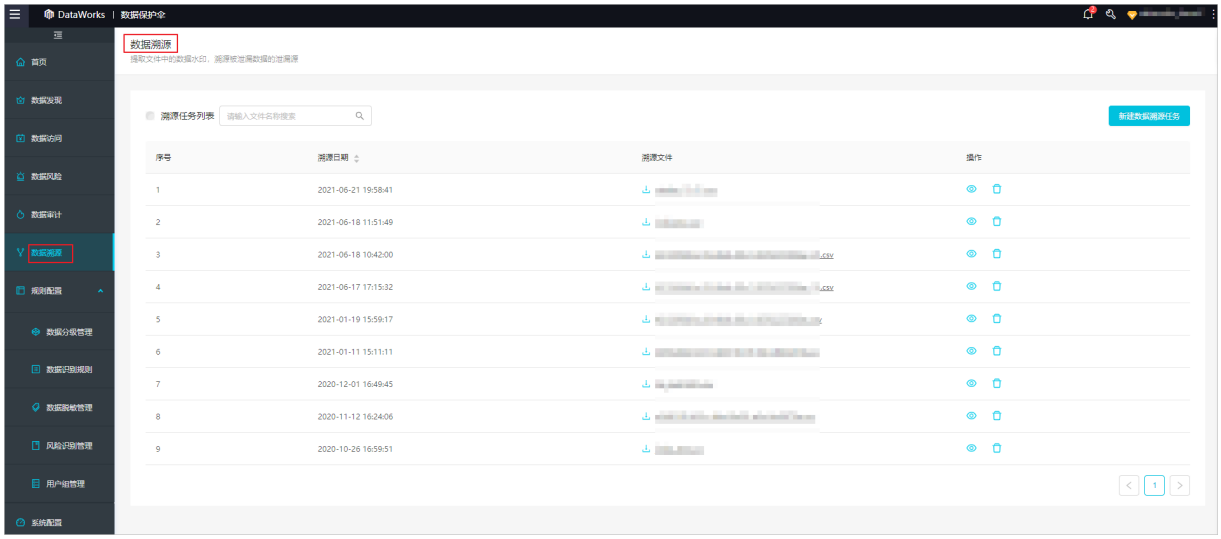
   

3. 单击**开始溯源**，启动目标溯源任务。

   > **说明** 执行溯源任务可能会花费一定时间，请您耐心等待。

查看可能的泄露源

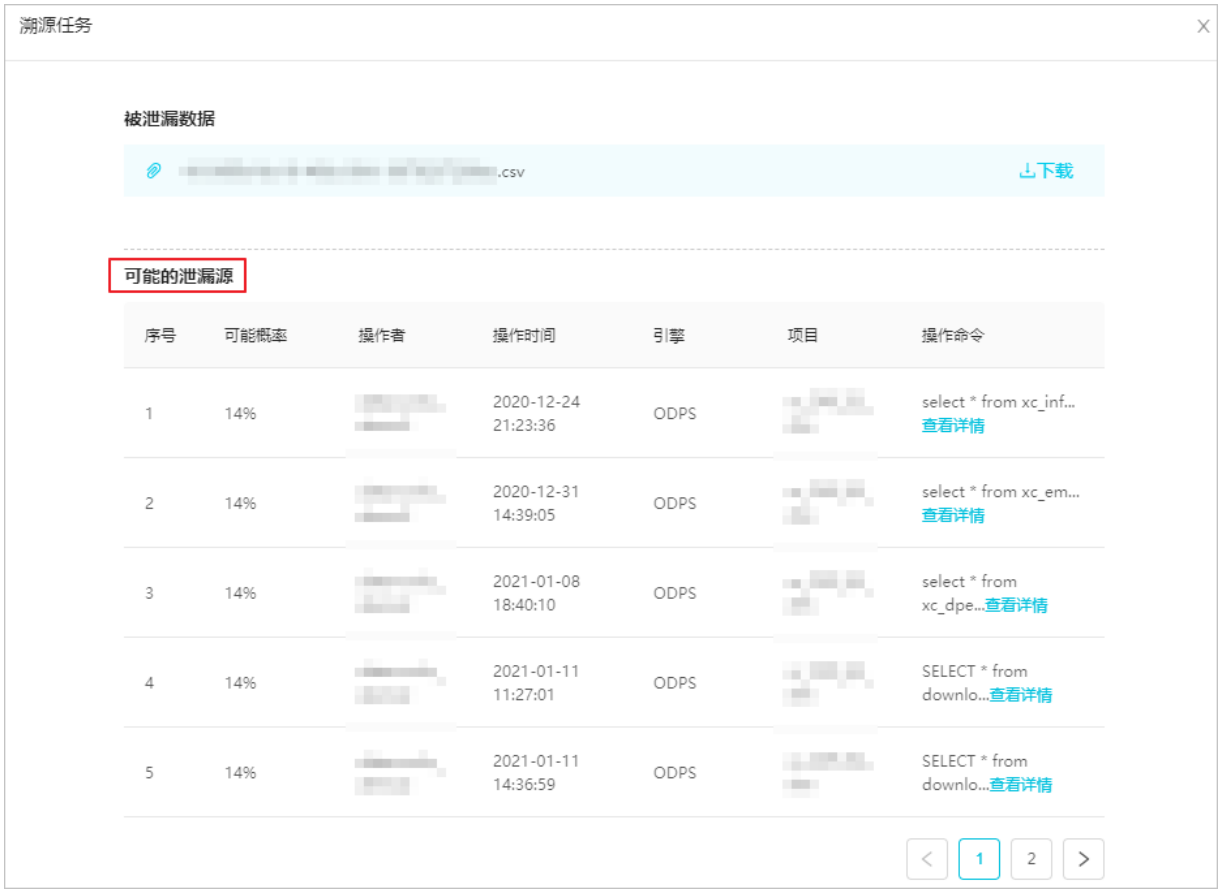
在**数据溯源**页面，您可以查看所有已执行溯源任务的**溯源日期**及**溯源文件**，并根据目标溯源任务的溯源详情，查看可能的数据泄漏源。



说明

- 您可以按照**溯源日期**的升降序对所有溯源任务进行排序，方便查找目标溯源任务。
- 您可以根据溯源文件的名称搜索目标溯源任务，并且溯源文件的名称搜索支持模糊匹配，输入关键词后，即可显示包含关键词的所有溯源任务。

单击目标溯源任务**操作**列的图标，即可查看该任务的溯源详情。您可以根据DataWorks分析的**可能概率**、**操作时间**及**操作命令**的内容，定位出最可能泄露数据的责任人。



常见问题

目标溯源任务执行结束后，**可能的泄漏源**显示**无结果**，则可能的原因及解决方案如下：

- 原因一：您所溯源的文件数据量不足，导致水印信息无法还原。

  解决方案：使用**数据水印**功能生成的水印信息需要提供充足的数据量，才能保证通过溯源任务还原出可靠的水印信息，进而定位出可能的数据泄露责任人。建议您使用数据量大于500条，并且不包含重复数据的文件进行溯源。

- 原因二：被泄漏的数据非本租户名下的数据。

  解决方案：您需要确认溯源数据的来源，确保溯源的数据为本租户名下的数据。

- 原因三：被溯源的文件中不包含水印信息。

  解决方案：

  - 您需要检查目标文件是否开启**数据水印**，DataWorks仅支持对开启**数据水印**功能之后所执行的数据访问操作进行溯源。查看并开启**数据水印**功能，详情请参见数据脱敏管理。
  - 您所溯源的文件不存在信息泄露，可能是其他外部系统的操作导致了数据泄露。

4.9. 数据血缘（公测）

DataWorks的数据血缘功能支持可视化展示敏感数据的血缘关系，自动分析字段之间的异常关联关系、敏感数据识别结果异常的字段，帮助您梳理敏感数据的扩散情况及影响面，提高数据识别效率。本文为您介绍如果查看血缘关系图。

背景信息

数据血缘为您提供如下功能：

- 可视化血缘图谱

  数据保护伞基于敏感字段之间的血缘关系，绘制成敏感数据血缘可视化图谱，帮助您清晰的了解数据的来龙去脉。

- 提升数据识别效率

  敏感数据自动识别任务可基于敏感字段血缘关系，将其中敏感字段类型相同的血缘关系进行识别结果扩散，极大提高识别效率。

- 异常血缘关系分析
  - 关联关系异常的字段分析

    系统根据敏感字段的血缘关系，自动分析字段之间的异常关联关系（例如，SELECT_CONCAT、SELECT_SUBSTRING等关系），避免相关人员通过字符拼接、拆解的方式绕过敏感数据的识别和使用审计。

  - 关联但识别结果不一致的字段分析

    帮助您识别出与查询字段有血缘关系，但敏感字段类型识别结果不一致的字段。例如，查询A字段，敏感数据类型为姓名，与其有血缘关系的字段有B（姓名）、C（省份），则识别结果不一致的字段是C。

使用限制

仅DataWorks企业版及以上版本用户才可以使用数据血缘功能。版本升级详情请参见版本服务计费说明。

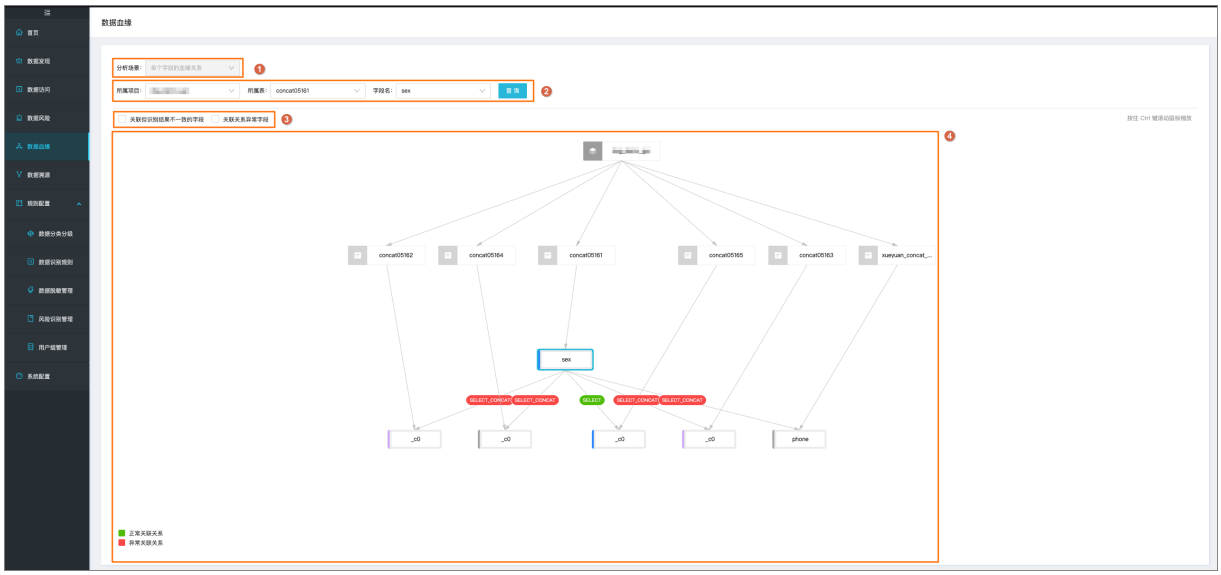
进入数据血缘

1. 进入数据保护伞。
   i. 登录**DataWorks控制台**后，进入**数据保护伞**页面，操作详情请参见概述。
   ii. 单击**开始体验**，默认进入数据保护伞的**首页**。
2. 进入数据血缘。

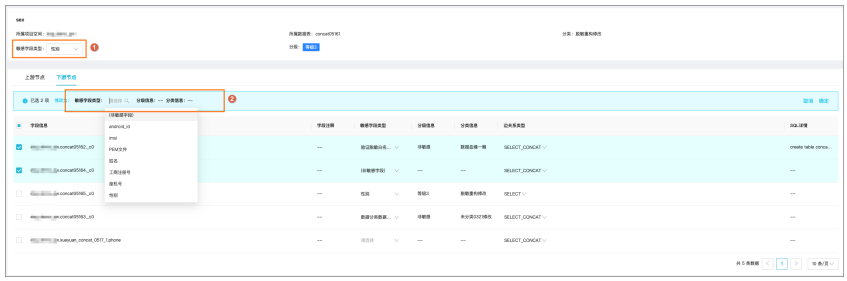
   您可以通过以下两种方式进入数据血缘页面。
   - 进入手动修正数据页面，找到需要查看血缘关系的字段，单击**操作**列的**血缘分析**跳转至数据血缘页面。
   - 进入**数据保护伞**后，在左侧导航栏单击**数据血缘**。进入数据血缘页面。

查看血缘关系图

数据血缘页面为您可视化展示敏感数据的血缘关系。



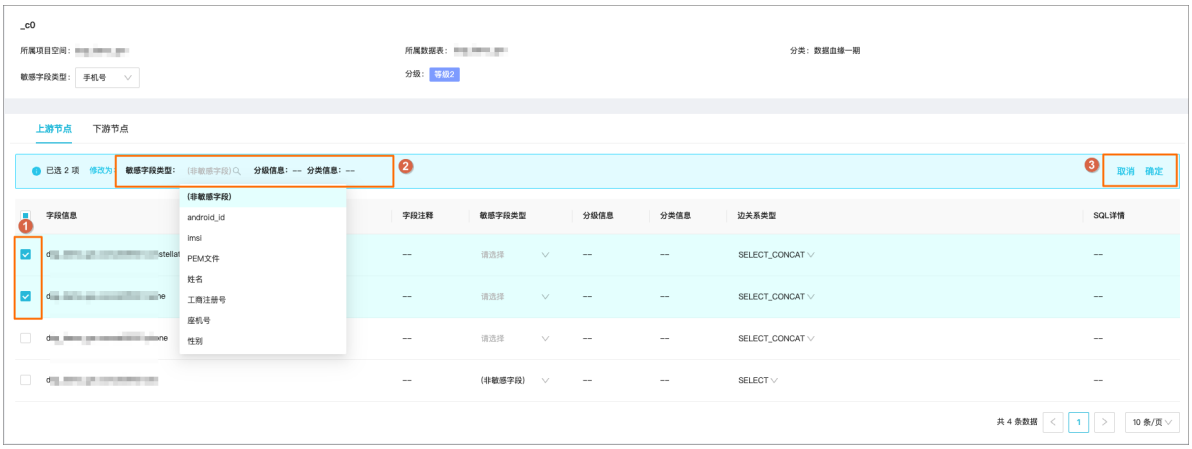
类别	描述
分析场景	当前默认分析场景为**单个字段的血缘关系**，后续会陆续上线其他场景，敬请期待。
筛选	在上图的模块②区域，支持您输入要查询的字段**所属项目**、**所属表**和**字段名**，单击**查询**，页面将为您展示对应字段的一层血缘关系，查询的字段会高亮显示。
过滤条件	在上图模块③区域，您可以根据需求选择过滤条件： • **关联但识别结果不一致的字段** 勾选后，将会自动刷新血缘关系图，为您展示与查询字段有血缘关系，但敏感字段类型识别结果不一致的字段，并展示字段之间的边关系。 **说明** 边关系为您创建字段时使用的SQL函数名称，例如，SELECT、SELECT_LTRIM等。 • **关联关系异常字段** 勾选后，将会自动刷新血缘关系图，为您展示关联关系异常（例如，SELECT_CONCAT、SELECT_SUBSTRING等关系）的字段，并展示字段之间的边关系。 • 同时勾选**关联但识别结果不一致的字段**和**关联关系异常字段**时，将为您展示与选中的查询字段，关联关系异常且敏感字段类型识别结果不一致的字段，并展示字段之间的边关系。
	在上图模块④区域，为您展示对应查询字段的一层血缘关系图，您可以单击对应字段或字段中间的边关系按钮，查看字段信息和边信息。 • 查看字段信息 单击字段，将为您打开字段详情页面，字段详情页面展示当前字段的数据位置信息、敏感字段类型，以及上、下游关联字段列表、字段之间的关联关系（例如，SELECT、SELECT_CONCAT、SELECT_REPEAT等关系）等。对于识别结果不准确的数据，您可以通过下图区域①修改当前字段的**敏感字段类型**；通过下图区域②修改上、下游关联字段的**敏感字段类型**、**分类**、**分级**等信息。

类别	描述 说明
血缘关系图	说明 ○ 当查询的字段没有上、下游关联字段时，列表将显示暂无数据。 ○ 当敏感字段类型为非敏感字段或未识别时，敏感字段类型、分类、分级等信息将展示为空。 ○ 修改敏感字段类型后，会同步更新数据发现和手动修正数据页面的数据。 ○ 每个字段最多显示一层上、下游关联字段信息。  • 查看边信息 单击两个字段中间的边关系按钮，将在右侧弹窗展示边关系详情。包括：**边关系**、**边关系类型**、**SQL详情**、**上游节点列表**、**下游节点列表**。对于识别结果不准确的数据，您可以单击**边关系类型**右侧的下拉框修改字段间关联关系；单击上下游节点**敏感字段类型**下拉框修改敏感字段类型。 说明 ○ 当边关系异常时，在边关系右侧将展示**异常关联**标签，若无异常右侧将不展示任何标签。 ○ 异常关联关系包括SELECT_CONCAT、SELECT_SUBSTRING等关系，即相关人员通过字符拼接、拆解的方式绕过敏感数据识别的情况。 ○ 边关系类型为您创建字段时使用的SQL函数名称，例如，SELECT、SELECT_LTRIM等。 ○ 当敏感字段类型为非敏感字段或未识别时，敏感字段类型将展示为空。 

批量修正数据

您可以通过以下两种方式，对敏感数据识别结果不准确的字段进行批量修正。

- 通过血缘关系图查看当前字段的上、下游关联字段详情，批量选中字段进行修正。



- 进入**手动修正数据**页面，批量选中字段进行修正。详情请参见：手动修正数据。

4.10. 系统配置

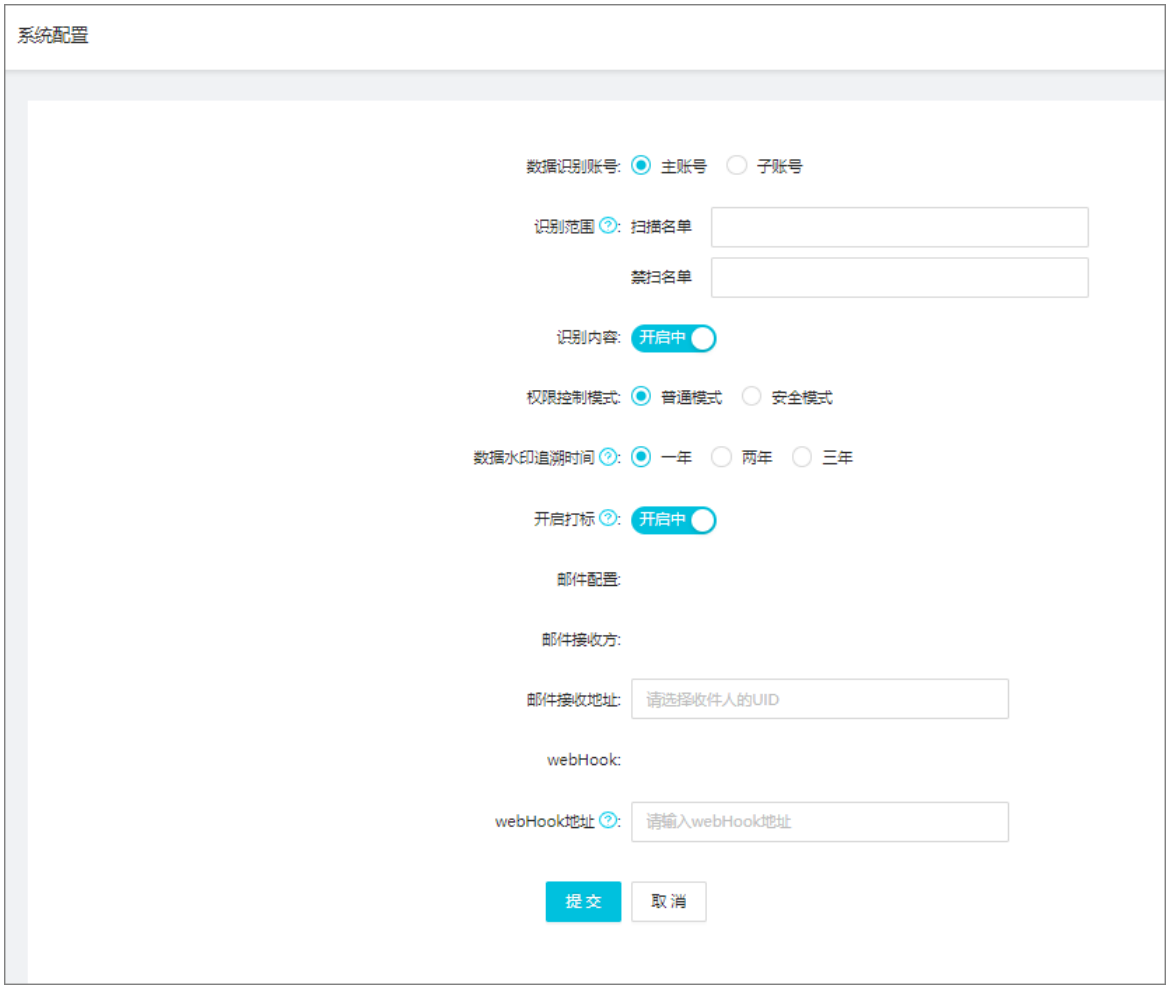
您可以在系统配置中控制登录数据保护伞的权限模式、数据水印追溯时间、数据识别管控的数据范围、风险识别结果的告警接收邮件及webHook地址等。

使用限制

DataWorks支持使用钉钉群、企业微信和飞书的WebHook地址。其中，仅企业版及以上版本支持推送告警信息至企业微信或飞书。

操作步骤

1. 进入系统配置。
   i. 进入数据保护伞，详情请参见：进入数据保护伞。
   ii. 单击左侧导航栏的**系统配置**，进入系统配置页面。
2. 配置相关信息。



类别	参数	描述
数据识别管控范围配置	**数据识别账号**	用于控制数据识别管控的数据范围。 ○ 主账号：表示管控主账号所属工作空间中的数据。 ○ 子账号：表示管控子账号所属工作空间中的数据。
	识别范围	支持您在**扫描名单**和**禁扫名单**中添加MaxCompute项目名称，以确定数据识别范围，如果您未填写相关名单，则默认扫描主账号或者子账号获取到的所有项目。 **说明** 识别范围仅支持配置MaxCompute project，多个project以英文逗号分隔。
	识别内容	您可以开启或关闭识别开关，关闭后，将无法按照数据识别规则进行内容识别。

类别	参数	描述
数据保护伞访问权限控制	**权限控制模式**	○ 安全模式：表示仅安全管理员可以登录使用数据保护伞。 ○ 普通模式：表示所有用户都可以登录使用数据保护伞。
数据水印追溯时间配置	**数据水印追溯时间**	您可以设置数据水印文件的存储时间，支持设置为一年、两年或三年。例如，您设置追溯时间为两年，那么当发现数据泄露时您可以追溯到是否是两年内的操作导致的数据泄露。
MaxCompute引擎中列安全等级控制	**开启打标**	开启打标后，对于MaxCompute引擎中的列，数据分类分级结果将直接打到对应列的Label（敏感等级标签）上，并在DataWorks数据地图表详情页面的字段信息中展示列级别安全等级。详情请参见：查看表详情。 **说明** ○ 如果系统配置中开启打标后，在数据地图中仍然看不到列级别安全等级，请确认是否开启**列级别访问控制**开关。详情请参见：MaxCompute高级配置。 ○ 开启打标后，MaxCompute引擎中的列分级结果将对您的权限管控产生影响，您需要在手动修正数据页面确认字段级别，当您在MaxCompute中配置的访问许可等级标签低于字段级别时，您将无法访问对应字段。设置访问许可等级标签请参见：为用户或角色设置访问许可等级标签。
风险识别告警配置	**邮件接收地址**	请选择告警信息收件人的UID。当识别到风险数据时，系统将根据您的配置，发送报警信息至指定邮箱，以便您对风险进行处理。如果您需要新增报警联系人信息，请参见查看和设置报警联系人。
	webHook接收地址	支持输入钉钉群、企业微信、飞书的WebHook地址。当识别到风险数据时，系统将根据您的配置，发送报警信息至指定群组，以便您对风险进行处理。 **说明** 仅企业版及以上版本支持推送告警信息至企业微信或飞书。

5.数据地图

5.1. 数据地图概述

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，涵盖全局数据检索、元数据详情查看、数据预览、数据血缘和数据类目管理等功能。数据地图可以帮助您更好地查找、理解和使用数据。

元数据采集与接入

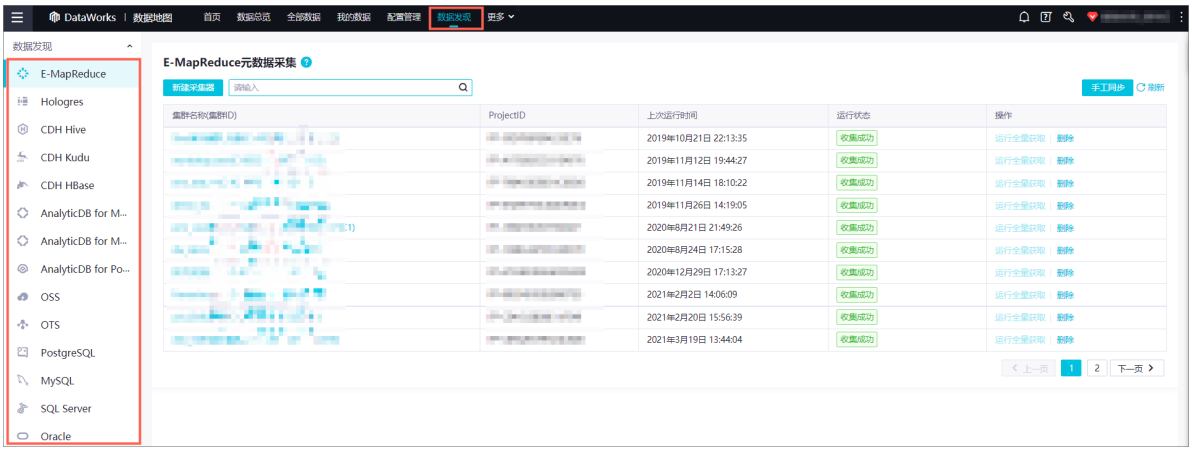
- MaxCompute数据

  如果您使用了MaxCompute引擎，您可以直接在数据地图进行MaxCompute表元数据管理的相关操作。

- 其他类型元数据

  除MaxComput外，您还可以通过元数据采集功能将不同数据源中的元数据导入数据地图进行统一管理。在**数据发现**页面通过**元数据采集器**将其他数据源中的元数据采集至DataWorks数据地图，采集完成后，您可以在数据地图搜索并查看各数据源的元数据信息。除MaxComput以外，目前数据地图支持的其他数据源类型有：**E-MapReduce**、**Hologres**、**CDH Hive**、**CDH Kudu**、**CDH Hbase**、**AnalyticDB for MySQL 2.0**、**AnalyticDB for MySQL 3.0**、**AnalyticDB for PostgreSQL**、**OSS**、**OTS**、**PostgreSQL**、**MySQL**、**SQL Server**、**Oracle**（持续扩充中），元数据采集配置详情请参见数据发现。

> **说明** 如果需要在**数据开发**中**表管理**进行可视化建表操作，请先在数据地图进行元数据采集，可视化建表操作仅支持绑定为引擎类型的数据源。详情请参见管理表。



网络连通

如果您需要将数据源中的元数据导入数据地图进行统一的元数据管理，需要先确保数据地图元数据采集器能正常访问您的数据库。如果您的数据库有白名单访问控制，您可以在数据库中根据如下说明配置对应白名单：

- 如果您需要进行元数据采集的数据库已开启白名单访问控制，请在数据库白名单列表中，添加您使用的DataWorks所在地域对应的IP网段。需要配置的白名单请参见元数据采集的数据源有白名单访问控制时需要配置的白名单。
- 如果MaxCompute项目未开启白名单访问控制，则DataWorks可以正常使用数据地图访问MaxCompute的数据表，如果MaxCompute项目开启了白名单访问控制，请在MaxCompute的白名单列表中，添加需要使用的DataWorks所在地域的IP网段。要配置的白名单请参见MaxCompute开启白名单访问控制时需要配置的白名单列表。

数据总览

- 您可以在数据总览页面查看当前地域（Region）下的MaxCompute总项目数，总表数、存储量、总API数、存储趋势图、项目占有率Top、表占有率Top和热门表。
- 您还可以查看元数据采集完成后的AnalyticDB MySQL 3.0、MySQL、E-MapReduce、Hologres、AnalyticDB PostgreSQL、OTS等的数据库总数，总表数、总API数等信息。

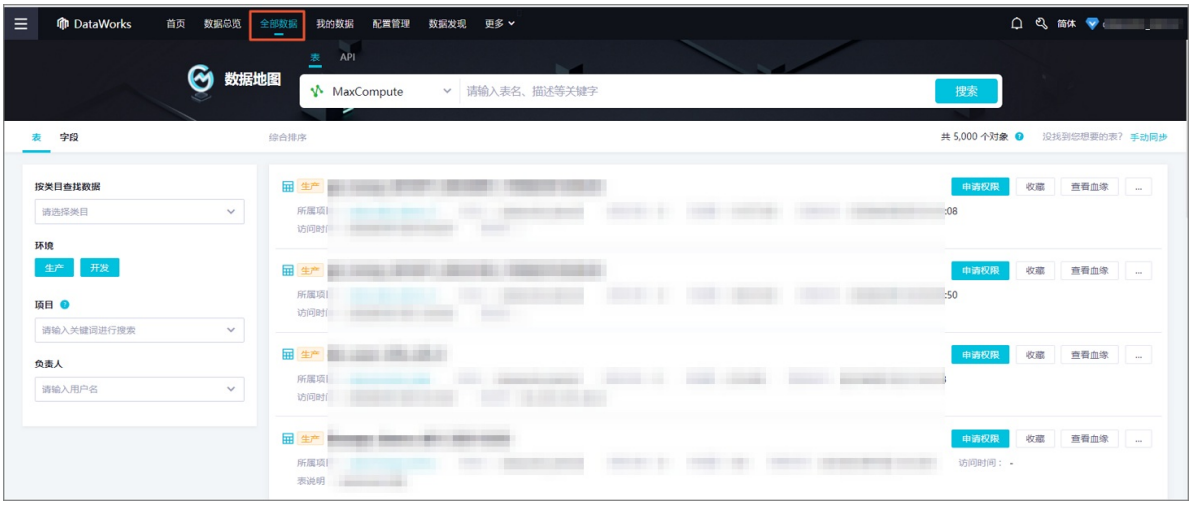
详情请参见数据总览。

查找表和API

数据地图支持您通过如下方式查找表和API。

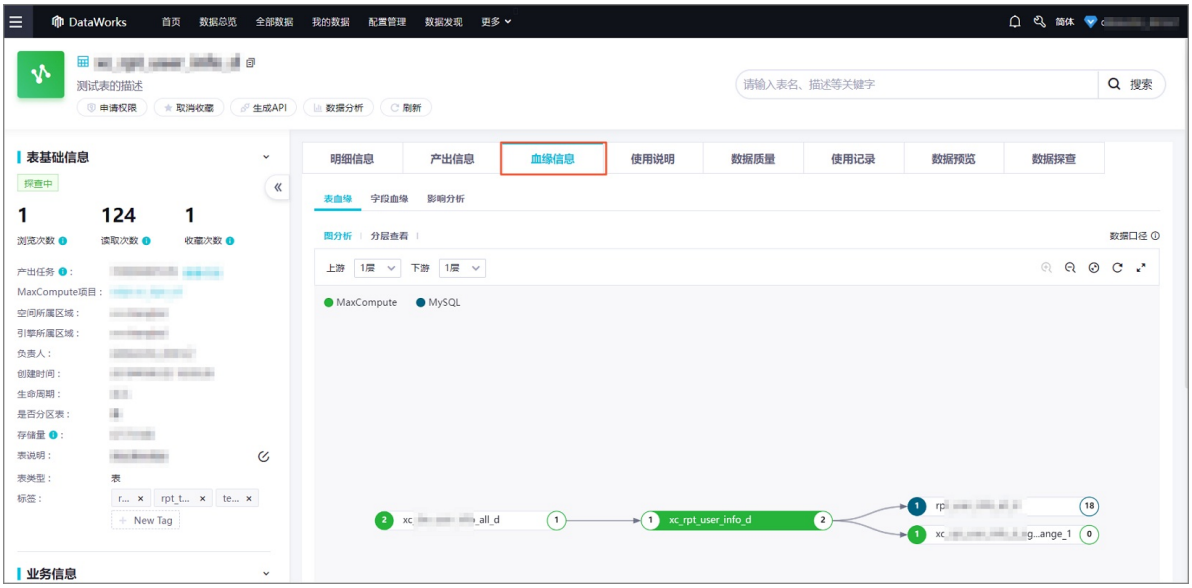
- 您可以通过**首页**中的**表**类目下展示的近期浏览的表、近期读取的表、热门浏览和热门读取等列表快速获取相应的表。或者在搜索框中输入关键字搜索目标表，详情请参见首页。
  - 您还可以通过**首页**中的**API**类目下展示的近期浏览的API、热门浏览的API、热门调用的API等列表快速获取相应的API，或者在搜索框中输入关键字搜索目标API，详情请参见首页。

- 您可以在**全部数据**界面中的**表**类目下对各数据源中元数据采集后的表通过表名，表描述及字段名，字段描述进行搜索。同时支持通过表所在类目，项目或数据库进行表过滤。此外，对于MaxCompute还支持根据表所在生产或开发环境及表负责人进行过滤，对于E-MapReduce还支持通过集群过滤表。详情请参见查找表。
  - 您还可以在**全部数据**界面中的**API**类目下对当前租户下所有空间中的API，通过输入API名称、API描述等关键字进行搜索，同时支持通过**API类型**、**工作空间**、**负责人**对搜索结果进行过滤。找到符合条件的API。详情请参见查找API。



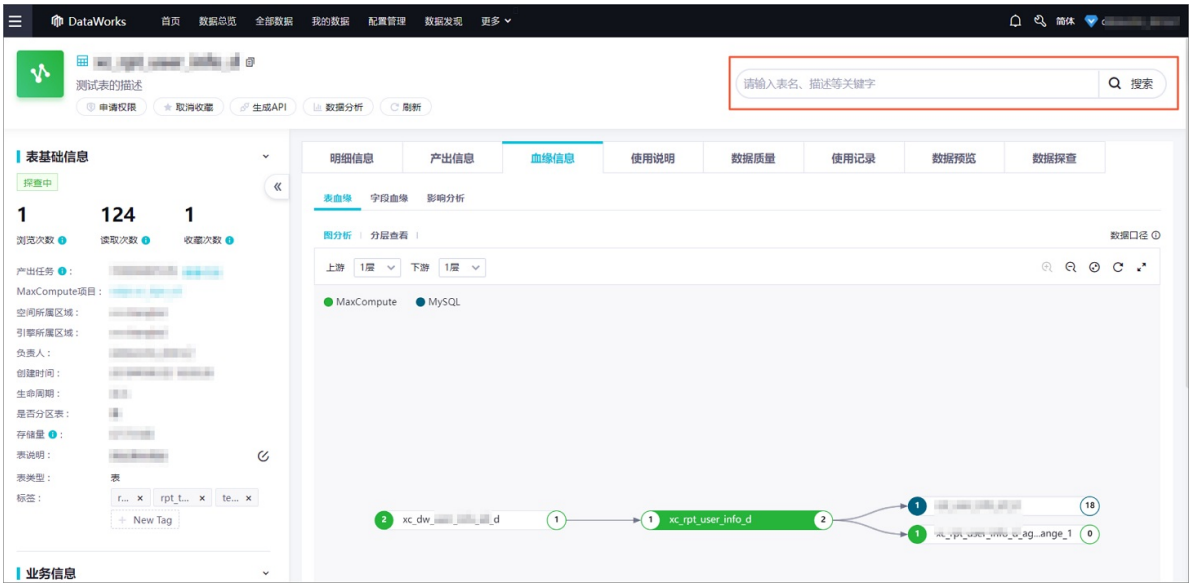
查看表详情和API详情

- 您可以单击目标表名称跳转至表详情页面，查看表的**基础信息**、**产出信息**和**血缘信息**等信息。请参见查看表详情。



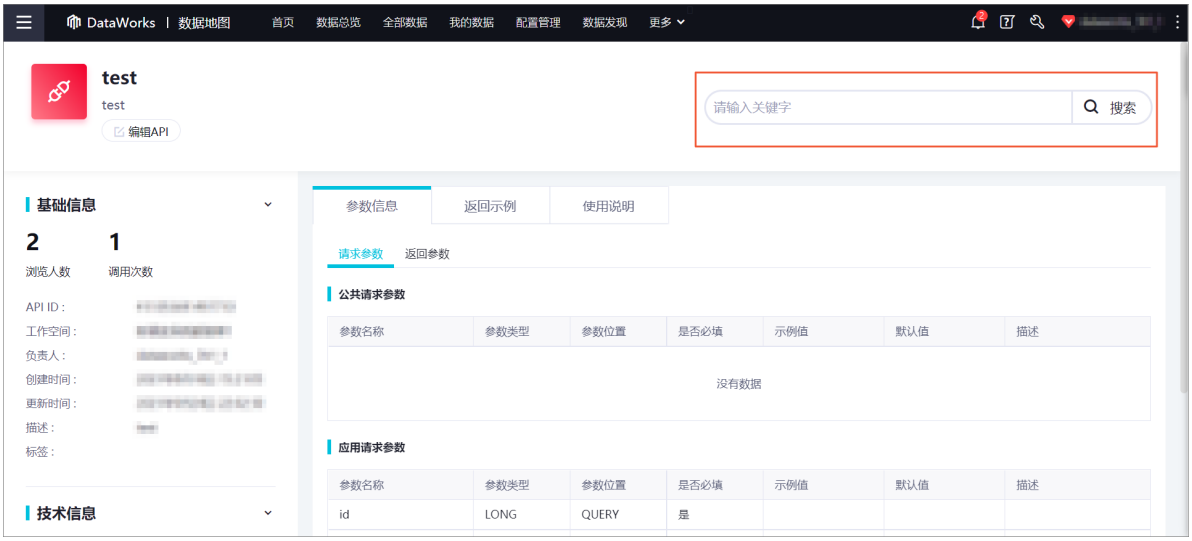
> **说明** **血缘信息**页面中您可以查看引擎节点内部血缘关系，具体引擎节点血缘支持情况以产品实际页面展示为准。此外，MaxCompute还支持基于离线同步的完整链路血缘查看。您可以查看MaxCompute表的上下游血缘，通过展开表血缘层级查看MaxCompute表的原始数据来源相关信息和MaxCompute表数据最终流向的数据库相关信息。

表详情页面右上角支持通过表名，表描述，字段名，字段描述及项目名等关键字进行搜索。详情请参见查看表详情。



- 您可以单击目标API名称跳转至API详情页面，查看API的**基础信息**、**技术信息**等信息。详情请参见查看API详情。

  API详情页面右上角支持通过**API ID**、**API Path**、**API名称**、**API描述**等关键字搜索API。



表的有序组织和管理

类目管理功能方便您通过类别有效地组织和管理表，表的类目管理配置完成后，您可以在查找表时，通过类目来过滤目标表。详情请参见配置管理，同时支持您管理表。

> **说明** 阿里云主账号及拥有AliyunDataWorksfullaccess权限的RAM用户可以编辑类目树。

- **类目管理**

  您可以通过如下方法将表添加至类目中：

  i. 通过**配置管理 > 类目导航配置**批量将表添加到指定类目。

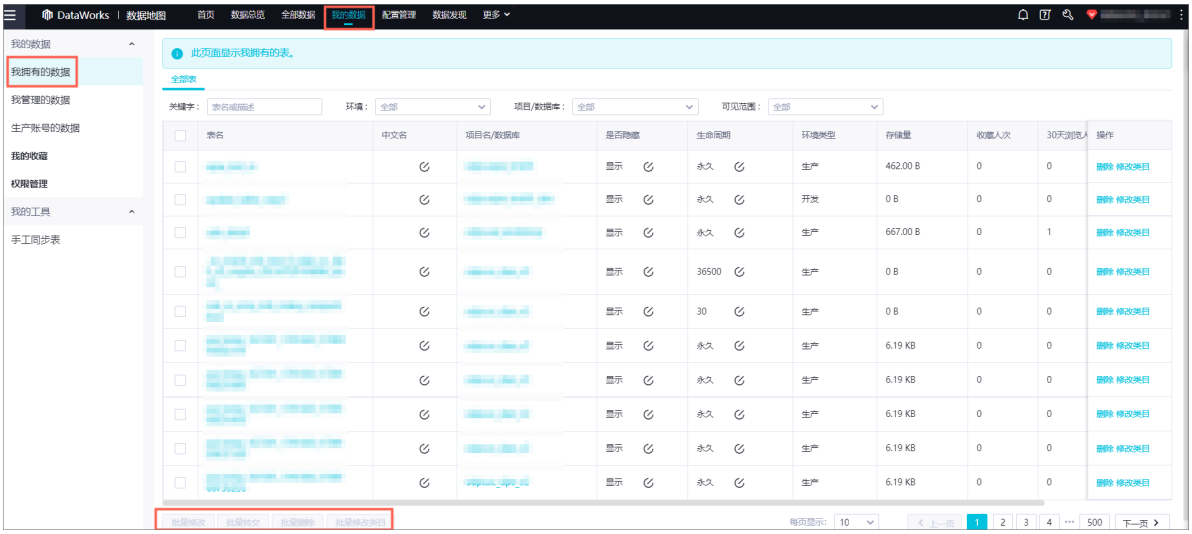
类目导航配置完成后，您可以选中最后一级类目，并通过界面的快速添加入口，快速将表某个项目下的某些表批量添加至该类目中。详情请参见配置管理。

ii. 通过**我的数据**页面批量将表添加到指定类目。

类目配置完成后，您可以在我的数据（我拥有的数据、我管理的数据）页面批量将表添加到指定类目。详情请参见我的数据。

- **表管理**

对于MaxCompute数据类型，数据地图支持批量修改中文名，生命周期、支持批量删除开发表或者生产表，批量转交表责任人。详情请参见我的数据。



- **个人收藏**

数据地图支持您将个人关注的表统一添加到个人收藏中进行管理，方便快速定位和查阅。您可以通过**我的数据**页面下的**我的收藏**分组中查看目前个人收藏的表。详情请参见我的数据。

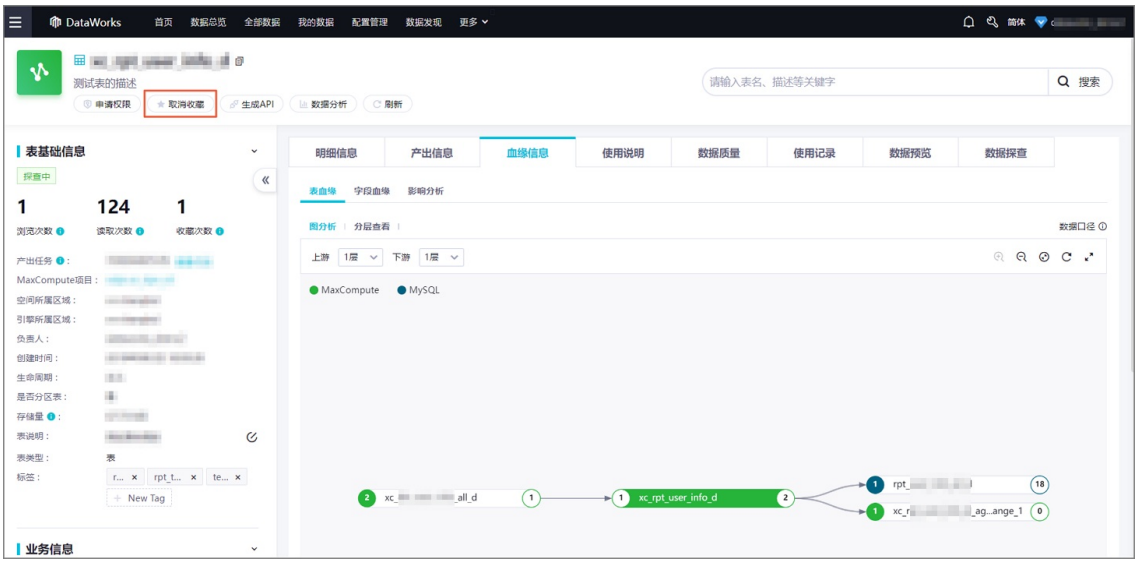
  - 将表添加入**我的收藏**

您在查看表详情时，可以通过表详情页的收藏按钮，快速将表加入的**我的收藏**，详情请参见查看表详情，添加收藏后，您可以通过**我的数据**页面下的**我的收藏**分组中进行查看，详情请参见收藏表。

- 将表从**我的收藏**列表移除

  您可以通过下面两种方式取消收藏表，取消收藏后，将不会展示在**我的数据**页面下的**我的收藏**列表中。

  - 您可以通过**我的数据**页面下的**我的收藏**分组中对目前收藏的表取消收藏。
  - 您可以通过已收藏的表详情页面中的**取消收藏**入口，快速取消收藏该表。



表权限管控

- 表操作权限申请

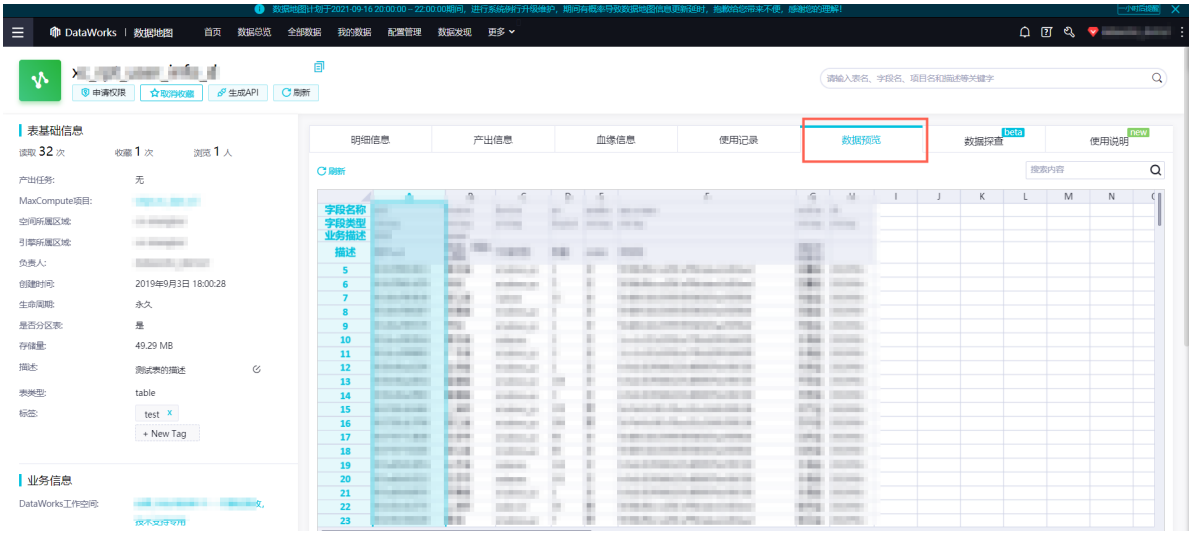
  标准模式工作空间下，RAM用户默认无法通过SQL命令直接操作生产表，如果您需要操作生产表或跨账号查询生产表，需要进行权限申请，您可以在表详情页中的**申请权限**入口申请表的相关权限。当您在表详情页单击申请权限时，将跳转到**安全中心**进行具体的权限申请操作，详情请参见申请表权限。

> **说明** 如果RAM用户无某张表的查询权限，默认情况下将无法通过数据地图表详情页中的数据预览功能来查看该表数据。

- MaxCompute表数据预览权限控制

您可以在配置管理，项目管理配置中对具体工作空间的MaxCompute开发表或生产表开启预览权限，开启后，该项目下的表无需申请访问权限，即可被工作空间中所有成员预览。详情请参见项目管理配置。

> **说明**
> - **此操作可能存在敏感数据泄露的风险，请谨慎评估后再开启。**
> - 所在工作空间的项目owner或者工作空间管理员可进行该操作。
> - 此权限仅控制数据地图中表详情页面的**数据预览**功能。



- 隐藏表

表隐藏后，搜索表时将无法搜索到该表。支持对所有人隐藏或者仅对表所在工作空间下用户可见。详情请参见我的数据。

您可以选择表状态为

- 隐藏：所有人都不可以通过搜索来访问到该表。
- 仅项目：仅对表所在工作空间下用户可见（可搜索到）。
- 显示：所有人都可以通过搜索访问到该表。

> **说明** 表owner和工作空间管理员默认不受上述权限控制。

其他

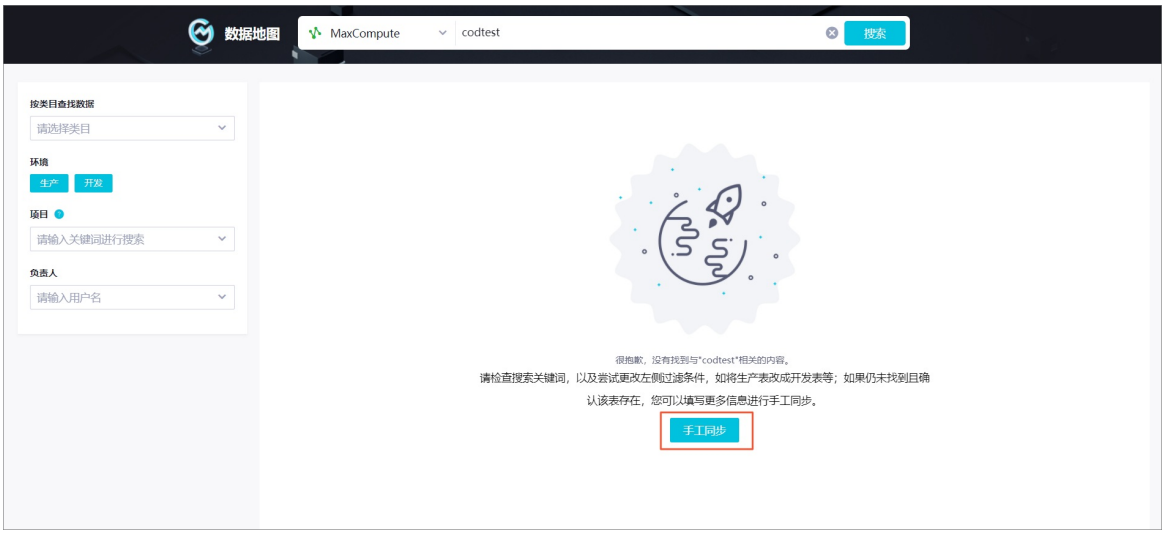
- 工作空间管理

  您可以通过**更多 > 工作空间**查看当前阿里云主账号下的所有工作空间详情列表，单击目标工作空间名称可以查看工作空间的详细信息。详情请参见工作空间列表。

- 手动同步工具

  如果表存在但是搜索不到或者表更新了但是数据地图显示还未更新，您需要手工同步表。

- 在搜索结果页面单击**手工同步**。



- 在数据地图的**我的数据 > 手工同步表**页面，输入格式为 `odps.项目名称.表名称` 的**表GUID**后，单击**手工同步**。

> **说明** 手工同步工具仅对MaxCompute有效。

完成上述操作后您可以在数据地图的全部数据中再次搜索关键词查询对应的表。

5.2. 首页

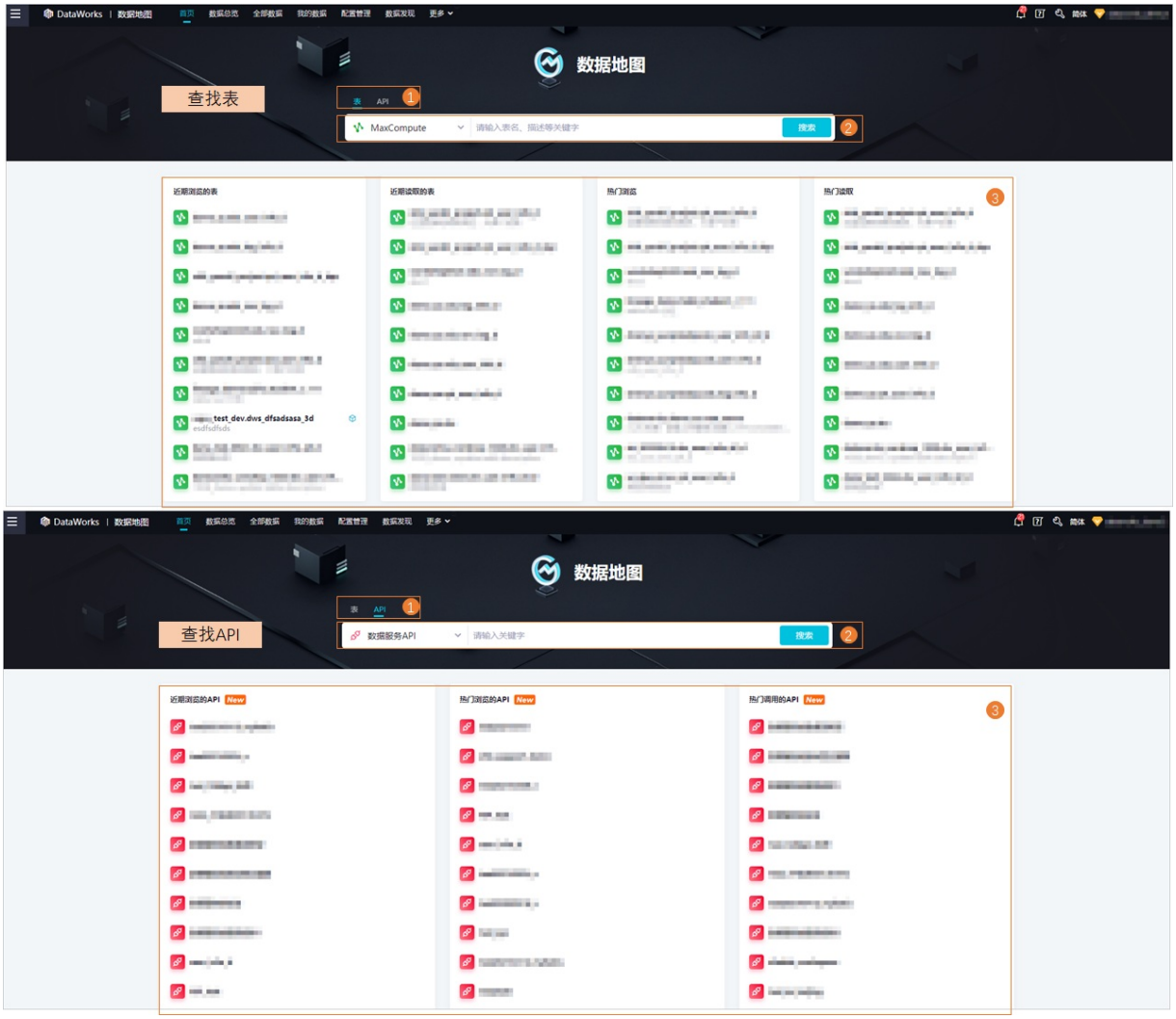
数据地图首页支持输入关键字搜索需要查看的表和API，并基于您的访问记录推荐热门浏览、热门读取的表及API。本文为您介绍数据地图首页的内容概况。

进入首页

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间**操作**列的**进入数据地图**，默认进入数据地图的**首页**。

查找表及API

数据地图 > 首页为您推荐了使用较为频繁的表或API，您可以直接查看；也可以通过关键词搜索需要查看的表或API。



区域	描述
①	在该区域，选择需要查找的表或**API**类目。

区域	描述
②	在该区域，输入关键字，即可检索名称或描述中包含关键字的表或API。 • **查找表** 单击图标，筛选数据源类型，在搜索框输入表名称、表描述、标签等关键字，单击**搜索**，跳转至全部数据页面查找该数据源中包含关键字的所有表，您还可以在全部数据页面选择**字段**分类，并在搜索框中输入字段名称、字段描述、智能建模指标名称和描述等关键字查找该数据源中包含关键字的所有字段，详情请参见：查找表。 **说明** ◦ 当前仅支持查看MaxCompute、E-MapReduce、Hologres、AnalyticDB for MySQL 2.0、AnalyticDB for MySQL 3.0、OSS、AnalyticDB for PostgreSQL、OTS、PostgreSQL、MySQL、SQLServer和Oracle等数据源下的表。 ◦ 输入关键字后，下拉列表最多支持展示符合条件的10个表。单击目标表名称，即可跳转至表详情页查看表的详细信息。查看表详情，请参见查看表详情。 ◦ 如果表后存在图标，则表示该表为DataWorks智能数据建模生成的表。如果您需要使用智能数据建模功能，请参见概述。 • **查找API** 在搜索框中输入关键字，即可查找包含关键字的所有API。您可以通过API ID、API Path、API名称、API描述等关键字进行搜索。 **说明** 输入关键字后，当前最多支持展示符合条件的10个API。单击目标API名称，即可跳转至API详情页查看API的详细信息。查看API详情，请参见查看API详情。
③	该区域为您展示了使用频率较高的表及API，方便您直接查看。 • **表类目** ◦ 您可以在表页签中查看**近期浏览的表**和**近期读取的表**，以及DataWorks基于当前阿里云账号的访问记录所推荐的**热门浏览**和**热门读取**列表。 ◦ 如果表后存在图标，则表示该表为DataWorks智能数据建模生成的表。如果您需要使用智能数据建模功能，请参见概述。 • **API类目** 您可以在API页签中查看**近期浏览的API**，以及DataWorks基于当前阿里云账号访问记录所推荐的**热门浏览的API**及**热门调用的API**列表。

5.3. 数据总览

本文为您介绍如何进入数据总览页面，查看当前Region阿里云主账号下所有的引擎资源情况。

操作步骤

1. 登录**DataWorks控制台**后，进入**数据地图**页面，操作详情请参见进入首页。
2. 在顶部菜单栏，单击**数据总览**。

数据总览页面对当前Region阿里云主账号下所有的引擎资源进行统计，离线产生整个页面的数据信息。

以MaxCompute引擎数据概览为例：

名称	描述
总项目数	为当前租户本地域下，所有MaxCompute项目总数。为实时统计的数据
总表数	○ 为当前租户本地域下，所有MaxCompute表总个数。 ○ 此数据为离线统计，有T+1的延迟。
占用存储量	○ 为当前租户本地域下，所有表的逻辑存储大小总和，包含调度任务的临时文件、删除表后尚未释放的存储空间。 ○ 此数据为离线统计，有T+1的延迟。
总API数	为当前租户本地域下，当前时间点，已发布至API网关的MaxCompute类型的API总数。
存储趋势图	○ 为当前租户本地域下，MaxCompute项目的逻辑存储总和趋势图，此数据包含调度任务的临时文件、删除表后尚未释放的存储空间。 ○ 此数据为离线统计，有T+1的延迟。
项目占用存储Top	○ 为您展示当前当前租户本地域下，MaxCompute项目的逻辑存储大小的排行。 ○ 此数据为离线统计，有T+1的延迟。

名称	描述
表占用存储Top	○ 按照MaxCompute表大小展示的排行榜。您可以单击具体表名跳转至表详情页。 ○ 此数据为离线统计，有T+1的延迟。 **说明** 项目存储及表占用的逻辑存储显示离线统计时间的用量，并且显示的为逻辑存储大小。项目存储量除表存储量外，还会计算包括资源存储量、回收站存储量及其它系统文件存储量等在内，因此会大于表存储量。 表的存储计费计算的是表的逻辑存储而非物理存储。
热门表	根据数据地图表详情页访问PV，为您展示表访问量的排行榜。为实时统计的数据。

5.4. 全部数据

5.4.1. 表详情

5.4.1.1. 查找表

数据地图支持通过表名，表描述、字段名、字段描述、标签、智能建模指标名称和描述等方式检索字段和表，同时还可以通过表所在类目，项目或数据库进行表过滤。

前提条件

如果您当前工作空间绑定了MaxCompute引擎，您可以直接搜索查看MaxCompute表，除MaxComput外，您需要通过元数据采集功能将不同数据源中的元数据导入数据地图后搜索查看相关表。详情请参见数据发现。

进入查找表界面

1. 登录**DataWorks控制台**后，进入**数据地图**页面，操作详情请参见进入首页。
2. 在顶部菜单栏，单击**全部数据**。
3. 在搜索框上方，选择**表**类目，进入查找表的页面。

查找表

区域	描述
①	单击图标，筛选数据源类型，并在输入框中输入表名称、表描述、标签查找该数据源中包含关键字的所有表，您还可以单击左侧**字段**类目，在输入框中输入字段名称、字段描述、智能建模指标名称和描述等关键字搜索该数据源中包含关键字的所有字段。 **说明** 当前仅支持查看MaxCompute、E-MapReduce、Hologres、AnalyticDB for MySQL 2.0、AnalyticDB for MySQL 3.0、OSS、AnalyticDB for PostgreSQL、OTS、PostgreSQL、MySQL、SQLServer和Oracle等数据源下的表。
②	您可以在该区域按**类目**、**数据库名**、**负责人**、**项目**对搜索结果中的表进行二次过滤。具体过滤选项以界面为准，类目配置请参见类目导航配置。 **说明** MaxCompute还支持根据表所在生成或开发环境及表负责人进行过滤，E-MapReduce、CDH等还支持通过集群过滤表。
③	为您展示当前数据源下的表详情列表，您可以对表执行如下的操作： • 申请权限：您可以在安全中心申请表权限，并在数据地图查看申请记录。详情请参见申请表权限。 • 收藏表：您可以**收藏**或者**取消收藏**不需要的表。 • 查看血缘：基于调度作业，数据流转信息，解析得出的表和表，字段和字段之间的血缘关系，暂不包含临时查询等手动操作产生的血缘关系。详情请参见查看表详情。 • 查看DDL：单击后，在**生成DDL语句**对话框中，查看或**复制**当前表的建表语句。

5.4.1.2. 查看表详情

本文为您介绍如何进入表详情页面，查看表的基础信息、产出信息和血缘信息等详情。

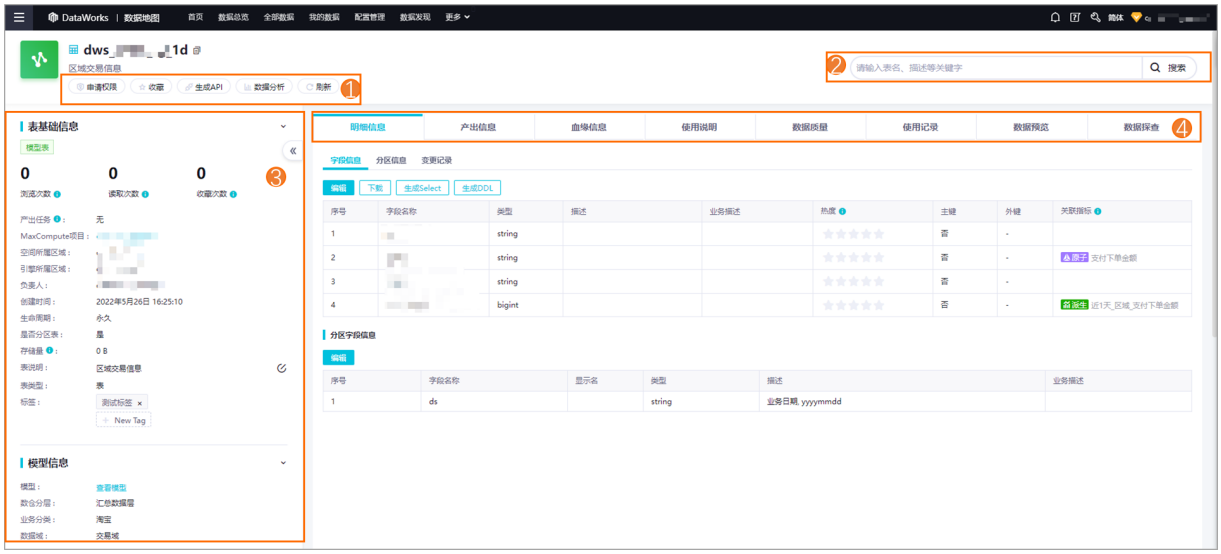
进入表详情页面

您可以通过如下两种方式进入表详情页面。

- 从**首页**进入
  i. 登录DataWorks控制台。
  ii. 在左侧导航栏，单击**工作空间列表**。
  iii. 选择工作空间所在地域后，单击相应工作空间**操作**列的**进入数据地图**，默认进入数据地图的**首页**。
  iv. 在表页签，筛选数据源类型并在搜索框输入关键字，查找目标表。单击目标表名称，即可进入表详情页面。
- 从**全部数据**进入
  i. 登录DataWorks控制台。
  ii. 在左侧导航栏，单击**工作空间列表**。
  iii. 选择工作空间所在地域后，单击相应工作空间**操作**列的**进入数据地图**，默认进入数据地图的**首页**。
  iv. 在**数据地图**顶部菜单栏，单击**全部数据**。
  v. 在表页签，筛选数据源类型并在搜索框输入关键字，查找目标表。单击目标表名称，即可进入表详情页面。

查看表详情

表详情页面为您展示表的**基础信息**、**业务信息**、**模型信息**、**权限信息**、**技术信息**、**明细信息**、**产出信息**、**血缘信息**、**使用说明**、**数据质量**、**使用记录**、**数据预览**及**数据探查**。



区域	描述
①	在该区域您可以执行如下的操作： • 申请权限：您可以在安全中心申请表权限，并在数据地图查看申请记录。 • 收藏表：您可以**收藏**或者**取消收藏**不需要的表。 • 生成API：您可以在数据服务页面生成和注册API。 • 数据分析：您可以在**SQL查询**页面通过编写SQL语句进行数据查询与分析。 • 刷新：刷新当前表的详细信息。
②	在该区域您可以输入表名、字段名、项目名等关键字搜索表。

区域	描述
③	在该区域您可以查看该表的如下信息： • **表基础信息**：用于查看表的**读取次数**、**收藏次数**、**浏览人数**等信息。单击**产出任务**后的**查看代码**，即可查看代码详情。 **说明** ○ **读取次数**：统计近30天内生产环境发起的读取MaxCompute表的任务计数，读取表的任务类型包括但不限于SQL、Tunnel Download、数据集成等。此数据为离线统计，有T+1的延迟。 ○ **收藏次数**：表被收藏的人次，为实时统计的数据。 ○ **浏览次数**：统计30天内在数据地图浏览此表的人数，此数据为离线统计，有T+1的延迟。 ○ **表存储量**：统计的为表的逻辑存储大小，此数据为离线统计，有T+1的延迟。 ○ **产出任务**：写入当前表的DataWorks周期调度任务ID。若表被周期更新，但没有展示任务ID，可能是非DataWorks周期调度任务写入，详可咨询表负责人。此数据为离线统计，有T+1的延迟。 • **业务信息**：用于查看表所在的DataWorks工作空间的详情、**环境类型**、**所属类目**等信息。 • **权限信息**：用于查看您当前拥有的表权限，您还可以单击**权限信息**区域右上方的**更多**，进入**表权限申请**页面申请权限。 • **技术信息**：用于查看**计算引擎信息**信息，单击**计算引擎信息**后的**点击查看**，即可查看或**复制**相关信息。 **说明** 最后数据查看时间： ○ 统计的为表的最后访问时间，其访问包括手动执行命令访问该表数据和任务调度场景下访问该表数据。 ○ 此数据仅供参考，不能百分之百精确反映该数据的真实访问时间。 ○ 此数据为离线统计，有T+1的延迟。
	在该区域您可以查看该表的如下信息： • **明细信息**：用于查看表的**字段信息**、**分区信息**和**变更记录**。详情请参见查看明细信息。 • **产出信息**：如果表的数据会随着对应的任务周期性发生变化，您可以单击**产出信息**，查看该表的变化情况、持续更新的数据等信息。此数据为离线统计，有T+1的延迟。 • **血缘信息**：用于查看引擎节点内部血缘关系，您也可以查看当引擎作为数据源时，与产出的数据接口API之间的血缘关系。此外，MaxCompute还支持基于离线同步的完整链路血缘查看。详情请参见查看血缘信息。此数据为离线统计，有T+1的延迟。 **说明** 如需从API视角查看上游（数据源）和下游（APP）的完整端到端血缘链路，请参考查看API详情。 • **使用说明**：您可以进行**编辑**、**查看历史版本**和**查看markdown语法**等操作，根据数据的业务说明了解相关的信息。 • **数据质量**：为您展示当前表配置的数据质量监控规则详情及DQC告警列表，您可以单击右侧的**配置规则**跳转至数据质量页面为表配置质量监控规则。详情请参见：按表配置监控规则。 • **使用记录**：用于查看表的**频繁关联**和**访问统计**：

区域	描述
④	○ **频繁关联**：为您展示有多少人在使用当前的表数据。 **说明** 统计为30天内作为关联条件参与计算的次数，此数据为离线统计，有T+1的延迟。 ○ **访问统计**：以图标方式为您展示表的使用记录。 ■ **读取趋势图**：折线图上日期对应的为日期当天的读取次数，区分是从开发环境还是生产环境进行读取；字段关联次数与任务执行次数和该字段在代码中出现的次数相关，此数据为离线统计，有T+1的延迟。 例如：如果在同一个任务中字段出现1次，如果任务执行2次，统计次数便为2次；如果字段在代码中出现2次，那么一次任务运行，其字段统计次数便为2次。 ■ **字段热度明细**：字段在SQL中的使用次数（where、select、join、groupBy）的统计信息。此数据为离线统计，有T+1的延迟。 ■ **读取Top人员**：统计近30天内，在SQL中对表的读取人员的统计信息（包含调度使用的生产账号和个人账号的访问），其读取内容包括对字段的where、select、join、groupBy等操作。此数据为离线统计，有T+1的延迟。 • **数据预览**：可以预览当前表的数据。 注意 ○ 您需要拥有权限，才可以预览生产环境的表。如果没有权限，请参见申请表权限进行申请。 ○ 如果表所在工作空间在项目管理配置开启了表预览权限，即使没有在安全中心申请表查询权限，同样可以在此处预览数据。 ○ 如果您已配置数据脱敏规则并设置数据脱敏规则为生效状态，那么数据脱敏规则也会在数据预览页面生效。关于数据脱敏规则配置方法，详情请参见数据脱敏管理。 ○ 暂不支持外部表数据预览。 • **数据探查**：数据探查通过分析数据的结构和取值，为您展示数据的统计信息和分布情况等探查结果。详情请参见数据探查。 **说明** 数据探查将会产生数据质量实例费，您可以在数据质量任务查询面板中，查看该表关于此次探查的日志。

查看明细信息

单击**明细信息**，查看表的**字段信息**、**分区信息**和**字段信息**：

- **字段信息**：您可以查看表的字段信息，如果该表为分区表，您还可以查看**分区字段信息**。

操作	描述
编辑	单击后，您可以编辑字段的**描述**、**业务描述**、**安全等级**和**主键**，并**保存**或**取消**编辑的内容。 您也可以选中多个字段，批量设置安全等级。 **说明** ○ 仅**空间管理员**及表Owner支持编辑表字段。如果目标用户需要编辑表字段，则可授权**空间管理员**权限，详情请参见角色及成员管理：全局级。 ○ 仅单独设置字段安全等级的表，会显示**安全等级**。 ○ 仅MaxCompute引擎支持该功能。 ○ 在MaxCompute引擎开启字段安全等级功能后，才可以在此处设置表字段的安全等级。开启安全等级功能，详情请参见Label权限控制。
批量编辑安全等级	用于批量设置表字段的安全等级，提升数据的安全性。 **说明** ○ 仅MaxCompute引擎支持该功能。 ○
上传	单击后，拖拽本地需要上传的数据至**批量上传字段信息**对话框中。 **说明** ○ 仅**空间管理员**及表Owner支持上传数据至目标表。如果目标用户需要上传数据，则可授权**空间管理员**权限，详情请参见角色及成员管理：全局级。 ○ 仅支持上传.xlsx（Excel 2007版本）格式的文件，您也可以**下载模板文件**。
下载	单击后，直接下载当前表的字段信息。
生成select	单击后，在**生成select语句**对话框中，查看或**复制**当前表的`select`语句。

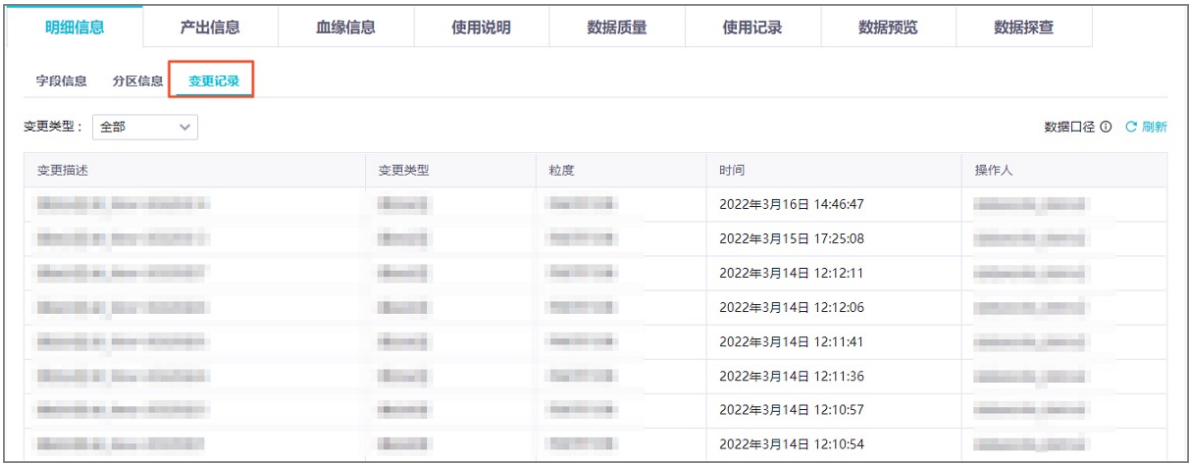
操作	描述
生成DDL	单击后，在**生成DDL语句**对话框中，查看或**复制**当前表的建表语句。

> **说明**
>
> - 字段热度：统计数据为前一天该字段在SQL中参与join的次数，次数按比例转换为星级，热度最高为5星，最小为0星。

- **分区信息**：查看当前表的**分区名**、**记录数**、**逻辑存储大小**等分区信息，仅MaxCompute分区表此部分展示有数据展示。

> **说明** 分区记录数和大小仅供参考。数据更新可能有延迟，实际以引擎侧为准。

- **变更记录**：查看当前表的**变更描述**、**变更类型**、**粒度**等变更记录。



您可以在**变更记录**页签的左上方，从变更类型列表中，选择需要查看其变更记录的变更类型。

变更类型包括**创建表**、**修改表**、**删除表**、**添加分区**、**删除分区**、**修改负责人**和**修改生命周期**。

查看血缘信息

表的血缘信息页面您可以查看引擎节点内部血缘关系，此外，MaxCompute还支持基于离线同步的完整链路血缘查看。您可以查看MaxCompute表的上下游血缘，通过展开表血缘层级查看MaxCompute表的原始数据来源相关信息和MaxCompute表数据最终流向的数据库相关信息。

> **说明**
>
> - 您需要购买DataWorks高级版本，才可以使用血缘信息功能。例如，MaxCompute和E-MapReduce计算引擎需要标准版及以上版本。
> - MaxCompute表血缘基于ODPS SQL调度作业，解析得出的表和表，以及字段和字段之间的血缘关系；暂不包含临时查询等手动操作产生的血缘关系。此部分为离线统计，有T+1的延迟。

单击**血缘信息**，查看表的**表血缘**、**字段血缘**和**影响分析**：

- **表血缘**包括**图分析**和**分层查看**：
  - **图分析**：为您展示中心节点的全部上游、下游的层级数，以及全部上游、下游的节点数量。

  - **分层查看**：默认以当前表为中心，展开其一级上游和一级下游的全部节点。您可以根据GUID搜索表的上下游表。

- **字段血缘**：从**字段名**列表中，选择需要查看的字段的血缘关系。

- **影响分析**：您可以根据**血缘层级**、**血缘字段**、**任务类型**、**表名称**、**项目名称**和**表负责人**等信息，查看血缘关系的**调度产出**和**完整链路**。

您可以单击**开始分析**，重新进行影响分析。分析完成后，您可以**下载**影响分析列表中的数据至本地，也可以通过**邮件**的方式，通知当前表的下游。

数据探查

数据探查通过分析数据的结构和取值，为您展示数据的统计信息和分布情况等探查结果。

数据探查的使用限制如下：

- 仅支持探查分区表。
- 仅支持探查生产环境的表。
- 仅表的所有者有权限开启自动探查功能。

单击**数据探查**，设置探查方式并查看探查记录。

数据探查提供**手动探查**和**自动探查**两种方式：

- **手动探查**

  > **说明** 探查任务运行在当前表所在的MaxCompute项目下，单表探查仅支持10列。为优化资源，请仅勾选需要探查的列。

  配置手动探查任务的操作如下：

  i. 在**数据探查**页签下，单击**手动探查**。

  ii. 在**手动探查**对话框中，配置各项参数。

参数	描述
表名	格式为工作空间名称.表名，默认不可以修改。
分区取值	从**分区取值**列表中，选择需要探查的分区。
详细配置	选中需要探查的列。

参数	描述
预估费用	根据上述配置，预估运行探查任务所需要的费用。 **注意** ■ 数据探查需要执行MaxCompute SQL语句，会带来一定的MaxCompute计算费用。该页面的预估费用仅为参考，实际费用受处理的数据量影响，会有波动，请以MaXCompute账单为准。 ■ 数据探查复用数据质量产品能力，将会同时产生数据质量实例费用，此部分费用由DataWorks收取，详情请参见：计费逻辑说明。

iii. 选中**我了解数据探查服务需要收费**。

iv. 单击**提交**。

v. 待探查结束，在**数据探查**页签下，查看探查结果。

您可以从**探查记录**列表中，选择需要查看的探查结果。其中**数据分布 > 值范围**是对某个字段的数据值分布的阶段进行统计。

- **自动探查**

配置自动探查的操作如下：

i. 打开**自动探查**开关。

ii. 在**自动探查（当分区信息发生变化时进行探查）**对话框中，配置各项参数。

参数	描述
表名	格式为工作空间名称.表名，默认不可以修改。
分区取值	默认为**触发数据探查时的最新分区值**，不可以修改。
详细配置	选中需要探查的列。

参数	描述
触发绑定	在**触发绑定**列表中，选择需要关联的调度节点触发自动探查。您可以在**运维中心**查找调度节点的ID，建议您选择当前表对应的产出任务。 选择需要探查的指标并提交自动探查后，探查任务会在关联的调度任务运行完成后再运行，针对最新的分区进行探查。
预估费用	根据上述配置，预估运行探查任务所需要的费用。 **注意** ■ 数据探查需要执行MaxCompute SQL语句，会带来一定的MaxCompute计算费用。该页面的预估费用仅为参考，实际费用受处理的数据量影响，会有波动，请以MaXCompute账单为准。 ■ 数据探查复用数据质量产品能力，将会同时产生数据质量实例费用，此部分费用由DataWorks收取，详情请参见：计费逻辑说明。

iii. 选中**我了解数据探查服务需要收费**。

iv. 单击**提交**。

v. 待探查结束，在**数据探查**页签下，查看探查结果。

您可以从**探查记录**列表中，选择需要查看的探查结果。

5.4.1.3. 申请表权限

本文为您介绍如何在安全中心申请表的查询与操作权限，并在数据地图查看申请记录。

进入表详情页面

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 单击相应工作空间后的**进入数据开发**。
4. 单击左上角的图标，选择**全部产品 > 数据地图**。
5. 在顶部菜单栏，单击**全部数据**。
6. 单击需要查看的表类型。
7. 在相应页签下，单击需要申请权限的**表名**。

通过新版安全中心申请表权限

1. 在表详情页面，单击**申请权限**。

> **说明** 如果表被隐藏，则不会显示**申请权限**按钮。

2. 默认进入新版安全中心的**权限申请**页面。详情请参见数据访问控制。
3. 选择需要申请的表。
   i. 在**申请内容**区域，选择目标**工作空间**及**项目**。

      目前仅支持通过**数据访问控制**申请MaxCompute表的权限。

      因此**申请类型**默认为**表**，**引擎类型**默认为**MaxCopmute**。

   ii. 在**待添加表**区域，勾选需要申请的目标表。

      勾选目标表后，右侧会显示目标表的相关信息。单击**表名称**前 + 图标，显示当前表的所有字段，您可以选择申请目标表的部分或全部字段的权限。默认申请目标表全部字段的权限。

> **说明**
>
> - MaxCompute项目开启Policy权限控制后，该项目下的表才可以定义并在安全中心单独对表中具体字段申请权限。详情请参见：MaxCompute高级配置。关于MaxCompute表中字段的安全等级说明，详情请参见：Label权限控制。
> - 目前支持申请表级别的Select、Describe、Drop、Alter、Update、Download权限。同时支持您针对单个字段单独申请字段权限。

4. 配置**申请信息**。

参数	描述
使用者	○ **当前登录账号**：表示为当前登录DataWorks工作空间的阿里云账号申请目标表权限。 ○ **调度访问账号**：表示为被设置为调度访问身份的云账号申请目标表权限。选择该选项时，需要配置**工作空间**参数。 ○ **代他人申请**：表示当前登录DataWorks工作空间的阿里云账号为其他阿里云账号申请目标表权限。选择该选项时，需要配置**用户名**参数。
工作空间	被设置为调度访问身份的云账号。
用户名	当前登录阿里云账号以外的其他阿里云账号。
申请时长	支持您按需自定义申请表权限的时长，过期权限将自动收回。 **说明** 使用此功能前需要表所在的MaxCompute项目开启Policy授权，详情请参见：MaxCompute高级配置，关于MaxCompute Policy的说明请参见：Policy权限控制。
申请原因	输入申请目标表权限的原因。

5. 单击**申请权限**，提交申请。

   您可以在**权限申请记录**页签，查看当前申请的审批详情及审批记录。

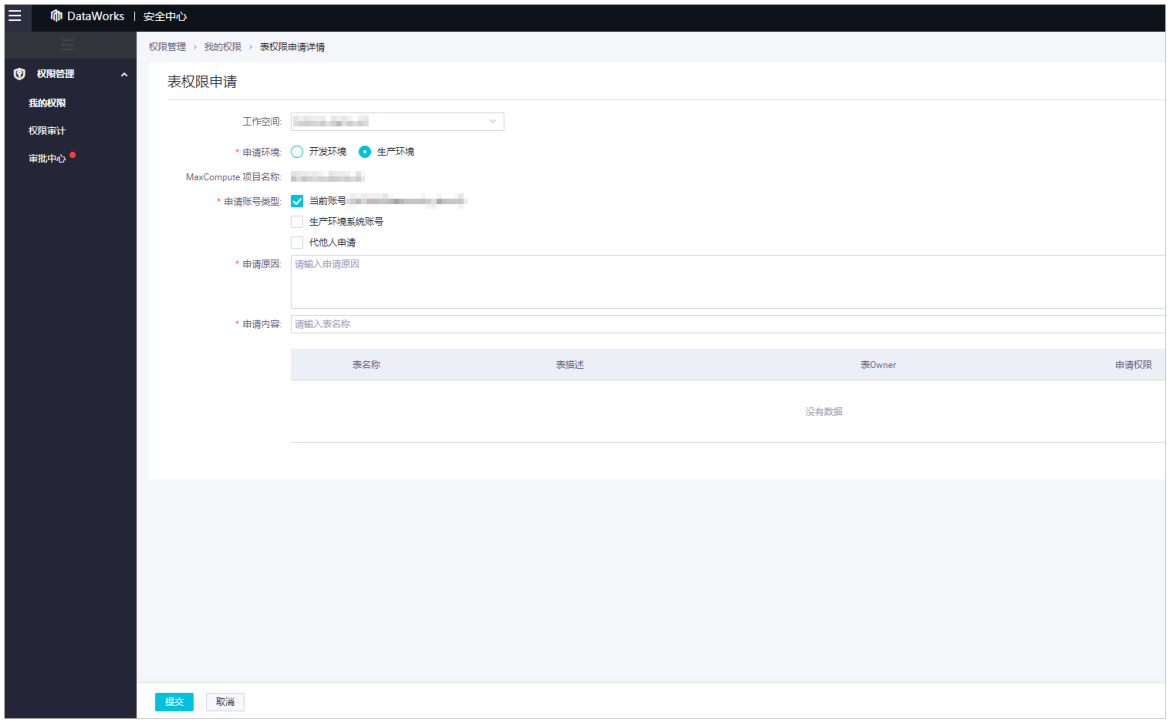
通过旧版安全中心申请表权限

在表详情页面，单击**申请权限**，默认进入新版安全中心的**权限申请**页面。您可以单击**安全中心**顶部菜单栏的**返回旧版**，即可进入旧版安全中心。

1. 在表详情页面，单击**申请权限**，进入新版安全中心的**权限申请**页面。

> **说明** 如果表被隐藏，则不会显示**申请权限**按钮。

2. 单击顶部菜单栏的**返回旧版**，进入旧版安全中心的**我的权限**页面。
3. 单击**申请权限**，进入**表权限申请**页面。
4. 在**表权限申请**页面，配置各项参数。



参数	描述
工作空间	需要申请权限的表所在的工作空间。
申请环境	标准模式的工作空间包括**开发环境**和**生产环境**，简单模式的工作空间仅支持**生产环境**。
MaxCompute项目名称	选择的DataWorks工作空间对应的MaxCompute项目名称，默认不可以修改。

参数	描述
申请账号类型	包括**当前账号**、**生产环境系统账号**和**代他人申请**。
申请原因	简要说明申请表权限的原因，以便更快地通过审批。
申请内容	选择需要申请的内容。

5. 单击**提交**。

查看申请状态

1. 单击左上角的☰图标，选择**全部产品 > 数据地图**。
2. 在顶部菜单栏，单击**我的数据**。
3. 在左侧导航栏，单击**权限管理**。
4. 在**权限管理**页面，单击**申请记录**。
5. 单击相应记录后的**查看**，确认申请状态。

5.4.2. API详情

5.4.2.1. 查找API

数据地图支持对当前租户下所有工作空间的API进行搜索和定位，实现API的高效查找。

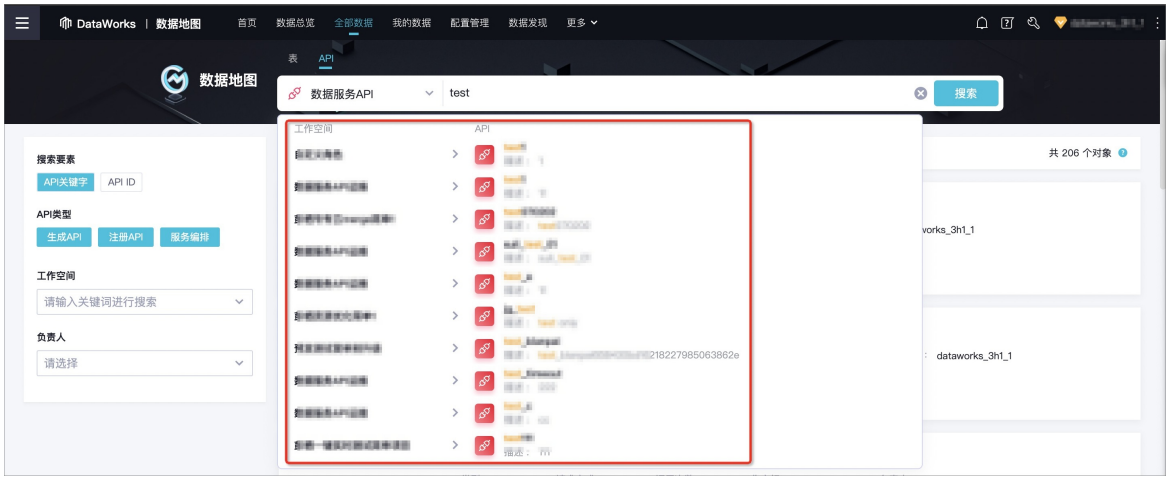
进入API查找界面

1. 登录DataWorks控制台后，进入**数据地图**页面，操作详情请参见进入首页。
2. 在顶部菜单栏，单击**全部数据**。
3. 在搜索框上方，选择**API**类目，进入查找API的页面。

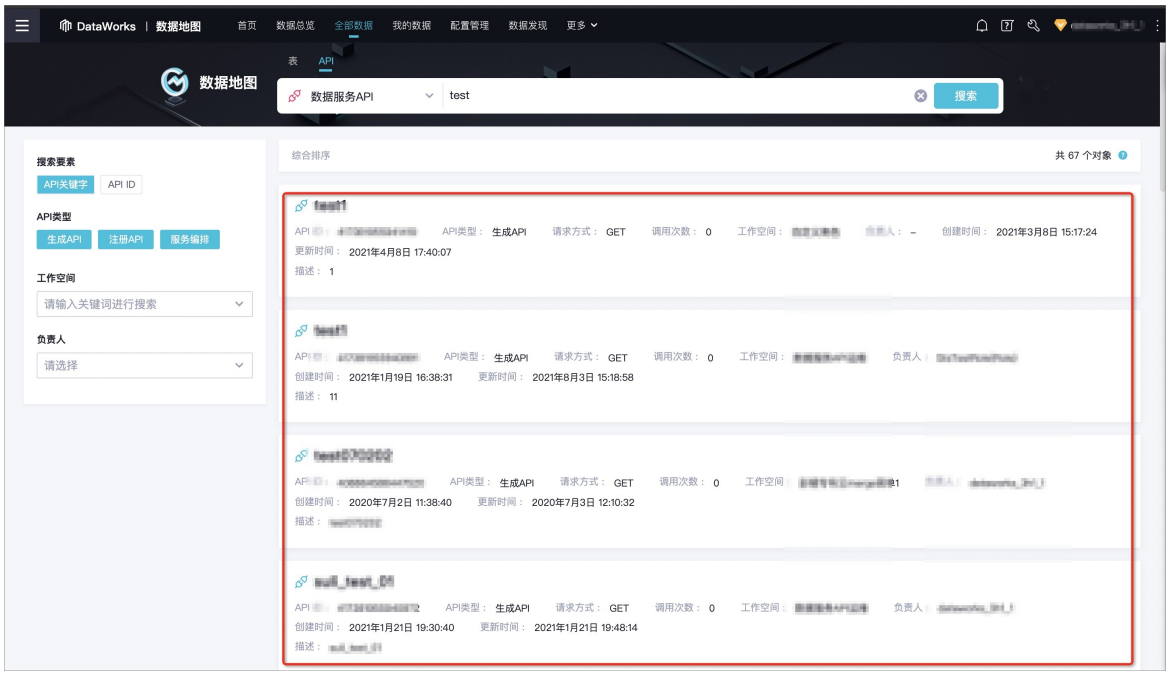
查找API

1. 在API类目下的搜索框中输入**API ID**、**API Path**、**API名称**、**API描述**等关键字进行搜索。

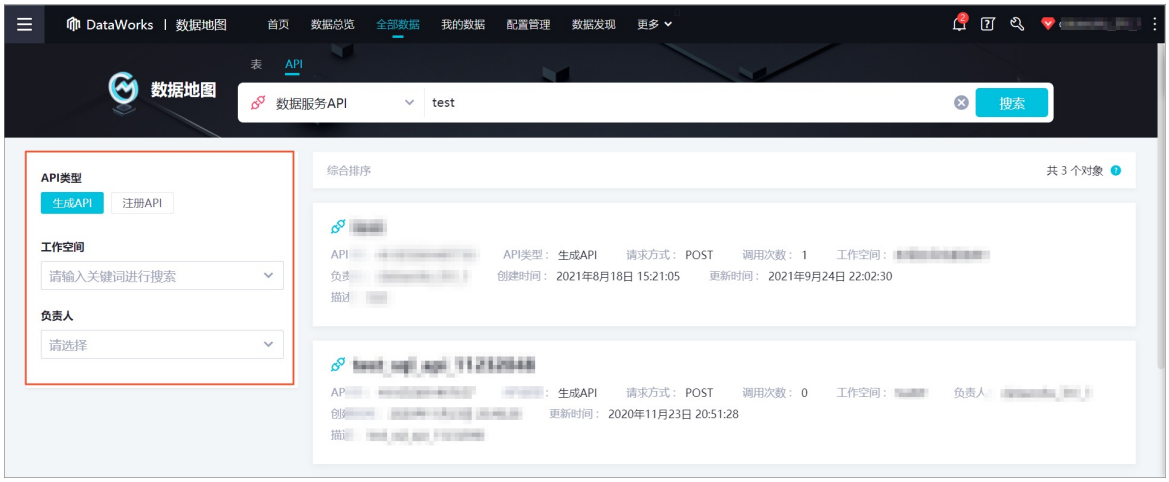
   在搜索框中输入关键字后，将通过关键字匹配，在搜索框的下拉列表中初步筛选，展示出符合条件的前10个API。



2. 如果在搜索框的下拉列表中无法定位到目标API，可以单击搜索，在列表中展示出所有符合条件的API。



3. 您可以在页面左侧根据**API类型**、**工作空间**、**负责人**对搜索结果中的API进行二次过滤。



4. 找到目标API后，如需查看API详情，可以单击目标API进入详情页。详情请参见查看API详情。

5.4.2.2. 查看API详情

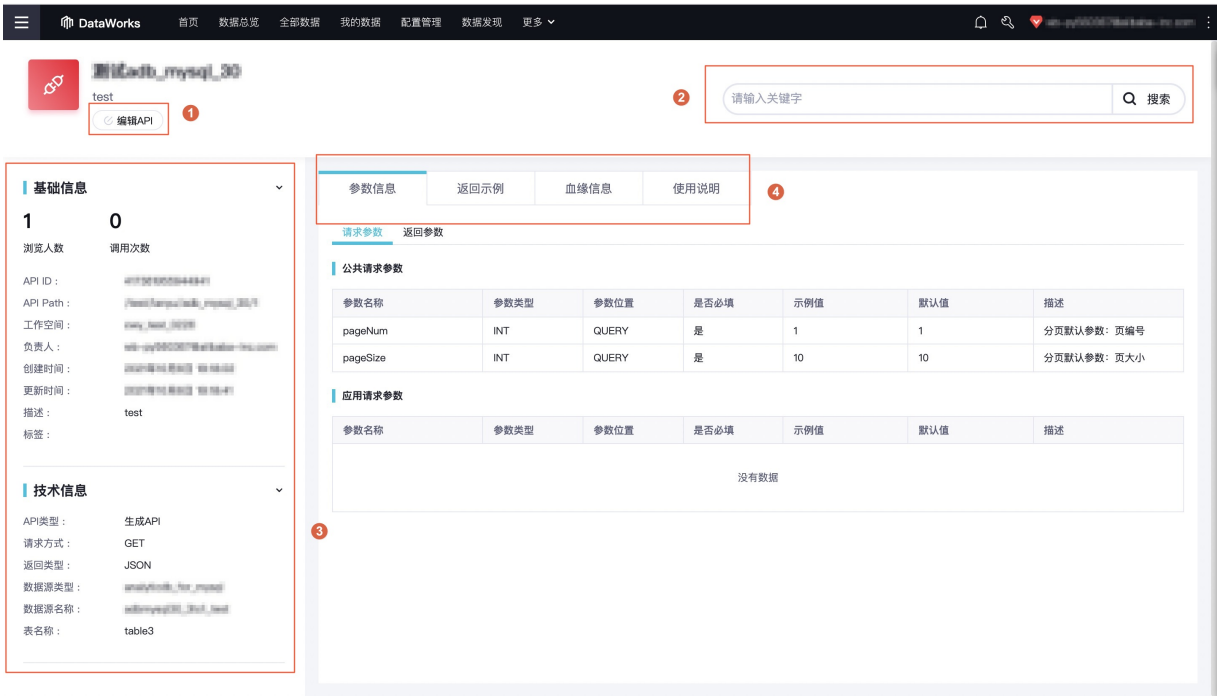
本文为您介绍如何进入API详情页面，查看API的基础信息、参数信息、返回示例等详情。

进入API详情页面

1. 登录**DataWorks控制台**后，进入**数据地图**页面，操作详情请参见进入首页。
2. 在顶部菜单栏，单击**全部数据**。
3. 在搜索框上方，选择**API**类目，单击需要查看的**API名称**。

查看API详情

API详情页面为您展示API的**基础信息**、**技术信息**、**参数信息**、**返回示例**和**使用说明**。



区域	描述
①	单击**编辑API**跳转至API所属工作空间的编辑窗口进行API编辑。详情请参见发布API和注册API。 **说明**：如果当前登录账号不在API所属的工作空间中，请联系目标工作空间对应管理员将您的账号添加到目标空间的成员列表中并赋予开发者角色，详情请参见角色及成员管理：空间级。
②	在该区域您可以输入**API ID**、**API Path**、**API名称**、**API描述**等关键字搜索API。
③	在该区域您可以查看该API的如下信息： • **基础信息**：您可以查看API的**浏览人数**、**调用次数**、**API ID**、**工作空间**、**负责人**、**创建时间**、**更新时间**、**描述**等信息。 • **技术信息**：您可以查看**API类型**、**请求方式**、**返回类型**、**数据源类型**、**数据源名称**、**表名称**等信息。

区域	描述

④	在该区域您可以查看该API的如下信息：
	• **参数信息**：您可以查看API的**请求参数**、**返回参数**。
	  ○ **请求参数**：您可以查看**公共请求参数**和**应用请求参数**的**参数名**、**参数类型**、**参数位置**、**示例值**、**默认值**、**描述**、**标签**等信息。
	  ○ **返回参数**：您可以查看**公共返回参数**和**应用返回参数**的**参数名**、**参数类型**、**示例值**、**描述**等信息。
	• **返回示例**：在API发布前，如果对API进行测试并在测试时勾选了**自动保存正常返回示例**，将根据测试结果更新API返回示例。详情请参见测试API，您可以根据**返回示例**的数据内容和数据结构了解API返回结果。
	• **血缘信息**：您可以查看数据表、API、APP之间的完整血缘关系。数据接口API的上游血缘支持的数据源类型包括：Hologres、MySQL、PostgreSQL、SQL Server、Oracle、AnalyticDB for MySQL 3.0、AnalyticDB for PostgreSQL。
	说明
	○ **生成API**是通过数据源或表封装的API，因此血缘链路将包括：数据表、API、APP。
	○ **注册API**是通过后端地址封装的API，因此血缘链路将仅包括：API和APP。
	○ 血缘信息的产出、更新的时效是**T+1**。
	• **使用说明**：您可以进行**编辑**、**查看历史版本**的操作，根据业务说明了解相关的信息。

5.5. 我的数据

本文为您介绍如何在我的数据页面，查看拥有的数据、管理的数据、生产账号的数据、收藏的数据，以及如何管理相应数据的权限。

背景信息

数据地图的部分数据是离线（T+1）更新的，会存在数据延迟的情况，建议您以SQL查询的结果为准。

进入我的数据

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 单击相应工作空间后的**进入数据开发**。
4. 单击左上角的图标，选择**全部产品 > 数据地图**。
5. 在顶部菜单栏，单击**我的数据**，默认进入**我的数据 > 我拥有的数据**页面。

查看我的数据

您可以查看当前账号拥有的表，您还可以根据关键字、环境、项目/数据库和可见范围等条件进行搜索，查看相应表的具体信息并进行操作。

- 我拥有的数据：显示当前用户是表owner时的所有表。

- 我管理的数据：显示当前用户是空间管理员时对应的工作空间中的所有表。
- 生产账号的数据：显示当前用户是成员时的工作空间对应的生产账号中的所有表。

进入上述页面后，您可以查看表的如下信息：

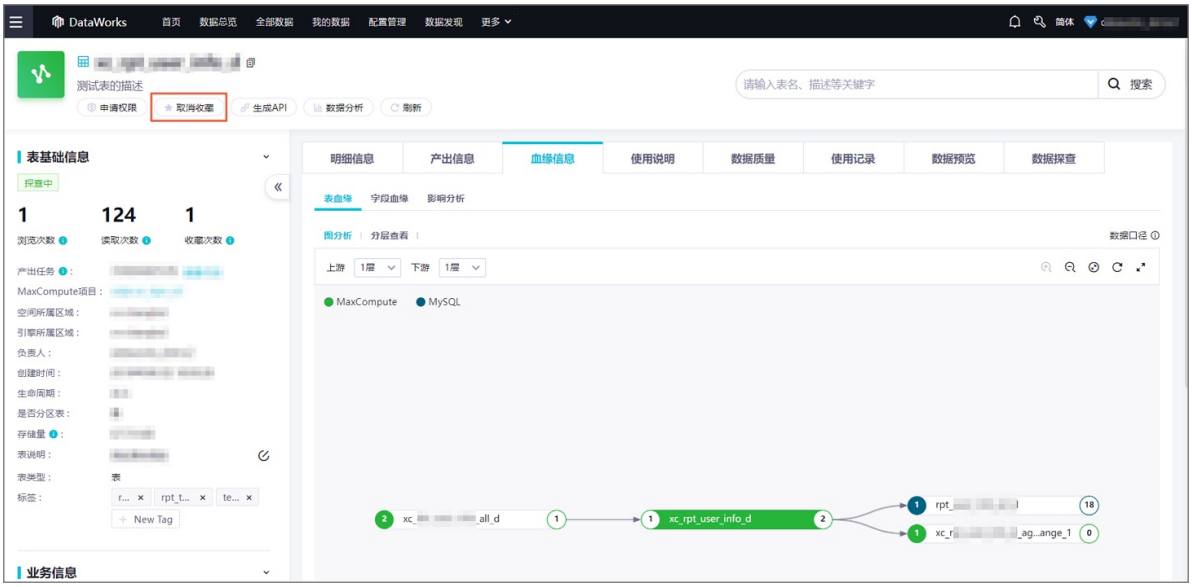
名称	描述
表名	单击相应的表名，进入该表的详情页面。
中文名	单击**中文名**下的图标，编辑表的中文名。
项目名/数据库	如果您的表是在不同的环境，会有相关的后缀。例如，_dev表示开发环境。
是否隐藏	单击**是否隐藏**下的图标，设置显示或隐藏表。 您可以选择表状态为 • 隐藏，即所有人都不可以通过搜索来访问到该表。 • 仅项目，即仅对表所在工作空间下用户可见（可搜索到）。 • 显示，默认显示，即所有人都可以通过搜索来访问到该表。 **说明** 表owner和工作空间管理员默认不受上述权限控制。
生命周期	和您创建表时设置的生命周期一致。
环境类型	包括**开发**和**生产**。
存储量	显示您存储的数据量。
收藏人次	收藏该表的人次。
30天浏览人次	30天内，浏览该表的人次。
创建时间	创建该表的时间。
操作	您可以在相应表后的操作栏下，进行**删除**和**修改类目**的操作。
批量操作	选中相应的表，可以进行**批量修改**、**批量转交**、**批量删除**和**批量修改类目**等操作。

查看我的收藏

您在查看表详情时，可以通过表详情页的收藏按钮，快速将表加入的**我的收藏**，详情请参见查看表详情，添加收藏后，您可以通过**我的数据**页面下的**我的收藏**分组中进行查看。

您可以通过下面两种方式取消收藏表，取消收藏后，将不会展示在我的数据页面下的我的收藏中列表中。

- 您可以单击相应表后的**取消收藏**，取消对该表的收藏。
- 您可以通过已收藏的表详情页面中的**取消收藏**入口，快速取消收藏该表。



查看和管理权限

在左侧导航栏，单击**权限管理**，查看和管理权限。

您可以在**权限管理**页面**申请函数和资源权限**，并查看**待我审批**、**申请记录**和**我已处理的**。

- **申请函数和资源权限**：
  i. 单击右上角的**申请函数和资源权限**。
  ii. 在**申请数据权限**对话框，配置各项参数。

参数	描述
申请类型	包括**函数**和**资源**两种类型。

参数	描述
权限归属人	包括**本人申请**和**代理申请**： ■ **本人申请**：如果选择本人申请，审批通过后权限归属于当前用户。 ■ **代理申请**：如果选择代理申请，请输入**对方用户名**，审批通过后权限归属于被代理人。
项目名	选择需要申请函数或者资源所在的项目名称（对应的MaxCompute项目名称），支持在本组织范围内模糊匹配查找。
函数名称或资源名称	输入项目中的函数或资源名称。请输入完整的资源名称，包括文件后缀，例如my_mr.jar。
权限有效期	申请表权限的时长，单位为天，不填则默认为永久。超过申请权限时长时，系统将自动回收该权限。
申请理由	请简要填写申请权限的理由。

- **待我审批**

  当前访问账号为管理员时，可以在**待我审批**页面，查看并审批所有项目下的表、资源和函数等待审批的权限申请记录。

- **申请记录**

  在**权限管理**页面，单击**申请记录**页签。

  在**申请记录**页签，当前访问账号可以查看其权限申请记录。

- **我已处理的**

  在**权限管理**页面，单击**我已处理的**页签。

  在**我已处理的**页签，当前账号为管理员时，可以查看其所有项目下已经处理过的表、资源和函数等权限申请记录。

手工同步表

如果表存在但是搜索不到或者表更新了但是数据地图显示还未更新，您需要手工同步一下。

在数据地图的**我的数据** > **手工同步表**页面，输入格式为 `项目名称.表名称` 的**表GUID**后，单击**手工同步**。完成操作后您可以在数据地图的全部数据中再次搜索关键词查询对应的表。

5.6. 配置管理

本文为您介绍如何在数据地图的配置管理页面，配置类目导航和工作空间下的MaxCompute表管理。

背景信息

类目管理功能方便您通过类别有效地组织和管理表，您需要添加相应权限才能使用该功能，如下所示：

- 如果您使用主账号来管理类目，则默认具有该功能的所有操作权限。
- 如果您使用子账号来管理类目，则需要为子账号添加AliyunDataWorksFullAccess权限。详情请参见给RAM子账号授权。

进入配置管理页面

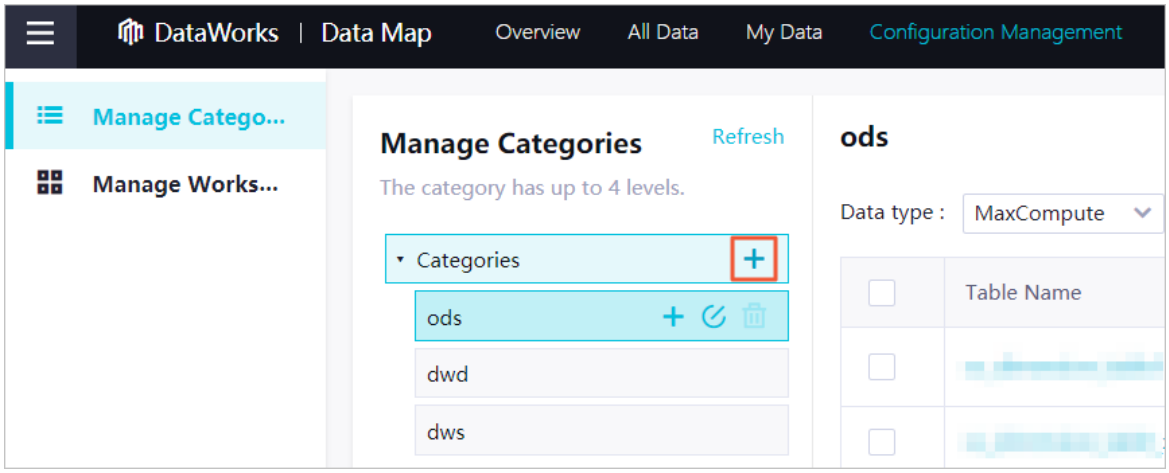
1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据地图**。
4. 在顶部菜单栏，单击**配置管理**，默认进入**配置管理 > 类目导航配置**页面。

   配置管理页面包括**类目导航配置**和**项目管理配置**两个模块。

类目导航配置

类目导航配置方便您通过类别有效地组织和管理表，表的类目管理配置完成后，您可以在查找表时，通过类目来过滤目标表。同时支持您将指定表加入到个人收藏，方便快速查看。配置完成的类目导航对阿里云当前region下所有工作空间成员生效。

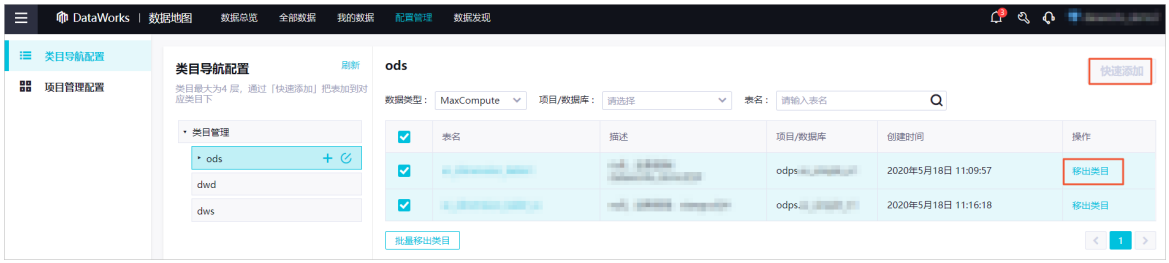
1. 在**类目导航配置**页面，鼠标悬停至**类目管理**，单击显示在**类目管理**后的+图标，即可添加一级类目。



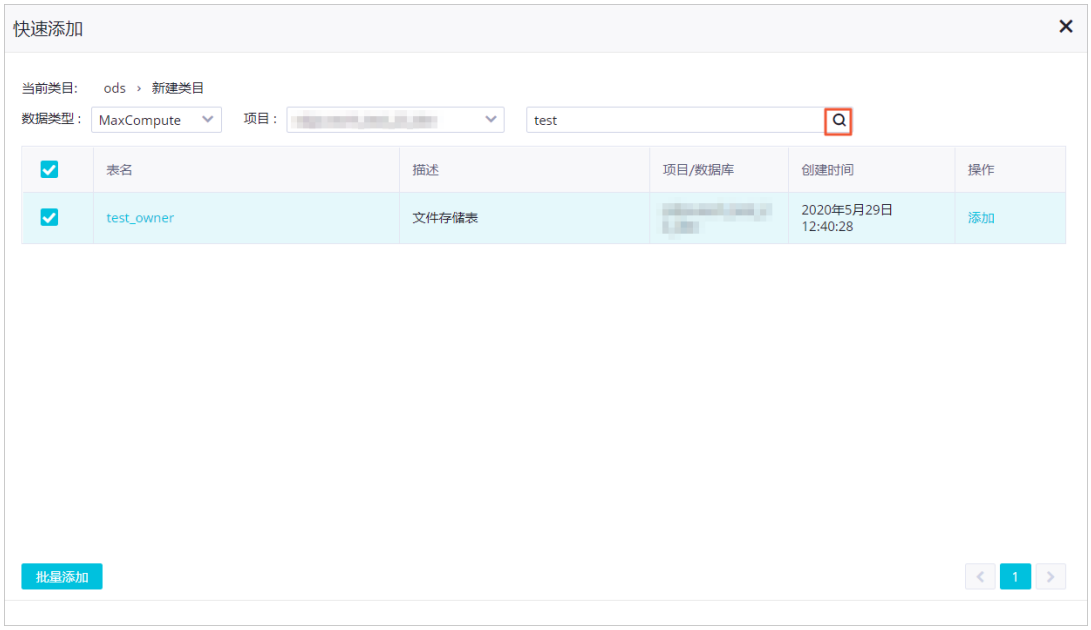
2. 鼠标悬停至相应的一级类目，单击显示在一级类目后的+图标，即可添加所属二级类目。

以此类推，最多支持添加四级类目。您可以单击图标编辑相应类目的名称，也可以单击图标删除相应类目。

3. 创建类目后，您可以**快速添加**表至该类目，也可以将表**移出类目**。



- **快速添加**表至当前类目：
  a. 在当前类目下，单击右上方的**快速添加**。
  b. 在**快速添加**对话框中，选择**数据类型**、**项目**，并输入表的名称或关键字，单击图标。



c. 单击相应表后的**添加**，即可添加表至当前类目。

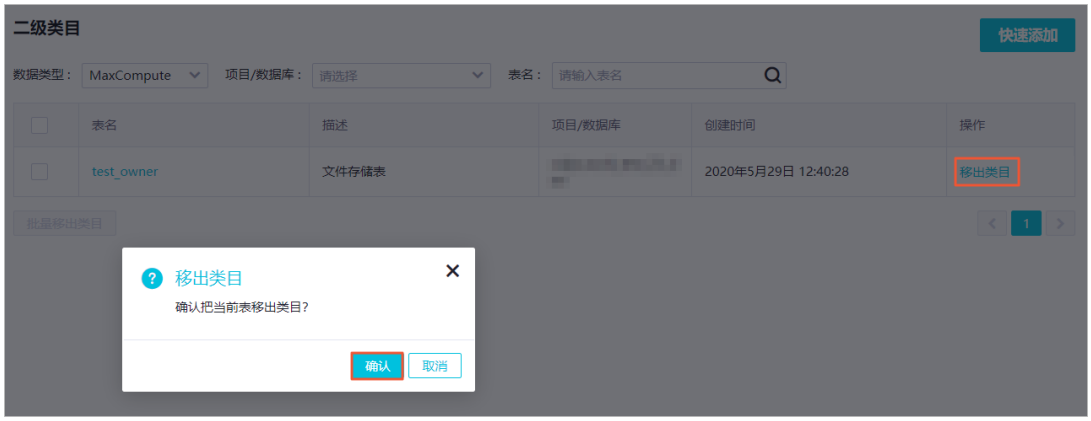
如果需要添加多张表，请全部选中后，单击**批量添加**。

- 将表**移出类目**：

a. 单击相应表后的**移出类目**。

如果您需要移出多张表，请全部选中后，单击**批量移出类目**。

b. 在**移出类目**对话框中，单击**确认**。



项目管理配置

您可以通过**项目管理配置**来控制是否可以在在数据地图表详情页的数据预览功能查看表数据。

1. 在左侧导航栏，单击**项目管理配置**。
2. 在**工作空间**区域，单击相应的工作空间。
3. 在右侧的**MaxCompute表管理**区域，根据业务需求开启或关闭**预览开发表**或**预览生产表**。

4. 开启**预览生产表**时，需要在**注意**对话框中单击**我已知悉，确定开启**。

> **说明**
>
> - 如果您使用的是简单模式的工作空间，仅可以设置是否开启**预览生产表**。
> - 如果您开启**预览生产表**，该项目生产环境的表，无需申请访问权限，即可被公司内的全部成员预览。可能存在敏感数据泄露的风险，请谨慎评估后再开启。

5.7. 数据发现

5.7.1. 元数据采集

5.7.1.1. 采集E-MapReduce元数据

本文为您介绍如何新建E-MapReduce采集器，采集E-MapReduce元数据至DataWorks。采集完成后，您可以在数据地图查看相关数据。

前提条件

在工作空间绑定EMR引擎后，您才可以进行EMR元数据采集操作，EMR引擎绑定详情请参见：绑定E-MapReduce计算引擎。

背景信息

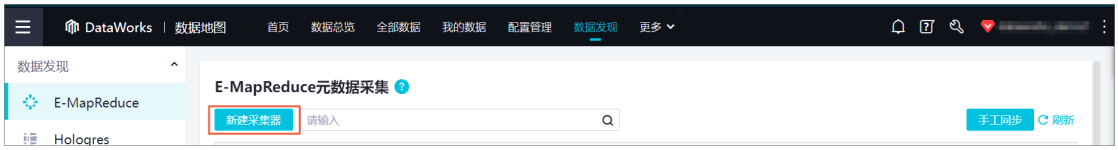
全量采集元数据后，系统会开启自动增量采集，自动同步表中新增的元数据。

使用限制

- 一个集群仅支持新建一个采集器，一个采集器中可以选择一个或多个需要进行元数据采集的DB。
- 仅阿里云主账号，拥有AliyunDataWorksFullAccess权限的子账号、元数据采集管理员可以进行采集。

新建采集器

1. 登录**DataWorks控制台**后，进入**数据地图**页面，操作详情请参见进入首页。
2. 在顶部菜单栏，单击**数据发现**。
3. 新建采集器。
   i. 在左侧导航栏，单击**元数据采集 > E-MapReduce**。
   ii. 在**E-MapReduce元数据采集**页面，单击**新建采集器**。



4. 配置采集器。
   i. 在**新建采集器**对话框中，单击**选择集群**下拉列表，选择目标集群。

> **说明**

   ii. （可选）单击**选择DB**下拉列表选择需要进行元数据采集的DB，如果不选择，默认采集该集群内所有DB的元数据。

iii. 单击**前往开启授权**，在E-MapReduce控制台中所选集群的**元数据**页面，单击**开启元数据收集**。

> **说明**
>
> - DataWorks新绑定的EMR集群，默认已开启元数据收集权限。
> - DataWorks已绑定但还未采集元数据的EMR集群，需要前往开启授权。

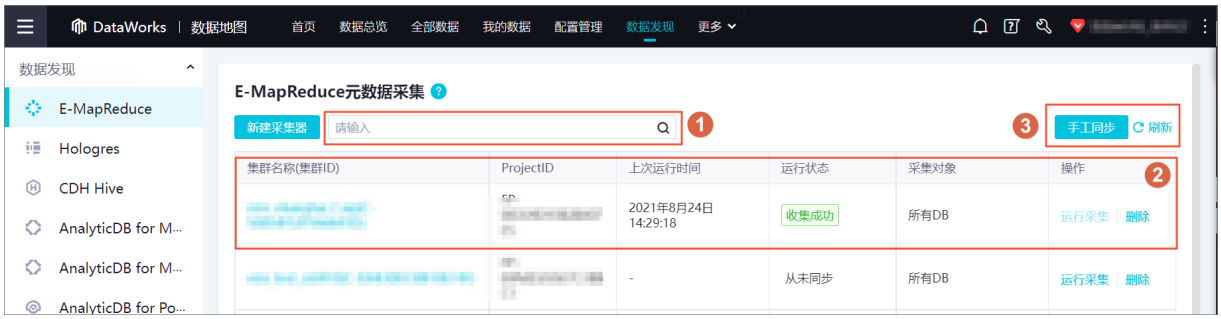
iv. 在弹出的**确认开关变更**对话框中，单击**确定**。

v. 成功开启元数据收集后，返回**数据地图**中的**新增采集器**对话框，单击**刷新**。

vi. **授权状态**刷新为**已授权**后，单击**确定**，即可完成采集器的创建。

管理采集器

您可以在**E-MapReduce元数据采集**页面，对已创建的采集器进行删除、运行采集等操作。



序号	描述
①	在该区域，您可以输入采集器名称，搜索目标采集器。 **说明** 搜索采集器支持模糊匹配，即输入关键词，即可显示名称中包含该关键词的所有采集器。

序号	描述

②	在该区域，您可以查看相应采集器的**运行状态**、**采集对象**、**上次运行时间**等信息。
	• **运行状态**：已创建的采集器的状态。
	  ○ 收集成功：表示采集器已成功完成元数据采集。
	  ○ 从未同步：表示您创建采集器后还未运行采集。
	  ○ 采集失败：表示运行采集器后元数据采集失败，您可以尝试重新运行采集，如果还未成功，请提交工单联系我们处理。
	• **采集对象**：展示已采集的DB信息。
	• **上次运行时间**：表示上次运行采集器的时间。
	您还可以对目标采集器执行如下操作：
	• **运行采集**：运行采集器，根据目标采集器的配置信息采集数据。
	  ○ 对未运行采集的集群，您可以单击操作列的**运行采集**，执行成功后，运行状态变为收集成功，完成元数据采集。
	  ○ 对已运行采集的集群，操作列的**运行采集**按钮无法单击。如果需要重新选择DB进行采集，您可以单击**删除**按钮，删除相应采集器后，重新创建采集器。
	• **删除**：如果您需要删除采集器，请单击相应采集器后的**删除**，在**删除实例**对话框中，单击**确定**。

③	在该区域，您可以执行如下操作： • **手工同步**：如果表存在但是搜索不到或者表更新了但是数据地图显示还未更新，您可以单击**手工同步**，选择目标**集群ID**、**数据库**、**表名**后，手工同步该表。 • **刷新**：刷新采集器运行的状态及结果。

后续步骤

采集E-MapReduce元数据成功后，您可以在数据地图的**全部数据**页签查看已采集的数据详情。详情请参见查找表。

5.7.1.2. 采集OTS元数据

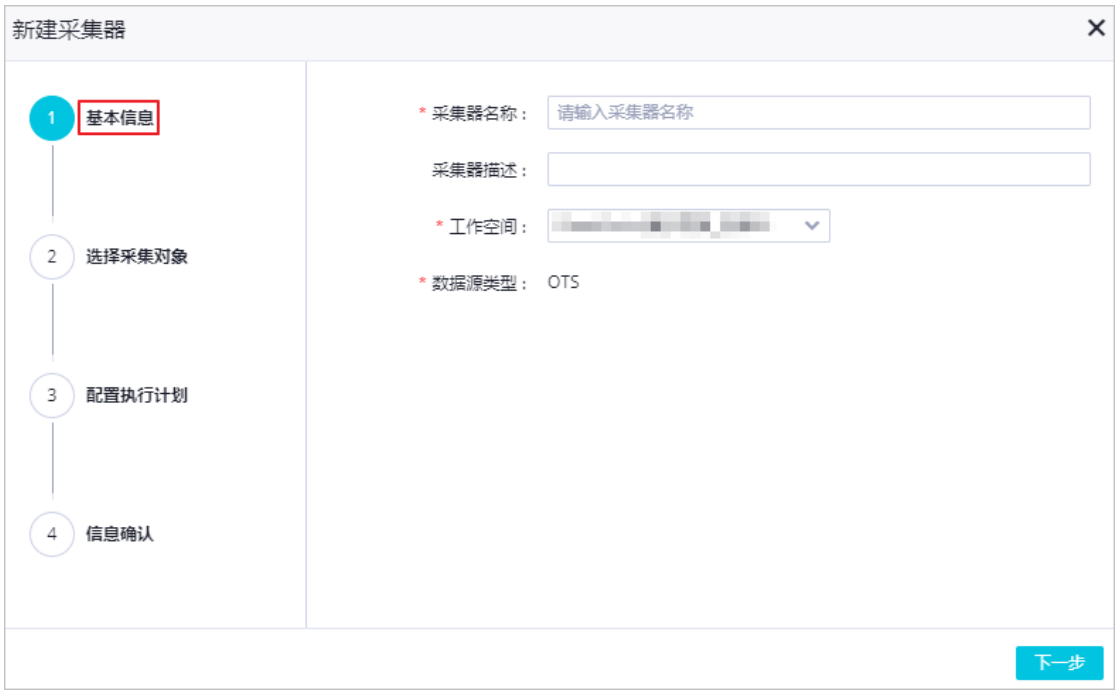
采集元数据是用于把表结构及血缘关系采集到数据地图中，清楚的为您展示表的内部结构及与表相关的关联关系。本文为您介绍如何新建采集器，并采集OTS元数据至DataWorks。采集完成后，您可以在数据地图查看数据。

背景信息

全量采集元数据后，系统会开启自动增量采集，自动同步表中新增的元数据。

1. 登录**DataWorks控制台**后，进入**数据地图**页面，操作详情请参见进入首页。
2. 在顶部菜单栏，单击**数据发现**。
3. 在左侧导航栏，单击**元数据采集 > OTS**。
4. 在**OTS元数据采集**页面，单击**新建采集器**。
5. 在**新建采集器**配置向导页面，完成以下操作。

i. 在**基本信息**页签，配置各项参数。



参数	描述
采集器名称	采集器的名称，必填且唯一。
采集器描述	对采集器进行简单描述。
工作空间	采集对象（数据源）所属的DataWorks工作空间。
数据源类型	采集对象的类型，默认为**OTS**。

ii. 单击**下一步**。

iii. 在**选择采集对象**页签，从**数据源**下拉列表中选择相应的数据源。

如果列表中没有您需要的数据源，请单击**去新建**，进入**工作管理空间 > 数据源管理**页面新建数据源，详情请参见配置OTS数据源。

iv. 单击**测试采集连通性**。

v. 待显示**测试成功**，单击**下一步**。

如果显示**测试连通性未通过**，请检查数据源是否配置正确。

vi. 在**配置执行计划**页签，配置**执行计划**。

执行计划包括**按需执行**、**每月**、**每周**、**每天**及**每小时**。根据不同的执行周期，生成不同的执行计划，在相应执行计划的时间内，对目标数据源进行元数据采集。具体如下：

- 按需采集：根据实际业务需求，在业务需要时才会采集OTS元数据。

- 月采集：即在每月的特定几天，在特定时间点自动采集一次OTS元数据。

  > **注意** 部分月份不包含29、30、31日，请您谨慎选择月末日期。

  如下图所示，在每月的1、11及21日的09:00，系统会自动采集一次OTS元数据。**CRON 表达式**会根据您的配置自动生成。

- 周采集：即在每周的特定几天，在特定时间点自动采集一次OTS元数据。

  如下图所示，在每周的星期一（MON）及星期天（SUN）的03:00，系统会自动采集一次OTS元数据。

  不输入**时间**时，则默认在每周指定几天的00:00:00采集。

- 天采集：即在每天特定的时间点自动采集一次OTS元数据。

  如下图所示，在每天的01:00，系统会自动采集一次OTS元数据。

- 小时采集：即在每小时的第 `N*5分钟` 自动采集一次OTS元数据。

  > **说明** 目前小时周期的采集任务，仅支持选择的周期时间为第5分钟的倍数。

  如下图所示，在每小时的第5分钟和第10分钟，系统会自动采集一次OTS元数据。

vii. 单击**下一步**。

viii. 在**信息确认**页签，确认配置信息无误后，单击**确认**。

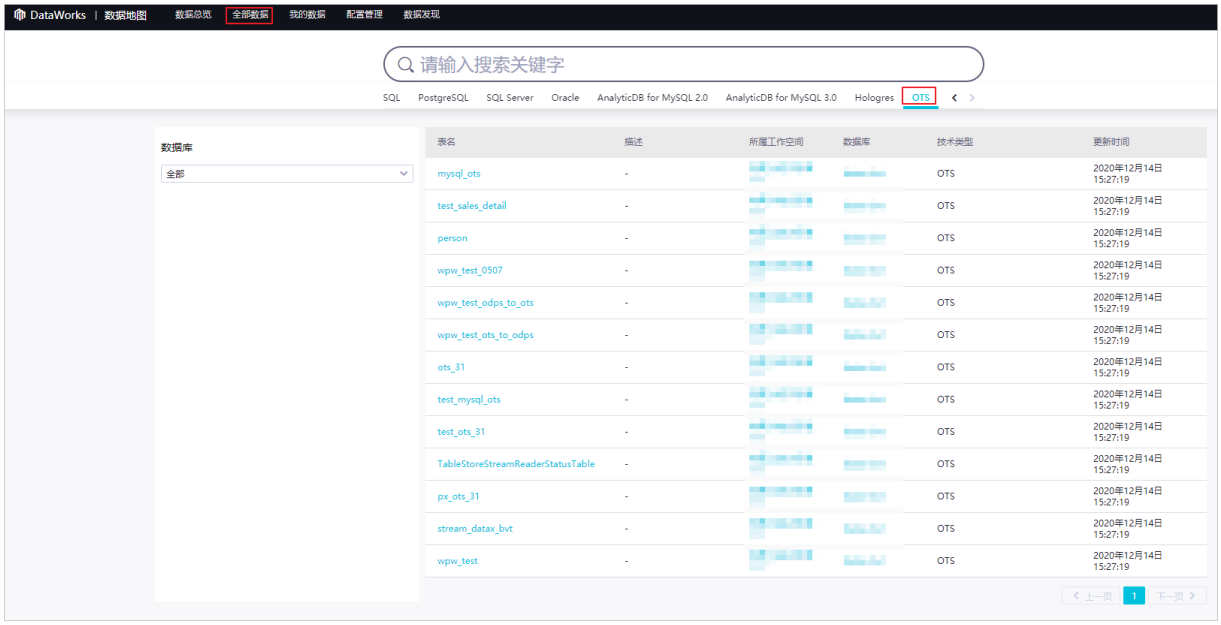
6. 在**OTS元数据采集**页面，您可以查看并管理目标采集器的相关信息。

主要操作说明如下：

- 您可以查看相应采集器的运行**状态**、**运行计划**、**上次运行时间**、**上次消耗时间**、**平均运行耗时**及上次运行时更新及添加的表数量。
- 单击目标采集器**操作**列的**详情**、**编辑**、**删除**、**运行中**及**停止**，执行相应操作：
  - **详情**：查看该采集器的**采集器名称**、**数据源类型**及**执行计划**。
  - **编辑**：修改该采集器的信息。
  - **删除**：删除该采集器。
  - **运行**：单击**运行**，即可根据该采集所配置的任务采集数据。仅当**执行计划**配置为**按需执行**时，才会生成**运行**操作，其他周期计划的任务不涉及该操作。
  - **停止**：停止运行该采集器。

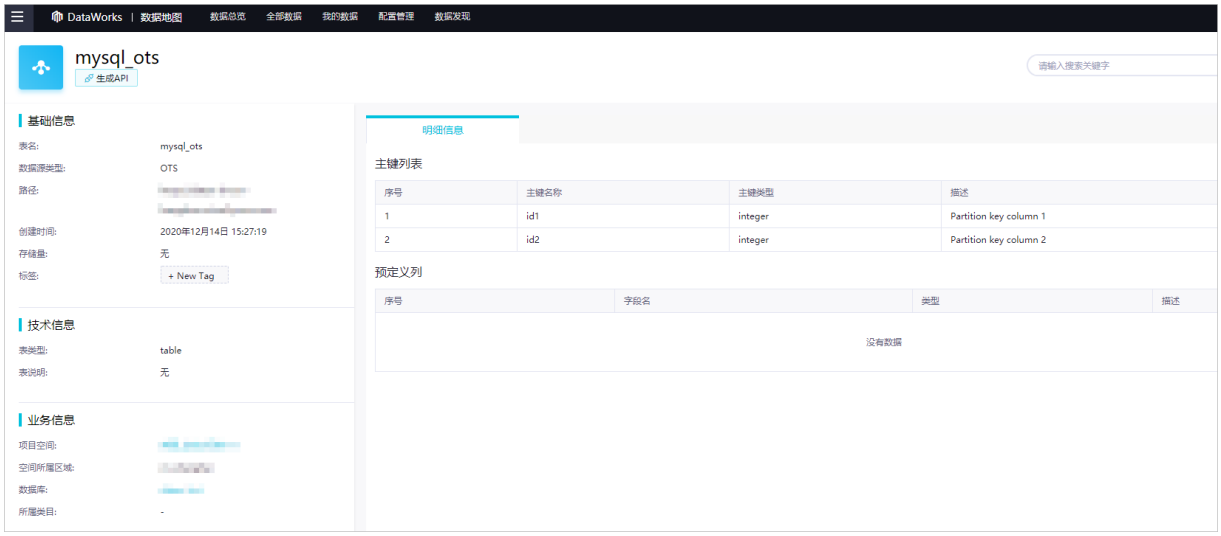
执行结果

采集OTS元数据成功后，您可以在**全部数据** > **OTS**页面查看已采集的表。

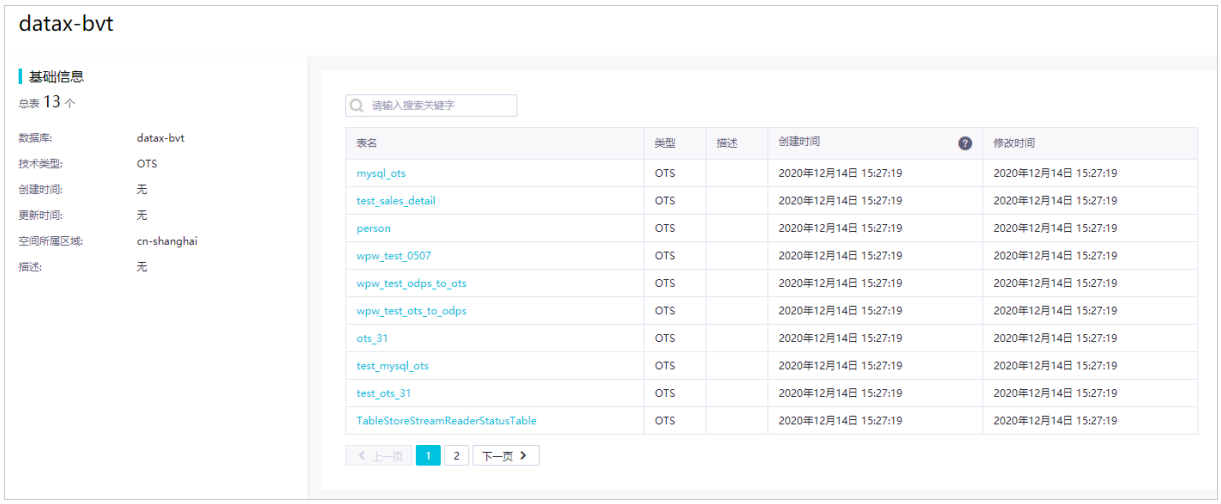


单击**表名**、**所属工作空间**及**数据库**，即可查看对应类目的详细信息。

示例一：查看*mysql_ots*表的详细信息。



示例二：查看*datax-bvt*数据库包含的所有表信息。



5.7.1.3. 采集MySQL元数据

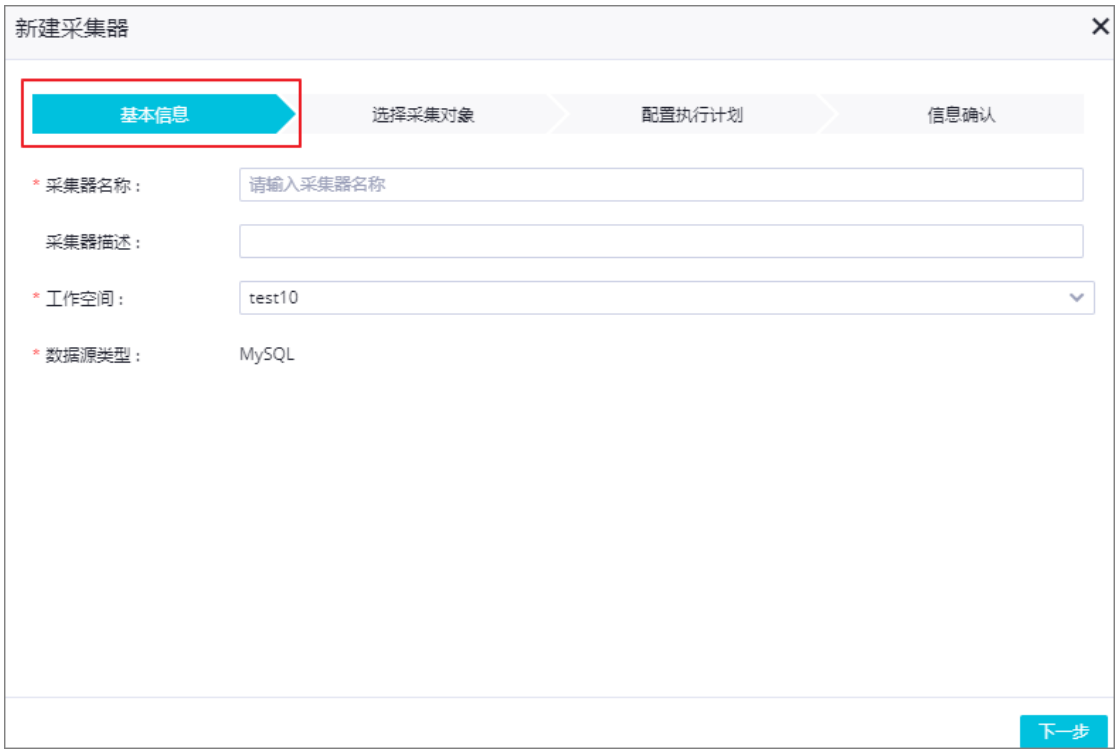
本文为您介绍如何新建采集器，以采集MySQL元数据至DataWorks。采集完成后，您可以在数据地图查看数据。

背景信息

全量采集元数据后，系统会开启自动增量采集，自动同步表中新增的元数据。

操作步骤

1. 登录**DataWorks控制台**后，进入**数据地图**页面，操作详情请参见进入首页。
2. 在顶部菜单栏，单击**数据发现**。
3. 在左侧导航栏，单击**元数据采集 > MySQL**。
4. 在**MySQL元数据采集**页面，单击**新建采集器**。
5. 在**新建采集器**配置向导页面，完成以下操作。
   i. 在**基本信息**页签下，配置各项参数。



参数	描述
采集器名称	采集器的名称，必填且唯一。
采集器描述	对采集器进行简单描述。
工作空间	采集对象（数据源）所属的DataWorks工作空间。
数据源类型	采集对象的类型，默认为**MySQL**。

   ii. 单击**下一步**。

iii. 在**选择采集对象**页签下，从**数据源**下拉列表中选择相应的数据源。

如果没有您需要的数据源，请单击**去新建**，进入**工作空间管理 > 数据源管理**页面新建，详情请参见配置MySQL数据源。

> **说明** 支持阿里云RDS实例模式和具备公网访问能力的JDBC连接串模式的MySQL数据源。

iv. 单击**测试采集连通性**。

v. 待显示**测试成功**，单击**下一步**。

vi. 在**配置执行计划**页签下，配置各项参数。

执行计划包括**按需执行**、**每月**、**每周**、**每天**和**每小时**。

vii. 单击**下一步**。

viii. 在**信息确认**页签下，确认配置信息无误后，单击**确认**。

6. 在**MySQL元数据采集**页面，单击相应采集器后的**运行**。

采集MySQL元数据成功后，您可以在**全部数据 > MySQL**页面查看已采集的表。

单击**表名**，即可查看该表的详情。

5.7.1.4. 采集SQL Server元数据

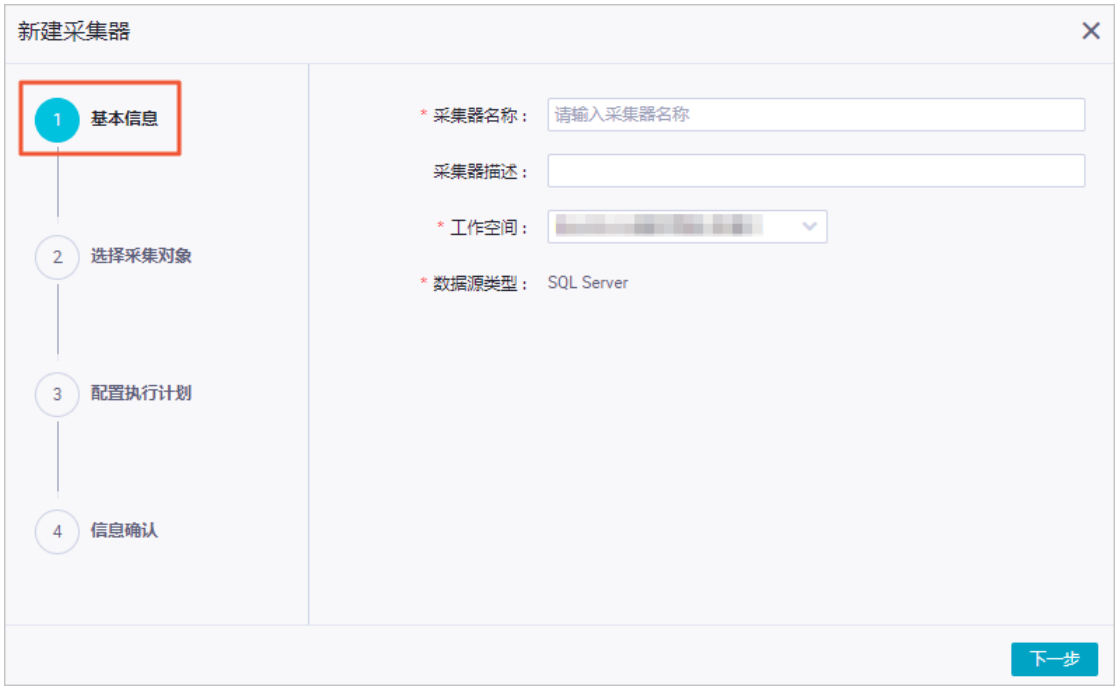
本文为您介绍如何新建采集器，以采集SQL Server元数据至DataWorks。采集完成后，您可以在数据地图查看数据。

背景信息

全量采集元数据后，系统会开启自动增量采集，自动同步表中新增的元数据。

操作步骤

1. 进入数据发现页面。
2. 在左侧导航栏，单击**元数据采集 > SQL Server**。
3. 在**SQL Server元数据采集**页面，单击**新建采集器**。
4. 在**新建采集器**配置向导页面，完成以下操作。
   i. 在**基本信息**页签下，配置各项参数。



参数	描述
采集器名称	采集器的名称，必填且唯一。
采集器描述	对采集器进行简单描述。
工作空间	采集对象（数据源）所属的DataWorks工作空间。
数据源类型	采集对象的类型，默认为**SQL Server**。

   ii. 单击**下一步**。

iii. 在**选择采集对象**页签下，选择相应的**数据源**。

如果没有您需要的数据源，请单击**去新建**，进入**工作空间管理 > 数据源管理**页面新建，详情请参见配置SQL Server数据源。

iv. 单击**测试采集连通性**。

v. 待显示**测试成功**，单击**下一步**。

vi. 在**配置执行计划**页签下，配置各项参数。

执行计划包括**按需执行**、**每月**、**每周**、**每天**和**每小时**。

vii. 单击**下一步**。

viii. 在**信息确认**页签下，确认配置信息无误后，单击**确认**。

5. 在**SQL Server元数据采集**页面，单击相应采集器后的**运行**。

5.7.1.5. 采集PostgreSQL元数据

本文为您介绍如何新建采集器，以采集PostgreSQL元数据至DataWorks。采集完成后，您可以在数据地图查看数据。

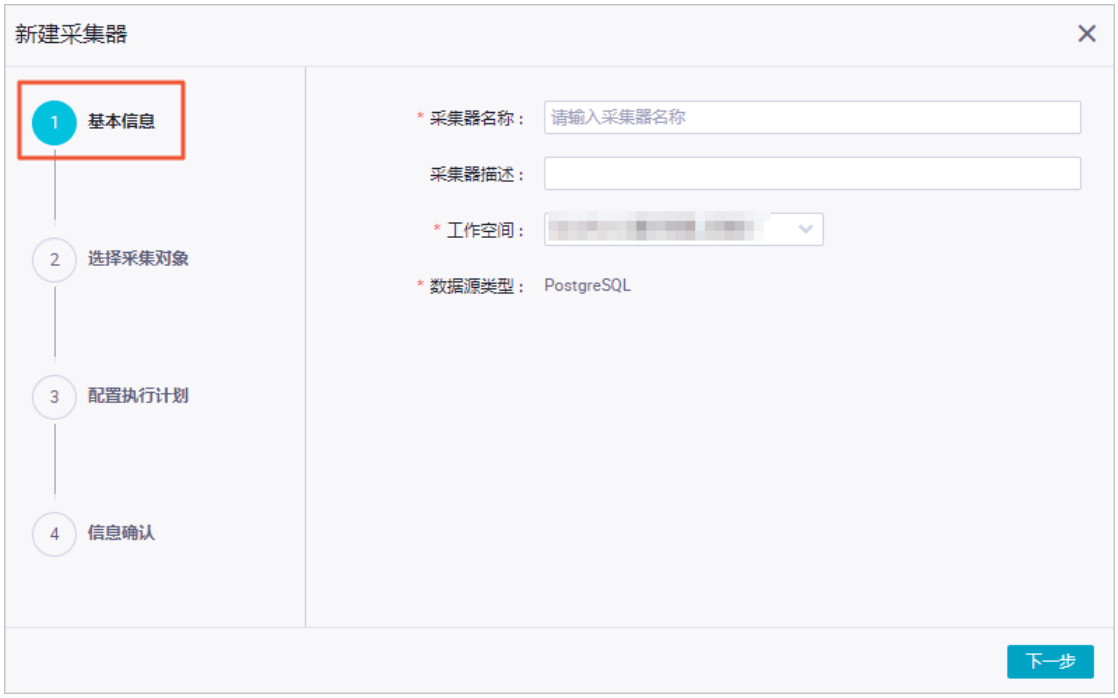
背景信息

全量采集元数据后，系统会开启自动增量采集，自动同步表中新增的元数据。

操作步骤

1. 进入数据发现页面。
2. 在左侧导航栏，单击**元数据采集 > PostgreSQL**。
3. 在**PostgreSQL元数据采集**页面，单击**新建采集器**。
4. 在**新建采集器**配置向导页面，完成以下操作。

i. 在**基本信息**页签下，配置各项参数。



参数	描述
采集器名称	采集器的名称，必填且唯一。
采集器描述	对采集器进行简单描述。
工作空间	采集对象（数据源）所属的DataWorks工作空间。
数据源类型	采集对象的类型，默认为**PostgreSQL**。

ii. 单击**下一步**。

iii. 在**选择采集对象**页签下，选择相应的**数据源**。

如果没有您需要的数据源，请单击**去新建**，进入**工作空间管理 > 数据源管理**页面新建，详情请参见配置PostgreSQL数据源。

iv. 单击**测试采集连通性**。

v. 待显示**测试成功**，单击**下一步**。

vi. 在**配置执行计划**页签下，配置各项参数。

执行计划包括**按需执行**、**每月**、**每周**、**每天**和**每小时**。

vii. 单击**下一步**。

viii. 在**信息确认**页签下，确认配置信息无误后，单击**确认**。

5. 在**PostgreSQL元数据采集**页面，单击相应采集器后的**运行**。

5.7.1.6. 采集Oracle元数据

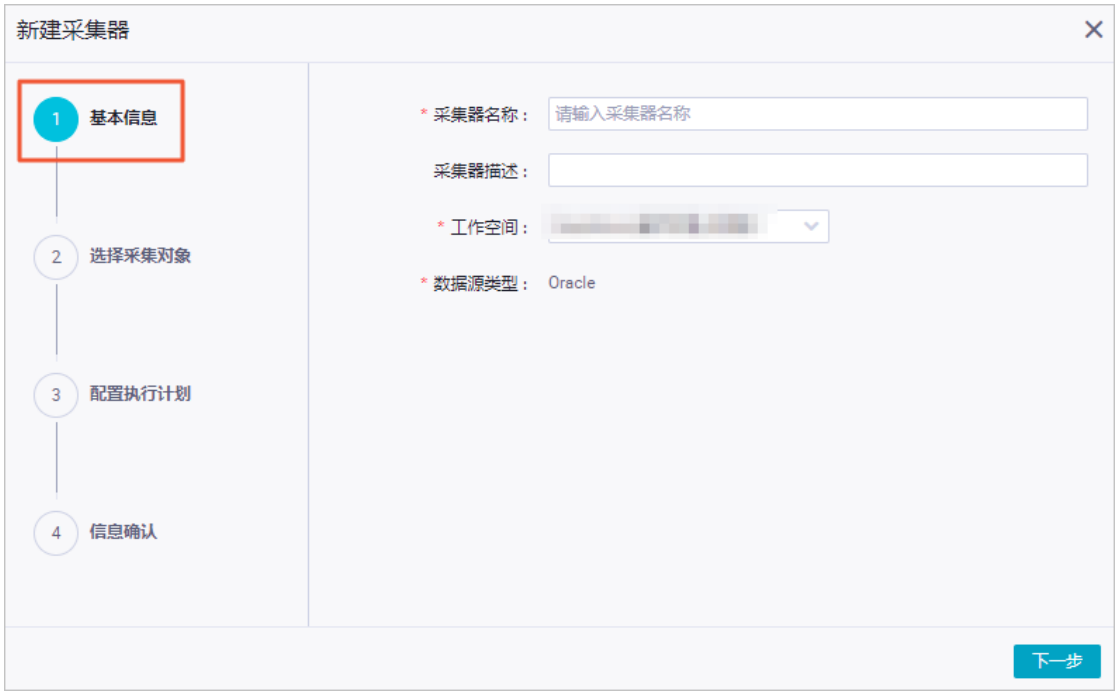
本文为您介绍如何新建采集器，以采集Oracle元数据至DataWorks。采集完成后，您可以在数据地图查看数据。

背景信息

全量采集元数据后，系统会开启自动增量采集，自动同步表中新增的元数据。

操作步骤

1. 登录**DataWorks控制台**后，进入**数据地图**页面，操作详情请参见进入首页。
2. 在顶部菜单栏，单击**数据发现**。
3. 在左侧导航栏，单击**元数据采集 > Oracle**。
4. 在**Oracle元数据采集**页面，单击**新建采集器**。
5. 在**新建采集器**配置向导页面，完成以下操作。
   i. 在**基本信息**页签下，配置各项参数。



参数	描述
采集器名称	采集器的名称，必填且唯一。
采集器描述	对采集器进行简单描述。
工作空间	采集对象（数据源）所属的DataWorks工作空间。
数据源类型	采集对象的类型，默认为**Oracle**。

   ii.
   iii. 在**选择采集对象**页签下，选择相应的**数据源**。

   如果没有您需要的数据源，请单击**去新建**，进入**工作空间管理 > 数据源管理**页面新建，详情请参见配置Oracle数据源。

   iv.
   v.

vi.

vii.

viii.

6. 在**Oracle元数据采集**页面，单击相应采集器后的**运行**。

5.7.1.7. 采集AnalyticDB for PostgreSQL元数据

本文为您介绍如何新建采集器，以采集AnalyticDB for PostgreSQL元数据至DataWorks。采集完成后，您可以通过数据地图管理AnalyticDB for PostgreSQL表。

前提条件

在工作空间绑定AnalyticDB for PostgreSQL引擎后，您才可以进行AnalyticDB for PostgreSQL元数据采集操作，AnalyticDB for PostgreSQL引擎绑定详情请参见：绑定AnalyticDB for PostgreSQL计算引擎。

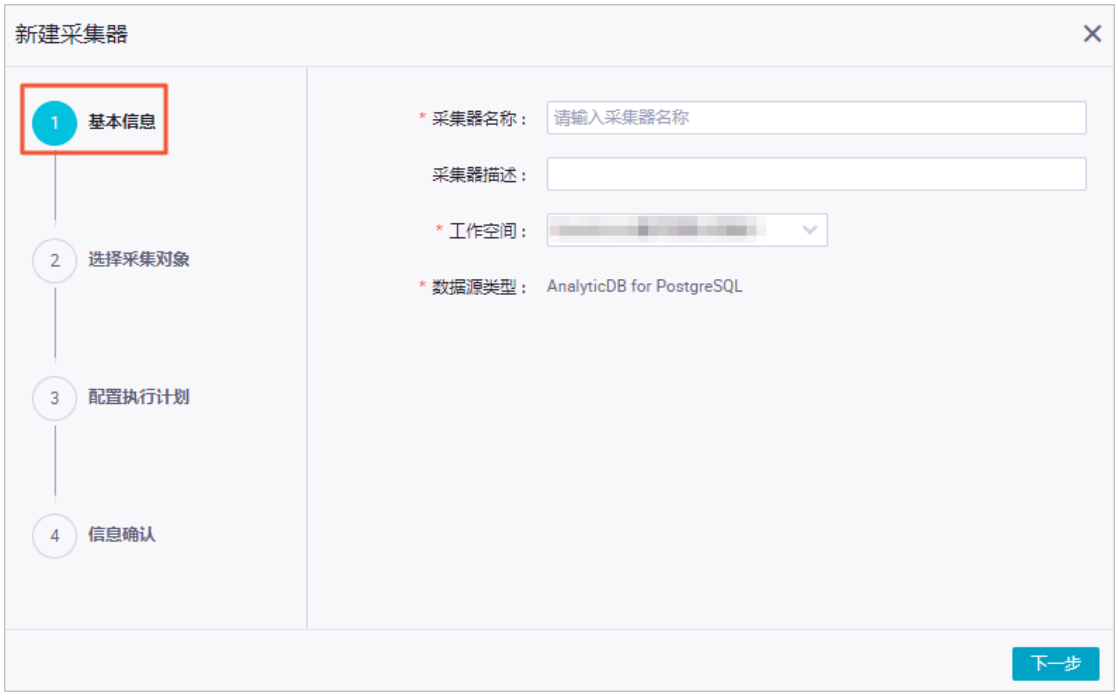
背景信息

全量采集元数据后，系统会开启自动增量采集，自动同步表中新增的元数据。

操作步骤

1. 登录**DataWorks控制台**后，进入**数据地图**页面，操作详情请参见进入首页。
2. 在顶部菜单栏，单击**数据发现**。
3. 在左侧导航栏，单击**元数据采集 > AnalyticDB for PostgreSQL**。
4. 在**AnalyticDB for PostgreSQL元数据采集**页面，单击**新建采集器**。
5. 在**新建采集器**配置向导页面，完成以下操作。

i. 在**基本信息**页签下，配置各项参数。



参数	描述
采集器名称	采集器的名称，必填且唯一。
采集器描述	对采集器进行简单描述。
工作空间	采集对象（数据源）所属的DataWorks工作空间。
数据源类型	采集对象的类型，默认为**AnalyticDB for PostgreSQL**。

ii.

iii. 在**选择采集对象**页签下，选择相应的**数据源**。

如果没有您需要的数据源，请单击**去新建**，进入**工作空间管理 > 数据源管理**页面新建，详情请参见配置AnalyticDB for PostgreSQL数据源。

iv. 单击**测试采集连通性**，待显示**测试成功**，说明已连通DataWorks元数据服务网络。

v. 单击**下一步**。

6. 在**AnalyticDB for PostgreSQL元数据采集**页面，单击相应采集器后的**运行**。

运行完成后，单击**上次运行更新表**或**上次运行添加表**列的数据，查看采集的表。

> 注意 仅手动触发的采集器后会显示**运行**。

您还可以在该页面进行以下操作：

- 单击相应采集器后的**详情**，在**采集器详情**对话框中，查看该采集器的详情。
- 单击相应采集器后的**编辑**，在**编辑采集器**对话框中，修改该采集器的信息。
- 单击相应采集器后的**删除**，在**请确认**对话框中，单击**确认**，删除该采集器。
- 单击处于**运行中**状态的采集器后的**停止**，停止运行该采集器。

5.7.1.8. 采集AnalyticDB for MySQL 2.0元数据

本文为您介绍如何新建采集器，以采集AnalyticDB for MySQL 2.0元数据至DataWorks。采集完成后，您可以在数据地图查看数据。

前提条件

在工作空间绑定AnalyticDB for MySQL引擎后，您才可以进行AnalyticDB for MySQL元数据采集操作，AnalyticDB for MySQL引擎绑定详情请参见：绑定AnalyticDB for MySQL计算引擎。

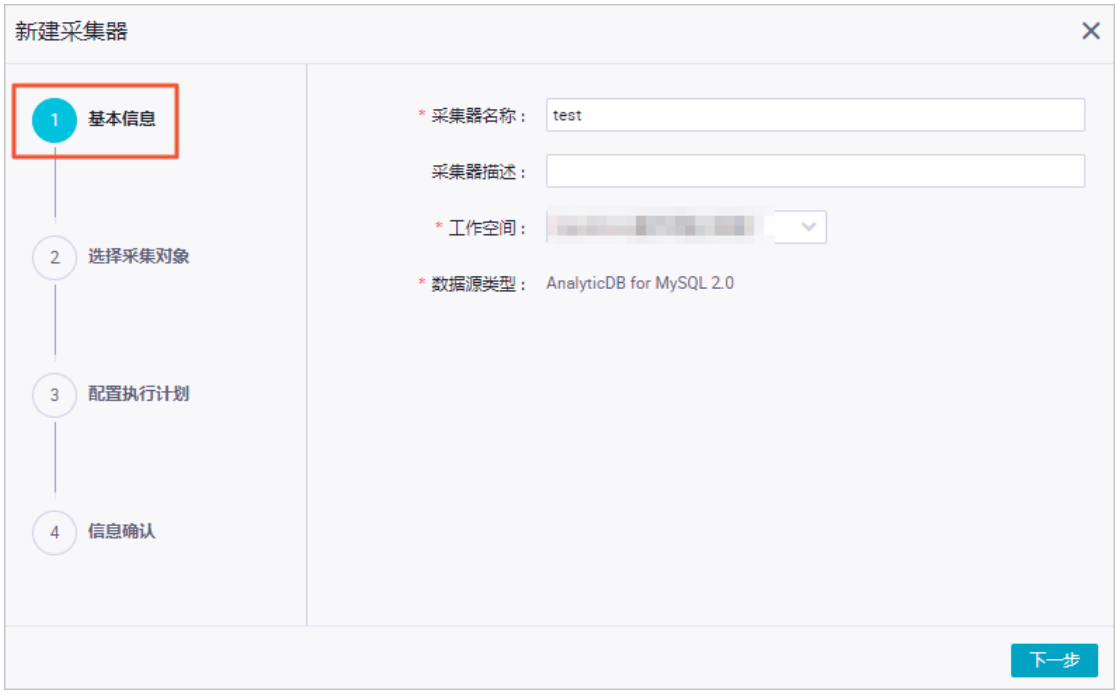
背景信息

全量采集元数据后，系统会开启自动增量采集，自动同步表中新增的元数据。

操作步骤

1. 登录**DataWorks控制台**后，进入**数据地图**页面，操作详情请参见进入首页。
2. 在顶部菜单栏，单击**数据发现**。
3. 在左侧导航栏，单击**元数据采集 > AnalyticDB for MySQL 2.0**。
4. 在**AnalyticDB for MySQL 2.0元数据采集**页面，单击**新建采集器**。
5. 在**新建采集器**配置向导页面，完成以下操作。

i. 在**基本信息**页签下，配置各项参数。



参数	描述
采集器名称	采集器的名称，必填且唯一。
采集器描述	对采集器进行简单描述。
工作空间	采集对象（数据源）所属的DataWorks工作空间。
数据源类型	采集对象的类型，默认为**AnalyticDB for MySQL 2.0**。

ii.

iii. 在**选择采集对象**页签下，选择相应的**数据源**。

如果没有您需要的数据源，请单击**去新建**，进入**工作空间管理 > 数据源管理**页面新建。

iv.

v.

vi.

vii.

viii.

6. 在**AnalyticDB for MySQL 2.0元数据采集**页面，单击相应采集器后的**运行**。

5.7.1.9. 采集AnalyticDB for MySQL 3.0元数据

本文为您介绍如何新建采集器，以采集AnalyticDB for MySQL 3.0元数据至DataWorks。采集完成后，您可以在数据地图查看数据。

前提条件

在工作空间绑定AnalyticDB for MySQL引擎后，您才可以进行AnalyticDB for MySQL元数据采集操作，AnalyticDB for MySQL引擎绑定详情请参见：绑定AnalyticDB for MySQL计算引擎。

背景信息

全量采集元数据后，系统会开启自动增量采集，自动同步表中新增的元数据。

操作步骤

1. 登录**DataWorks控制台**后，进入**数据地图**页面，操作详情请参见进入首页。
2. 在顶部菜单栏，单击**数据发现**。
3. 在左侧导航栏，单击**元数据采集 > AnalyticDB for MySQL 3.0**。
4. 在**AnalyticDB for MySQL 3.0元数据采集**页面，单击**新建采集器**。
5. 在**新建采集器**配置向导页面，完成以下操作。
   i. 在**基本信息**页签下，配置各项参数。

参数	描述
采集器名称	采集器的名称，必填且唯一。
采集器描述	对采集器进行简单描述。
工作空间	采集对象（数据源）所属的DataWorks工作空间。
数据源类型	采集对象的类型，默认为**AnalyticDB for MySQL 3.0**。

   ii.

   iii. 在**选择采集对象**页签下，选择相应的**数据源**。

   如果没有您需要的数据源，请单击**去新建**，进入**工作空间管理 > 数据源管理**页面新建，详情请参见配置AnalyticDB for MySQL 3.0数据源。

iv.

v.

vi.

vii.

viii.

6. 在AnalyticDB for MySQL 3.0元数据采集页面，单击相应采集器后的运行。

5.7.1.10. 采集Hologres元数据

本文为您介绍如何新建采集器，采集Hologres元数据至DataWorks。采集完成后，您可以在数据地图查看数据。

前提条件

在工作空间绑定Hologres引擎后，您才可以进行Hologres元数据采集操作，Hologres引擎绑定详情请参见：绑定Hologres计算引擎。

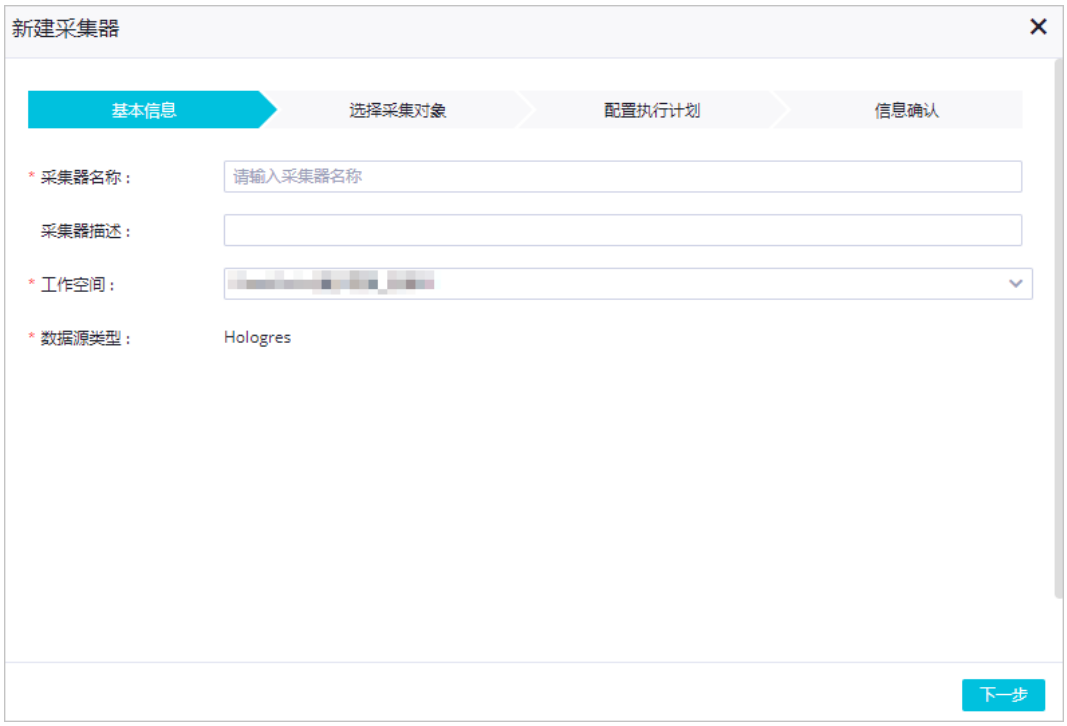
背景信息

全量采集元数据后，系统会开启自动增量采集，自动同步表中新增的元数据。

操作步骤

1. 登录DataWorks控制台后，进入数据地图页面，操作详情请参见进入首页。
2. 在顶部菜单栏，单击数据发现。
3. 在左侧导航栏，单击元数据采集 > Hologres。
4. 在Hologres元数据采集页面，单击新建采集器。
5. 在新建采集器配置向导页面，完成以下操作。
   i. 配置基本信息。

a. 在**基本信息**页签下，配置各项参数。



参数	描述
采集器名称	采集器的名称，必填且唯一。
采集器描述	对采集器进行简单描述。
工作空间	采集对象所属的DataWorks工作空间。
数据源类型	采集对象的类型，默认为**Hologres**。

b. 单击**下一步**。

ii. 选择采集对象。

a. 在**选择采集对象**页签，选择相应的**数据源**。

目前仅支持采集已绑定的Hologres实例的元数据。如果没有您需要的数据源，请单击**去新建**，创建新的数据源。详情请参见配置Hologres数据源。

b. 单击**测试采集连通性**后的**开始测试**，待显示**测试成功**，说明已连通DataWorks元数据服务网络。

> **说明** 如果显示**测试连通性未通过**，则您需要查看具体原因解决相关问题。

c. 单击**下一步**。

iii. 配置执行计划。

在**配置执行计划**页签，配置**执行计划**。

执行计划包括**按需执行**、**每月**、**每周**、**每天**及**每小时**。根据不同的执行周期，生成不同的执行计划，在相应执行计划的时间内，对目标数据源进行元数据采集。具体如下：

- 按需采集：根据实际业务需求，在业务需要时才会采集Hologres元数据。
- 月采集：即在每月的特定几天，在特定时间点自动采集一次Hologres元数据。

> **注意** 部分月份不包含29、30、31日，请您谨慎选择月末日期。

如下图所示，在每月的1、11及21日的09:00，系统会自动采集一次Hologres元数据。**CRON 表达式**会根据您的配置自动生成。

- 周采集：即在每周的特定几天，在特定时间点自动采集一次Hologres元数据。

  如下图所示，在每周的星期一（MON）及星期天（SUN）的03:00，系统会自动采集一次Hologres元数据。**CRON 表达式**会根据您的配置自动生成。

  不输入**时间**时，则默认在每周指定几天的00:00:00采集。

- 天采集：即在每天特定的时间点自动采集一次Hologres元数据。

  如下图所示，在每天的01:00，系统会自动采集一次Hologres元数据。**CRON 表达式**会根据您的配置自动生成。

- 小时采集：即在每小时的第 `N*5分钟` 自动采集一次Hologres元数据。

  > **说明** 目前小时周期的采集任务，仅支持选择的周期时间为第5分钟的倍数。

  如下图所示，在每小时的第5分钟和第10分钟，系统会自动采集一次Hologres元数据。**CRON 表达式**会根据您的配置自动生成。

- 单击**下一步**。

iv. 确认信息。

在**信息确认**页签，确认新建采集器的内容。

6. 确认配置信息无误后，单击**确认**，成功创建采集器。
7. 在**Hologres元数据采集**页面，您可以查看并管理目标采集器的相关信息。

主要操作说明如下：

- 您可以查看相应采集器的运行**状态**、**运行计划**、**上次运行时间**、**上次消耗时间**、**平均运行耗时**及上次运行时更新及添加的表数量。
- 单击目标采集器**操作**列的**详情**、**编辑**、**删除**、**运行**、**停止**，执行相应操作：
  - **详情**：查看该采集器的**采集器名称**、**数据源类型**及**执行计划**。
  - **编辑**：修改该采集器的信息。
  - **删除**：删除该采集器。
  - **运行**：单击**运行**，即可根据该采集所配置的任务采集数据。仅当**执行计划**配置为**按需执行**时，才会生成**运行**操作，其他周期计划的任务不涉及该操作。
  - **停止**：停止运行中的采集器。仅运行中状态的采集器会显示该操作按钮。

执行结果

采集Hologres元数据成功后，您可以在**全部数据**页签，查看已采集的Hologres表。

5.7.1.11. 采集CDH Hive元数据

您可以通过DataWorks的采集元数据功能，将表结构及血缘关系采集到数据地图中，清楚的查看表的内部结构及表间的关联关系。本文为您介绍如何新建CDH Hive采集器，采集CDH Hive元数据至DataWorks。采集完成后，您可以在数据地图查看相关数据。

前提条件

DataWorks工作空间绑定CDH引擎，详情请参见绑定CDH计算引擎。

背景信息

全量采集元数据后，系统会开启自动增量采集，自动同步表中新增的元数据。

使用限制

- DataWorks目前不支持跨地域采集数据，即DataWorks采集器所在的地域需要与元数据所在的地域相同。
- DataWorks目前仅支持使用公网访问元数据。

新建采集器

1. 进入**数据发现**页面。
   i. 登录DataWorks控制台。
   ii. 在左侧导航栏，单击**工作空间列表**。
   iii. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
   iv. 单击左上方的图标，选择**全部产品 > 数据治理 > 数据地图**。
   v. 在顶部菜单栏，单击**数据发现**，进入**数据发现**页面。
2. 新建采集器。
   i. 在左侧导航栏，单击**元数据采集 > CDH Hive**。
   ii. 在**CDH Hive元数据采集**页面，单击**新建采集器**。
3. 配置采集器。
   i. 选择CDH集群。

      在**新建采集器**对话框选择需要采集数据的CDH**集群**。
   ii. 配置执行计划。

      在**新建采集器**对话框选择**执行计划**。

      执行计划包括**按需执行**、**每月**、**每周**、**每天**、**每小时**及**自定义**。您需要根据业务需求配置合适的执行计划，不同的执行计划会生成不同的周期任务，系统会在该执行计划的时间内，对目标数据源进行元数据采集。具体如下：
      - 按需采集：根据实际业务需求，在业务需要时，您需要手动启动采集任务进行数据采集。

- 月采集：即在每月的特定几天，在特定时间点自动采集一次元数据。

  > 注意 部分月份不包含29、30、31日，请您谨慎选择月末日期。

  如下图所示，在每月的1、11及21日的09:00，系统会自动采集一次元数据。**CRON 表达式**会根据您的配置自动生成。

- 周采集：即在每周的特定几天，在特定时间点自动采集一次元数据。

  如下图所示，在每周的星期一（MON）及星期天（SUN）的03:00，系统会自动采集一次元数据。**CRON 表达式**会根据您的配置自动生成。

  不配置**时间**时，则默认在每周指定几天的 `00:00:00` 时间采集数据。

- 天采集：即在每天特定的时间点自动采集一次元数据。

  如下图所示，在每天的01:00，系统会自动采集一次元数据。**CRON 表达式**会根据您的配置自动生成。

- 小时采集：即在每小时的第 `N*5分钟` 自动采集一次元数据。

  > **说明** 目前小时周期的采集任务，仅支持选择的周期时间为第5分钟的倍数。

  如下图所示，在每小时的第5分钟和第10分钟，系统会自动采集一次元数据。**CRON 表达式**会根据您的配置自动生成。

- 自定义采集时间：您可以根据业务需求，配置采集时间的CRON 表达式，系统会根据您的配置采集数据。

iii. 单击**确认**，采集器创建完成。

管理采集器

您可以在**CDH Hive元数据采集**页面，对已创建的采集器进行查看、编辑、删除等管理操作。

序号	描述
1	在该区域，您可以输入采集器名称，搜索目标采集器。 > **说明** 搜索采集器支持模糊匹配，即输入关键词，即可显示名称中包含该关键词的所有采集器。
2	在该区域，您可以查看相应采集器的运行**状态**、**运行计划**、**上次运行时间**、**上次消耗时间**及**平均运行耗时**等信息。 同时，您还可以对目标采集器执行如下操作： • **详情**：查看目标采集器所配置的**集群**及**执行计划**。 • **编辑**：修改目标采集器所配置的**集群**及**执行计划**。 • **删除**：删除目标采集器。 • **运行**：运行采集器，根据目标采集器的配置信息采集数据。 • **停止**：停止运行中的采集器。 > **说明** 仅当**执行计划**配置为**按需执行**时，采集器的**操作**列才会显示**运行**及**停止**操作，其他周期计划的任务不涉及该操作。

后续步骤

采集CDH Hive元数据成功后，您可以在数据地图的**全部数据**页签查看已采集的数据详情。

5.7.2. 数据抽样采集器

5.7.2.1. CDH Hive数据抽样采集器

您可以通过DataWorks的数据抽样采集器功能，从CDH Hive表中随机抽取表的部分数据用于数据保护伞的敏感数据识别。如果您在数据保护伞中配置了脱敏规则，那么在数据地图表详情页面进行数据预览时，命中的敏感字段将会被脱敏。本文为您介绍如何新建CDH Hive数据抽样采集器。

前提条件

- 已购买并创建DataWorks的独享调度资源组。详情请参见：新增和使用独享调度资源组。
- 在工作空间绑定CDH引擎后，您才可以进行CDH数据抽样采集操作，详情请参见绑定CDH计算引擎。
- 已经开通数据保护伞服务，并配置数据识别规则，详情请参见开通数据保护伞、敏感数据识别。

使用限制

- 目前仅上海和成都地域可以使用数据抽样采集器功能。
- 支持基于集群按照数据库进行数据抽样采集。一个集群仅支持新建一个采集器，一个采集器中可以选择一个或多个需要进行数据抽样采集的数据库。
- 选择集群后，如果不选择数据库，默认对所有数据库下的表进行数据抽样。
- 阿里云主账号，拥有AliyunDataWorksFullAccess权限的子账号可以进行采集。
- CDH Hive新增、变更、删除表后需要重新进行数据抽样采集。
- 目前仅支持按需采集。

新建采集器

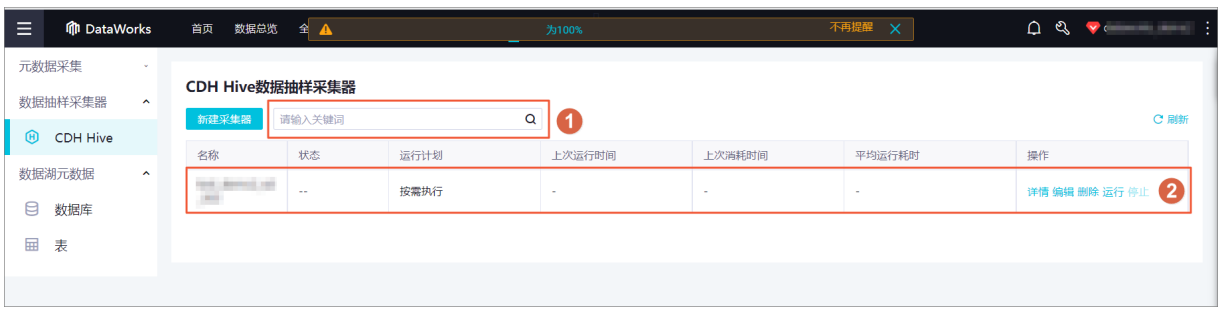
1. 登录**DataWorks控制台**后，进入**数据地图**页面，操作详情请参见进入首页。
2. 在顶部菜单栏，单击**数据发现**。
3. 新建采集器。
   i. 在左侧导航栏，单击**数据抽样采集器 > CDH Hive**。
   ii. 在**CDH Hive数据抽样采集器**页面，单击**新建采集器**，弹出**新建数据抽样采集器**对话框。
4. 配置数据抽样采集器。

参数	描述
集群	下拉列表中展示当前Region下DataWorks已进行引擎绑定的CDH集群。您可以选择需要采集数据的CDH集群。详情请参见：对接使用CDH。
数据库	选择需要进行数据抽样采集的数据库。如果不选择，默认对该集群内所有数据库的表数据进行抽样采集。
独享资源组	选择在绑定CDH引擎时，网络已连通的独享调度资源组。
抽样采集服务	选择需要进行数据抽样采集的服务。详情请参见：对接使用CDH。
采集账号	为您展示用于此次数据抽样采集的账号，该账号将自动根据工作空间引擎绑定页面配置的账号映射关系进行读取。详情请参见：绑定CDH计算引擎
执行计划	定义该采集器多久进行一次数据抽样采集，目前仅支持按需采集。

5. 单击**确认**，采集器创建完成。

管理采集器

您可以在**CDH Hive数据抽样采集器**页面，对已创建的采集器进行查看、编辑、删除等管理操作。



序号	描述
1	在该区域，您可以输入采集器名称，搜索目标采集器。 **说明** 搜索采集器支持模糊匹配，即输入关键词，即可显示名称中包含该关键词的所有采集器。
2	在该区域，您可以查看相应采集器的运行**状态**、**运行计划**、**上次运行时间**、**上次消耗时间**及**平均运行耗时**等信息。 同时，您还可以对目标采集器执行如下操作： • **详情**：查看目标采集器所配置的详细信息。 • **编辑**：修改目标采集器所配置的**集群**、**独享资源组**等信息。 • **删除**：删除目标采集器。 • **运行**：运行采集器，根据目标采集器的配置信息采集数据。运行后，识别出的敏感字段会展示在数据保护伞页面，当您在数据保护伞中配置脱敏规则后，命中的敏感字段在数据地图中预览时将会被脱敏。 • **停止**：停止运行中的采集器。

后续步骤

CDH Hive数据抽样采集成功，如果您已在数据保护伞中配置脱敏规则，那么在数据地图表详情页面进行表数据预览时，命中脱敏规则的敏感字段将会被脱敏。详情请参见：数据保护伞、查看表详情。

5.8. 更多

5.8.1. 工作空间列表

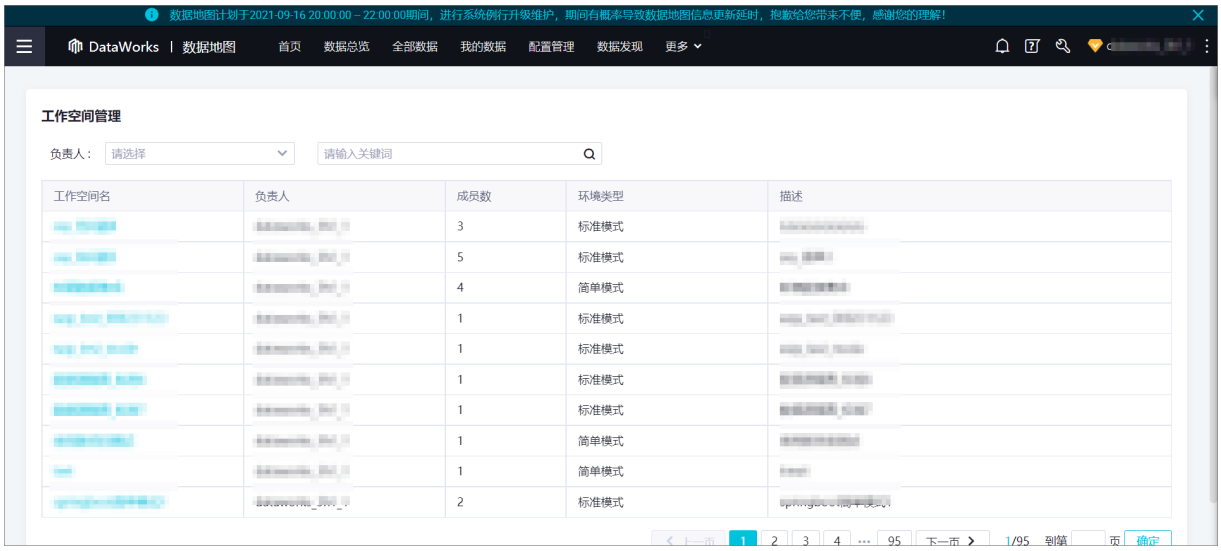
您可以通过工作空间列表查看当前阿里云主账号下的所有工作空间详情列表，单击目标工作空间名称可以查看工作空间的详细信息。

进入工作空间

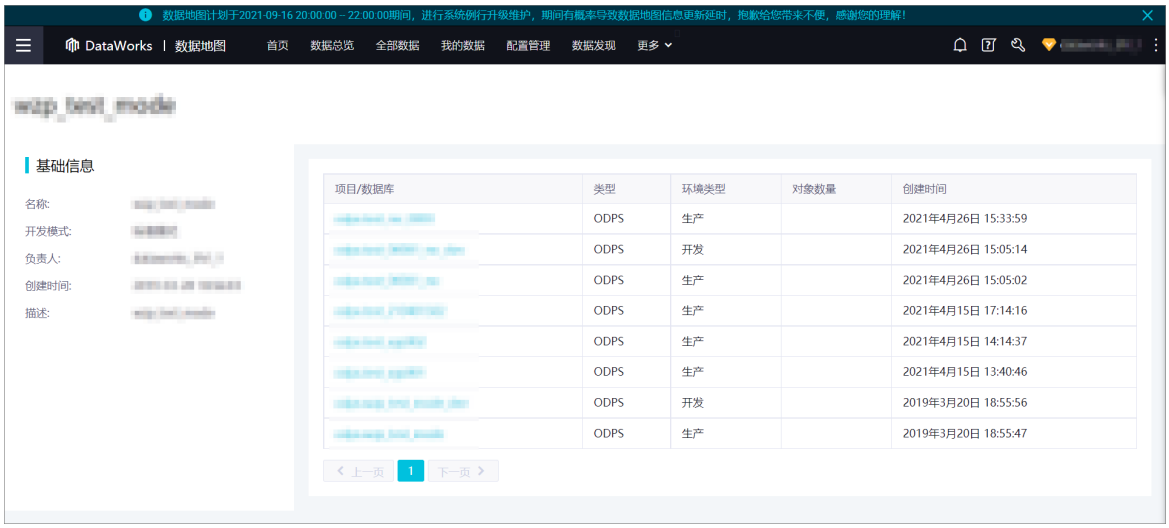
1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据地图**。
4. 在顶部菜单栏，单击**更多 > 工作空间**进入工作空间页面。

工作空间管理

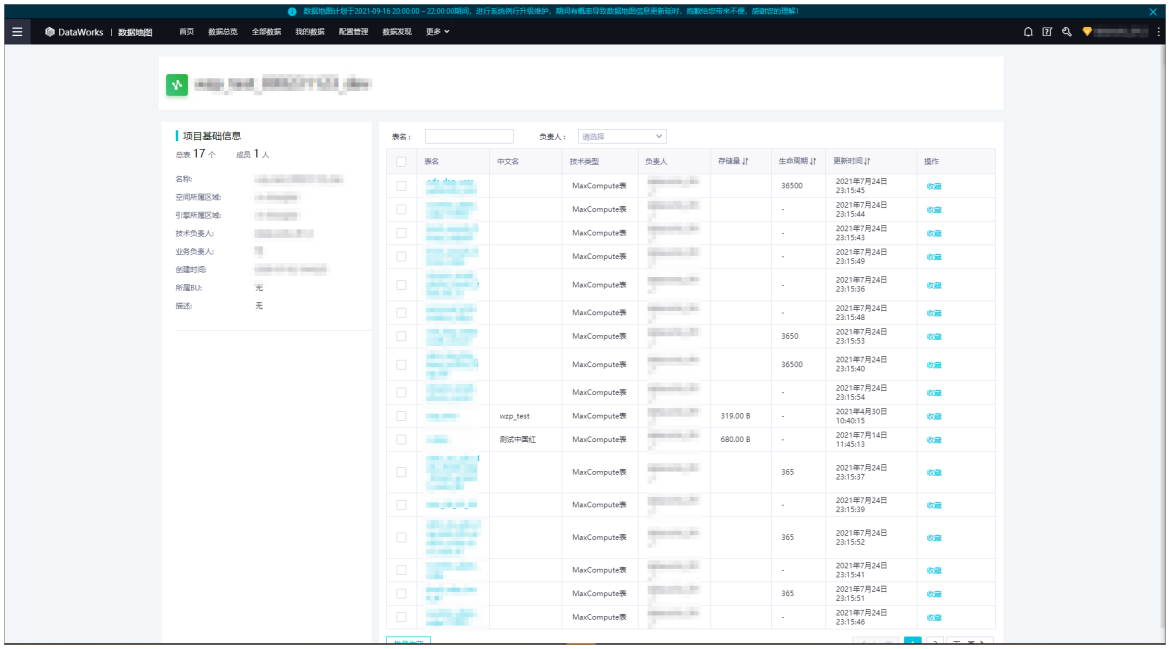
工作空间列表为您展示当前主账号下的所有**工作空间名称**、**成员数**、**环境类型**、**空间描述**。



1. 单击相应的空间名称可以跳转至空间详情页面，为您展示空间的**基础信息**及该空间下的**项目**或**数据库**详情。



2. 单击**项目**或**数据库**名称，跳转至项目或数据库详情页面，为您展示项目或数据库的**基础信息**及该项目或数据库中的表详情，您可以单击表名跳转至**表详情**页面，您也可以在操作列收藏表至**我的收藏**中。



5.9. 其他

5.9.1. 元数据采集的数据源有白名单访问控制时需要配置的白名单

为保证您能正常使用数据地图的元数据采集和类目管理功能，您需要提前配置好白名单，将使用的DataWorks所在地域的IP网段添加至目标项目或数据库中，并为所使用的账号配置类目管理权限。本文为您介绍如何配置白名单及类目管理权限。

背景信息

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。采集完成后，您可以在数据地图查看各数据源的元数据信息。当您需要采集已开启白名单访问控制的元数据时，则需要提前配置好数据库的白名单权限。

使用限制

目前只支持实例形式配置的数据源，或者连接串形式使用公网地址配置的数据源进行元数据采集，暂不支持通过内网地址配置的数据源进行元数据采集，即**连接串模式**配置数据源的时候暂不支持JDBC URL设置为VPC的地址。如果您在元数据采集时，测试连通性失败，请您确认元数据采集的数据源配置。详情请参见配置数据源。

配置元数据采集白名单

1. 查看目标元数据是否开启白名单访问控制。

   DataWorks目前已支持采集的元数据包括：

   - E-MapReduce
   - Hologres
   - CDH Hive
   - AnalyticDB for MySQL 2.0

- AnalyticDB for MySQL 3.0
- AnalyticDB for PostgreSQL
- OTS
- OSS
- PostgreSQL
- MySQL
- SQL Server
- Oracle

不同元数据查看白名单访问控制的方式不同，您可以提交工单咨询技术支持同学。

如果目标数据库未开启白名单访问控制，则DataWorks可以正常使用数据地图采集数据库的元数据；如果数据库开启了白名单访问控制，则需要进行下一步的添加白名单操作。

2. 数据库添加白名单。

   如果您需要进行元数据采集的数据库已开启白名单访问控制，请在数据库白名单列表中，添加需要使用的DataWorks所在地域的IP网段。DataWorks的地域与白名单IP网段对应关系如下表所示。不同数据库的添加位置可能不同，您可以提交工单咨询技术支持同学。

地域	白名单
华东2（上海）	100.104.189.64/26,11.115.110.10/24,11.115.109.9/24,47.102.181.128/26,47.102.181.192/26,47.102.234.0/26,47.102.234.64/26,100.104.38.192/26
华东1（杭州）	100.104.135.128/26,11.193.215.233/24,11.194.73.32/24,118.31.243.0/26,118.31.243.64/26,118.31.243.128/26,118.31.243.192/26,100.104.242.0/26
华南1（深圳）	100.104.46.128/26,11.192.91.119/24,120.77.195.128/26,120.77.195.192/26,120.77.195.64/26,47.112.86.0/26,100.104.138.128/26
华北2（北京）	100.104.37.128/26,11.193.82.20/24,11.197.254.171/24,39.107.223.0/26,39.107.223.64/26,39.107.223.128/26,39.107.223.192/26,100.104.152.128/26
西南1（成都）	100.104.88.64/26,11.195.57.28/24,47.108.46.0/26,47.108.46.64/26,47.108.46.128/26,47.108.46.192/26,100.104.248.128/26
华北3（张家口）	100.104.197.0/26,11.193.236.121/24,47.92.185.0/26,47.92.185.64/26,47.92.185.128/26,47.92.185.192/26,100.104.75.64/26
英国（伦敦）	8.208.84.22, 100.104.161.0/26
亚太东南1（新加坡）	11.193.8.90, 11.193.8.93

云产品白名单配置注意事项

以阿里云云数据库RDS为例，您需要在元数据采集时将相应的IP地址段添加到数据库白名单列表中，在配置白名单前您可以先了解以下问题。

目前云产品支持通用模式IP白名单和高安全模式IP白名单配置，您添加白名单时配置白名单分组可能会影响元数据采集时的网络连通：

- 如果您目前数据库设置的为通用模式IP白名单：
  - 通用模式IP白名单不区分经典网络和专有网络白名单分组。

- 公共资源组、独享调度资源组使用同样的白名单分组配置。

  > **说明** 在通用白名单模式下，设置的IP地址，既可通过经典网络，也可通过专有网络访问RDS实例。

- 如果您目前数据库设置的为高安全模式IP白名单模式：
  - 高安全模式区分经典网络和专有网络白名单分组。

    > **说明** 在高安全白名单模式下，白名单分组需指定网络隔离模式，例如设置在经典网络的白名单IP地址，不可从专有网络访问RDS实例，反之亦然。

  - 使用独享调度资源组VPC内网直接连接数据库，使用专有网络白名单分组。
  - 使用公共资源组访问VPC网络数据源（例如，实例模式配置的专有网络 RDS MySQL），使用专有网络白名单分组。
  - 使用公网连接地址、经典网络地址直接访问数据库，走的是经典网络白名单分组。
- 如果您在数据库将白名单模式从通用模式IP白名单模式切换为高安全模式IP白名单模式：

  RDS会将通用模式IP白名单复制分为2份，分别放到经典网络和专有网络白名单分组类型里面。

其他白名单配置注意事项：

- 设置白名单不会影响RDS实例的正常运行。
- 默认的IP白名单分组（default ）不能删除，只能清空。
- 请勿修改或删除系统自动生成的分组，避免影响相关产品的使用。例如**ali_dms_group**（DMS产品IP地址白名单分组）、**hdm_security_ips**（DAS产品IP地址白名单分组）。

  > **说明** 建议您在数据库配置白名单时，单独为DataWorks新建一个白名单分组。

- 默认的IP白名单只包含127.0.0.1，表示任何IP均无法访问该RDS实例。

RDS白名单配置详情可参见通过客户端、命令行连接RDS MySQL实例。其他类型的数据源类似，可参考各数据源数据库的白名单配置步骤，分别添加对应的白名单。

后续步骤

配置完成白名单及类目管理权限后，您可以进行元数据采集或类目管理。详情请参见采集元数据或配置管理类目。

5.9.2. MaxCompute开启白名单访问控制时需要配置的白名单列表

当使用数据地图查看MaxCompute表数据时，如果MaxCompute项目空间开启了白名单访问控制，则您可能无法访问该项目空间的相关内容。为保证DataWorks能顺利访问MaxCompute的项目空间，则需要提前配置好MaxCompute的白名单权限。

操作步骤

1. 查看目标MaxCompute项目空间是否开启了白名单访问控制，详情请参见查看白名单访问控制。

   如果MaxCompute项目空间未开启白名单访问控制，则DataWorks可以正常使用数据地图访问MaxCompute的数据表；如果项目空间开启了白名单访问控制，则需要进行下一步的添加白名单操作。

2. MaxCompute项目添加白名单。

   如果您的MaxCompute项目空间开启了白名单访问控制，请在MaxCompute的白名单列表中，添加需要使用的DataWorks所在地域的IP网段。DataWorks的地域与白名单IP网段对应关系如下表所示。MaxCompute添加白名单的详细操作请参见管理IP白名单。

地域	白名单
华东1（杭州）	100.64.0.0/10,11.193.102.0/24,11.193.215.0/24,11.194.110.0/24,11.194.73.0/24,118.31.157.0/24,47.97.53.0/24,11.196.23.0/24,47.99.12.0/24,47.99.13.0/24,114.55.197.0/24,11.197.246.0/24,11.197.247.0/24
华东2（上海）	11.193.109.0/24,11.193.252.0/24,47.101.107.0/24,47.100.129.0/24,106.15.14.0/24,10.117.28.203,10.143.32.0/24,10.152.69.0/24,10.153.136.0/24,10.27.63.15,10.27.63.38,10.27.63.41,10.27.63.60,10.46.64.81,10.46.67.156,11.192.97.0/24,11.192.98.0/24,11.193.102.0/24,11.218.89.0/24,11.218.96.0/24,11.219.217.0/24,11.219.218.0/24,11.219.219.0/24,11.219.233.0/24,11.219.234.0/24,118.178.142.154,118.178.56.228,118.178.59.233,118.178.84.74,120.27.160.26,120.27.160.81,121.43.110.160,121.43.112.137,100.64.0.0/10,10.117.39.238
华南1（深圳）	100.106.46.0/24,100.106.49.0/24,10.152.27.0/24,10.152.28.0/24,11.192.91.0/24,11.192.96.0/24,11.193.103.0/24,100.64.0.0/10,120.76.104.0/24,120.76.91.0/24,120.78.45.0/24,47.106.63.0/26,47.106.63.128/26,47.106.63.192/26,47.106.63.64/26
西南1（成都）	11.195.52.0/24,11.195.55.0/24,47.108.22.0/24,100.64.0.0/10
华北3（张家口）	11.193.235.0/24,47.92.22.0/24,100.64.0.0/10
中国（香港）	10.152.162.0/24,11.192.196.0/24,11.193.11.0/24,100.64.0.0/10,47.89.61.0/24,47.91.171.0/24,11.193.118.0/24,47.75.228.0/24,47.56.45.0/25,47.244.92.128/25,47.101.109.0/24
亚太东南1（新加坡）	100.106.10.0/24,100.106.35.0/24,10.151.234.0/24,10.151.238.0/24,10.152.248.0/24,11.192.153.0/24,11.192.40.0/24,11.193.8.0/24,100.64.0.0/10,47.88.147.0/24,47.88.235.0/24,11.193.162.0/24,11.193.163.0/24,11.193.220.0/24,11.193.158.0/24,47.74.162.0/24,47.74.203.0/24,47.74.161.0/24,11.197.188.0/24
亚太东南2（悉尼）	11.192.100.0/24,11.192.134.0/24,11.192.135.0/24,11.192.184.0/24,11.192.99.0/24,100.64.0.0/10,47.91.49.0/24,47.91.50.0/24,11.193.165.0/24,47.91.60.0/24
华北2（北京）	100.106.48.0/24,10.152.167.0/24,10.152.168.0/24,11.193.50.0/24,11.193.75.0/24,11.193.82.0/24,11.193.99.0/24,100.64.0.0/10,47.93.110.0/24,47.94.185.0/24,47.95.63.0/24,11.197.231.0/24,11.195.172.0/24,47.94.49.0/24,182.92.144.0/24,39.99.77.0/26,39.99.77.64/26,39.99.77.128/26,39.104.220.192/26,39.107.7.0/26,39.107.7.64/26,182.92.32.128/26,182.92.32.192/26
美国西部1（硅谷）	10.152.160.0/24,100.64.0.0/10,47.89.224.0/24,11.193.216.0/24,47.88.108.0/24
美国东部1（弗吉尼亚）	47.88.98.0/26,47.88.98.64/26,47.88.98.128/26,47.88.98.192/26,47.252.91.0/26,47.252.91.64/26,47.252.91.128/26,47.252.91.192/26,10.128.134.0/24,11.193.203.0/24,11.194.68.0/24,11.194.69.0/24,100.64.0.0/10

地域	白名单
亚太东南3（吉隆坡）	11.193.188.0/24,11.221.205.0/24,11.221.206.0/24,11.221.207.0/24,100.64.0.0/10,11.214.81.0/24,47.254.212.0/24,11.193.189.0/24
欧洲中部1（法兰克福）	11.192.116.0/24,11.192.168.0/24,11.192.169.0/24,11.192.170.0/24,11.193.106.0/24,100.64.0.0/10,11.192.116.14,11.192.116.142,11.192.116.160,11.192.116.75,11.192.170.27,47.91.82.22,47.91.83.74,47.91.83.93,47.91.84.11,47.91.84.110,47.91.84.82,11.193.167.0/24,47.254.138.0/24
亚太东北1（日本）	100.105.55.0/24,11.192.147.0/24,11.192.148.0/24,11.192.149.0/24,100.64.0.0/10,47.91.12.0/24,47.91.13.0/24,47.91.9.0/24,11.199.250.0/24,47.91.27.0/24,11.59.59.0/24,47.245.51.128/26,47.245.51.192/26,47.91.0.128/26,47.91.0.192/26
中东东部1（迪拜）	11.192.107.0/24,11.192.127.0/24,11.192.88.0/24,11.193.246.0/24,47.91.116.0/24,100.64.0.0/10
亚太南部1（孟买）	11.194.10.0/24,11.246.70.0/24,11.246.71.0/24,11.246.73.0/24,11.246.74.0/24,100.64.0.0/10,149.129.164.0/24,11.194.11.0/24,11.59.62.0/24,147.139.23.0/26,147.139.23.128/26,147.139.23.64/26,149.129.165.192/26
英国（伦敦）	11.199.93.0/24,100.64.0.0/10
亚太东南5（雅加达）	11.194.49.0/24,11.200.93.0/24,11.200.95.0/24,11.200.97.0/24,100.64.0.0/10,149.129.228.0/24,10.143.32.0/24,11.194.50.0/24,11.59.135.0/24,147.139.156.0/26,147.139.156.128/26,147.139.156.64/26,149.129.230.192/26
华北2（政务云）	11.194.116.0/24,100.64.0.0/10,39.107.188.202如果IP地址段添加不成功，请添加下述IP地址： 11.194.116.160,11.194.116.161,11.194.116.162,11.194.116.163,11.194.116.164,11.194.116.165,11.194.116.167,11.194.116.169,11.194.116.170,11.194.116.171,11.194.116.172,11.194.116.173,11.194.116.174,11.194.116.175,39.107.188.0/24
华东2（上海）金融云	140.205.46.128/25,140.205.48.0/25,140.205.48.128/25,140.205.49.0/25,140.205.49.128/25,11.192.156.0/25,11.192.157.0/25,11.192.164.0/25,11.192.165.0/25,11.192.166.0/25,11.192.167.0/25,106.11.245.0/26,106.11.245.128/26,106.11.245.192/26,106.11.245.64/26,140.205.39.0/24,106.11.225.0/24,106.11.226.0/24,106.11.227.0/24,106.11.242.0/24,100.104.8.0/24

6.通过操作审计查询行为事件日志

DataWorks已集成至操作审计（ActionTrail）中，您可以在ActionTrail中查看及检索阿里云账号最近90天的DataWorks行为事件日志。后续可以通过ActionTrail将事件日志投递至日志服务LogStore或指定的OSS Bucket中，实现对事件的监控和告警，满足及时审计、问题回溯分析等需求。本文为您介绍如何在ActionTrail中查询DataWorks的行为事件日志。

背景信息

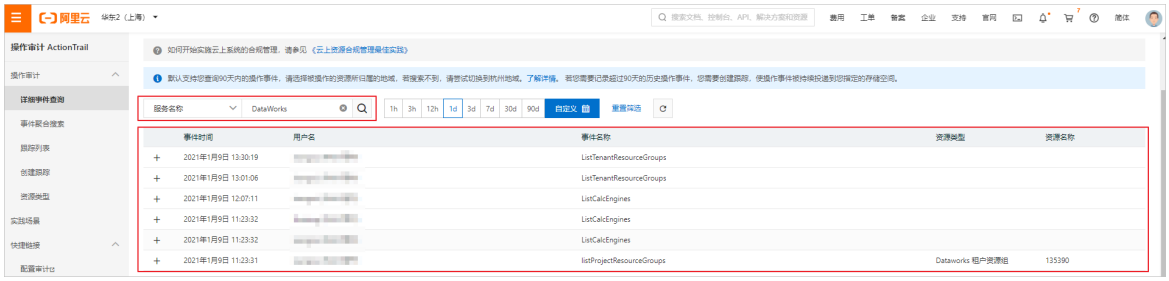
操作审计（ActionTrail）是阿里云提供的云账号资源操作记录的查询和投递服务，帮助您监控并记录包括通过阿里云控制台、OpenAPI、开发者工具对云上产品和服务的访问和使用行为。您可以将这些行为事件下载或保存到日志服务LogStore或OSS存储空间，然后进行行为分析、安全分析、资源变更行为追踪和行为合规性审计等操作。详情请参见操作审计。

注意事项

- 从发生DataWorks操作行为到生成行为事件日志，存在大约5~10分钟的延时，请您耐心等待。
- 您可以为重要事件配置跟踪告警，以便及时发现并处理异常行为。

查询DataWorks行为事件日志

1. 登录ActionTrail管理控制台。
2. 在左侧导航栏单击**详细事件查询**，并选择相应地域。
3. 在**详细事件查询**页面的下拉列表，选择**服务名称**为**DataWorks**，查看已进行操作审计的DataWorks事件列表。



该列表为您展示了事件的基本信息，包括**事件时间**、**用户名**、**事件名称**、**资源类型**以及**资源名称**。

您可以使用**事件名称**通过区分目标事件是否为OpenAPI类行为调用事件，来查询事件含义：

> **说明** OpenAPI类行为调用事件包括涉及调用OpenAPI的页面操作事件以及通过程序代码调用OpenAPI的事件。

- 目标事件为OpenAPI类行为调用事件。

  该类事件的**事件名称**与OpenAPI的名称一一对应。您可以使用**事件名称**在OpenAPI列表进行搜索，查询事件含义。

- 目标事件为非OpenAPI类行为调用事件。

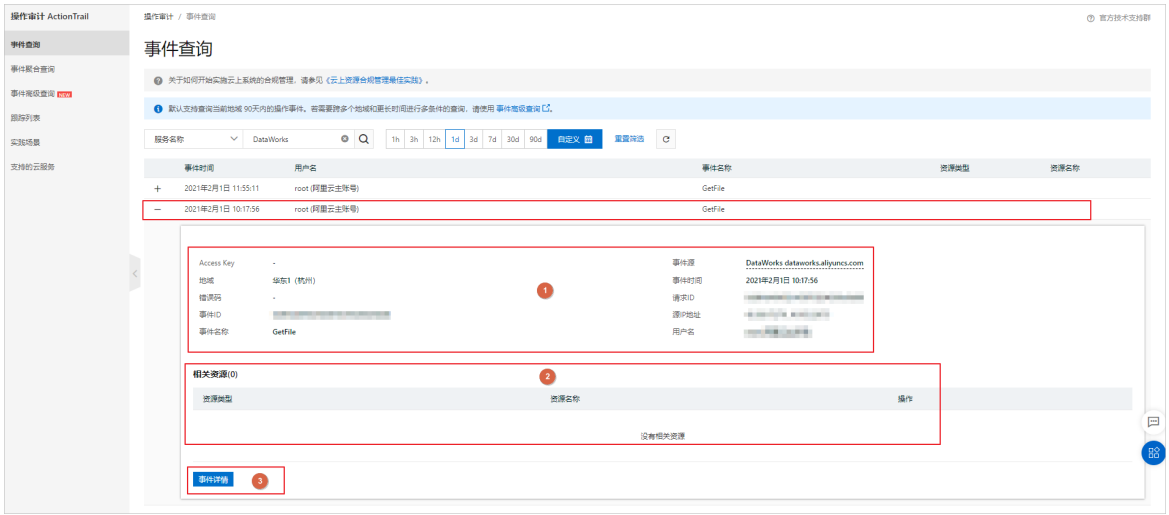
  您可以通过如下表格查询事件含义。

事件名称	事件含义	服务模块
DownloadExecutionResult	下载查询结果。	

事件名称	事件含义	服务模块
CreateBusiness	创建业务流程。	数据开发
DestroyRelationTableFromBusiness	删除业务流程中的所有表。	
DeleteBusiness	删除业务流程。	
ExecuteFile	将文件作为临时任务执行。	
LockFile	偷锁编辑。	
RecoverFile	恢复回收站中的文件。	
CloneFile	克隆文件。	
DeleteFolder	删除文件夹。	
DeleteDeployment	删除发布包。	
ListCodingProjects	当前用户可以查看的代码工程列表。	AppStudio

> **说明** 如果您通过上述方式均未查询到目标**事件名称**，请提交工单咨询目标事件的相关信息。

4. 单击目标事件查看**事件详情**。

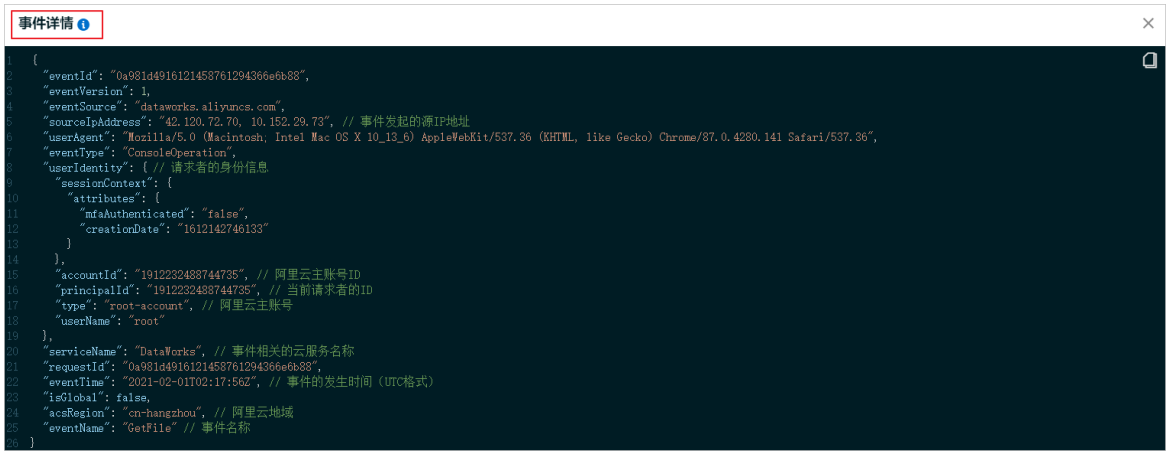


事件详情描述如下表所示。

序号	描述
①	显示目标事件的详细信息。 鼠标悬停至用户名详情，单击**详情**，即可跳转至**RAM访问控制**页面，查看对应用户的详细信息。

序号	描述
②	显示目标事件所使用资源的**资源类型**、**资源名称**及相关**操作**。
③	单击**事件详情**即可查询目标事件的代码记录。

示例查询的listProjectResourceGroups事件代码记录如下。



您可以在**事件详情**对话框，单击右上角的图标，即可复制事件的代码记录。

后续步骤

您可以通过查询到的事件日志进行行为分析、安全分析、资源变更行为追踪和行为合规性审计等操作。