阿里云
DataWorks

数据治理

文档版本：20190911

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
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	**禁止:** 重置操作将丢失用户配置数据。
	该类警示信息可能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，恢复业务所需时间约10分钟。
	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。	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。	单击 确定。
`courier` 字体	命令。	执行 `cd /d C:/windows` 命令，进入Windows系统文件夹。
*## *	表示参数、变量。	`bae log list --instanceid` *`Instance_ID`*
[]或者[a|b]	表示可选项，至多选择一个。	`ipconfig` *`[-all|-t]`*
{}或者{a|b}	表示必选项，至多选择一个。	`swich` *`{stand | slave}`*

目录

法律声明.....I
通用约定.....I
1 安全中心..... 1
1.1 安全中心概述.....1
1.2 快速入门.....2
1.3 我的权限.....11
1.4 权限审计.....15
1.5 审批中心.....18
1.6 常见问题.....21
2 数据质量.....24
2.1 数据质量概述.....24
2.2 功能介绍.....25
2.2.1 首页概览.....25
2.2.2 我的订阅.....25
2.2.3 规则配置.....26
2.2.4 任务查询.....36
2.3 使用指南.....38
2.3.1 DataHub监控..... 38
2.3.2 MaxCompute监控..... 45
3 数据保护伞.....54
3.1 进入数据保护伞.....54
3.2 数据发现.....56
3.3 数据访问.....57
3.4 数据风险.....60
3.5 数据审计.....60
3.6 规则配置.....61
3.7 分级信息管理.....66
3.8 手动修正数据.....68
3.9 风险识别管理.....69
3.10 设置并查询自定义脱敏..... 70
4 数据地图.....79
4.1 数据管理升级为数据地图..... 79
4.2 数据地图概述.....81
4.3 数据总览.....83
4.4 我的数据.....84
4.5 表详情页介绍.....87
4.6 权限管理.....97
4.7 申请数据权限.....97
4.8 配置管理..... 102

5 数据管理.....106
5.1 数据管理概述.....106
5.2 全局概览.....107
5.3 表详情页介绍.....108
5.4 权限管理.....118
5.5 数据权限申请.....118
5.6 管理配置.....122
5.7 查找数据.....124
5.8 数据表管理.....125
5.9 创建表.....129
6 资源优化.....137
6.1 资源优化概述.....137
6.2 个人资产优化.....137
6.3 工作空间资产优化.....141

1 安全中心

1.1 安全中心概述

安全中心为您提供便捷的权限管控能力，提供可视化申请审批流程，并可以进行权限的审计和管理。提高数据安全的同时，还可以方便您进行数据权限管控。

> 说明:
> 安全中心目前仅华东2区域支持，已开放内测邀请。

1. 登录DataWorks控制台，单击相应工作空间后的进入数据开发。
2. 单击左上角的DataWorks图标，选择全部产品 > 数据地图（数据管理），进入数据地图页面。
3. 单击左上角的DataWorks图标，选择全部产品 > 安全中心，即可进入安全中心页面。

安全中心模块包括我的权限、权限审计和审批中心三大模块。目前安全中心具备的功能如下所示：

- 权限自助申请：您可以选择自己需要权限的数据表，在线上快速发起申请，改变原有线下联系管理员的模式，提高工作效率。
- 权限审计/交还：管理员可以快速方便地查看数据库表权限对应人员，进行审计管理，用户也可以主动交还不再需要的权限。
- 权限审批管理：将以前管理员直接授权的模式改为审批授权模式，提供可视化、流程化的管理授权机制，并可以对审批流程进行事后追溯。

您可在安全中心模块进行组织内全局数据表权限的查看、表权限管理、数据表权限申请/审批等操作。

安全中心的各项操作支持同一租户下的所有工作空间，包括标准模式（开发环境及生产环境）及简单模式。

1.2 快速入门

本文将通过一个简单示例，为您介绍安全中心各功能的使用。

前提条件

使用安全中心前请注意以下事项。

- 关于字段级别授权与LabelSecurity

  未开启LabelSecurity机制的工作空间将无法进行字段级别的授权，只能进行整表授权，同时无法在申请时设置有效期。LabelSecurity的详情请参见#unique_6。

- 关于权限有效期

  为保证字段级别的授权在有效期正常运行，除开启LabelSecurity外，还请务必保证各字段的LabelSecurity大于账号的LabelSecurity。

  如果部分字段的LabelSecurity为空或不大于账号的LabelSecurity，账号申请该表权限时虽然可以选择有效期，但申请审批通过后将自动获取这部分字段的权限，同时有效期为永久，并且无法单独交还/回收这部分字段权限。

- 关于申请及展示权限

  目前安全中心仅展示通过ACL授权方式所获取的权限，不展示通过用户角色等其他方式获取的权限。例如作为项目开发角色的账号可以访问项目中所有的表，但在安全中心中，并不显示对项目中的表具有权限。当发现安全中心未展示具有权限但实际可以访问的情况，请与系统管理员确认是否通过用户角色等其他方式获取到了相应权限。

用户角色对应功能介绍

- 子账号-普通用户
  - 我的权限：支持查看权限、申请权限、交还操作权限和交还字段权限等操作。
  - 审批中心 > 我的申请：支持查看申请单进度和查看历史申请等操作。
- 子账号-表Owner
  - 我的权限：支持查看权限、申请自己不是表Owner的表权限、交还自己不是表Owner的表操作权限、交还自己不是表Owner的表字段权限等操作。
  - 审批中心 > 我的申请：支持查看申请单进度和查看历史申请等操作。
  - 审批中心 > 待我审批：支持审批等待自己处理的申请单。
  - 审批中心 > 我已审批：支持查看自己处理过的申请单。
- 子账号-项目管理员
  - 权限审计：支持查看项目成员和回收成员权限等操作。
  - 审批中心 > 我的申请：支持查看申请单进度和查看历史申请等操作。
  - 审批中心 > 待我审批：支持审批等待自己处理的申请单。
  - 审批中心 > 我已审批：支持查看自己处理过的申请单。
- 主账号
  - 权限审计：支持查看项目成员和回收成员权限等操作。
  - 审批中心 > 我的申请：支持查看申请单进度和查看历史申请等操作。
  - 审批中心 > 待我审批：支持审批等待自己处理的申请单。
  - 审批中心 > 我已审批：支持查看自己处理过的申请单。

示例介绍

本示例的角色分类有：普通用户、表Owner和项目管理员。

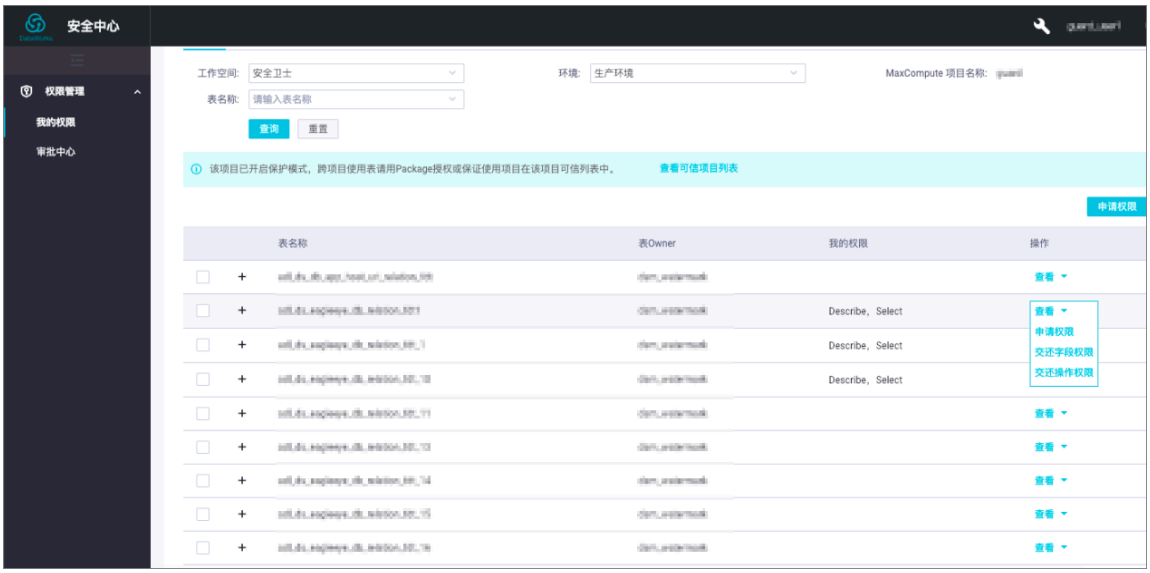
本示例需要实现以下场景。

- 子账号甲作为普通用户查看自己当前具有的权限，然后申请A、B这两张目前无权限的数据表的权限。
- 子账号乙作为A表的表Owner进行审批，通过A表申请。
- 主账号作为项目管理员进行审批，通过B表申请。
- 子账号甲无需使用A表部分字段，交还部分字段权限。
- 子账号甲不再需要使用A表，交还A表权限。
- 主账号回收子账号甲B表的权限。

普通用户

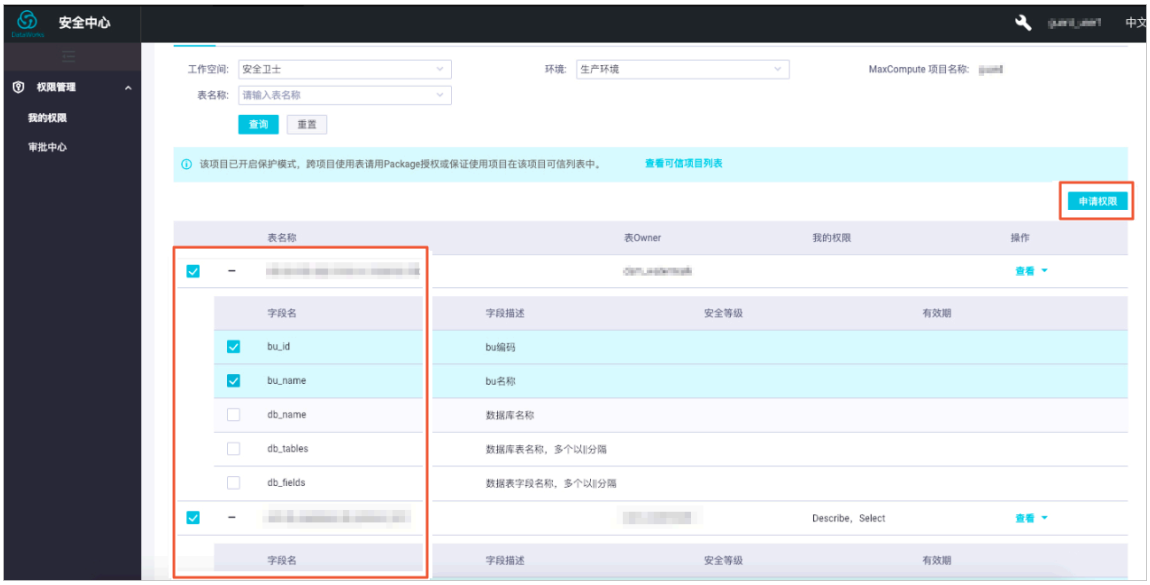
- 子账号甲查看自己在项目中的权限

  1. 子账号登录控制台，进入安全中心 > 我的权限 > 表权限页面，对本组织范围内（多工作空间）的数据表进行查看和搜索。
  2. 在表权限页面，通过选择工作空间 + 环境（标准模式下）展示该工作空间对应环境下所有表， 查看自己对哪些表具有权限。

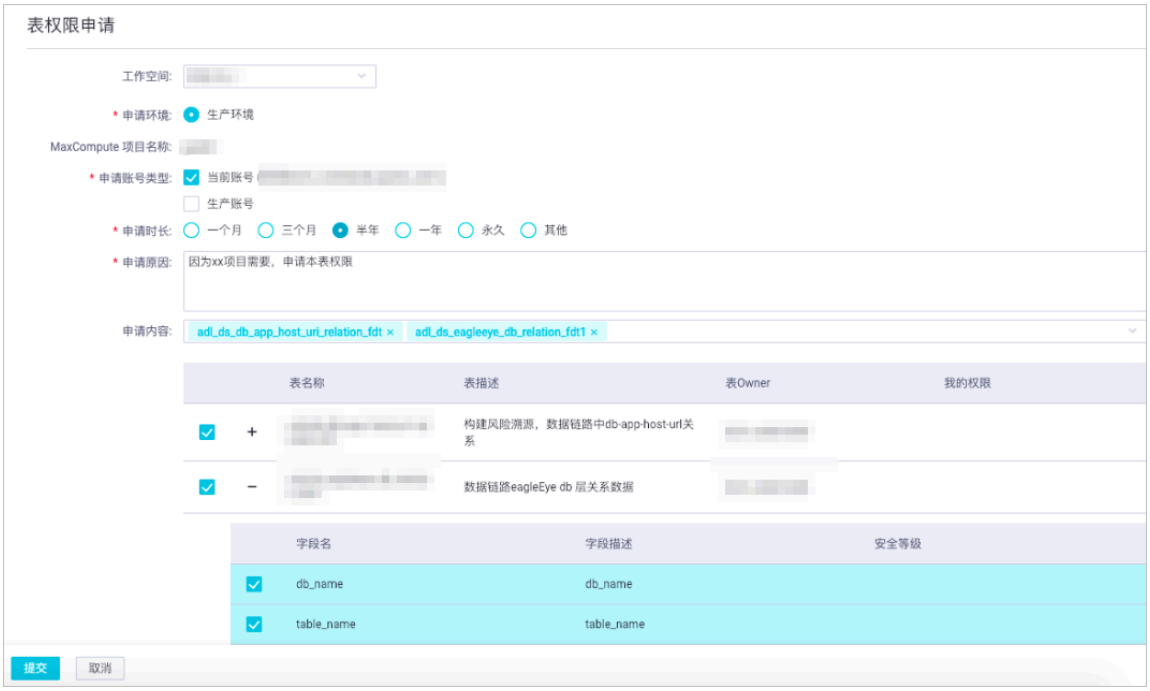


- 申请A、B两张表的权限

1. 子账号登录控制台，进入安全中心 > 我的权限 > 表权限页面。
2. 勾选A、B两张表需要申请权限的字段，单击申请权限。



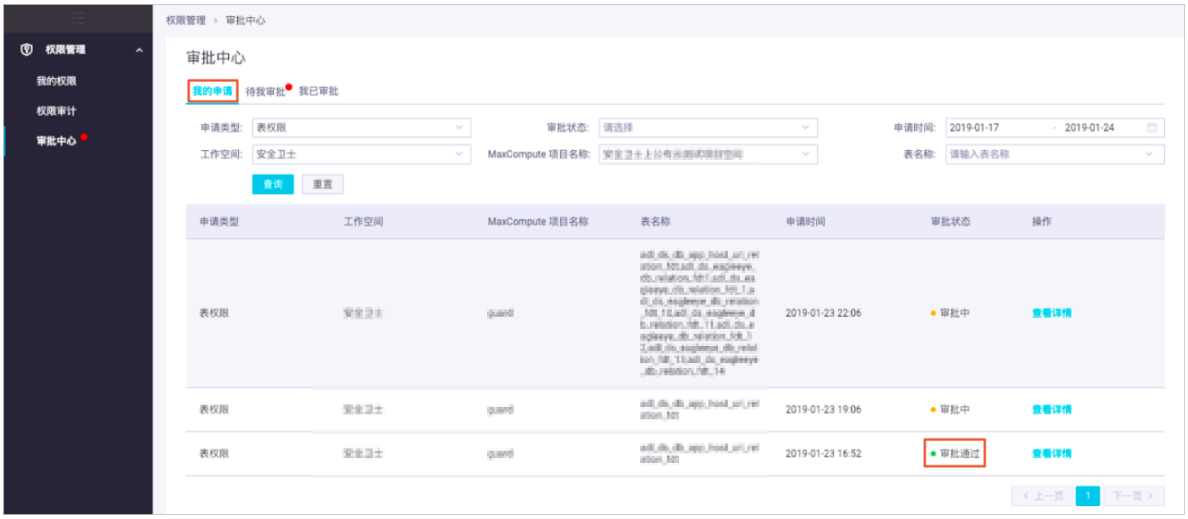
3. 填写表权限申请对话框中的信息。



4. 单击提交。

- 等待审批人审批通过

  进入安全中心 > 审批中心 > 我的申请页面，查看之前申请的进度。当审批状态为审批通过时，代表已具有相应的表权限，可以进行相关操作。

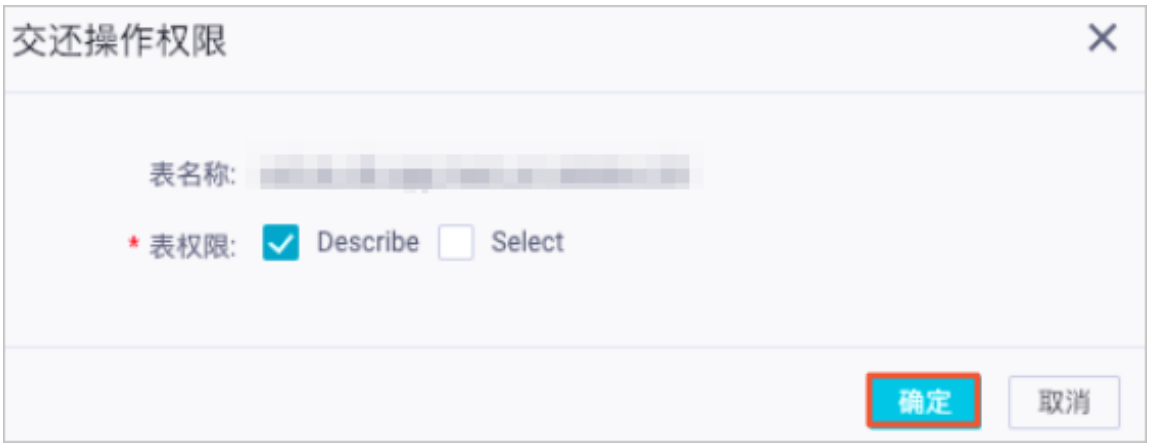


- **交还A表部分字段的权限**

  1. **子账号登录安全中心 > 我的权限 > 表权限页面。**
  2. **单击A表操作栏中的查看，选择交还字段权限。**
  3. **勾选交还字段权限对话框中需要交还的字段，单击确定。**

- 交还A表操作权限

  1. 子账号登录安全中心 > 我的权限 > 表权限页面。
  2. 单击A表操作栏中的查看，选择交还操作权限。
  3. 勾选交还操作权限对话框中需要交还的权限，单击确定。

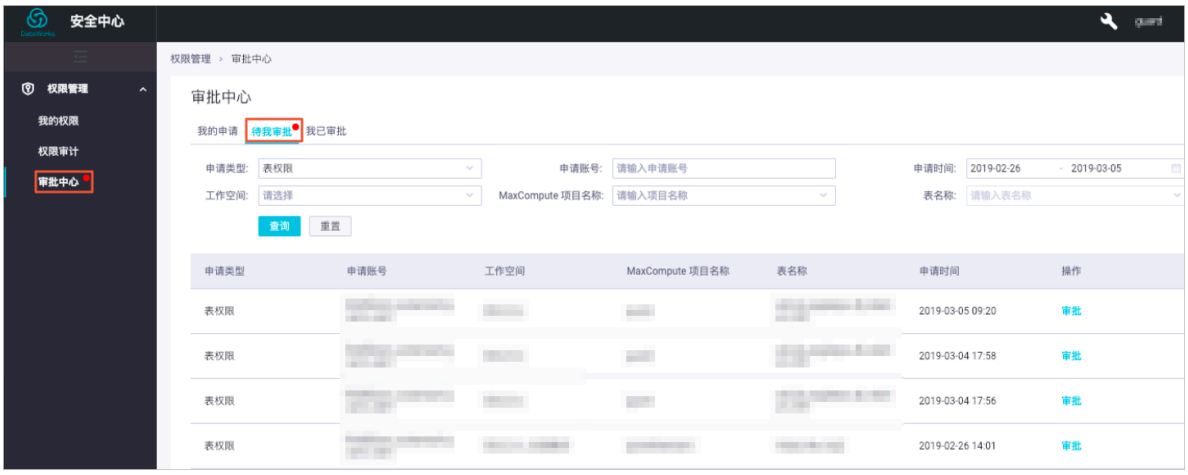


表Owner

子账号乙作为A表的表Owner进行审批，通过A表的申请。

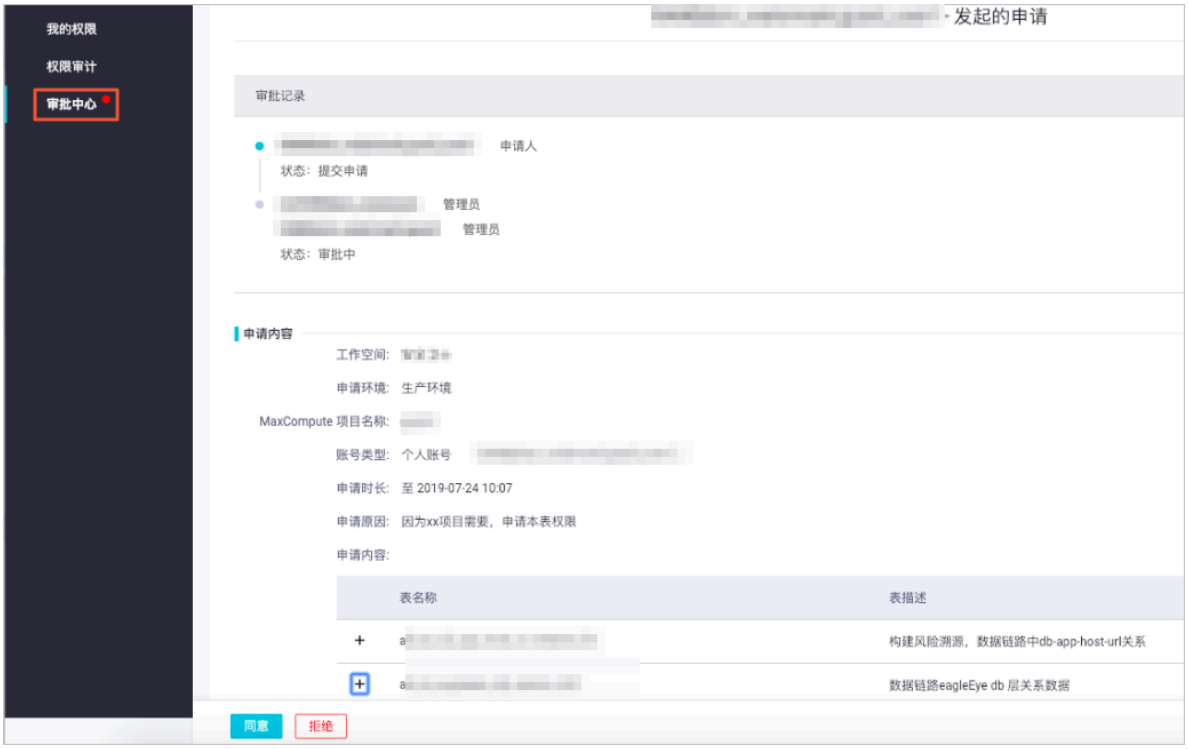
子账号乙作为表Owner，除能使用子账号甲作为普通用户的功能外，还可进行自己作为表Onwer的相关表的审批。

1. 进入安全中心 > 审批中心 > 待我审批页面。



2. 单击子账号刚刚提交的申请单后的审批，即可进入申请单详情页面查看审批记录和申请内容。

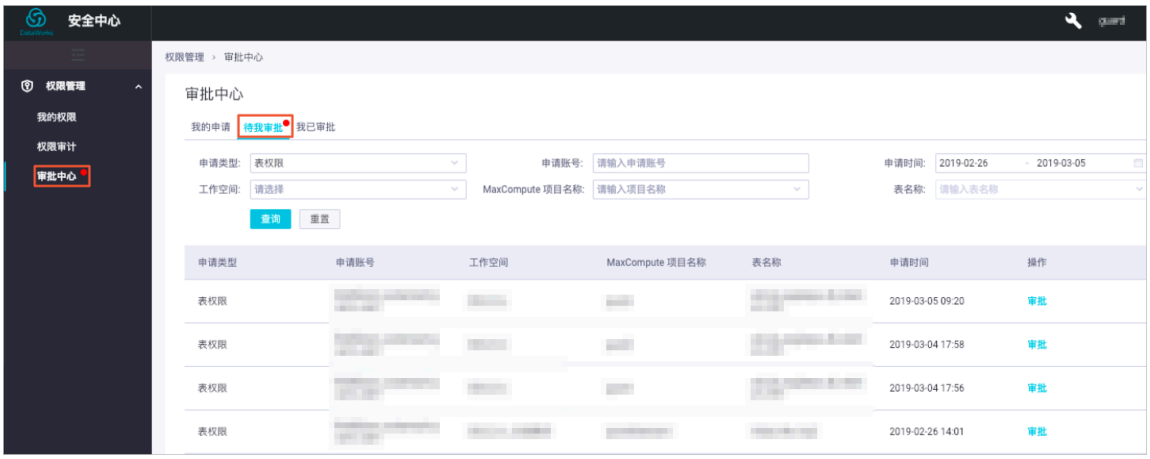
3. 填写审批意见，并单击同意审批通过该申请。



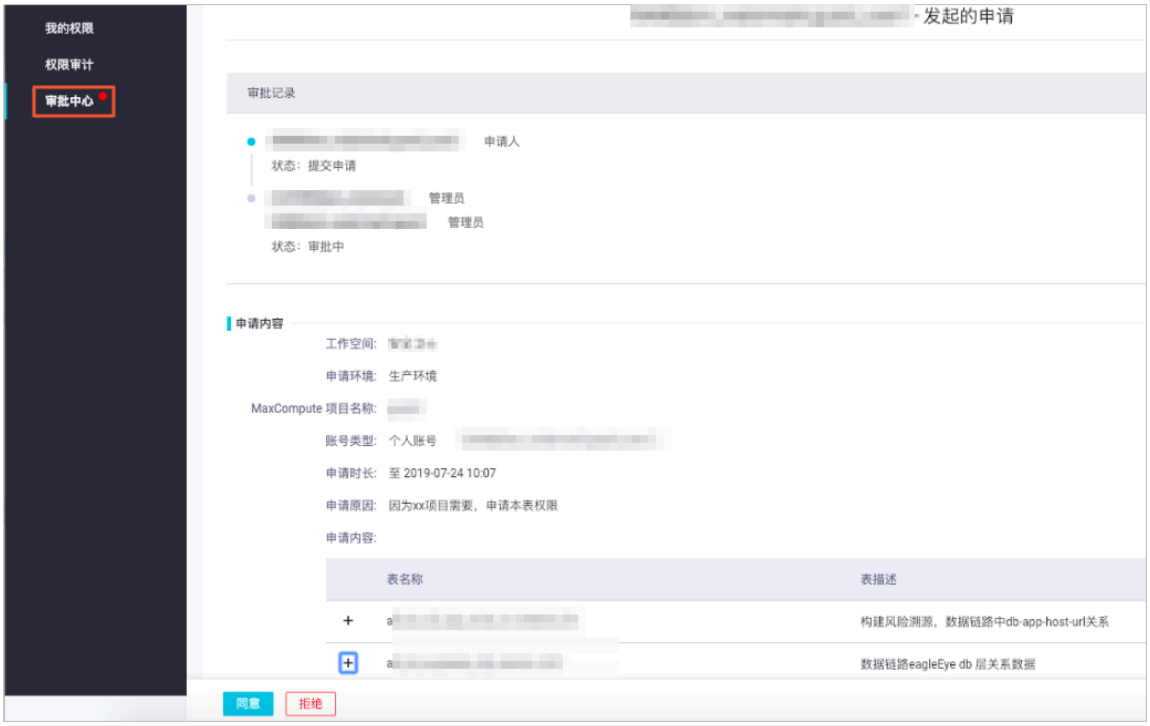
项目管理员

- 主账号作为项目管理员进行审批，通过B表申请。

1. 主账号进入安全中心 > 审批中心 > 待我审批页面。

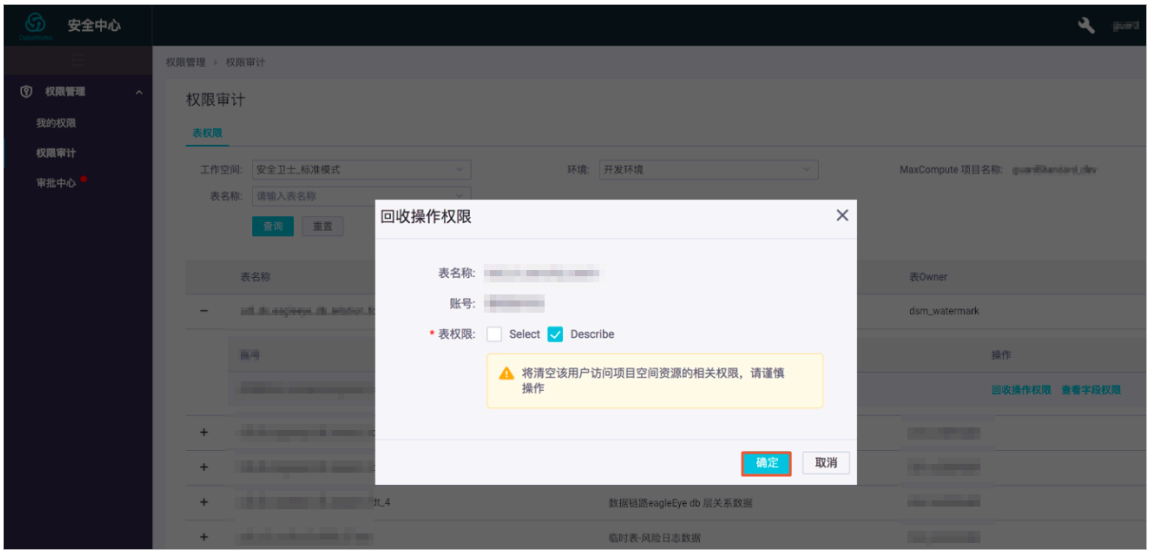


2. 单击子账号刚刚提交的申请单后的审批，即可进入申请单详情页面查看审批记录和申请内容。
3. 填写审批意见，并单击同意审批通过该申请。



- 回收子账号B表的权限

  1. 主账号进入安全中心权限审计页面。
  2. 找到子账号的B表，单击前面的加号展开按钮，展开表详情信息。
  3. 单击子账号操作栏中的回收操作权限。
  4. 勾选回收操作权限对话框中需要回收的权限，单击确定，即可回收该账号拥有的当前数据表的权限。

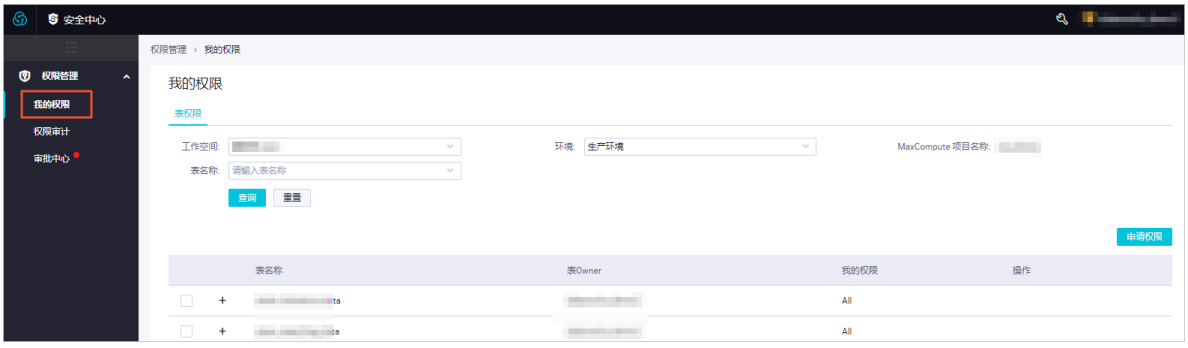


1.3 我的权限

您可以在我的权限页面，查看工作空间内自己拥有的表/字段权限，并对表/字段的权限进行申请或交还。

查看表/字段权限

1. 进入安全中心 > 我的权限 > 表权限页面。
2. 在表权限页面，您可以通过选择工作空间+环境（标准模式下）查看该工作空间对应环境下所有的表，也可以在搜索框中输入表名快速查找需要的表。



您可以在该页面查看当前工作空间的表名称、表Owner和我的权限，并进行申请权限、交还字段权限和交还操作权限的操作。

申请表/字段权限

1. 选择需要申请的表/字段权限。

   · 申请单张表/字段权限

   您可以勾选自己当前无权限但需要使用的表的具体字段后，选择相应操作栏中的查看 > 申请权限。

   您也可以不勾选具体字段，直接选择相应表后操作栏中的查看 > 申请权限，则默认申请该表的所有字段权限。

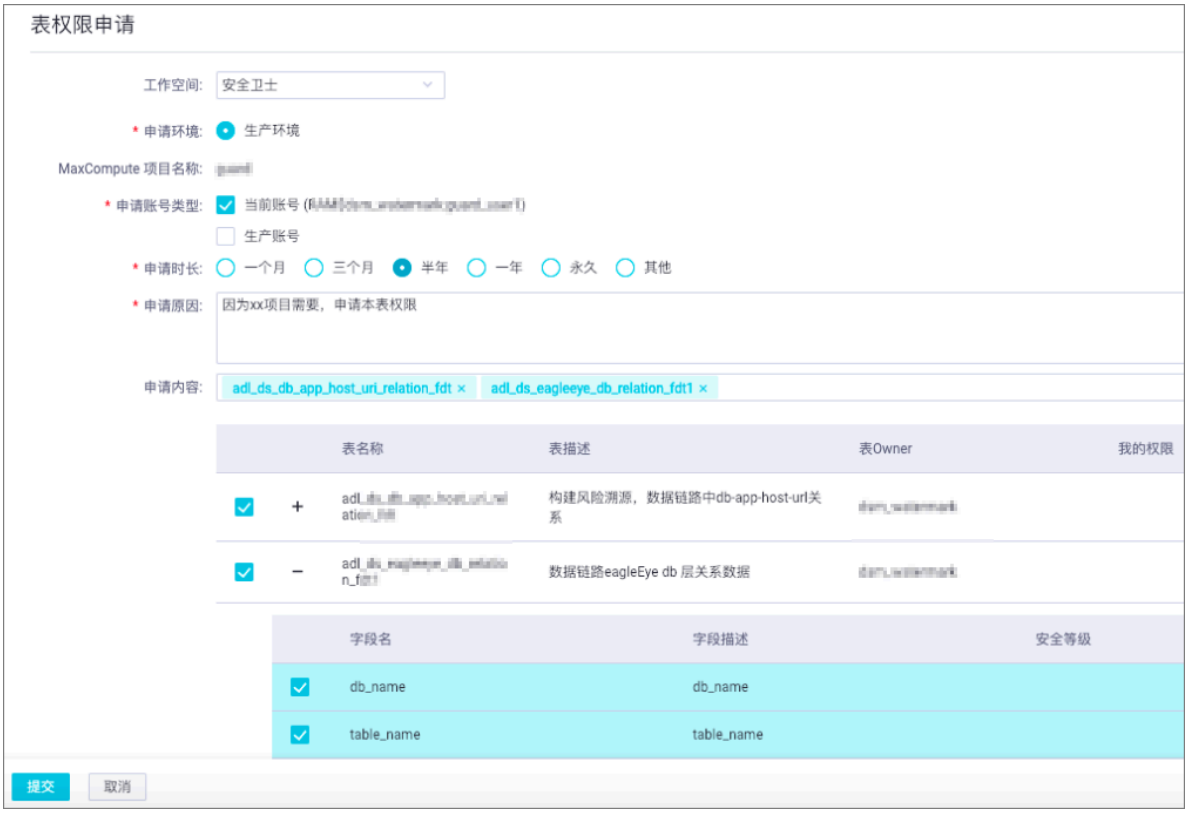
   说明:
   未开启LabelSecurity机制的工作空间将无法进行字段级别的授权，只能进行整表授权。

   · 申请多张表/字段权限

   勾选需要申请的所有表/字段后，单击申请权限进行批量申请。

   说明:
   您也可以直接单击申请权限，然后在申请详情中填写需要申请的内容。

2. 填写表权限申请对话框中的配置。



配置	申请
工作空间	工作空间会根据我的权限页面中的信息自动填写，您也可以进行修改。
申请环境	申请的工作空间的环境。
MaxCompute项目名称	申请的MaxCompute项目名称。
申请账号类型	您可以为当前账号申请权限，也可以替自己加入的其他工作空间的生产账号申请权限。
申请时长	您可选择一个月、三个月、半年、一年、永久或其他。
申请原因	简单说明申请权限的原因。
申请内容	填写表名称，可在原有基础上进行增加和删除。

3. 配置项填写完成后，单击提交。如果不想申请，可单击取消。

交还权限

交还权限包括交还字段权限和交还操作权限。

- 交还字段权限。

> 说明:

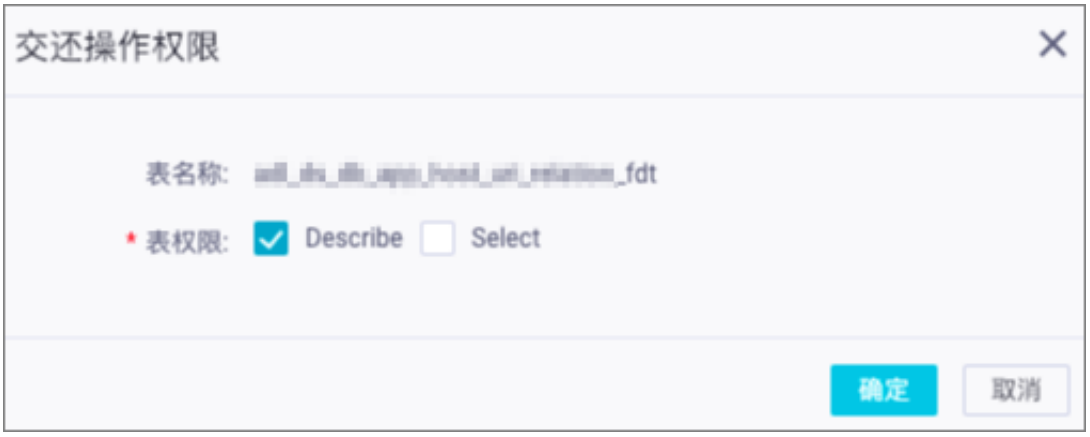
- 交还字段权限仅在已开启LabelSecurity的工作空间中可见。
- 如果希望交还整张表的权限，请使用交还操作权限功能释放所有权限。

1. 选择想要交还权限的表后的操作 > 查看 > 选择字段权限。
2. 在交还字段权限对话框中，勾选需要交还权限的字段。

3. 单击确定。

- 交还操作权限。

  1. 选择想要交还权限的表后的操作 > 查看 > 交还操作权限。
  2. 在交还操作权限对话框中，勾选需要交还的表权限。



  3. 单击确定。

1.4 权限审计

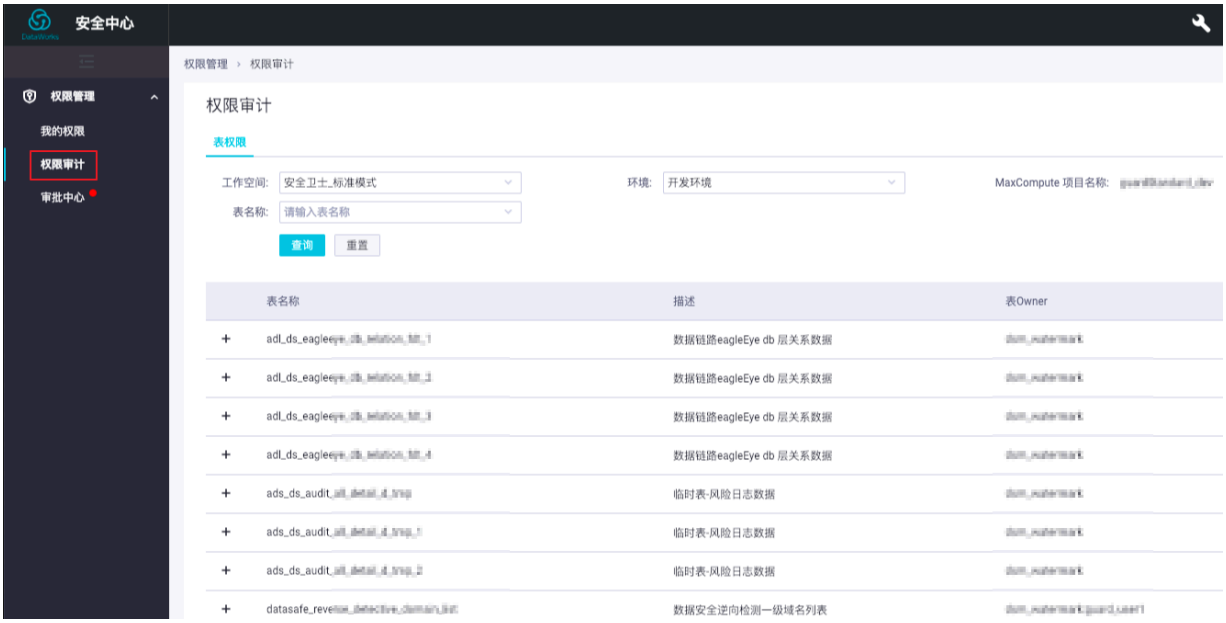
项目管理员可在权限审计页面查看各个工作空间内，分别有哪些账号拥有表和字段的权限，并可回收不必要的表/字段权限。

您可进入安全中心 > 权限审计 > 表权限页面，对本组织范围内（多工作空间）的数据表进行查看和搜索。

在表权限页面中，您可通过选择工作空间+环境（标准模式下）展示该工作空间对应环境下的所有表， 也可以通过搜索框中输入表名进行模糊匹配的方式快速查找需要的表。

查看拥有表权限的账号

您可通过权限审计 > 表权限查看拥有该表权限的所有账号。

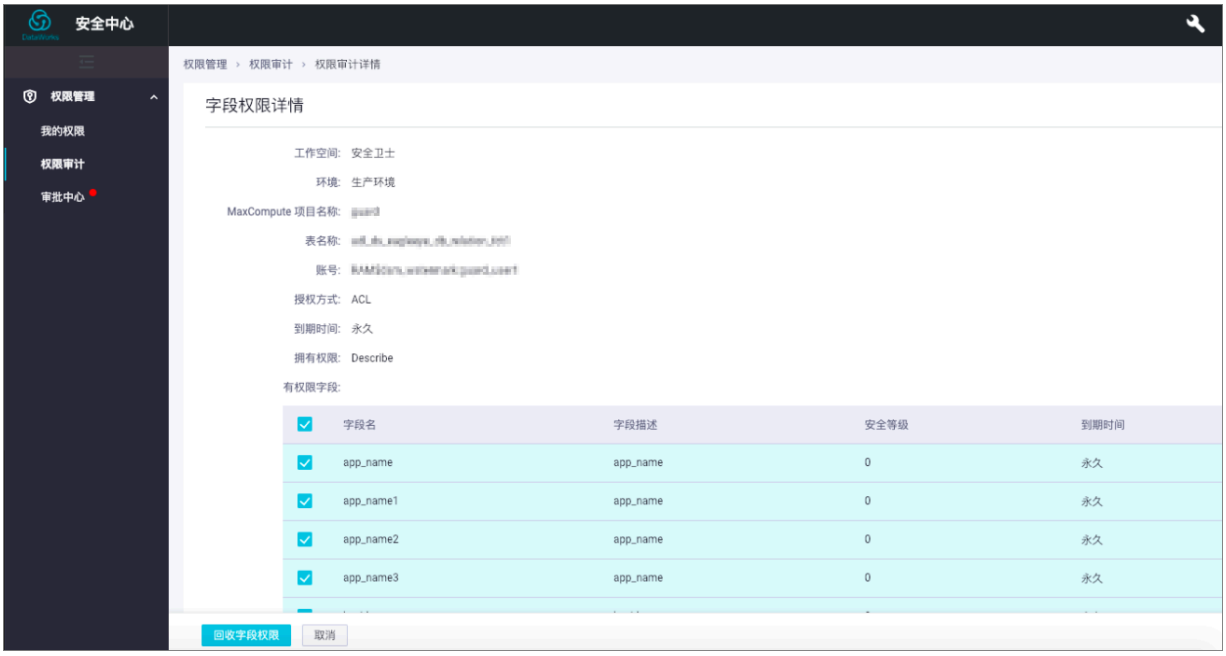


回收表权限

单击相应账号后操作栏中的回收操作权限，即可回收该账号拥有的当前数据表的权限。

查看字段权限

单击相应账号后操作栏中的查看字段权限，即可查看该账号拥有的当前数据表的字段权限。



回收字段权限

对于已经开启LabelSecurity的项目，可以在字段权限详情页面，勾选需要回收的字段，单击回收字段权限，即可进行回收。

1.5 审批中心

您可以在审批中心页面，查看自己提交的申请及进度、查看待自己审批的申请并进行审批、查看自己以前完成的审批任务。

我的申请

1. 进入安全中心 > 审批中心 > 我的申请页面。

   您可在此页面查看自己提交的申请，包括申请类型、工作空间、MaxCompute项目名称、表名称、申请时间、审批结果等信息。

说明:

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

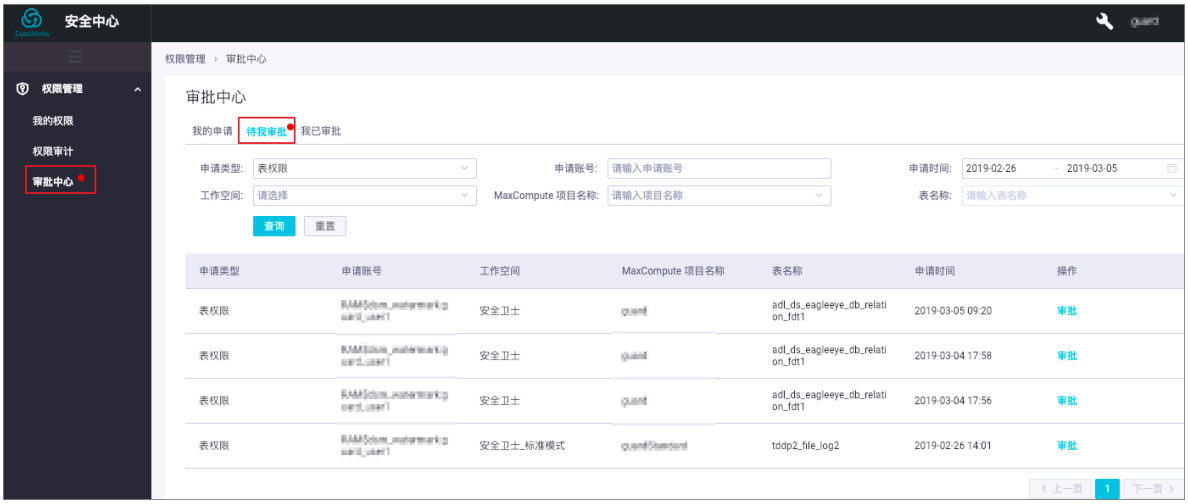
2. 单击操作栏中的查看详情，可以查看申请单的详细信息。

待我审批

1. 进入安全中心 > 审批中心 > 待我审批页面。

   您可在此页面查看当前需要您进行审批的申请单。如果存在需要您审批的申请，则审批中心和待我审批页面会显示红点进行提醒。

   您可查看待审批的申请类型、申请账号、工作空间、MaxCompute项目名称、表名称、申请时间等信息。



2. 单击操作栏中的审批，可以查看申请单的详细信息并进行审批，审批详情页面会为您展示审批记录和申请内容。

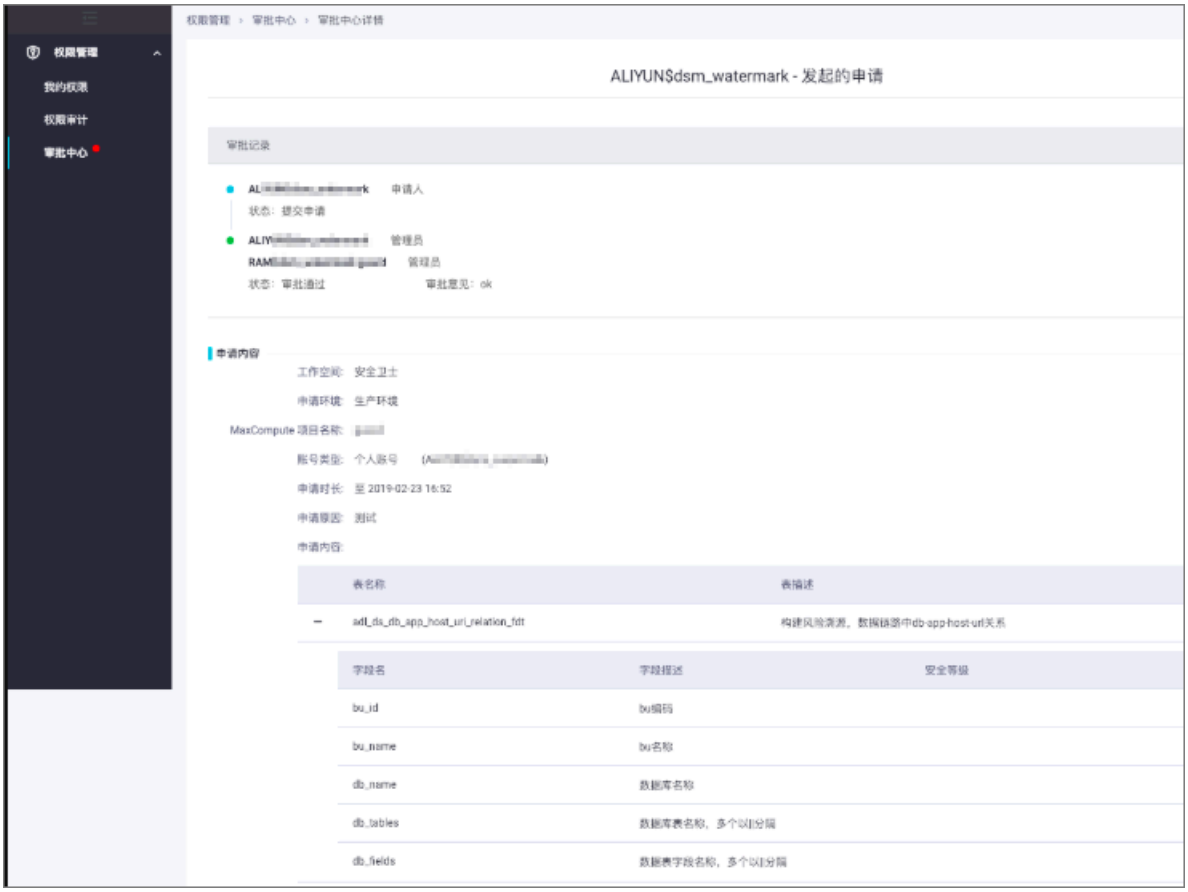
3. 填写审批意见后，您可根据实际情况同意或拒绝该申请。

我已审批

1. 进入安全中心 > 审批中心 > 我已审批页面。

   您可在此页面查看以前审批过的申请单，包括已审批的申请类型、申请账号、工作空间、MaxCompute项目名称、表名称、申请时间等信息。

2. 单击操作栏中的查看详情，即可查看申请单的审批记录和申请内容等详细信息。



1.6 常见问题

本文将为您介绍DataWorks安全中心模块的常见问题和解决方法。

- Q：通过安全中心可以申请什么权限?

  A：您可通过安全中心页面申请DataWorks工作空间内的表权限，包括开发环境和生产环境。

- Q：数据管理和安全中心是什么关系?

  A：安全中心是数据管理中与权限和安全相关功能的升级替代产品。此前已经在数据管理模块中申请的权限和通过`odpscmd grant`命令授权的权限，仍可在安全中心 > 我的权限中显示。

  如果您需通过可视化的方式进行新的权限申请、审批操作，请进入安全中心进行操作，数据管理模块后续将不再支持权限的申请和审批。

- Q：为什么在申请时，有时可以选择字段，有时不可以选择?

  A：如果该工作空间开启了LabelSecurity，即可在申请时选择字段，未开启则只能整表申请。

- Q：提交申请后，需要谁进行审批?

  A：提交的申请需要项目管理员或表Owner进行审批，其中任何一个审批通过/拒绝，则审批完成。

- Q：为什么提交了一个申请，在我的申请中却看到两个申请单？

  A：因为您的申请单中包含的数据表的表Owner不同，安全中心会按照表Owner对于申请单自动进行拆分。
- Q：为什么有的字段只申请1个月权限，审批完成后查看变为永久？

  A：说明字段的安全等级为0或者小于等于您账号的安全等级。
- Q：为什么有的表和字段没有申请权限，但能看到有权限？

  A：出现此情况有以下两种可能：

  - 除安全中心外，管理员还可通过控制台命令行给您授权。
  - 如果您是通过安全中心进行了申请，则说明字段的安全等级为0或者小于等于您账号的安全等级。
- Q：为什么并没有审批某个待我审批中的申请单，却没有了？

  A：因为申请单由项目管理员或表Owner进行审批，其他项目管理员或表Owner在您之前完成了审批，因此该申请单已成为完结状态，便从您的待我审批中消失了。
- Q：查询某个工作空间和环境，提示MaxCompute项目异常，无法进行操作，该如何处理？

  A：请将提示框及框内的错误编码发给项目管理员，由他进行问题的排查及解决。
- Q：为什么交还/回收某个字段权限却无效？

  A：只能交还/回收字段的安全等级大于账号安全等级的字段，对于安全等级为0或者小于等于账号安全等级的字段，无法进行字段权限的交还/回收。
- Q：为什么主账号不能申请权限？

  A：主账号默认具有所有权限，无需单独申请权限，因此对主账号无需的操作，如申请权限等进行隐藏，不会影响到主账号的正常使用。
- Q：是否可以在安全中心页面查看以前在数据管理页面的申请/审批记录？

  A：目前安全中心和数据管理中的申请/审批记录没有进行关联，如果需要查看在数据管理进行申请/审批的历史记录，请跳转至数据管理页面进行查看。
- Q：是否可以通过安全中心的申请记录来回收权限？

  A：目前安全中心并非唯一的授权渠道，为了最大程度地支持权限回收，权限审计中列出的是所有用户的ACL权限，不区分授权渠道。您可以基于目前的权限现状进行回收，不必通过申请记录来操作。
- Q：之前在数据管理提交的权限申请，仍未通过审批，是否需要重新申请？

  A：安全中心和数据管理中的申请审批记录未进行关联，您需要重新申请。

- **Q：如何设置字段的LabelSecurity?**

  A：您需前往数据地图设置字段的LabelSecurity。

2 数据质量

2.1 数据质量概述

数据质量是支持多种异构数据源的质量校验、通知、管理服务的一站式平台。

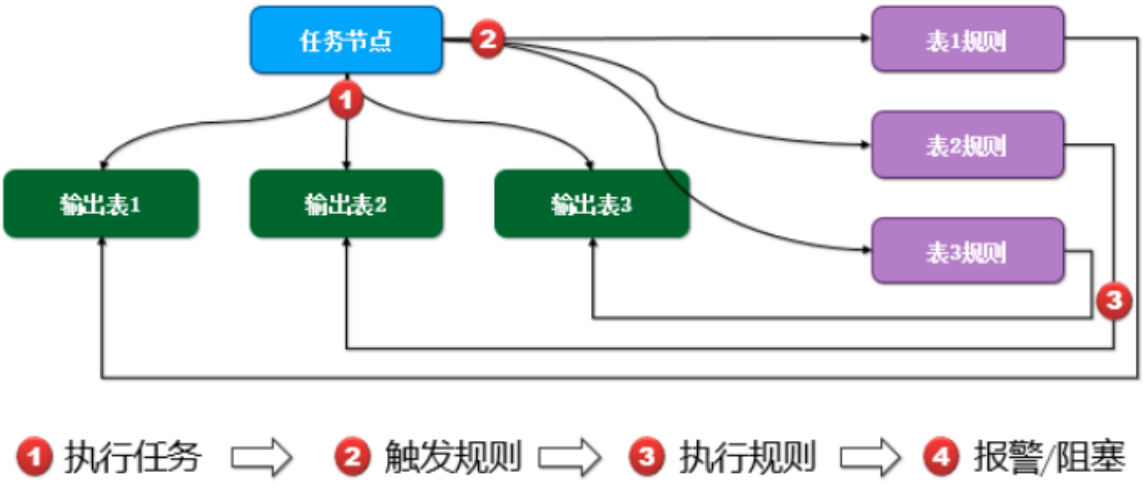
数据质量依托DataWorks平台，为您提供全链路的数据质量方案，包括数据探查、数据对比、数据质量监控、SQL扫描和智能报警等功能。

数据质量监控可以全程监控数据加工流水线，根据质量规则及时发现问题，并通过报警通知负责人及时处理。

数据质量以数据集（DataSet）为监控对象。目前，数据质量支持MaxCompute数据表和DataHub实时数据流的监控。当MaxCompute离线数据发生变化时，数据质量会对数据进行校验，并阻塞生产链路，以避免问题数据污染扩散。同时，数据质量支持管理历史校验结果，您可以对数据质量进行分析和定级。

在流式数据场景下，数据质量能够基于Datahub数据通道进行监控和断流，第一时间告警给订阅用户。数据质量支持设置橙色、红色告警等级和告警频次，最大限度地减少冗余报警。

数据质量监控的流程如下图所示。



说明:

数据质量主要对MaxCompute和DataHub数据集的质量进行监控。因此，您需要先创建表，并在表中写入数据后，方可使用数据质量功能。

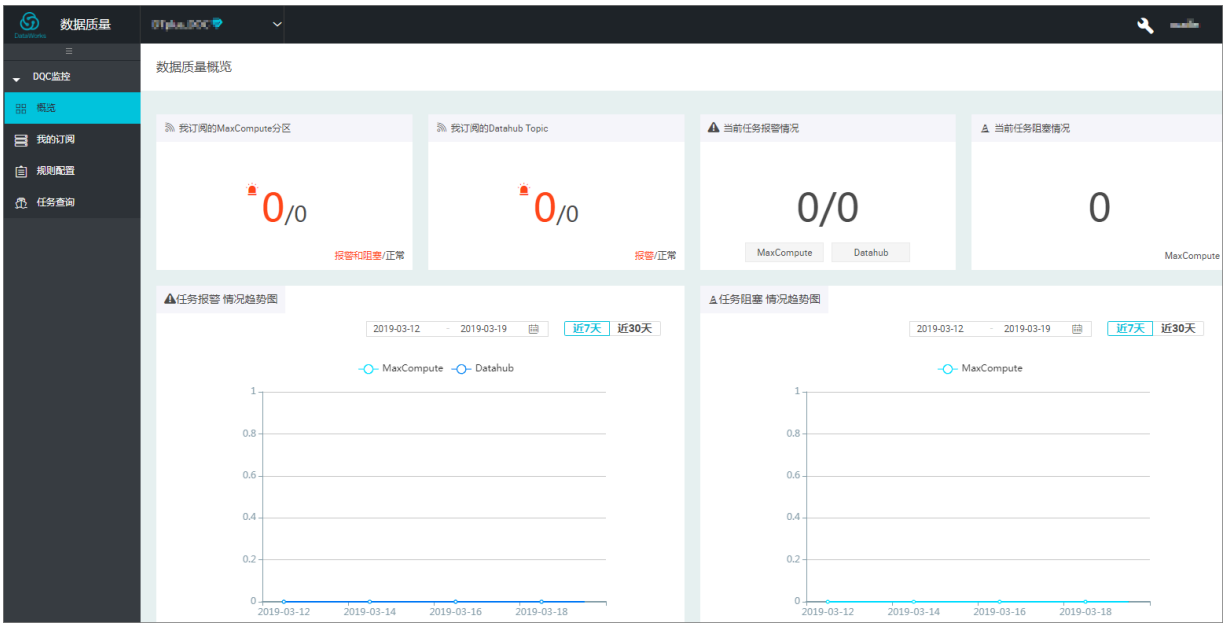
您可以通过MaxCompute客户端或DataWorks控制台创建MaxCompute表并写入数据。

数据质量管理原则请参见#unique_13。

2.2 功能介绍

2.2.1 首页概览

数据质量首页的概览页面为您展示订阅数据的报警以及任务的阻塞情况，您可以查看经过筛选的结果。



模块	说明
我订阅的MaxCompute分区	显示当天订阅的MaxCompute分区的报警和阻塞、正常两种情况。单击此模块可快速跳转至MaxCompute数据源的任务查询页面，查看报警详情。
我订阅的DataHub Topic	显示当天订阅的DataHub数据源报警、正常两种情况，单击此模块可快速跳转至DataHub数据源的任务查询页面，查看报警详情。
当前任务报警情况	显示当天、当前应用下的MaxCompute和DataHub两种数据源的任务报警情况。
当前任务阻塞情况	显示当天、当前应用下MaxCompute数据源的任务阻塞情况。
任务报警情况趋势图	可选7天、30天以及自定义时间区间，支持日期范围为近三个月内的MaxCompute和DataHub数据源的任务报警趋势图。
任务阻塞情况趋势图	可选7天、30天以及自定义时间区间，支持日期范围为近三个月内的MaxCompute的任务阻塞情况。

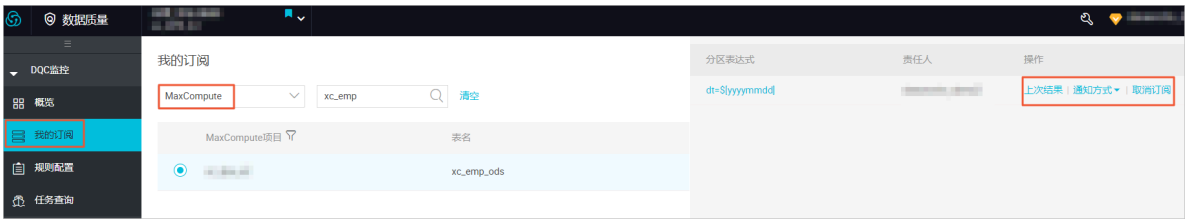
2.2.2 我的订阅

我的订阅页面为您展示当前账号订阅的所有任务。

当前，数据质量支持MaxCompute监控和Datahub监控，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。

您可以选择以下两种数据源：

- MaxCompute数据源



  - 单击右侧相应的分区表达式，即可进入规则配置页面，详情请参见MaxCompute监控。
  - 单击上次结果，可跳转至任务查询页面。
  - 目前支持邮件通知、邮件和短信通知、钉钉群机器人和钉钉群机器人@ALL四种通知方式。
  - 单击取消订阅，即可删除相应的订阅信息。

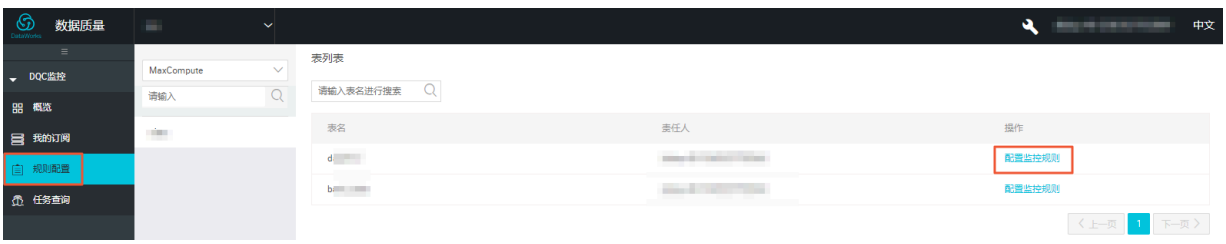
- DataHub数据源

  - 单击对应Topic操作栏下的配置监控规则，即可进入规则配置页面，详情请参见DataHub监控。
  - 单击对应Topic操作栏下的取消订阅，即可取消已订阅的Topic。

2.2.3 规则配置

目前数据质量支持离线的MaxCompute监控和DataHub监控，本文将为您介绍如何配置MaxCompute规则。

- 选择MaxCompute数据源，即可显示当前数据源下所有的表。您也可以使用搜索功能，快速定位至其他数据源下查看表。
- 选择Datahub数据源，即可显示当前数据源下所有的Topic。您也可以使用搜索功能，快速定位至其他数据源下查看Topic。



单击配置监控规则，即可进入规则配置页面。

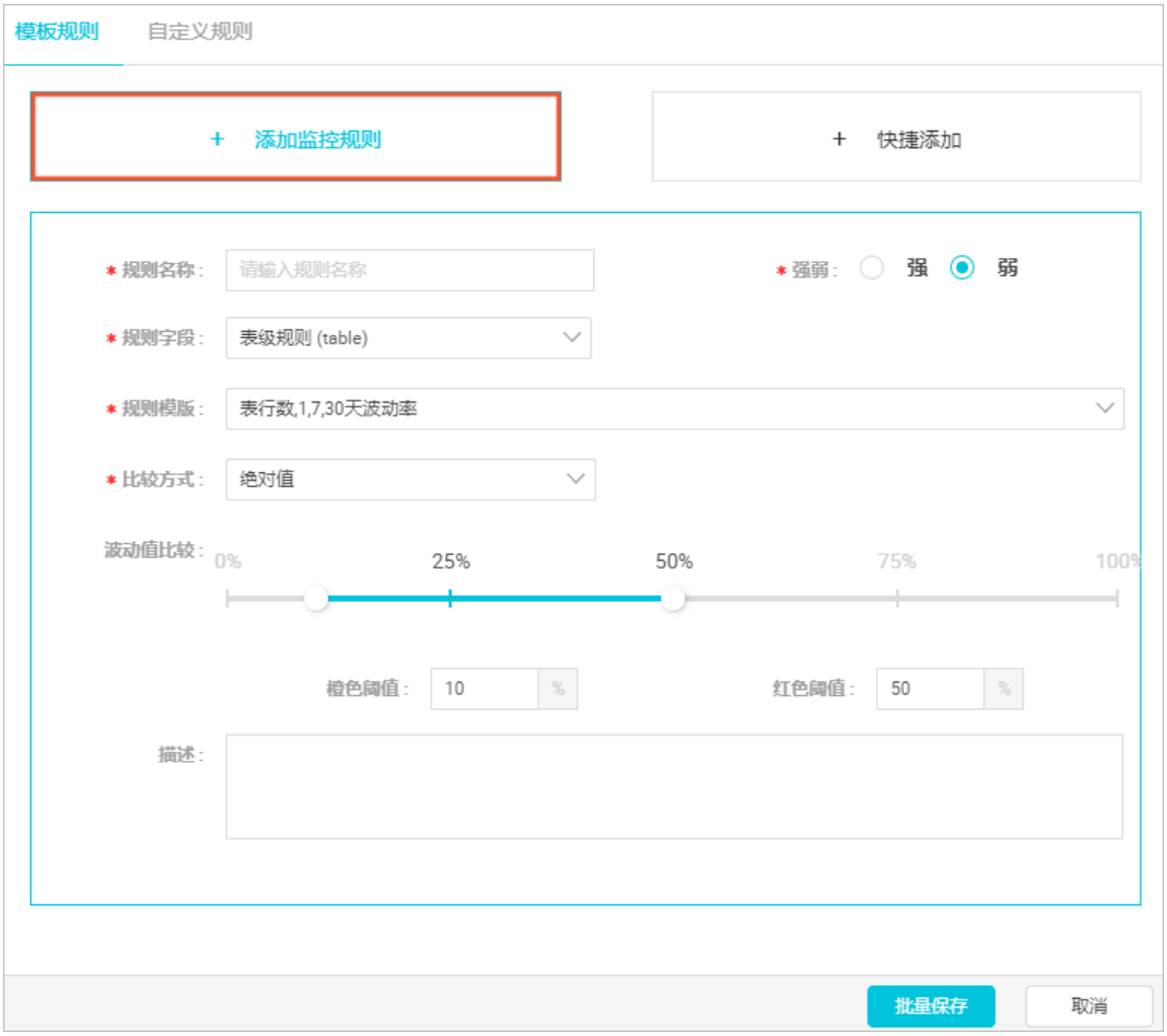
目前数据质量规则配置包括模板规则和自定义规则。

> **说明:**
> 配置模板规则前，需首先配置分区表达式，详情请参见#unique_20/unique_20_Connect_42_section_bby_bb3_r2b。

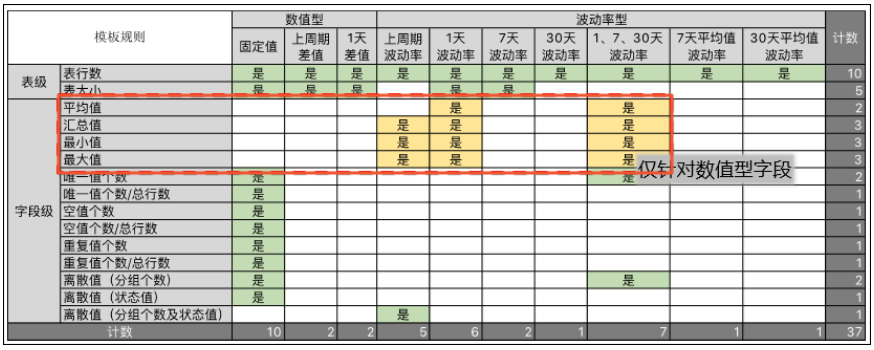
模板规则

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



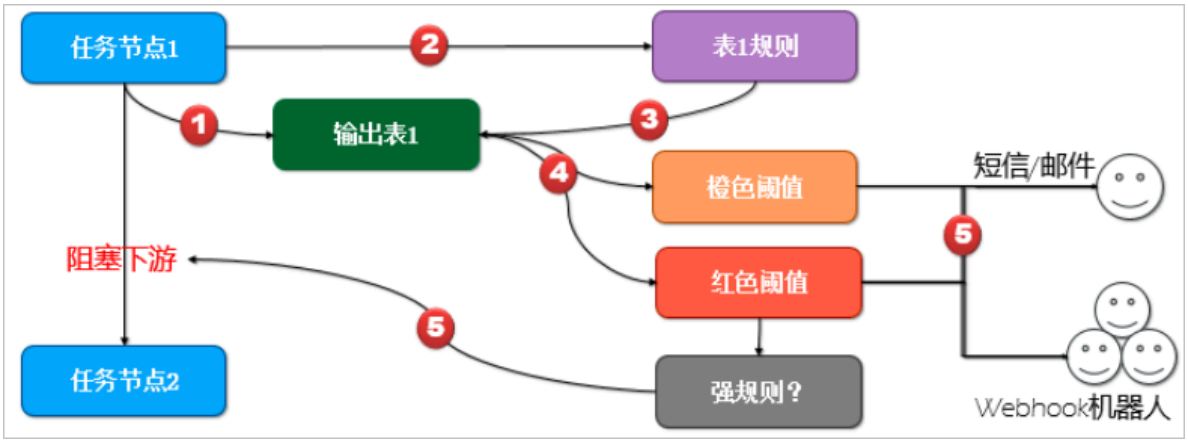
配置	说明
规则名称	请输入规则名称。
强弱	设置强规则或弱规则： - 如果设置强规则，红色异常报警并阻塞下游任务节点，橙色异常报警不阻塞。 - 如果设置弱规则，红色异常报警不阻塞下游任务节点，橙色异常不报警不阻塞。
规则字段	包括表级规则和字段级规则，字段级规则包括数据类型和非数据类型。

配置	说明
规则模板	目前共有37种规则，不支持的规则模板将不能被选择。 您可以单击下拉框选择相关的模板，支持的模板详情请参见下表。  说明: 目前平均值、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。

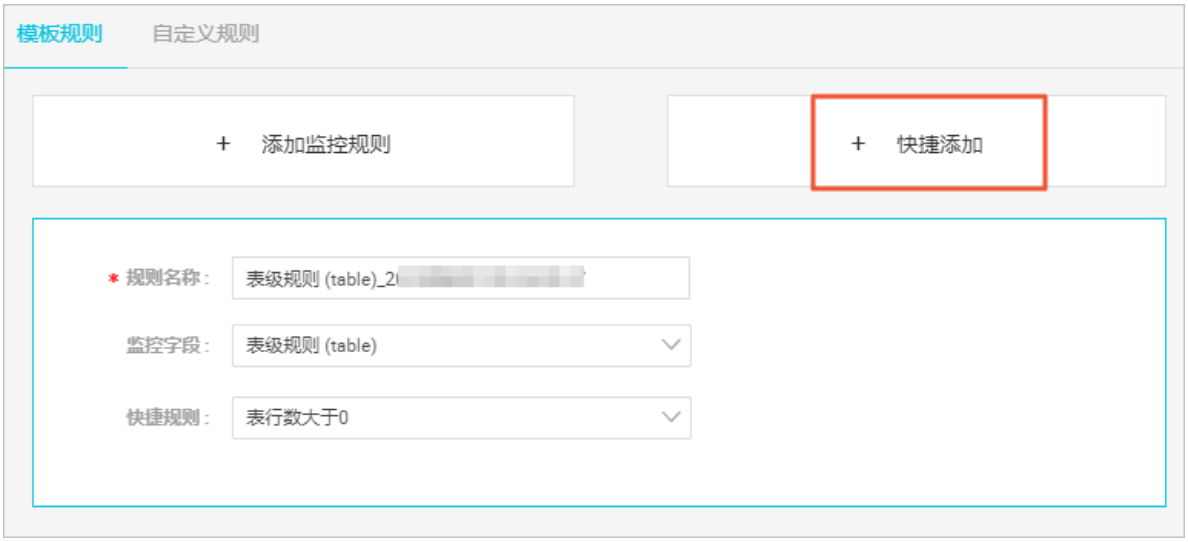
模板规则		数值型			波动率型							计数
		固定值	上周期差值	1天差值	上周期波动率	1天波动率	7天波动率	30天波动率	1、7、30天波动率	7天平均值波动率	30天平均值波动率	
表级	表行数	是	是	是	是	是	是	是	是	是	是	10
	表大小	是	是	是		是	是					5
字段级	平均值					是			是			2
	汇总值				是	是			是			3
	最小值				是	是			是			3
	最大值				是	是			是			3
	唯一值个数	是							是			2
	唯一值个数/总行数	是										1
	空值个数	是										1
	空值个数/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	5	6	2	1	7	1	1	37

配置	说明
波动值比较	- 计算波动率，您可以根据波动率计算公式计算结果。 波动率=（样本-基准值）/基准值。 - 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。 说明: 样本和基准值的名词解释如下： - 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1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。 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。 - 如果没有橙色阈值，则只会出现红色报警和正常两种情况。 - 如果没有红色阈值，则只会出现橙色报警和正常两种情况。 - 两个都不填，则红色报警（通常禁止两个阈值都不填，阈值校验会默认橙色10%，红色50%）。

下图为报警与阻塞的实现逻辑。



- 快捷添加



配置	说明
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，快捷规则仅支持表行数大于0。 - 选择字段级规则，快捷规则可以选择字段重复值或字段空值。

自定义规则

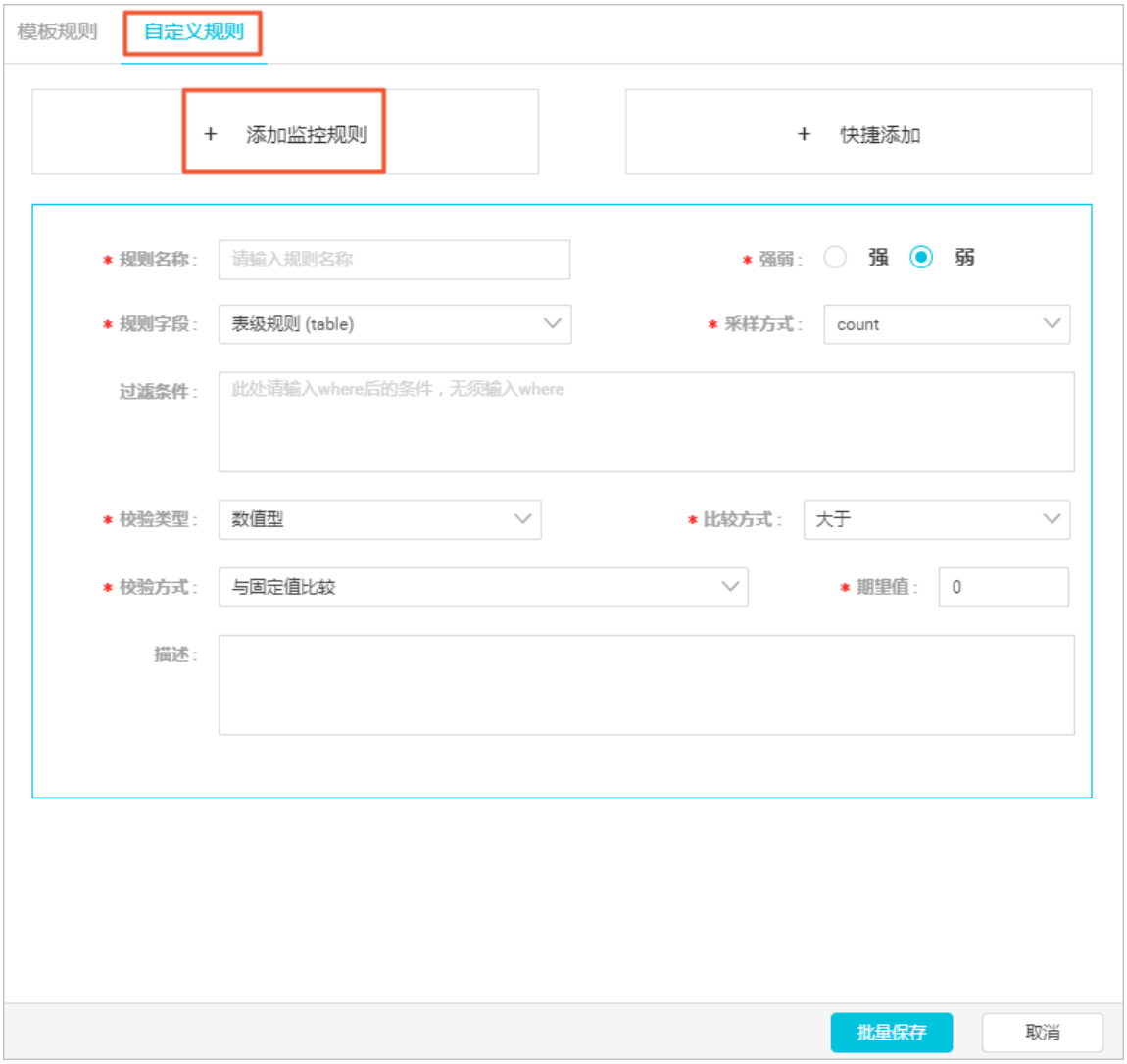
如果模板规则不能满足您对分区表达式中数据质量的监控需求，您还可以通过创建自定义规则来满足个性化的监控需求。

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和自定义SQL。

  - 表级规则

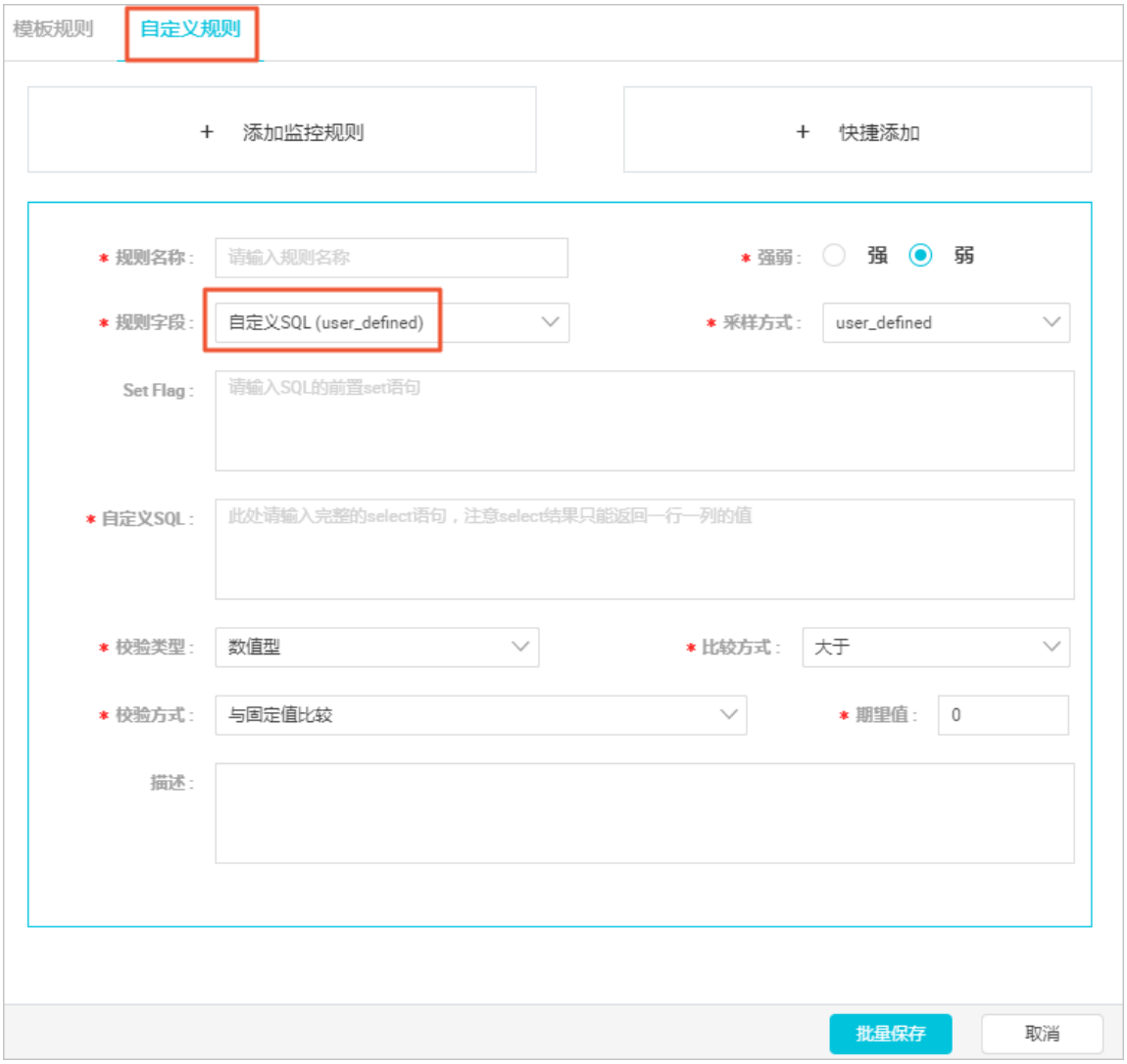


配置	说明
规则名称	请输入规则名称。
规则字段	此处选择表级规则。表级自定义规则，支持根据业务属性自定义where过滤条件。
强弱	设置强规则或弱规则： ■ 如果设置强规则，红色异常报警并阻塞下游任务节点，橙色异常报警不阻塞。 ■ 如果设置弱规则，红色异常报警不阻塞下游任务节点，橙色异常不报警不阻塞。
采样方式	支持count和count/table_count两种方式。

配置	说明
过滤条件	例如：您需要查询业务日期下表的分区，可以将where条件设置为`pt=$[yyyymmdd-1]`。
校检类型	支持数值型和波动率型两种类型。
比较方式	■ 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	■ 如果选择校检类型为数值型，目前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。

配置	说明
描述	对创建的自定义规则进行描述。

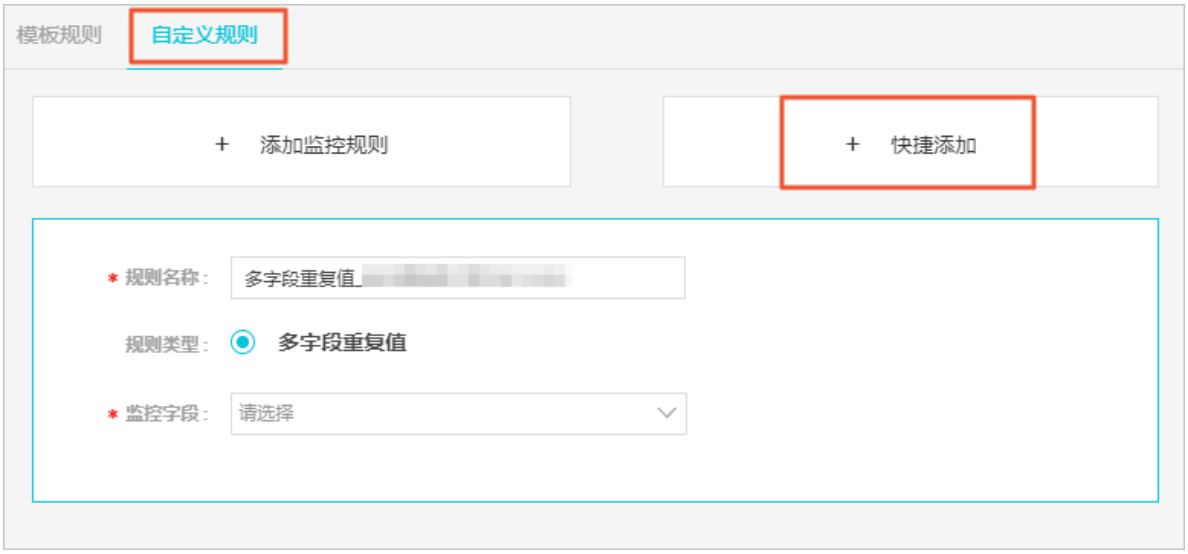
- 自定义SQL



配置	说明
规则名称	请输入规则名称。
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）。
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。
采样方式	支持count和count/table_count两种方式。
Set Flag	输入SQL的前置set语句。

配置	说明
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。
校检类型	支持数值型和波动率型两种类型。
比较方式	■ 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	■ 如果选择校检类型为数值型，目前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



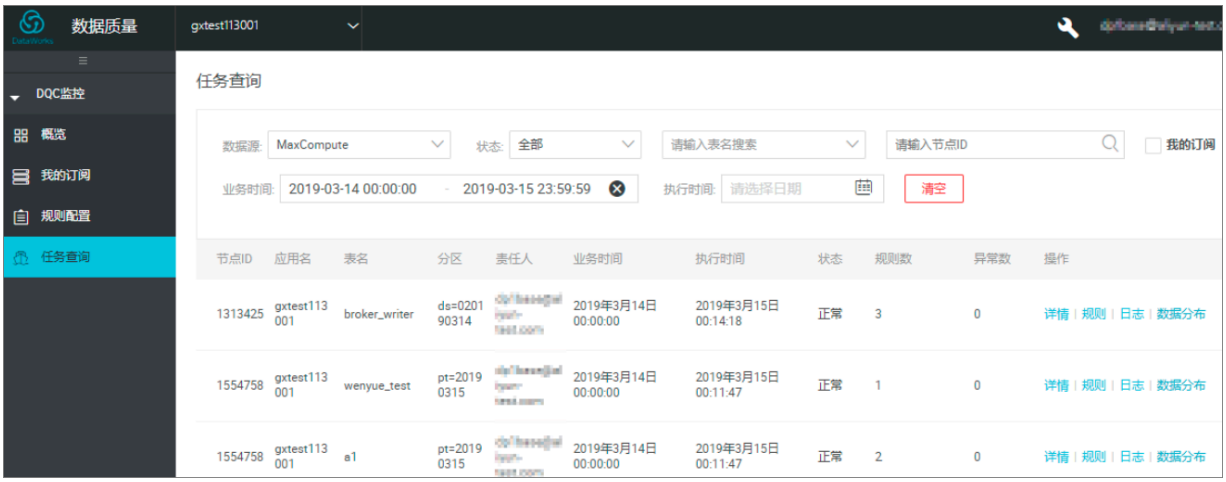
配置	说明
规则名称	请输入规则名称。
规则类型	仅支持多字段重复值。
监控字段	设置监控字段。

2.2.4 任务查询

任务查询模块主要显示规则校验结果，规则运行后，即可在任务查询模块查看运行记录。

查看MaxCompute监控任务

您可以在任务查询页面，根据状态、表名、节点ID等信息进行搜索，查看任务运行情况。右侧对应的蓝色字体可以链接到其他页面，方便您进行查看和修改。

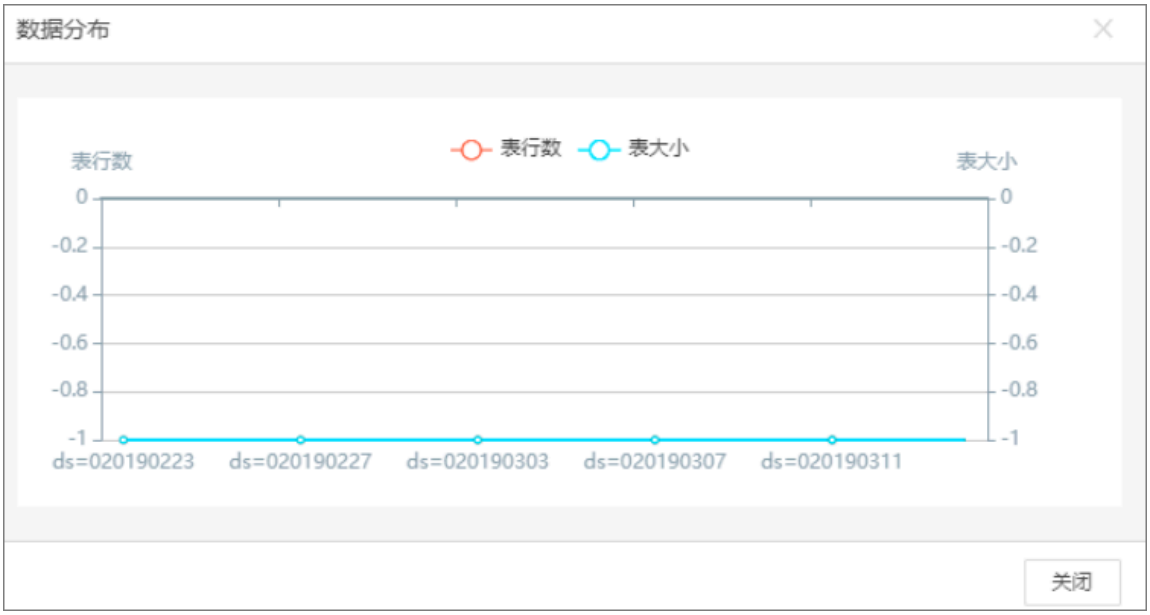


- 节点ID：触发规则的任务节点。
- 执行时间：规则执行时间。
- 状态：执行结果的状态，如果是报警或阻塞状态，需要多加注意。

- 操作：
  - 详情

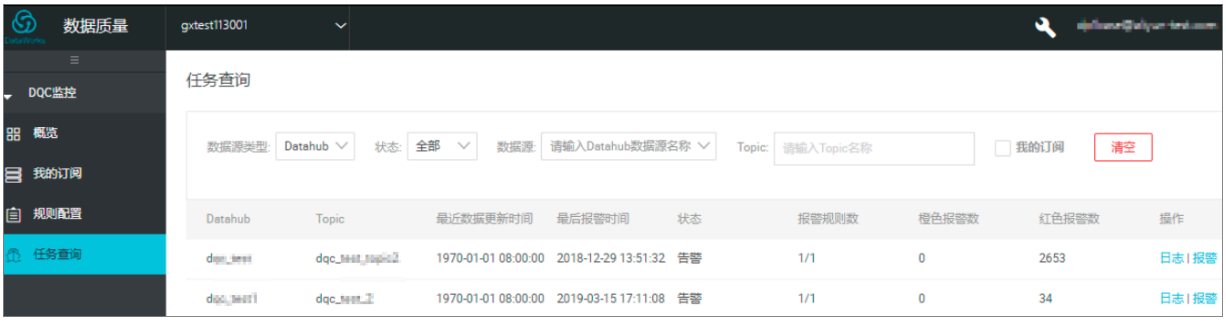
    ■ 单击更多，即可查看数据源、应用名、节点ID和责任人等信息。

    ■ 单击对应字段后的查看历史记录，即可查看每次调度后的运行记录。
  - 规则：单击即可跳转至规则配置页面。可在此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#unique_17。
  - 日志：查看规则运行日志（高级）。
  - 数据分布：一键探查数据量、表行数，单击对应任务后的数据分布，即可查看该任务从创建至今，每次运行的情况。



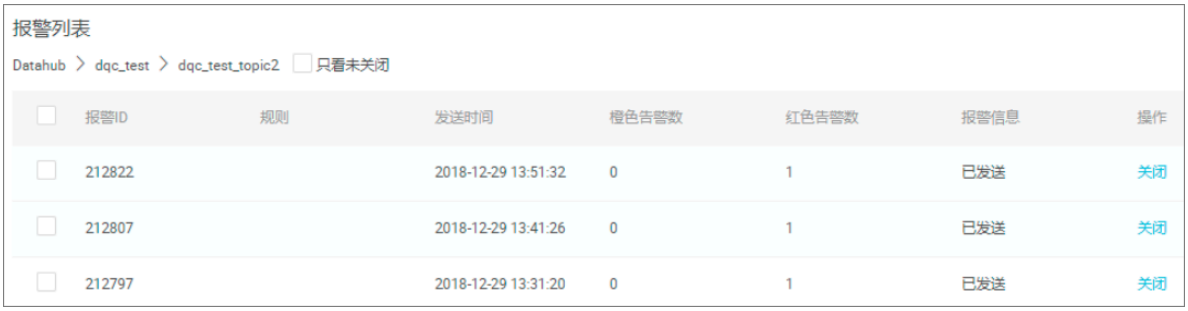
查看DataHub监控任务

您可以在任务查询页面，根据状态、表名、节点ID等信息进行搜索，查看任务运行情况。右侧对应的蓝色字体可以链接到其他页面，方便您进行查看和修改。



操作：

- **日志：查看规则运行日志（高级）。**
- **报警：单击相应Topic右侧的报警，可查看任务运行报警详情，您也可以在详情页面关闭报警。**



2.3 使用指南

2.3.1 DataHub监控

规则配置模块是数据质量（DQC）中最核心的部分，当前支持MaxCompute监控和DataHub监控。本文将为您介绍如何配置DataHub监控。

DataHub实时数据监控当前支持以下功能：

- **支持数据断流、数据延迟两种监控模板。**
- **自定义Flink SQL、维表join、多流join以及窗口函数等流计算特性。**

添加数据源

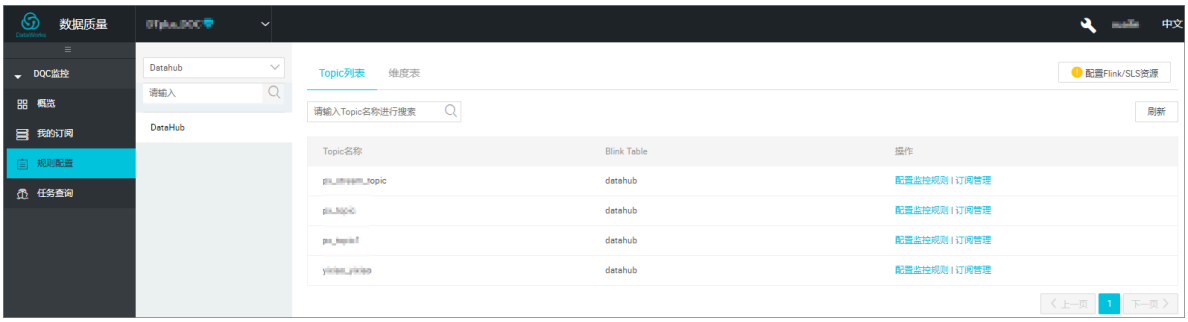
您需要首先进入数据集成页面添加数据源，详情请参见配置Datahub数据源。新建数据源成功后，即可进入数据质量（DQC）页面进行规则的配置。

选择数据源

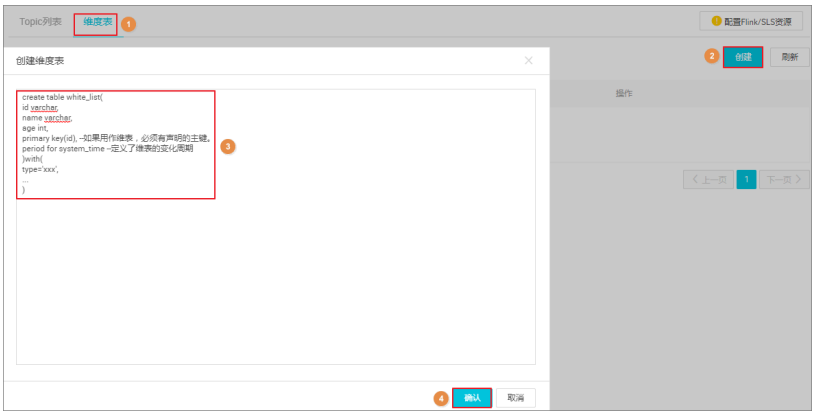
1. 单击左侧导航栏的规则配置，进入规则配置页面。

2. 选择DataHub数据源，即可显示当前数据源下所有的Topic。

   您也可以通过搜索功能，快速定位到其他数据源下查看Topic。



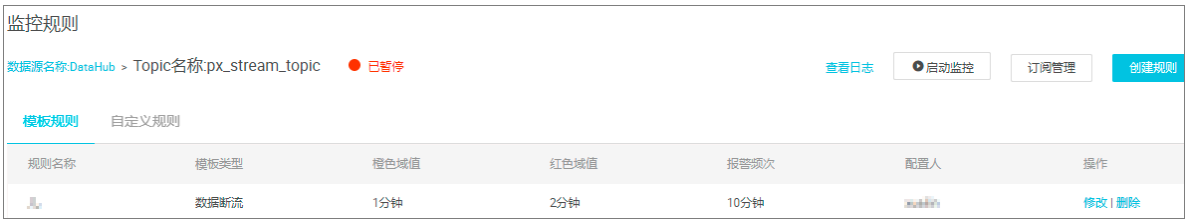
配置	说明
配置Flink/SLS规则	添加数据源后Flink/SLS资源会根据数据源拉取相关的信息。
Topic列表	DataHub数据源下所有Topic的名称，您可以在相应Topic后进行下述操作。 · 配置监控规则：对当前Topic创建规则，支持创建模板规则和自定义规则。 · 订阅管理：查看当前Topic的订阅人，可以快捷修改订阅人和报警方式，也可以配置钉钉群报警，这里修改的报警方式对所有订阅人有效。

配置	说明
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、ODPS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。

3. 在Topic列表页面选择需要配置的Topic，单击右侧的配置监控规则。

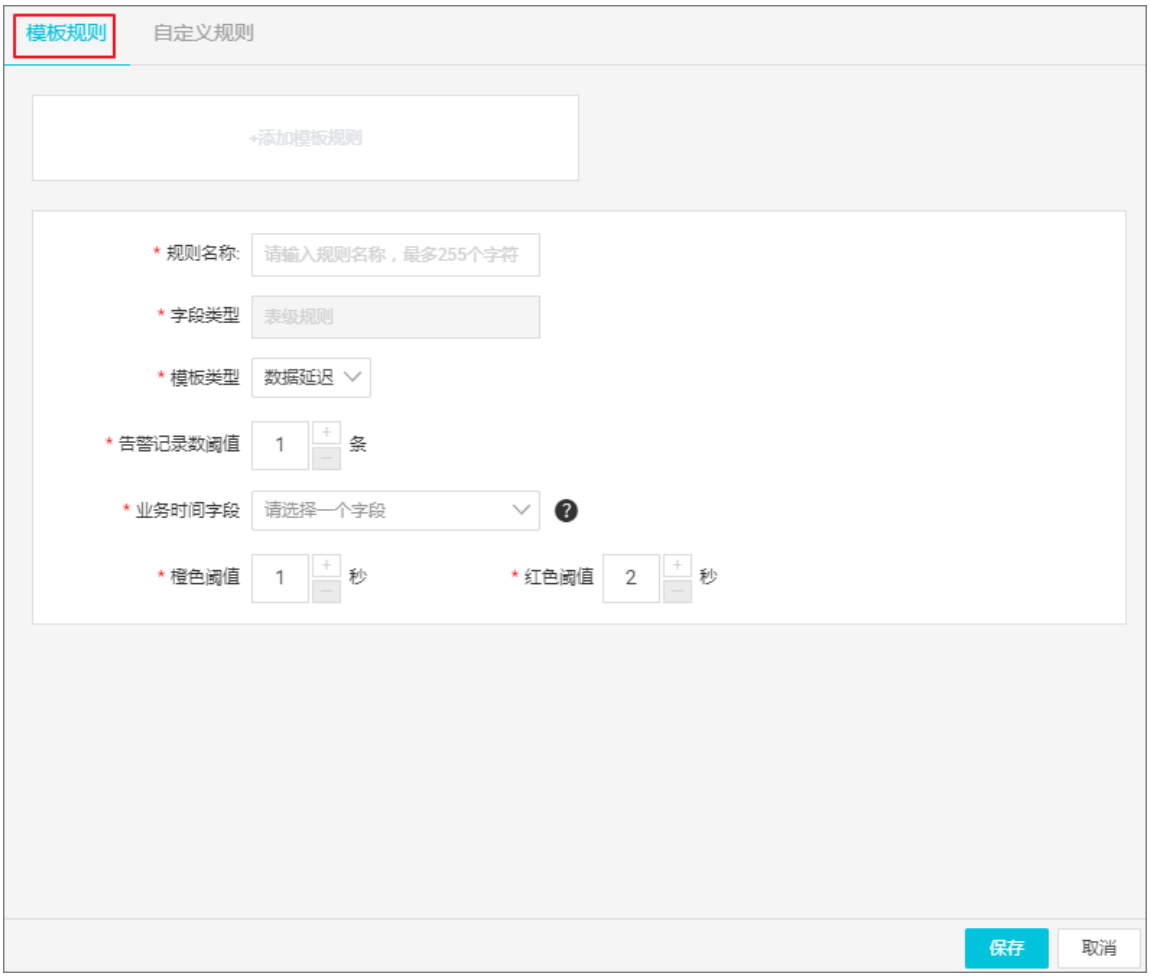
配置监控规则

1. 单击监控规则页面右上角的创建规则，当前支持模板规则和自定义规则两种类型。



- 单击添加模板规则，目前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。

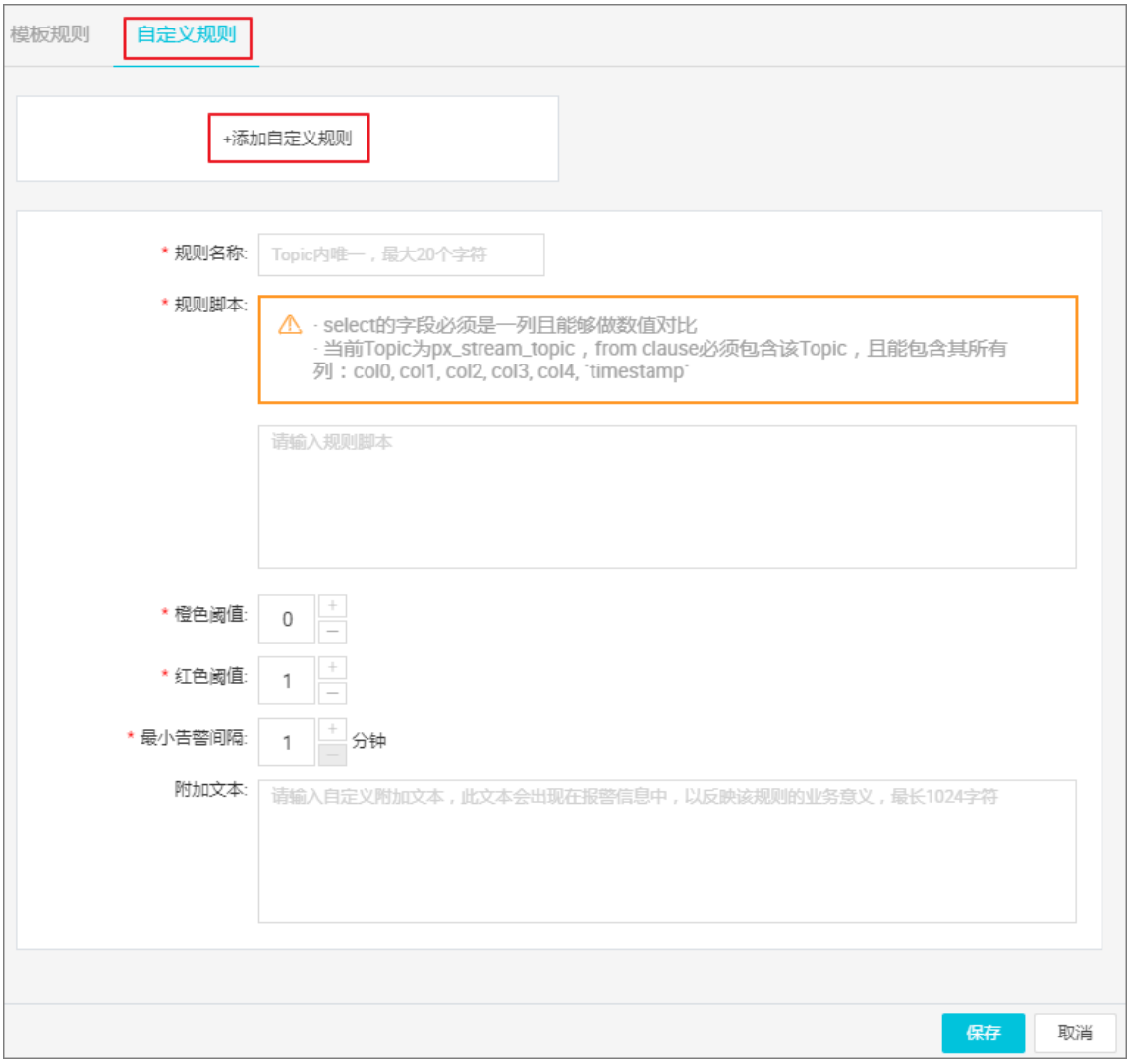


配置	说明
规则名称	输入规则名称，最多255个字符。
字段类型	默认为表级规则。

配置	说明
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 H 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。 说明: 配置数据断流前，请首先在Flink中购买服务并创建项目。
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才会有此参数。
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。

配置	说明
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规则有其他的使用方式，可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，且包含此Topic中所有的列。

配置	说明
规则名称	输入规则名称，需在Topic内唯一，最多支持20个字符。

配置	说明

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。
	- 示例一：简单SQL。
	`select id as a from zmr_tst02;`
	- 示例二：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`
	- 示例三：两个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'`

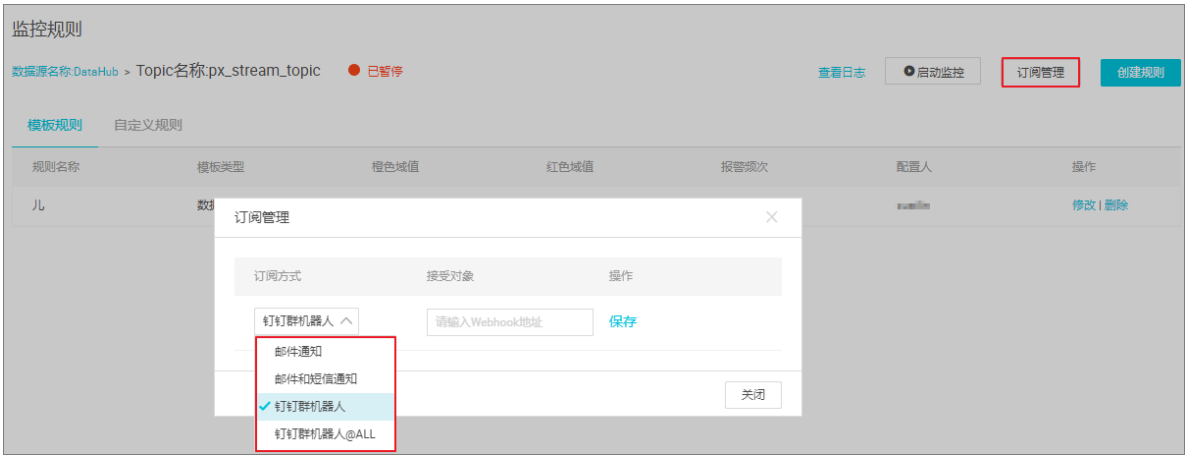
橙色阈值	以分钟为单位，仅支持输入整数，且必须小于红色阈值。
红色阈值	以分钟为单位，仅支持输入整数，且必须大于橙色阈值。
最小告警间隔	允许告警的最小时间差，以分钟为单位。
附加文本	对当前自定义Topic的描述。

2. 配置完成后，单击保存，将创建的规则添加到Topic中。

更多操作

- 单击查看日志，查看当前规则的运行日志。

- 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四种方式。

钉钉群机器人：先添加一个机器人到钉钉群中，通过详情页获取Webhook，然后将Webhook地址复制到订阅管理中，即可添加成功。

2.3.2 MaxCompute监控

规则配置模块是数据质量（DQC）的核心，目前数据质量支持离线的MaxCompute监控和DataHub监控，本文将为您介绍如何配置离线MaxCompute监控。

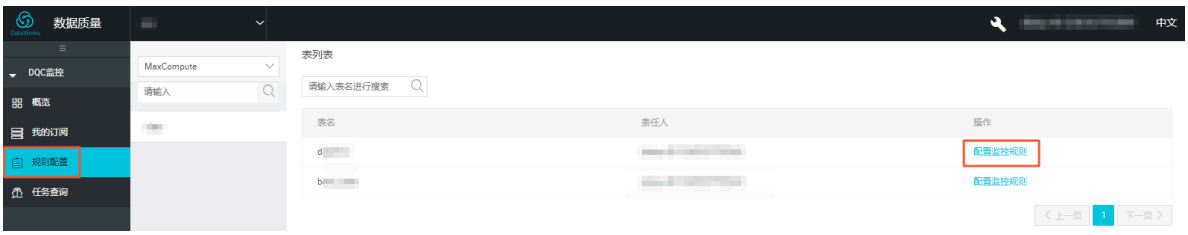
添加数据源

您需要首先进入数据集成页面添加数据源，详情请参见配置MaxCompute数据源。新建成功后，即可进入数据质量（DQC）页面进行规则的配置。

选择数据源

1. 单击左侧导航栏的规则配置，进入规则配置页面。
2. 选择MaxCompute，即可显示当前数据源下所有的表。

   您也可以输入对象表名（支持表名首字母模糊搜索），找到对应的表。
3. 单击右侧的配置监控规则。

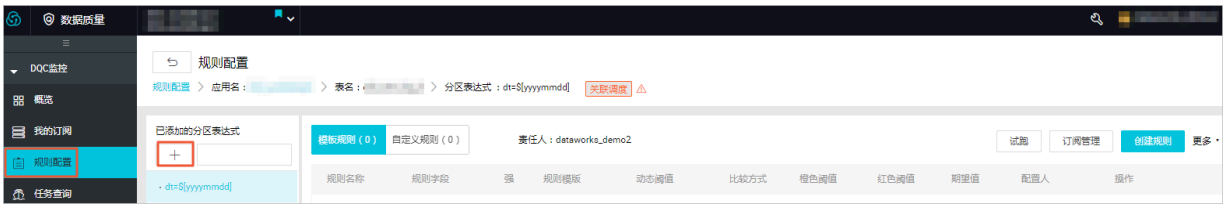


配置分区表达式

数据质量用分区表达式来确定需要配置哪条规则。

> **说明:**
>
> - 如果您的检查对象为非分区表，则此处可填写为NOTAPARTITIONTABLE 。
> - 如果您的表为分区表，则可以配置为业务日期的表达式（如$[yyyymmdd]），同时也可以配置为正则表达式。

进入数据表的规则配置页面，单击左上角的+，添加分区表达式。



- 新建分区的表达式：单击左上角的+，会弹出分区配置窗口，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NOTAPARTITIONTABLE。
  - 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]。

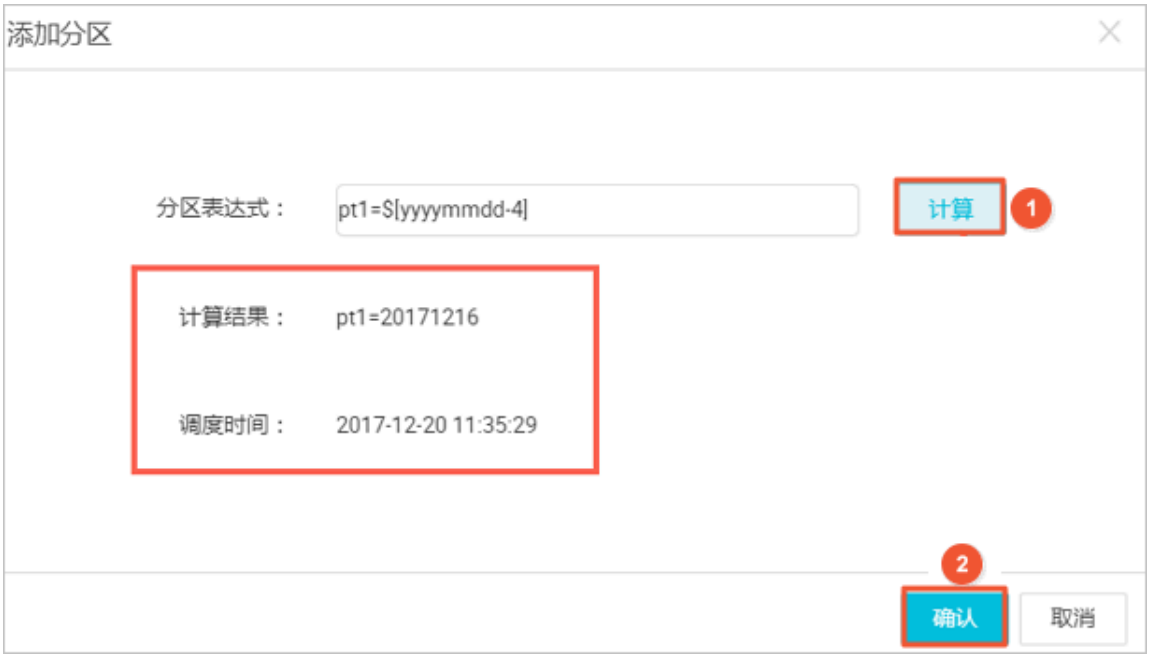
  分区表达式周期由配置的业务日期决定，例如配置运行时间为前5天，则周期为每5天调度一次。支持的分区表达式如下表所示。

分区表达式	说明
dt=$[yyyymmdd-N]	代表前N天。
dt=$[yyyymm01-1]	代表每月1日。
dt=$[yyyymm01-Nm]	代表N月前1日。
dt=$[yyyymmld-1]	代表每月最后一天。
dt=$[yyyymmld-1m]	代表N月前最后一天。
dt=$[hh24miss-1/24]	代表一个小时前。
dt=$[hh24miss-30/24/60]	代表半个小时前。
$[yyyymmdd]	调度日期。
$[yyyymmdd-1]	格式为yyyymmddmiss-1，默认为当前实例运行的业务日期的前一天。

分区表达式	说明
$[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
NOTAPARTITIONTABLE	非分区表可以选择此分区表达式。

- 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含有正则的分区表达式。

  1. 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，则会自动同步到输出窗口。
     - 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。
  2. 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。



  3. 单击确认。

- 删除已添加分区表达式：您可以删除不需要的分区表达式。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。

关联调度

如果要在生产链路上监控离线数据质量，需要将数据质量关联调度。

> 说明:
>
> · 关联界面仅能找到已经提交的节点，且关联调度支持跨工作空间的关联。
>
> · 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。

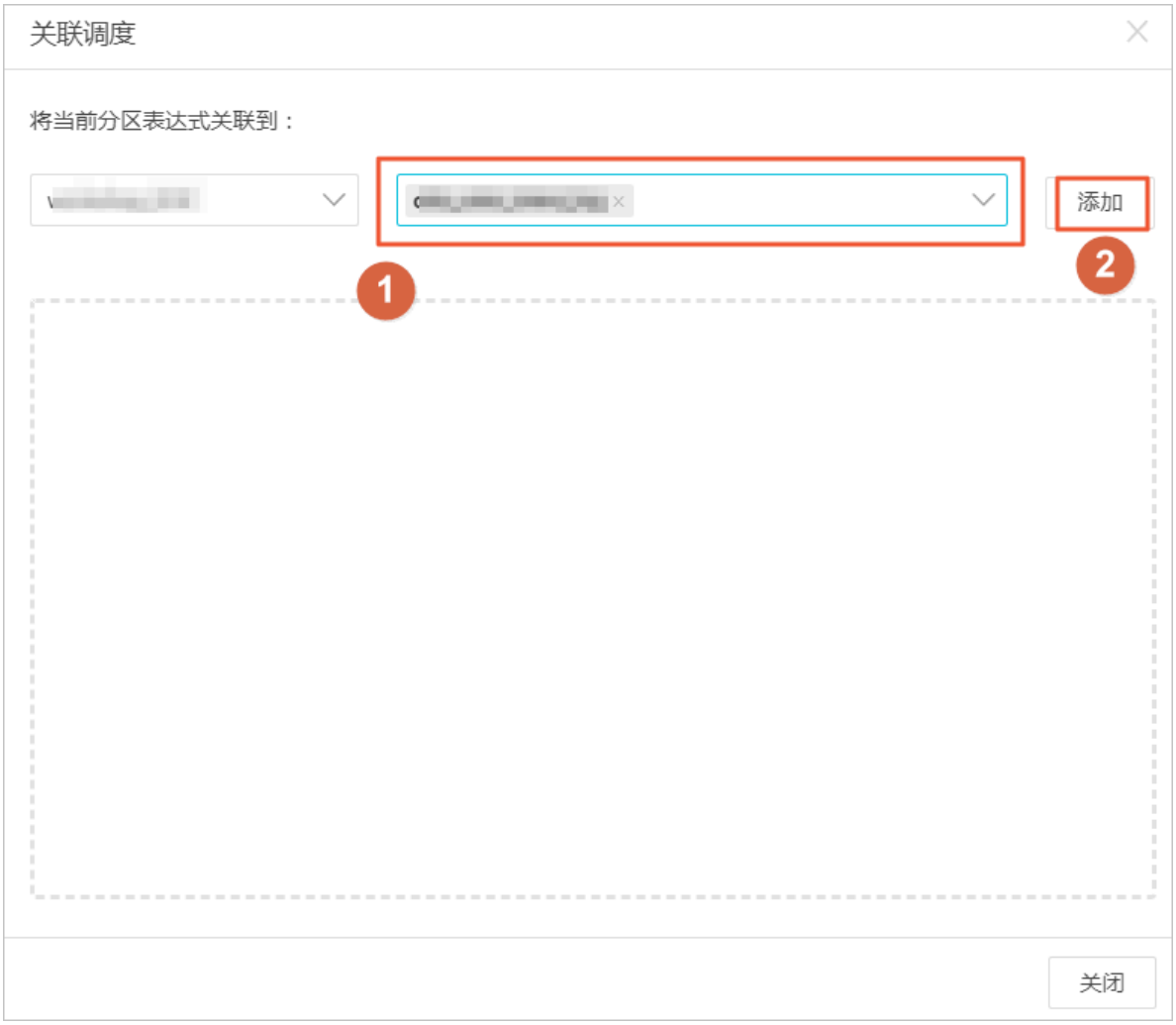
数据质量的关联调度可以关联单个或多个节点任务，关联调度完成后，离线数据质量监控任务可以自动运行。

> 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要与您的表有关系。

关联配置的操作步骤如下：

1. 在规则配置页面，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，单击添加。



3. 关联调度后，表名后的关联调度图标会变为蓝色。

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，详情请参见规则配置。

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	配置	说明
添加监控规则	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	系统内置的表级监控规则模块。

添加方式	配置	说明
	比较方式	比较方式包括绝对值、上升和下降三种类型。
	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。
	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。
	监控字段	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，快捷规则仅支持表行数大于0。 · 选择字段级规则，快捷规则可以选择字段重复值或字段空值。

试跑

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，可以测试规则配置的正确性、测试订阅发送渠道，它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。

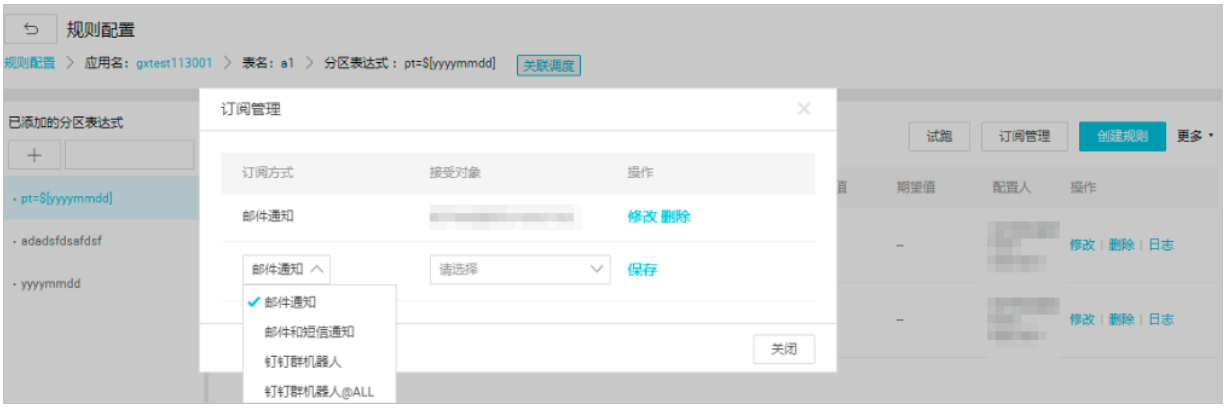
1. 选择需要试跑的调度日期，单击试跑，即可进行试跑。

配置	说明
试跑分区	实际分区会随着业务日期变化而改变。如果为NOPARTITIONTABLE，则会自动添加实际分区。
调度时间	默认为当前时间。

2. 单击试跑成功！点击查看试跑结果，即可跳转至任务查询页面，查看校验结果。

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。目前支持邮件通知、邮件和短信通知、钉钉群机器人和钉钉群机器人@ALL。



转交责任人

当责任人离职或者转岗，可以将分区表达式负责人转交给其它项目成员。默认分区表达式负责人为创建人。

鼠标悬停至责任人上，会显示一个隐藏按钮，单击即可修改责任人。在对话框中输入交接人的名称，单击确认，即可提交成功。

更多

更多选项中包括分区操作日志、上一次校验结果和复制规则。

操作	说明
分区操作日志	显示对当前分区表达式所有的规则设置的记录。
上一次校验结果	跳转至任务查询页面，查看当前分区表达式下的运行结果情况，您也可以在此查看历史结果。
复制规则	可以复制当前设置的规则至目标表达式中，也可以同步订阅人。

3 数据保护伞

3.1 进入数据保护伞

本文将为您介绍如何进入、授权并开通数据保护伞功能。

进入引导页面

登录DataWorks控制台，单击左上角导航栏中的数据保护伞，即可进入数据保护伞页面。

如果您是首次登录数据保护伞，会出现引导页面为您介绍数据保护伞的核心功能及使用流程，帮助您对数据保护伞进行初步了解。

单击立即体验即可进入数据保护伞授权页面。

说明:

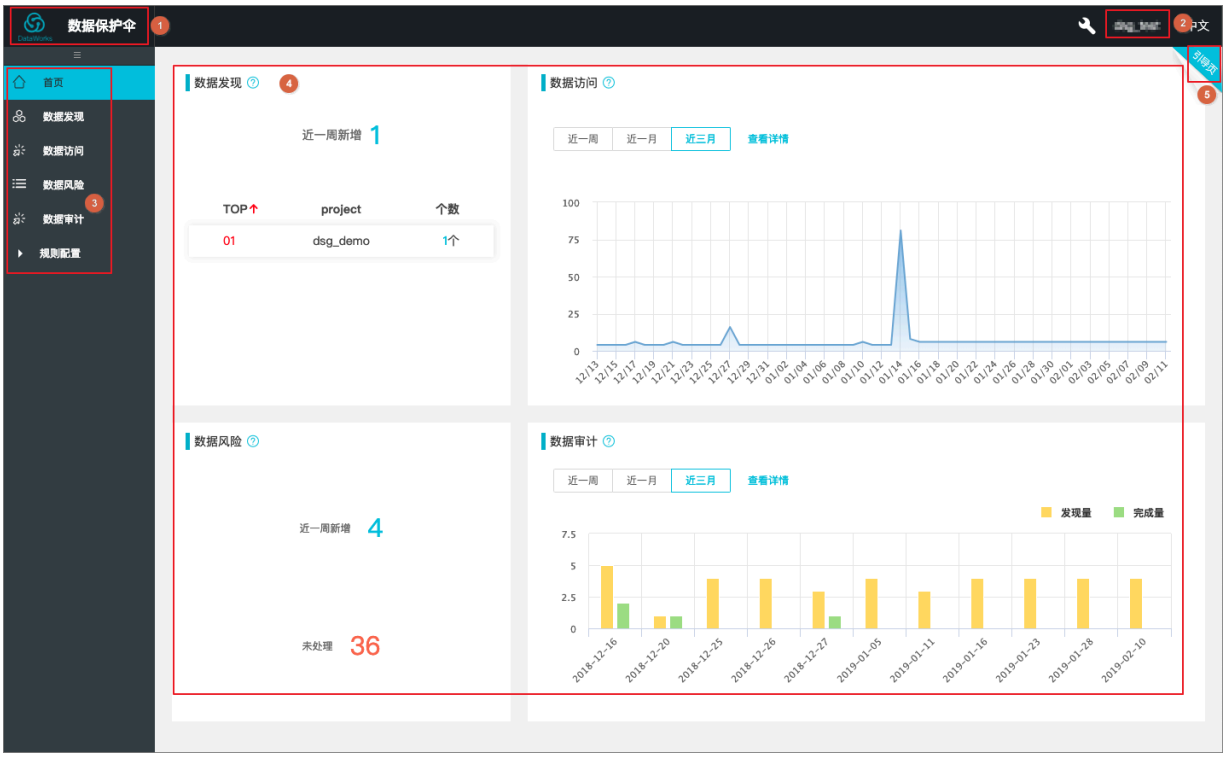
如果租户管理员已授权，则直接进入数据保护伞首页。

进入授权页面

只有租户管理员才能进行授权，开通数据保护伞。

登录数据保护伞

登录数据保护伞平台，首页如下所示：



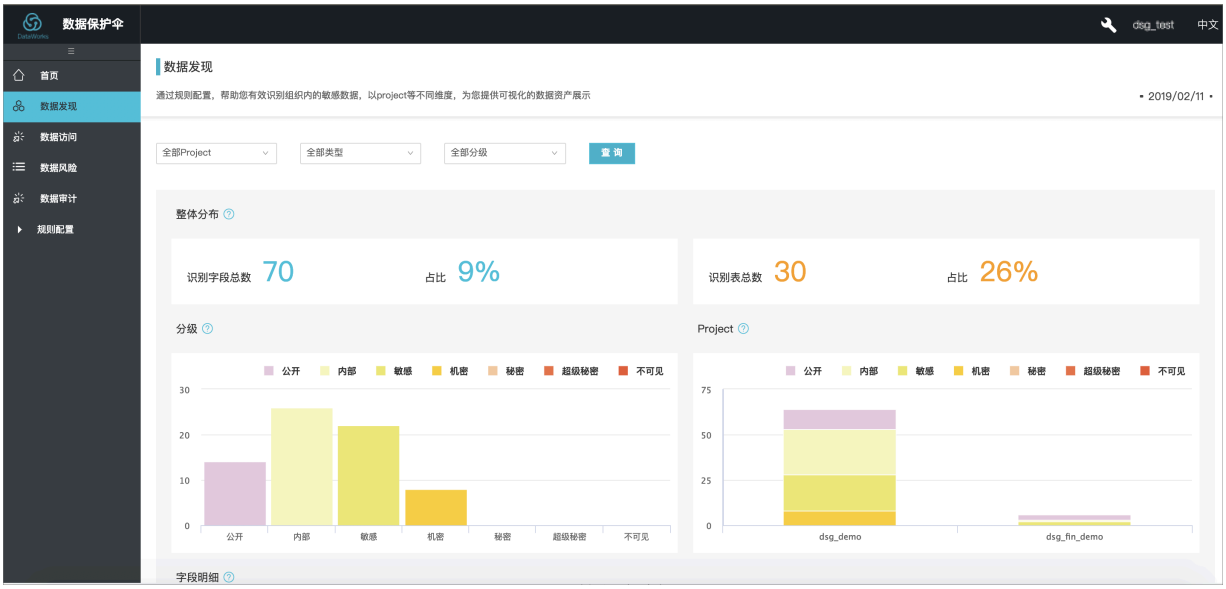
序号	名称	说明
1	功能菜单栏	当前用户有权可见的功能模块，包括数据开发、数据质量、数据集成、数据服务、运维中心、数据管理和数据保护伞。

序号	名称	说明
2	用户信息	当前登录用户，可查看并编辑用户信息，包括邮箱、手机、AccessKeyID和AccessKeySecret。
3	导航栏	对应功能菜单的导航栏，不同的功能模块对应不同的左侧导航栏。
4	数据保护伞首页	· 该租户近一周新识别数据。 · 所有识别数据近一周、近一月、近三月的访问趋势。 · 近一周新增数据风险量。 · 所有风险近一周、近一月、近三月的发现量以及完成量。
5	引导页切换	单击引导页即可切换到引导页面，查看产品的信息。

3.2 数据发现

数据安全管理员在完成敏感数据规则配置T+1后，即可在识别数据分布中查看数据分布情况。

数据分布分为整体分布、按等级分布和字段明细，您可根据自己的查询需要，按照Project、规则名、规则类型、风险等级（即分级）等进行过滤选择。



字段明细

Project	规则名(全部)	类型(全部)	分级(全部)	表数量	字段数量	操作
dsg_demo	姓名	姓名	内部	8	14	明细
dsg_demo	邮箱	邮箱	敏感	7	11	明细
dsg_demo	手机号	手机号	内部	8	10	明细
dsg_demo	地址	地址	敏感	6	8	明细
dsg_demo	身份证号	身份证号	机密	4	6	明细
dsg_demo	mac地址	mac地址	公开	4	4	明细
dsg_demo	车牌号	车牌号	公开	3	3	明细
dsg_demo	IP	IP	公开	2	2	明细
dsg_fin_demo	邮箱	邮箱	敏感	1	2	明细
dsg_demo	交易金额	交易金额	机密	2	2	明细

1 2

3.3 数据访问

本文将为您介绍数据访问的访问行为和导出行为。

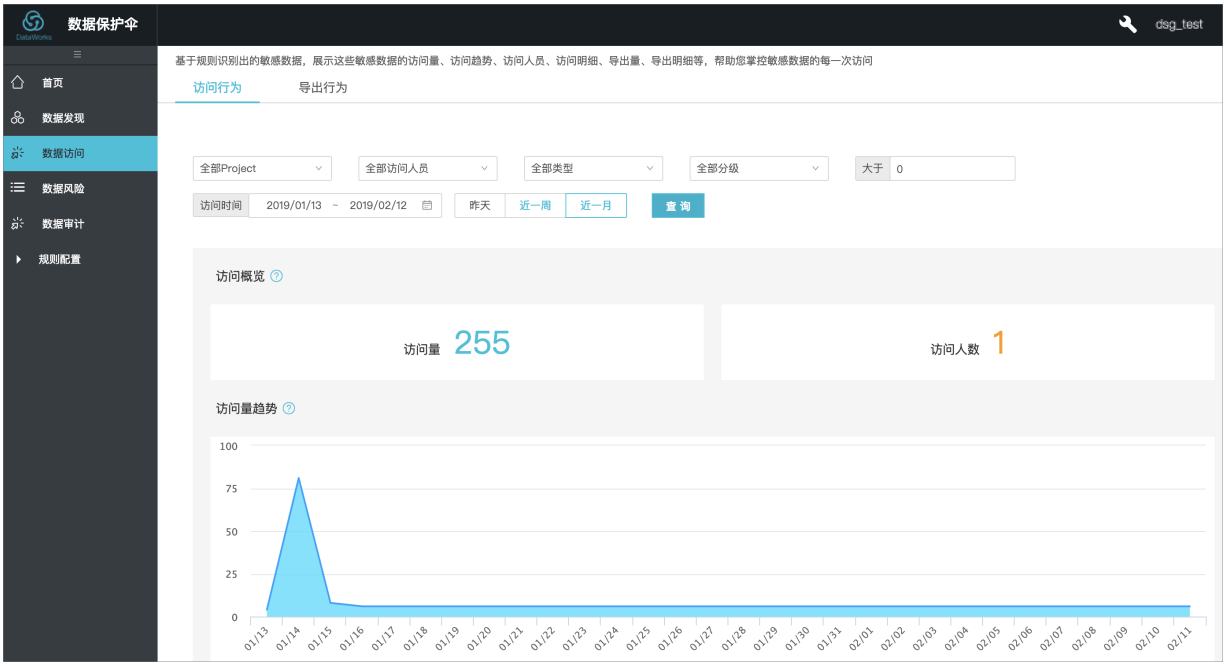
数据访问包括访问行为和导出行为两个模块。

- 访问行为：包含Create、Insert操作，但不包括访问失败的行为。
- 导出行为：数据从MaxCompute导出的行为。

访问行为

数据安全管理员在完成敏感数据规则配置T+1后，即可在数据访问行为中查看数据使用情况，包括访问概览、访问趋势以及访问明细。

您可根据自己的查询需要，按照Project、规则名、规则类型、风险等级（即分级）、访问人员等进行过滤选择。



访问记录 2019-01-13 ~ 2019-02-12

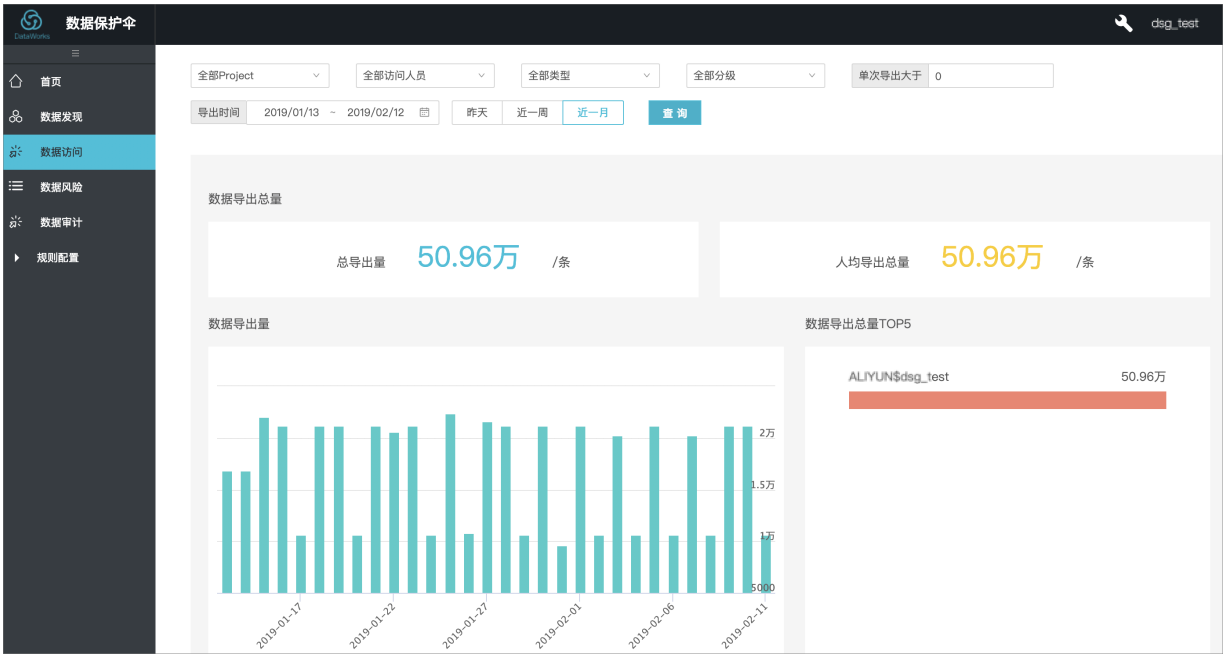
Project	规则名(全部)	规则类型(全部)	分级(全部)	访问人员(全部)	访问次数	操作
dsg_demo	手机号	手机号	内部	ALIYUN$dsg_test	142	明细
dsg_demo	邮箱	邮箱	敏感	ALIYUN$dsg_test	86	明细
dsg_demo	姓名	姓名	内部	ALIYUN$dsg_test	10	明细
dsg_demo	供应ID	供应ID	公开	ALIYUN$dsg_test	9	明细
dsg_demo	mac地址	mac地址	公开	ALIYUN$dsg_test	8	明细

1

风险识别_数据导出

数据安全管理员在完成敏感数据规则配置T+1后，即可在数据导出中查看用户从MaxCompute中把数据导出到外部的情况，包括数据导出总量、TOP导出用户和导出明细。

您可根据自己的查询需要，按照规则名、规则类型、导出量等进行过滤选择。

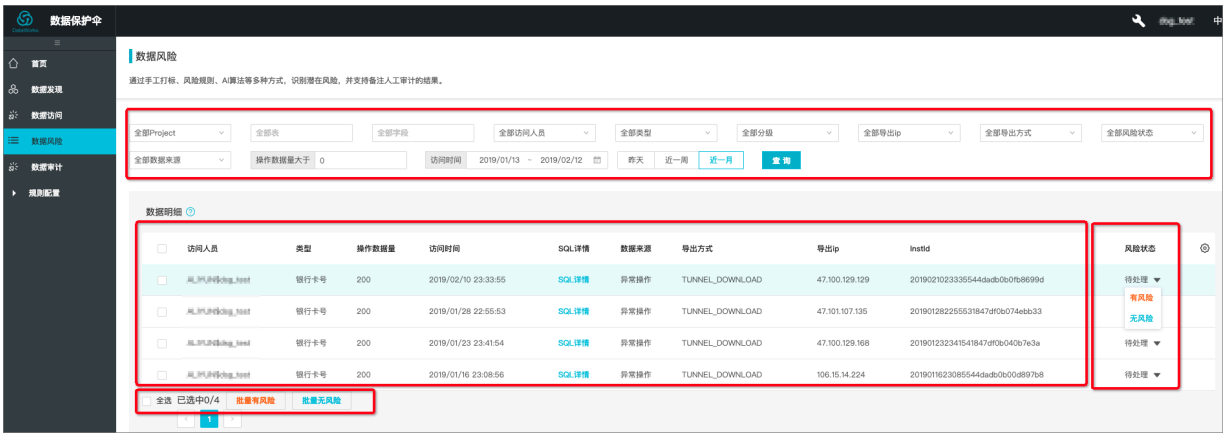


数据导出明细

导出人(全部)	导出账号(全部)	导出ip(全部)	IP所在地(全部)	导出量	导出方式	导出时间	操作
ALIYUN$dsg_test	ALIYUN$dsg_test	47.100.129.87	数据保护伞扫描任务	200	Tunnel下载	2019/02/11 15:19:55	明细
ALIYUN$dsg_test	ALIYUN$dsg_test	47.100.129.87	数据保护伞扫描任务	200	Tunnel下载	2019/02/11 15:19:54	明细
ALIYUN$dsg_test	ALIYUN$dsg_test	47.100.129.87	数据保护伞扫描任务	200	Tunnel下载	2019/02/11 15:19:54	明细
ALIYUN$dsg_test	ALIYUN$dsg_test	47.100.129.87	数据保护伞扫描任务	200	Tunnel下载	2019/02/11 15:19:54	明细
ALIYUN$dsg_test	ALIYUN$dsg_test	47.100.129.87	数据保护伞扫描任务	200	Tunnel下载	2019/02/11 15:19:54	明细
ALIYUN$dsg_test	ALIYUN$dsg_test	47.100.129.87	数据保护伞扫描任务	200	Tunnel下载	2019/02/11 15:19:54	明细
ALIYUN$dsg_test	ALIYUN$dsg_test	47.100.129.87	数据保护伞扫描任务	200	Tunnel下载	2019/02/11 15:19:53	明细
ALIYUN$dsg_test	ALIYUN$dsg_test	47.100.129.87	数据保护伞扫描任务	200	Tunnel下载	2019/02/11 15:19:53	明细

3.4 数据风险

数据风险页面提供通过手工风险数据识别、风险识别管理（风险规则配置识别、AI识别）产生的风险数据清单，同时可对这些风险数据进行审计备注。

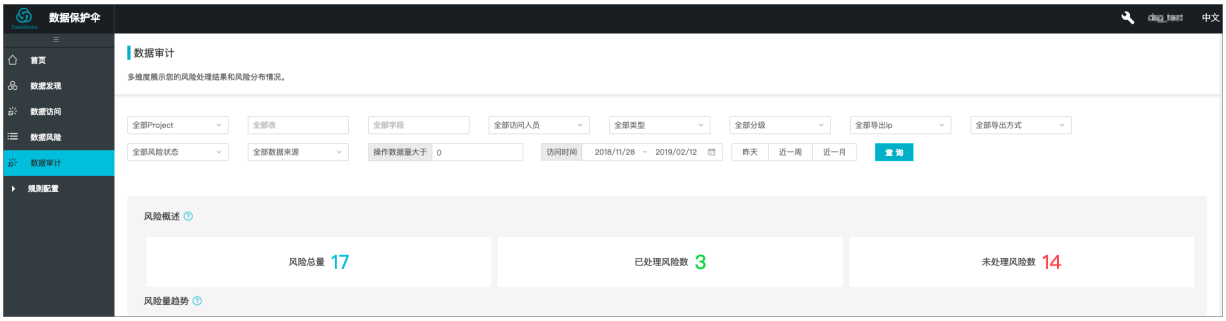


页面说明如下：

- 查询风险数据条件：可供筛选的条件包括Project、表名、字段、访问人员、规则类型、规则名、分级、导出IP、导出风险、风险状态和风险数据类型。
- 风险数据明细：可根据查看指标的需要在标题栏处的设置按钮中选择审计备注，对风险数据进行有/无风险标注，支持增加标签、增加详细备注信息。
- 批量审计处理：分为批量有/无风险标注以及详细信息备注。

3.5 数据审计

数据审计页面是数据风险统计信息的汇总，包括风险数据概览、每日风险趋势和风险维度分析。



3.6 规则配置

本文将为您介绍如何定义敏感数据，新建规则并进行配置。

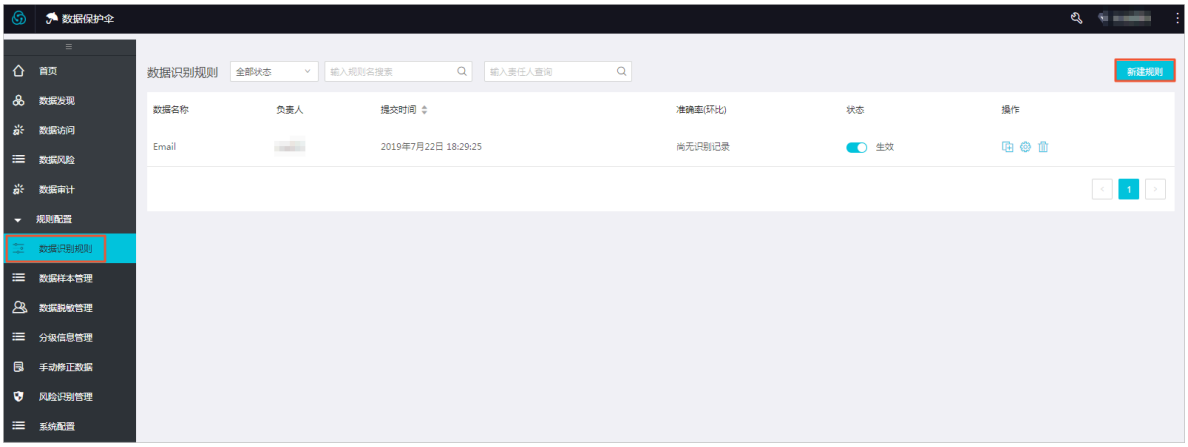
定义敏感数据

数据安全管理员的操作步骤如下所示：

1. 单击左上角的图标，选择全部产品 > 数据保护伞。

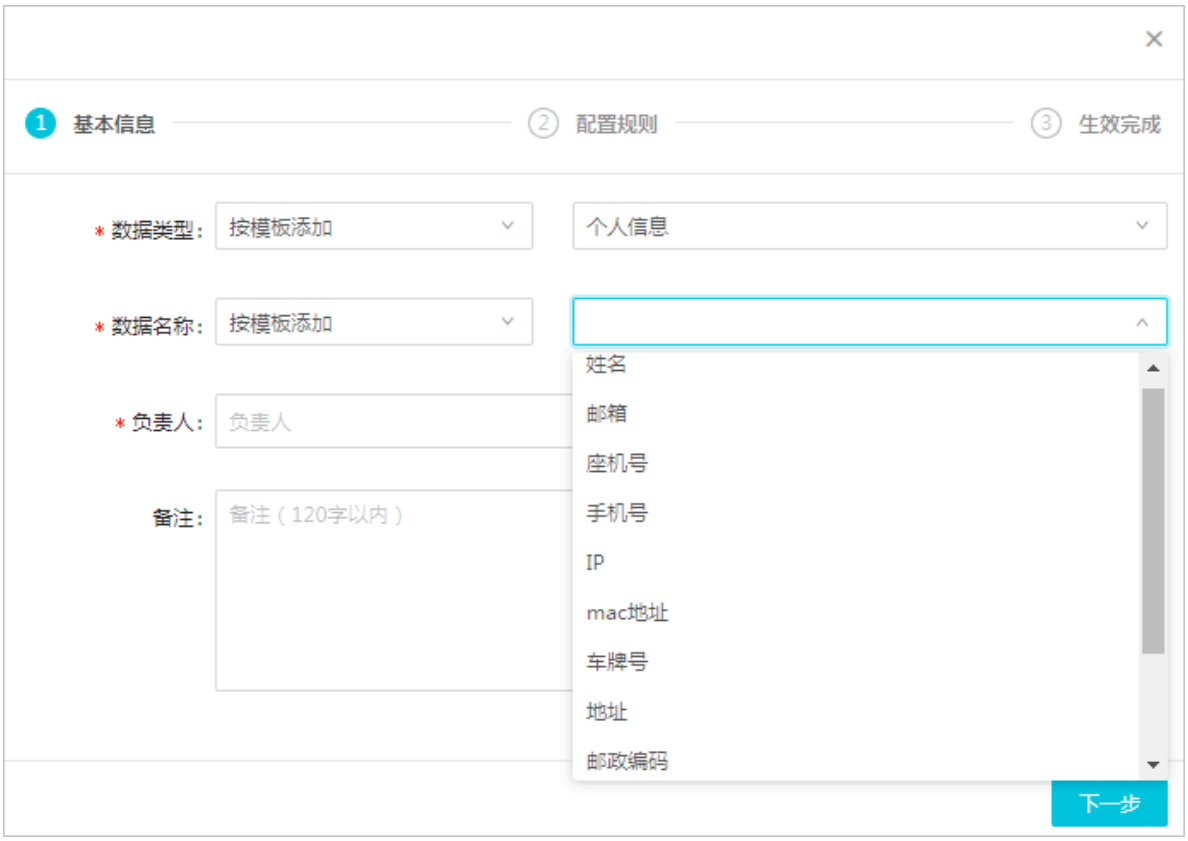
2. 单击数据保护伞产品页面的立即体验，即可进入数据保护伞控制台。

3. 导航至规则配置 > 识别数据规则，单击新建规则。



4. 填写对话框中的基本信息，单击下一步。

您可以通过按模板添加和自定义添加两种方式新建规则。



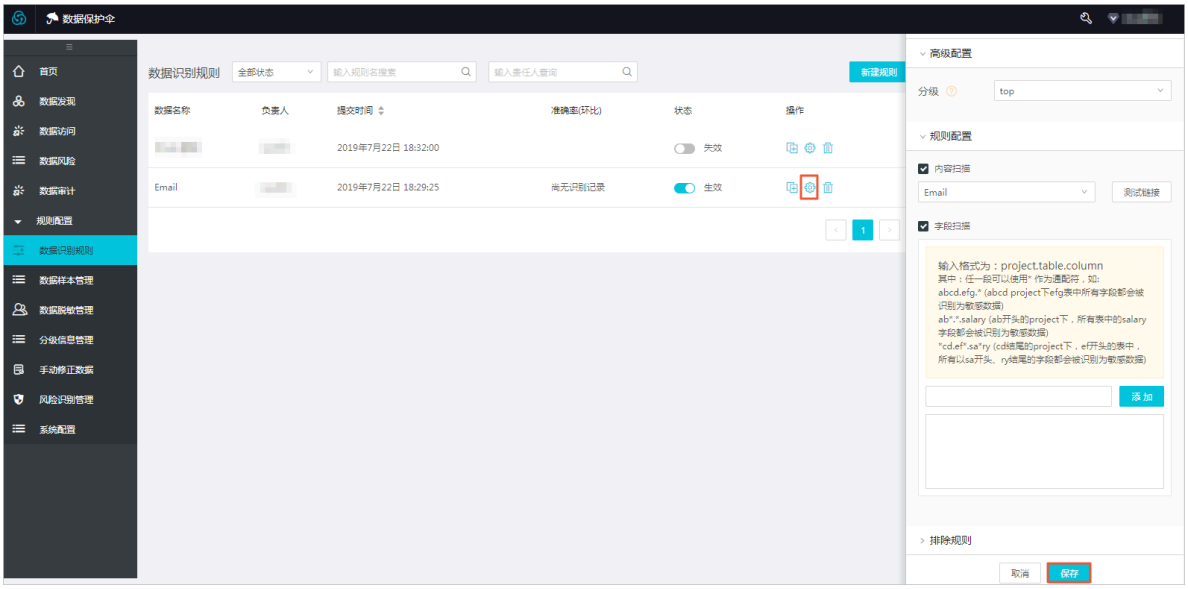
配置	说明
数据类型	即规则所属分类，支持按模板添加或自定义添加。 · 如果选择按模板添加，可以选择个人信息、商户信息和公司信息。 · 如果选择自定义添加，您可以自行填写数据类型。
数据名称	· 如果选择按模板添加，系统内置姓名、邮箱、座机号、手机号、IP、mac地址、车牌号、地址、邮政编码、身份证号、银行卡号和公司名12种敏感数据识别定义模板。 · 如果选择自定义添加，您可以自行填写数据名称。
责任人	规则设置人员信息。
备注	对当前规则进行简单描述。

5. 单击对话框中的配置规则，单击下一步。

配置	说明
分级	对配置的数据进行等级划分，如果现有等级不满足需求，请在分级信息管理处进行设置。
内容扫描	提供的数据识别方式之一，系统内部12种数据识别模板皆为内容扫描。 · 如果选择模板则无法更改识别规则，但提供规则准确性验证的通道，同时对识别不准的情况可进行手动修正。 · 如果选择正则式匹配，则自定义识别规则。
字段扫描	提供字段名精确匹配和模糊匹配方式，支持多个字段匹配，各字段间为或关系。

6. 确认配置无误后，单击保存并生效。

如果需要修改已有规则，需要将该规则置为失效状态，然后单击操作下的规则配置，在右侧对话框进行高级信息的配置和修改。



规则配置修改完成后，单击保存。确认规则无误后，更改状态为生效。

说明:

定义敏感数据时，需遵循如下规则：

- 规则名必须唯一。

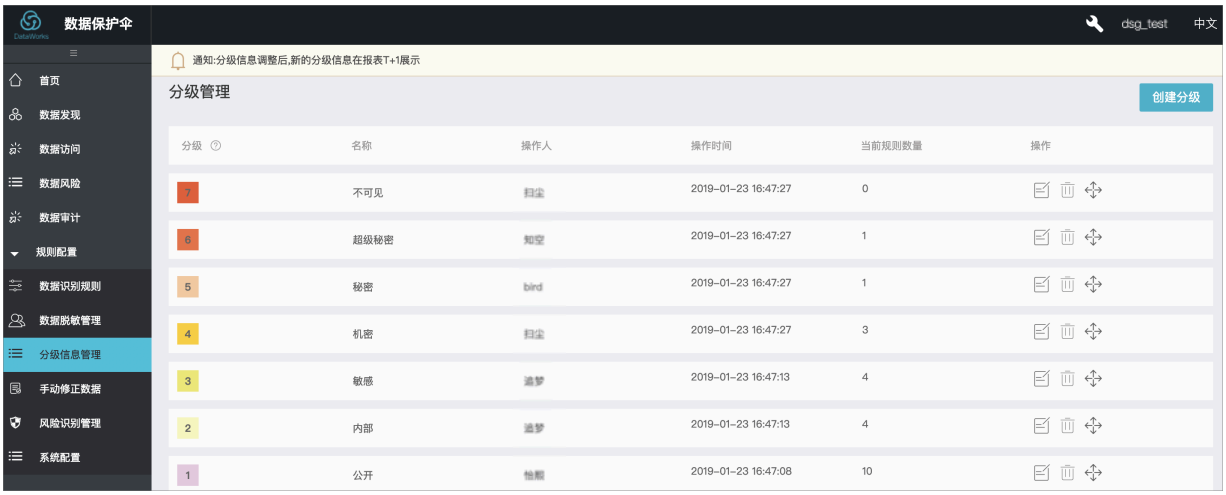
- 不同规则的内容扫描或者字段扫描必须唯一。
- 规则识别数据，T+1在报表展现。

已定义敏感数据

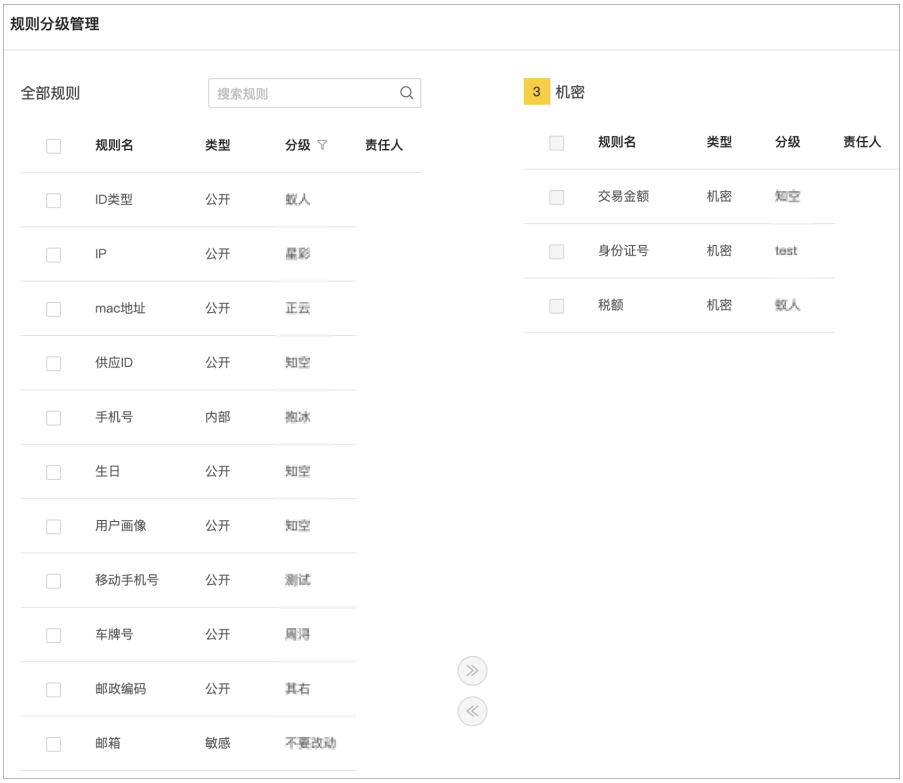
如果您已经定义敏感数据，请直接跳转至数据发现、数据访问和数据风险模块进行查看。

3.7 分级信息管理

当规则配置中的分级选择无法满足您的需求时，可在分级页面管理中进行设置，该页面提供新建分级、删除分级、分级优先级调整和规则分级调整的功能。

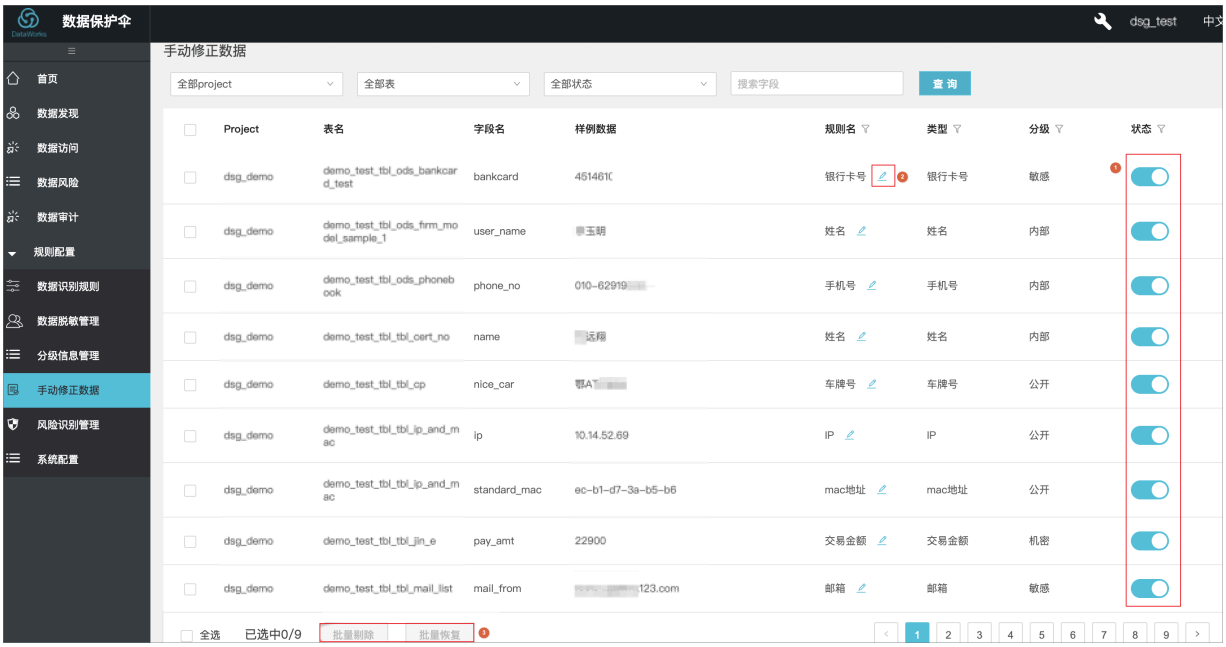


配置	说明
创建分级	单击创建分级增加新的分级，填写名称和操作人。

配置	说明
	调整规则分级，单击后可进行规则的选择与调整。 
	调整分级优先级，单击后向下（降低优先级）或者向上拖动（提高优先级）。
	删除分级，单击后可删除不需要的分级。

3.8 手动修正数据

手动修正页面提供对规则识别的敏感数据不准确的情况进行手动修正的功能，包括剔除识别错误数据、更改识别数据类型和批量处理。



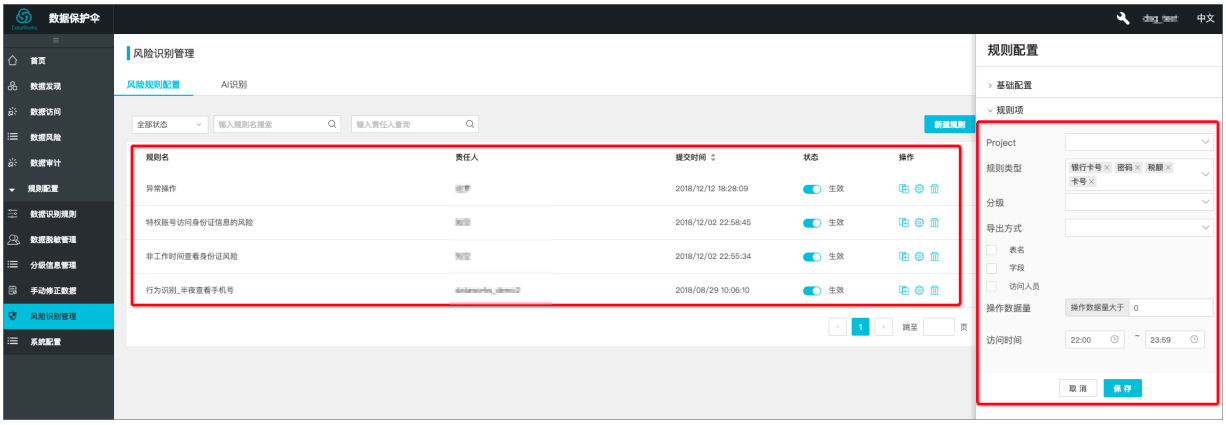
操作	说明
剔除识别错误数据	滑动状态一栏下的按钮，即可更改为已剔除状态，已剔除的数据可恢复。
	更改识别数据类型。如果识别为邮箱，实际为车牌号，则单击邮箱右边的 进行更改，只能选择已配置的规则名称。
批量处理	包括批量剔除和批量恢复，选择需操作的数据，单击数据左方的复选框，再单击相应的操作。

说明:

手动修正数据时需要遵循剔出以及更改数据名称类型均T+1生效于识别数据分布、数据访问行为、数据导出页面的规则。

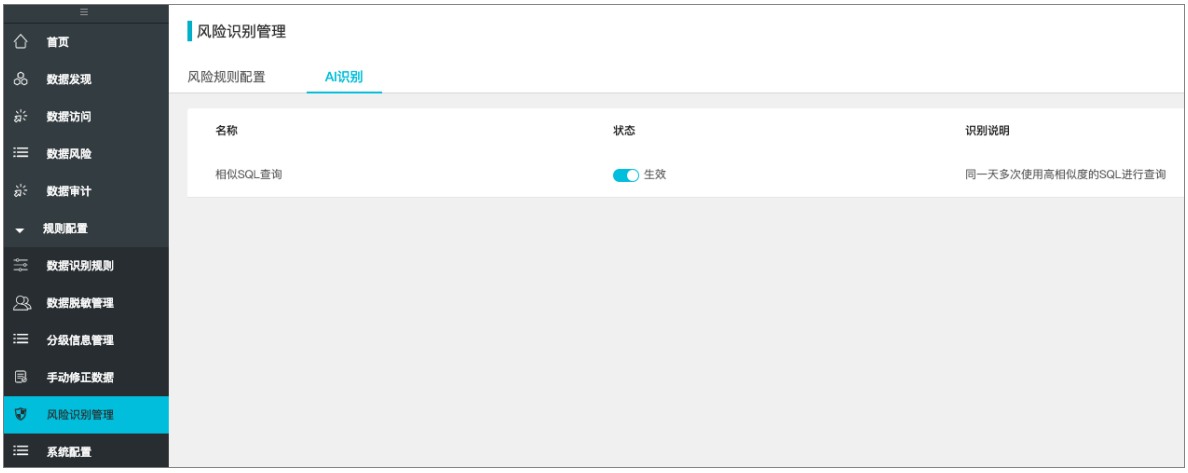
3.9 风险识别管理

风险识别管理页面提供风险数据规则配置，您可以识别日常访问中的风险以及启动AI识别自动识别数据风险，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，同时也会在数据访问页面处的数据打上识别标志。



页面说明如下：

- 风险识别管理：分为风险规则配置和AI识别。AI识别的页面包括个人信息查询、相似SQL查询以及这两块内容的识别介绍，您只需在状态列启动即可，同时启动后也可关闭（不删除之前识别的数据）。



- 风险规则配置_新建规则：在弹出框中录入规则名称、责任人和规则备注信息后，规则基本信息创建完成。

- 风险规则配置_操作：提供复制规则、编辑风险规则项和删除规则功能。
- 风险规则配置_规则项配置：提供Project（支持多选）、类型（支持多选）、规则（支持多选）、分级（支持多选）、导出方式（支持多选）、表（支持模糊/精确匹配）、字段（支持模糊/精确匹配）、访问人员（支持模糊/精确匹配）、操作数据量和访问时间条件配置。
- 风险规则配置_状态：您在配置完规则后，需在状态列启动规则后生效。

说明:
风险识别管理数据需要遵循规则配置以及AI识别启动后，数据均T+1生效于数据风险页面的规则。

3.10 设置并查询自定义脱敏

DataWorks目前支持动态脱敏，本文将为您介绍如何设置数据保护伞自定义脱敏，并在DataWorks中进行脱敏查询。

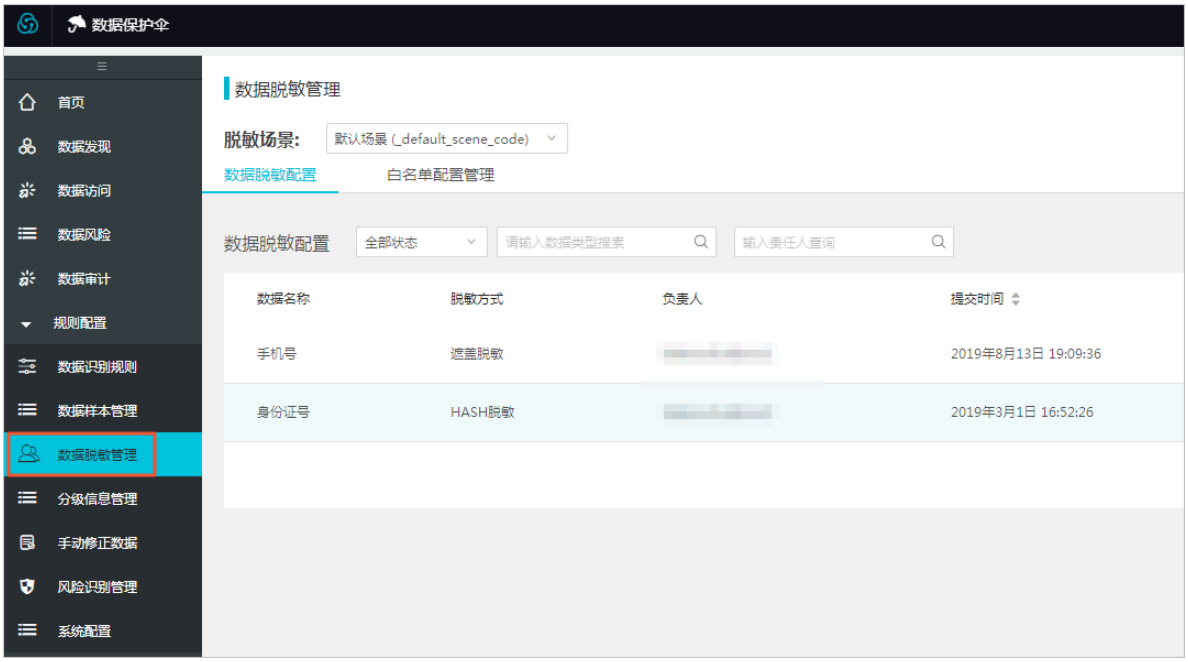
数据保护伞自定义脱敏设置

说明:
您需要首先开通数据保护伞服务，方可使用数据保护伞自定义脱敏。

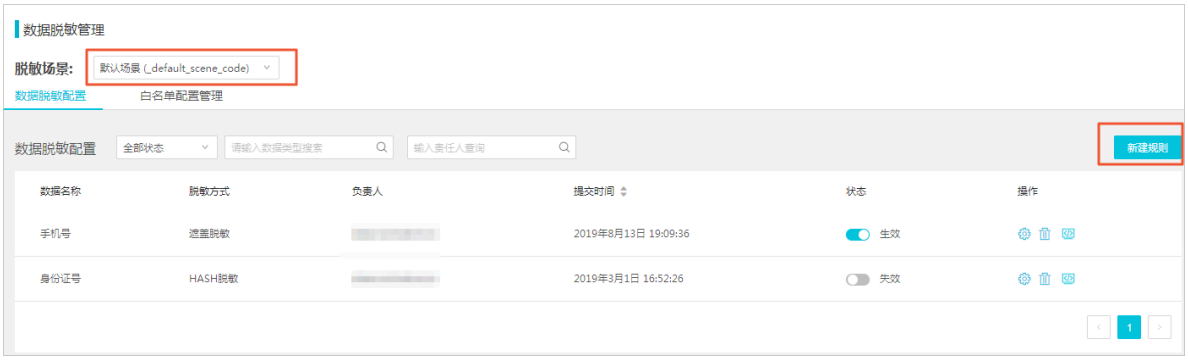
1. 登录DataWorks控制台，单击相应工作空间后的进入数据开发。
2. 单击左上角的DataWorks图标，选择全部产品 > 数据保护伞。

> **说明：**
> 如果是第一次进入数据保护伞平台，则会跳转至产品页面，单击立即体验，即可进入数据保护伞页面。

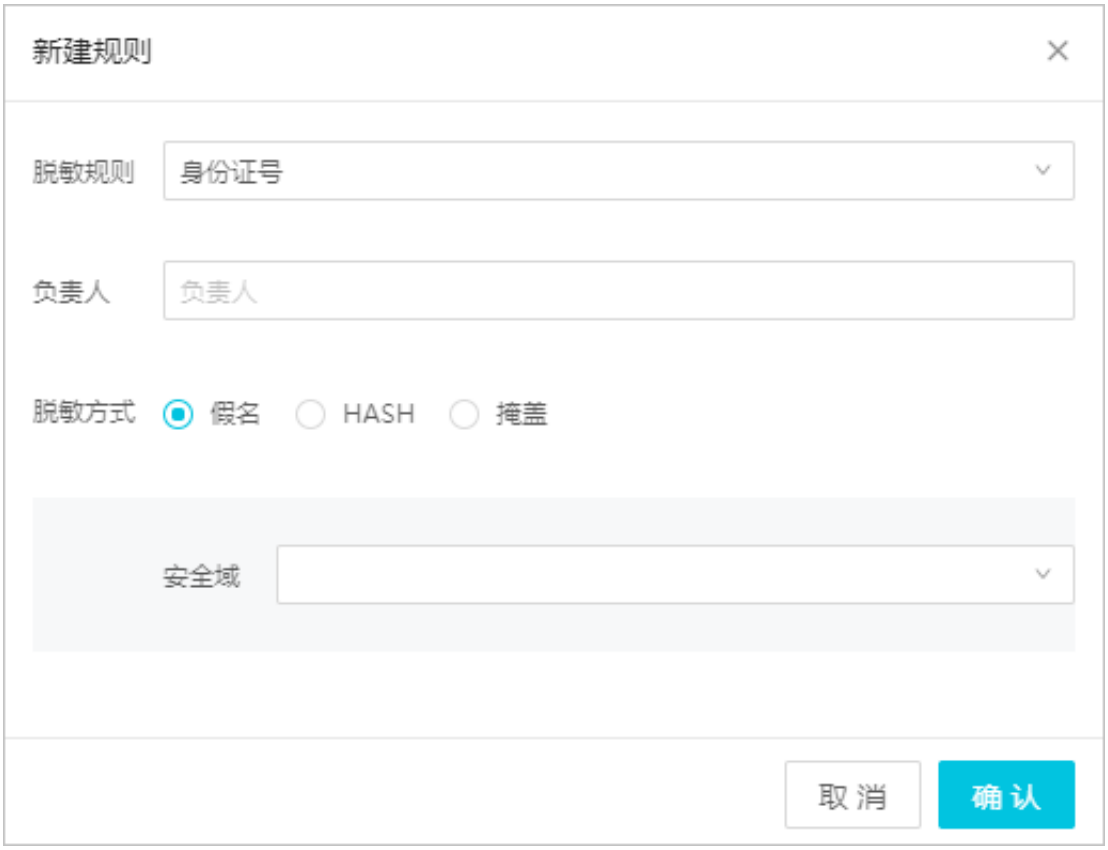
3. 选择左侧菜单栏中的规则配置 > 数据脱敏管理。



4. 在脱敏场景中选择默认场景（_default_scene_code），然后单击新建规则，添加需要进行自定义脱敏的规则。



5. 在新建规则对话框中，选择需要设置的脱敏规则、责任人和脱敏方式。



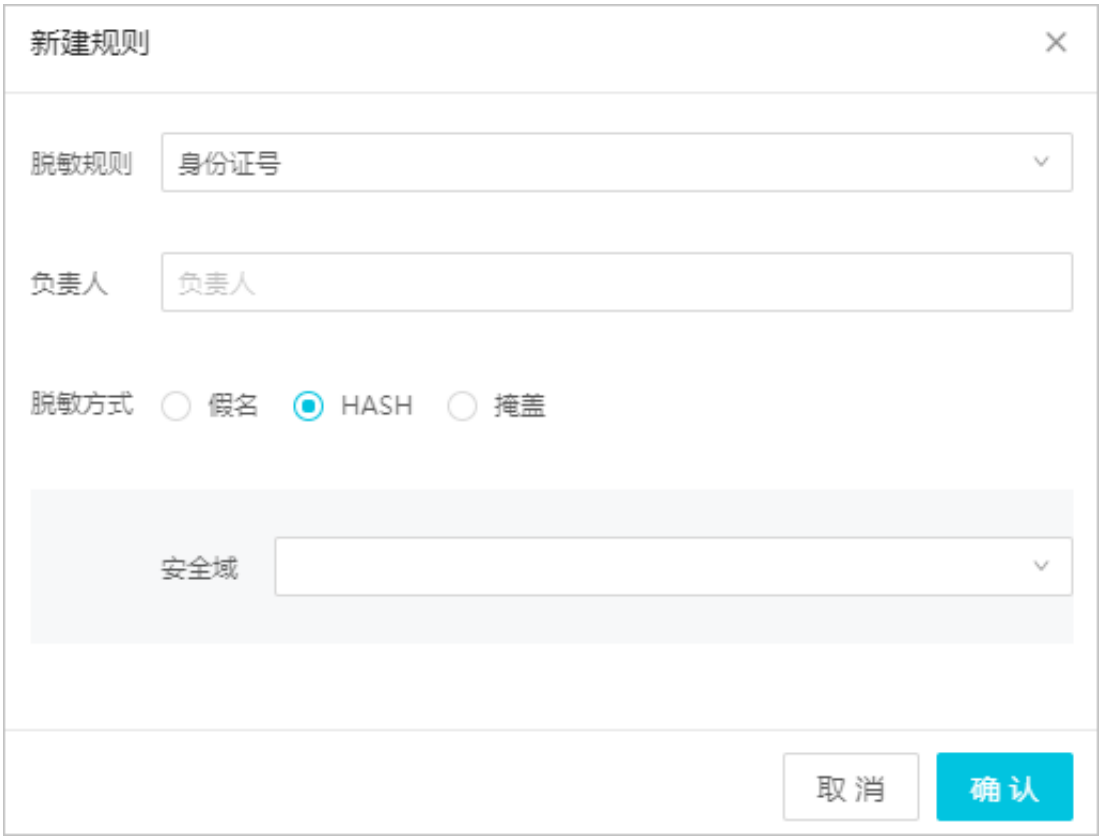
目前数据保护伞仅为身份证号脱敏规则提供假名、HASH和掩盖三种脱敏方式，其它脱敏规则仅提供HASH和掩盖两种脱敏方式。

- 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。使用假名脱敏时，需要选择安全域，相同的值在不同的安全域脱敏出来的假名信息不同。

- HASH

  HASH脱敏需要选择一个安全域，相同的值在不同的安全域HASH脱敏后的值不一致。



- **掩盖**

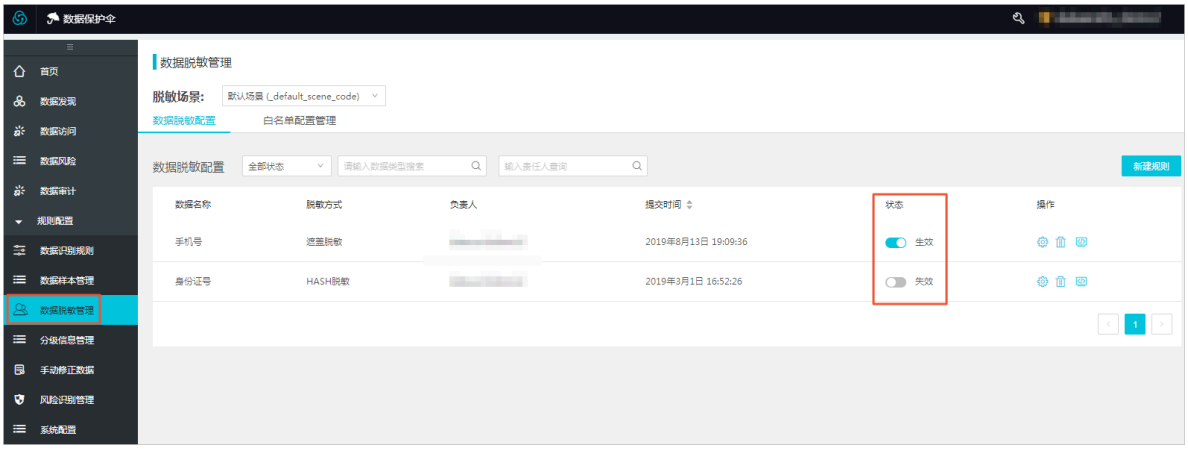
  掩盖脱敏是使用*对部分信息进行掩盖，达到脱敏的效果，是一种比较常用的脱敏方式。

配置	说明
推荐方式	为身份证、银行卡等常用的数据类型提供掩盖脱敏策略。

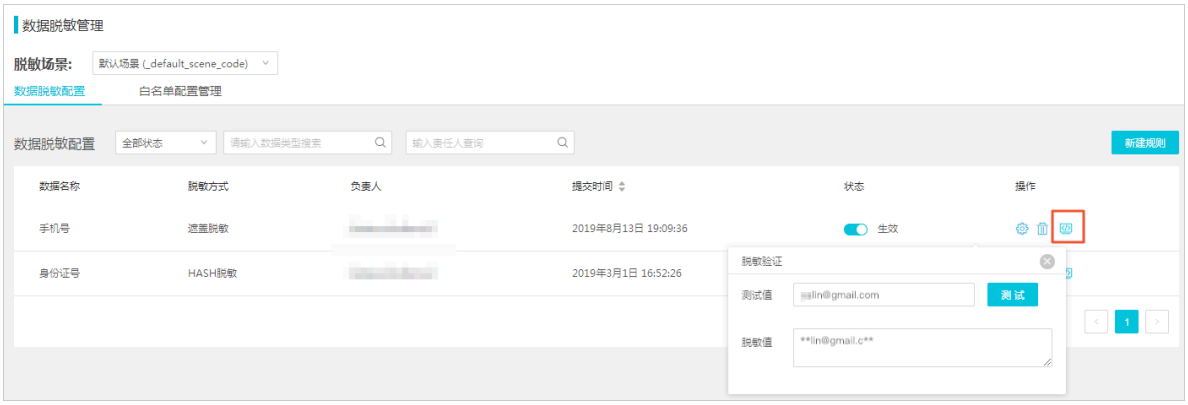
配置	说明
自定义	自定义设置提供了更加灵活的设置方式，可以在前中后三段设置是否脱敏，以及需要脱敏（或者不脱敏）的字符长度。

6. 设置成功后，单击确认，跳转至数据脱敏管理页面。

7. 在数据脱敏配置下，对脱敏策略进行生效或失效操作。



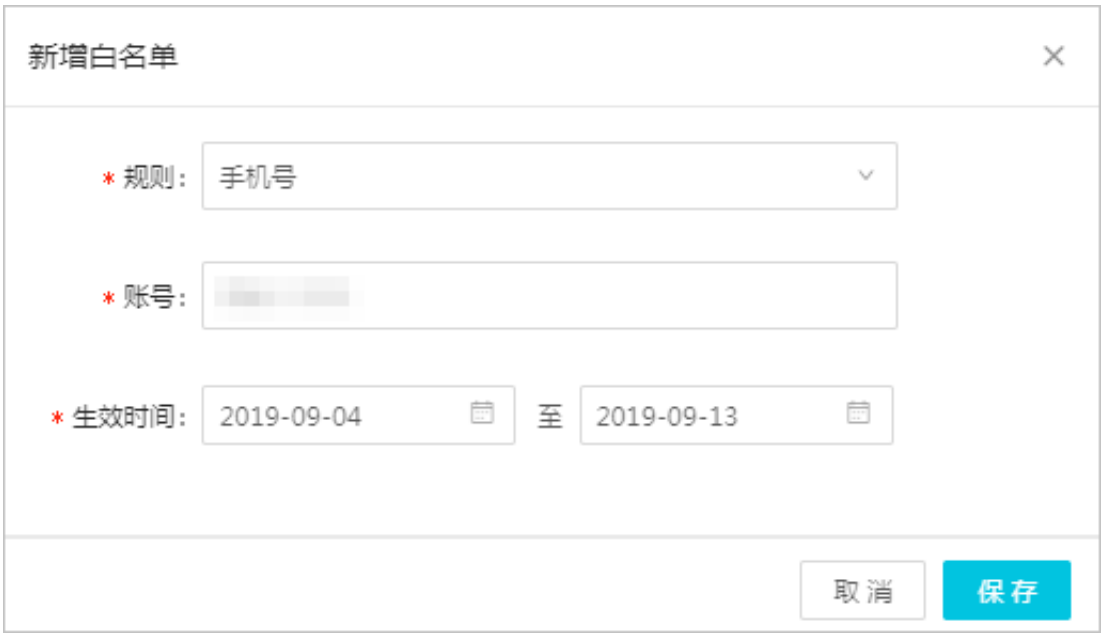
设置成功后，您可以单击相应脱敏规则后的 图标，输入测试值进行脱敏验证。



8. 进入白名单配置管理页面，单击新增白名单。

9. 在新增白名单对话框中，设置规则、账号和生效时间，并单击保存。

   此处的账号支持主账号和RAM子账号。



> 说明:
> 设置白名单生效时间后，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。

DataWorks中进行脱敏查询

成功新建脱敏规则并进行配置后，您可以在DataWorks中进行脱敏查询。

> 说明:
> 您需要首先开启DataWorks项目空间的查询脱敏功能，详情请参见#unique_40。

7
8
select '**lin.jyl@gmail.com'
运行日志
结果[1]
结果[2]
A
1
_c0
2
lin.jyl@gmail.c

4 数据地图

4.1 数据管理升级为数据地图

本文将为您及时同步数据管理升级到数据地图的进展和计划。

DataWorks数据地图模块第一批发布计划

- 发布版本：发布DataWorks数据地图模块，取代数据管理模块。
- 发布时间：
  - 2019年6月25日~6月28日20:00~20:20，发布数据地图至各区域。
  - 2019年7月1日20:00，下线所有区域的数据管理模块。
- 发布内容：DataWorks数据地图模块在DataWorks数据管理模块的基础上，根据角色区分对应的功能，控制数据预览权限、新建表权限等，帮助您更好地构建企业级数据信息知识库。
- 已发布：华东2（上海）、华北2（北京）、华东1（杭州）、华南1（深圳）、中国（香港）和亚太东南1（新加坡）。
- 待发布：
  - 2019年6月25日20:00~20:20：发布数据地图至亚太东南2（悉尼）、亚太东南3（吉隆坡）和亚太南部1（孟买）。
  - 2019年6月26日20:00~20:20：发布数据地图至亚太东南5（雅加达）、欧洲中部1（法兰克福）和英国（伦敦）。
  - 2019年6月27日20:00~20:20：发布数据地图至美国东部1（弗吉尼亚）、美国西部1（硅谷）和中东东部1（迪拜）。
  - 2019年7月1日20:00~20:20：下线所有区域的数据管理模块。

DataWorks数据地图与数据管理功能对比一览表

DataWorks数据地图相比于数据管理，在整体视觉交互和角色区分等方面提升使用体验，详情如下表所示。

对比项	数据管理	数据地图	改进效果
页面功能和角色关联	按功能类型组织，分为总览和分页面如下： · 全局概览 · 查找数据 · 数据表管理 · 权限管理 · 管理配置	· 按角色和功能的关系组织页面如下： - 数据总览 - 数据地图首页和全部类目：针对数据搜索。 - 我的数据： 针对个人的表管理、收藏与权限申请和审批。 - 配置管理：针对工作空间管理或类目管理。 · 专为使用实际数据的使用者强化搜索能力、提供按类目检索的能力。	· 用户群体中，人数占比最多的数据实际使用者，能够更方便快捷地查找自己想要的数据。 · 数据管理者，包括个人作为表负责人、工作空间管理者、类目管理者，根据角色选择页面，权责更加清晰，相关功能更易用。
个人操作更集中	表管理、收藏管理、权限管理、资源和函数的权限申请等分散在各处。	· 资源和函数的权限申请入口以及表、资源、函数的审批入口，全部归属到我的数据页面。 · 个人对表的操作入口、表收藏功能等，也移动到我的数据页面。	个人对数据的操作更加集中，方便查找和使用。
表的预览权限收拢	任意用户可以看到表的数据预览，表的负责人或工作空间管理员无法控制。	· 表的预览权限归属于表的负责人所有，其他用户需要申请表的权限才能查看数据。 · 工作空间管理员可以通过开关配置开发表、生产表的数据预览权限对其他用户开放或关闭。 - 开发表默认开放对其他用户的数据预览权限。 - 生产表默认关闭对其他用户的数据预览权限。	· 使表的负责人对于自己所拥有的表的权限控制粒度更细致。 · 工作空间管理员可以方便地控制空间内表的数据预览权限。
新建表的入口收拢	可以在数据管理页面直接新建表，没有权限控制。	· 新建表的入口暂时关闭。 · 建议您使用数据开发的表管理或手动任务来新建和修改表结构，可以复用数据开发模块的开发、运维角色控制。 · 如果需要类目关联，您可以在数据地图 > 我的数据页面快速关联自己所拥有的表所属的类目。	· 使表的新建和修改的入口统一在数据开发模块，并有角色权限控制。 · 控制表的结构变化归属于数据开发者，避免无权限人员随意创建表或关联类目。

数据地图问题反馈途径

如您对DataWorks数据地图模块存在任何产品使用方面的问题，请随时搜索钉钉群号（23182329）或通过扫码进入DataWorks产品交流群，我们将竭诚为您答疑。

4.2 数据地图概述

数据地图主要包括全局查找数据、个人账号管理数据和管理员配置。

单击左上角的图标，选择全部产品 > 数据地图，即可进入数据地图页面。

DataStudio
全部产品
数据集成
DataStudio(数据开发)
运维中心(工作流)
任务发布
数据地图(数据管理)
DataStudio(数据开发)
数据服务
Function Studio
New
数据分析
任务运维
运维中心(工作流)
任务发布
跨项目克隆
数据治理
数据质量
数据保护伞
数据地图(数据管理)

数据地图
数据总览
全部类目
我的数据
配置管理
请输入搜索关键字
表
biao
近期浏览的表
近期读取的表
热门浏览
热门读取

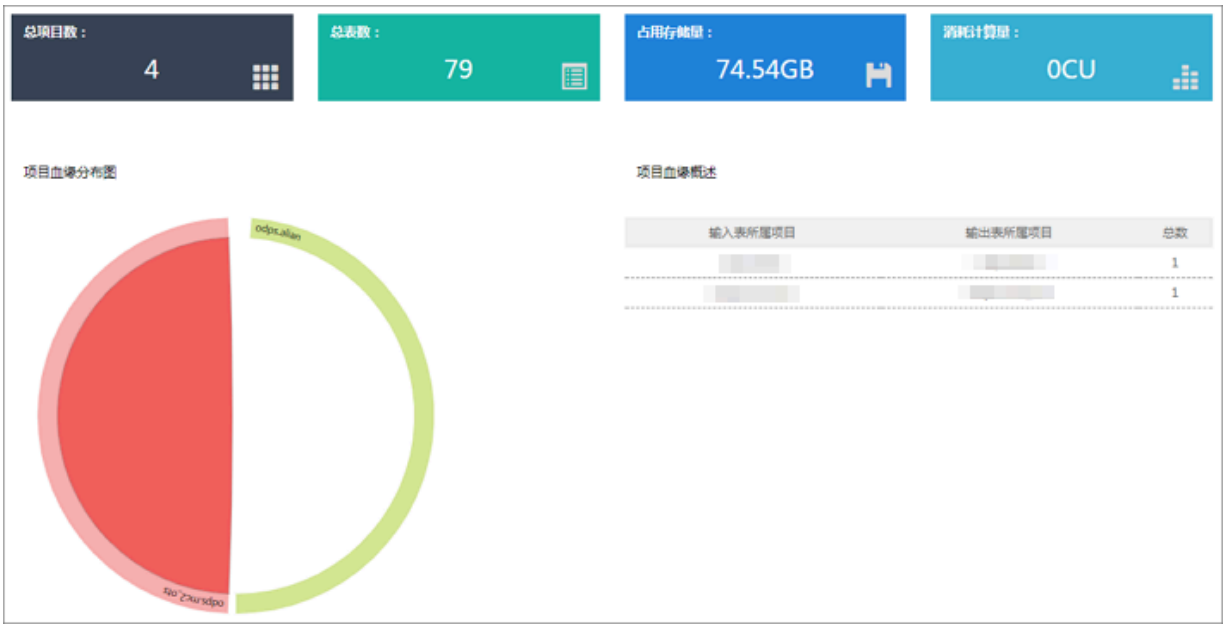
- 如果您更习惯使用强大的搜索引擎，请单击左上方的数据地图，进入首页开始搜索。

  目前输入关键字匹配更准确，同时还提供更多的搜索对象。例如，如果您经常使用数据地图，在数据地图的首页，为您提供了近期浏览和近期读取的表，以及基于您的访问记录推荐的热门浏览和热门读取。
- 如果您需要根据路径找到您想要的表，请单击全部类目。该页面将按照类目结构，为您展示各级类目及对应的表数量。
- 如果您有一些个人事务需要处理，例如修改您拥有的表或者使用小工具，请单击我的数据。
- 如果您是类目管理者或项目管理者，需要修改项目级配置或全局类目，请单击配置管理。

4.3 数据总览

您可以通过数据总览页面，查看工作空间的整体情况。

选择数据地图 > 数据总览，进入数据总览页面。该页面在整个组织的前提下进行统计，同时离线产生整个页面的数据信息，即页面中的数据信息为昨天的统计数据。



列表项	说明
总项目数	在组织视角下，所有工作空间的数量。
总表数	在组织视角下，所有表的数量。
占用存储量	在组织视角下，所有表存储量的总和。
消耗计算量	在组织视角下，该组织累计一天的计算消耗。一个CPU核满载运行一天为1CU。
项目血缘分布图	在组织视角下，以网络来刻画工作空间之间的血缘关系，图中弧形代表工作空间。如果存在血缘关系，则两个工作空间之间进行连线。
项目血缘概述	在组织视角下，左侧为上游表所在工作空间，右侧为下游表所属工作空间，总数表示两个工作空间存在的血缘关系的数量。
项目占用存储Top	在组织视角下，各个工作空间所占用的存储排行榜，取前十名。
表占用存储Top	在组织视角下，展示数据表占用存储量前十名，并可单击具体表名跳转至表详情页。 说明: 项目存储及表占用的逻辑存储显示T+1用量，并且显示的为逻辑存储大小。项目存储量除了表存储量外，还会计算包括资源存储量、回收站存储量及其他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	在组织视角下，被引用次数最多的数据表排行，显示前十名。并可以单击具体表名跳转至表详情页。

4.4 我的数据

本文将为您介绍数据地图中我的数据模块的功能。

进入数据地图模块，单击顶部菜单栏中的我的数据，即可进入我的数据模块。

您可以在该模块查看我拥有的数据、我管理的数据、生产账号的数据和我的收藏，并可以对相应数据进行权限管理。

我拥有的数据

单击左侧菜单栏中的我拥有的数据，即可根据表名、环境、项目/数据库和可见范围等信息进行搜索，查看相应表的具体信息并进行相关操作。

配置	说明
表名	单击表名，即可跳转至表详情页面。
中文名	您可以编辑表的中文名。
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。
环境类型	环境类型通常包括开发和生产。
存储量	显示您存储的数据量。
生命周期	和您创建表时设置的生命周期一致。
操作	您可以在相应表后的操作栏下，进行删除、修改类目、隐藏和自动探查等操作。 **说明：** 如果您隐藏了表，在表详情页将看不到申请权限的入口。

我管理的数据

单击左侧菜单栏中的我管理的数据，即可根据表名、环境和项目/数据库等信息进行搜索，查看相应表的具体信息并进行相关操作。

配置	说明
表名	单击表名，即可跳转至表详情页面。
中文名	您可以编辑表的中文名。

配置	说明
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。
环境类型	环境类型通常包括开发和生产。
存储量	显示您存储的数据量。
生命周期	和您创建表时设置的生命周期一致。
操作	您可以在相应表后的操作栏下，进行删除和修改类目等操作。

生产账号的数据

单击左侧菜单栏中的生产账号的数据，即可根据表名、环境和项目/数据库等信息进行搜索，查看相应表的具体信息。

配置	说明
表名	单击表名，即可跳转至表详情页面。
中文名	您可以编辑表的中文名。
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。
环境类型	环境类型通常包括开发和生产。
存储量	显示您存储的数据量。
生命周期	和您创建表时设置的生命周期一致。
收藏人次	用户收藏该表的次数。
30天浏览人次	用户30天内浏览该表的次数。
创建时间	该表的创建时间。

我的收藏

收藏表后，您可以单击左侧菜单栏中的我的收藏进行查看。

您可以单击相应表后的取消收藏，取消对该表的收藏。

权限管理

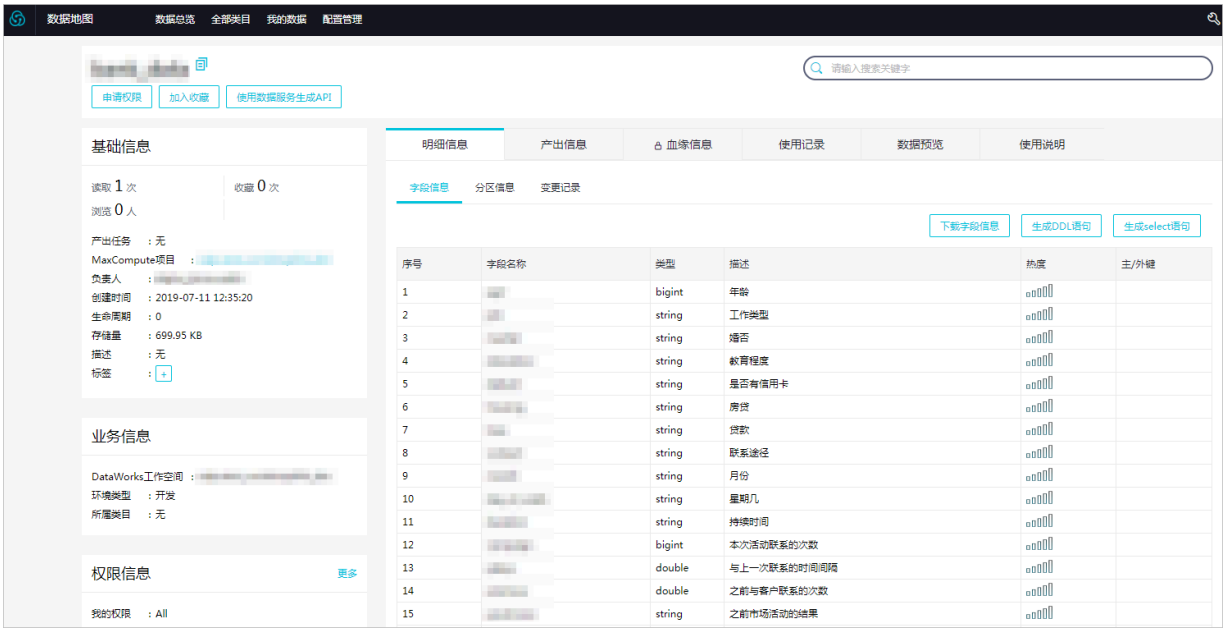
单击左侧菜单栏中的权限管理，即可进入权限管理功能模块。

- 权限管理模块的功能详情请参见权限管理。
- 单击右上方的申请数据权限按钮，可以申请函数和资源权限，详情参见#unique_47/unique_47_Connect_42_section_cjr_fz5_q2b。

4.5 表详情页介绍

您可以通过表详情页面，查看表的具体信息。

单击数据地图模块任意列表中的数据表名称，即可跳转至表详情页面。



表详情页面为您展示表的基础信息、业务信息、权限信息、技术信息、明细信息（包括字段信息、分区信息和变更记录）、产出信息、血缘信息、使用记录、数据预览和使用说明。

申请表权限

申请表权限的详情请参见申请表权限。

收藏表

单击表名称下方的加入收藏，即可收藏该表。表被收藏后，您可以进入数据地图 > 我的数据 > 我的收藏页面进行查看。

使用数据服务生产API

单击表名称下方的使用数据服务生产API，直接跳转至数据服务页面，详情请参见#unique_49。

基础信息

表的基础信息包括表的读取次数、收藏次数、浏览人数、产出任务、MaxCompute项目、负责人、创建时间、生命周期、存储量、描述和标签等信息。

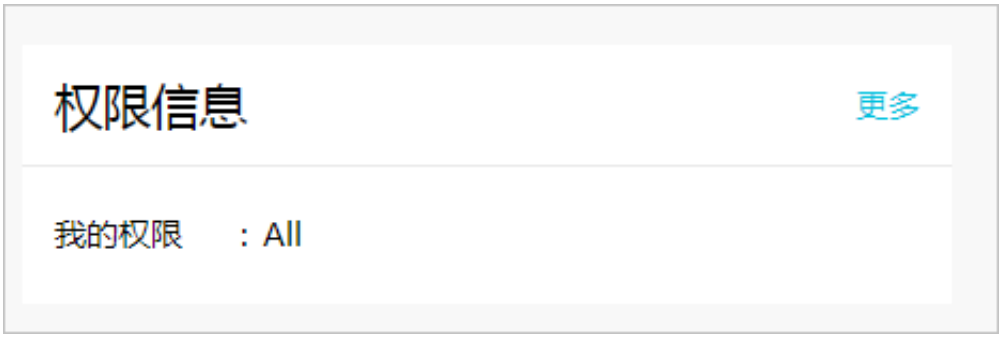
- 单击MaxCompute项目名称，即可跳转至项目详情页。
- 单击标签后面的 + ，即可为该表添加标签。

业务信息

表的业务信息主要包括DataWorks工作空间、环境类型和所属类目。

权限信息

表的权限信息为您展示您当前拥有的权限。单击右侧的更多，即可跳转至表权限申请页面。



技术信息

表的技术信息主要包括技术类型、DDL最后变更时间、最后数据变更时间、最后数据查看时间和计算引擎信息等信息。

- 时间格式默认为`yyyy-mm-dd hh:ss:mm`。
- 单击计算引擎信息后的点击查看按钮，即可查看计算引擎信息弹出框中的信息。

明细信息

表的明细信息是以表的元数据信息为主，显示字段定义、热度、等级、主外键及表结构变更情况。

- 字段信息

  表的字段信息包括字段名称、类型、描述、热度和主/外键等信息。

操作	说明
下载字段信息	单击后，即可直接下载相应的字段信息。
生成DDL语句	单击后，即可在弹出框中显示相应的建表语句。
生成select语句	单击后，即可在弹出框中显示相应的select语句。

- 分区信息

  通过表的分区信息模块，您可以查看表当前的分区，包括分区名、记录数、存储量、创建时间和最后更新时间等信息。

- 变更记录

  表的变更记录包括分区名、变更类型、粒度、时间和操作人等信息。

  您可以在下拉框中选择不同的变更类型，来过滤、筛选变更记录。

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，您可以在该模块查看变化的情况、持续更新的数据等信息。

任务名称	操作	节点ID	执行时间(秒)	等待时间(秒)	开始时间	结束时间
[redacted]	代码 日志	700000624	8	6	2019-04-17 00:06:07	2019-04-17 00:06:15
[redacted]	代码 日志	700000624	103	3	2019-04-17 00:29:05	2019-04-17 00:30:48
[redacted]	代码 日志	700000624	10	25	2019-04-18 00:06:27	2019-04-18 00:06:37
[redacted]	代码 日志	700000624	100	8	2019-04-18 00:29:10	2019-04-18 00:30:50
[redacted]	代码 日志	700000624	11	5	2019-04-19 00:06:06	2019-04-19 00:06:17
[redacted]	代码 日志	700000624	100	3	2019-04-19 00:29:05	2019-04-19 00:30:45
[redacted]	代码 日志	700000624	10	57	2019-04-20 00:07:26	2019-04-20 00:07:36
[redacted]	代码 日志	700000624	100	7	2019-04-20 00:29:09	2019-04-20 00:30:49
[redacted]	代码 日志	700000624	12	4	2019-04-21 00:06:05	2019-04-21 00:06:17
[redacted]	代码 日志	700000624	97	6	2019-04-21 00:29:08	2019-04-21 00:30:45

血缘信息

表的血缘信息可以为您清晰地展示数据的来源和去向，并提供便捷的交互。

说明:

仅购买DataWorks标准版及以上版本的用户，可以查看表的血缘信息。

- **表血缘：您可以根据GUID搜索表的上下游表。**

- **字段血缘：您可以根据字段进行过滤。**

使用记录

- **频繁关联：频繁关联模块为您展示有多少人在使用当前的数据。**

- **访问统计：访问统计模块，以线型方式为您展示表的使用记录。**

数据预览

您可以通过表的数据预览模块，预览当前表的数据信息。

说明:

您需要拥有权限，方可预览生产环境的表。如果没有权限，单击去申请，前往申请页面进行申请。

使用说明

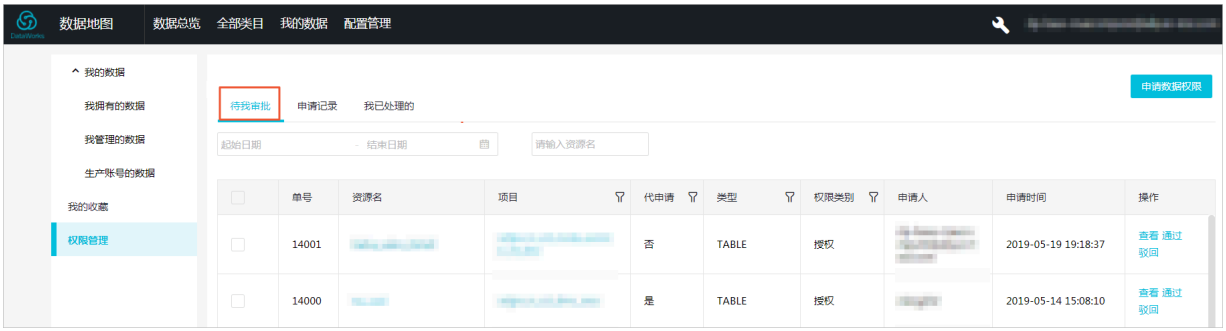
您可以编辑此界面，查看相关的历史版本和Markdown语法。您也可以根据数据的业务说明了解相关的信息。

4.6 权限管理

权限管理模块主要对待我审批、申请记录、我已处理、权限回收等相关表、资源、函数的权限申请进行管理。

待我审批

当前访问账号为管理员时，可通过待我审批模块，查看并审批其所有项目下的待审批记录，包括表、资源和函数等权限申请。



申请记录

当前访问账号可通过申请记录模块，查看其权限申请记录。

我已处理的

当前账号为管理员时，可通过我已处理的模块，查看其所有项目下已经处理过的权限申请记录，包括表、资源和函数等权限申请。

4.7 申请数据权限

本文将为您介绍如何申请数据权限。

DataWorks中有以下三种数据类型。

- 表：即数据表。
- 函数：即UDF，可以在SQL中使用的函数。
- 资源：例如文本文件、MapReduce的Jar文件等。

上述三种数据类型具有严格的权限控制机制，您需要申请权限后才能使用。

申请表权限（表数据预览权限）

1. 单击左上角的图标，选择全部产品 > 数据地图，即可进入数据地图页面。

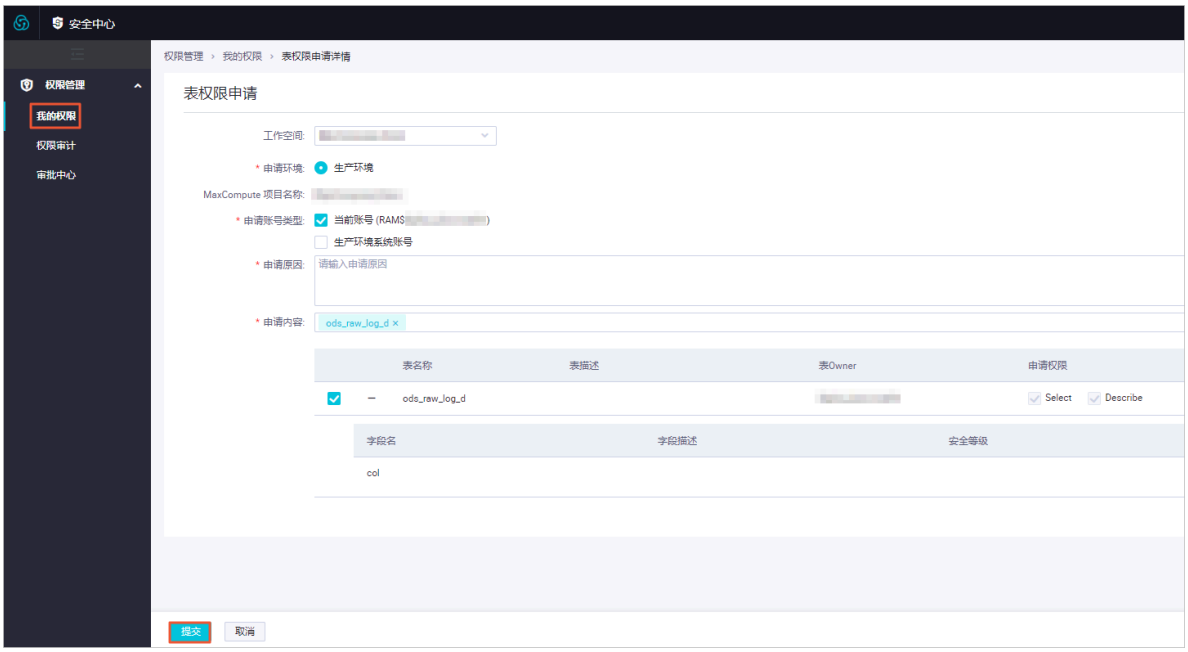
2. 在首页搜索需要申请权限的数据表，单击数据表名称进入表详情页面。

3. 在表详情页面中，单击申请权限，即可跳转至安全中心 > 我的权限页面。

说明：

如果您将表隐藏，便不会显示申请权限按钮。关于如何隐藏表，详情参见#unique_51/unique_51_Connect_42_section_6uh_f1o_4n9。

4. 填写表权限申请页面的各配置项。



配置	说明
工作空间	需要申请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生产环境。

配置	说明
MaxCompute项目名称	选择的DataWorks工作空间对应的MaxCompute项目名称，默认不可修改。
申请账号类型	包括当前账号和生产环境系统账号。
申请原因	请简要填写申请原因，以便更快地通过审批。
申请内容	勾选需要申请的内容。

5. 配置完成后，单击提交，待审批通过后即可预览表数据。

> 说明:
> 权限申请提交后，您可以进入数据地图 > 我的数据 > 权限管理 > 申请记录页面，查看申请状态。

申请函数和资源权限

1. 进入数据地图 > 我的数据 > 权限管理页面。
2. 单击右上方的申请数据权限。

3. 填写申请授权对话框中的配置。

配置	说明
申请类型	支持函数和资源两种类型。
权限归属人	支持本人申请和代理申请。 · 本人申请：选择该项，审批通过后权限归属于当期用户。 · 代理申请：选择代理申请，需填写对方用户名。审批通过后权限归属于被代理人。
项目名	选择所需申请函数或者资源所在的项目名称（对应的MaxCompute项目名称），支持本组织范围内模糊匹配查找。
函数名称/资源名称	输入项目中的函数或资源名称。资源名称请填写完整，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，该权限将被系统自动回收。
申请理由	请简要填写申请理由，以便更快地通过审批。

4. 填写完成后单击提交，然后等待审批。您可以通过数据地图 > 我的数据 > 权限管理 > 申请记录查看申请状态。

4.8 配置管理

本文为您介绍数据地图的配置管理功能模块。

1. 以开发者身份进入DataWorks管理控制台，单击项目列表下对应项目后的进入数据开发。
2. 单击左上方菜单栏中的数据地图。
3. 在数据地图页面，单击顶部菜单栏中的配置管理。

   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

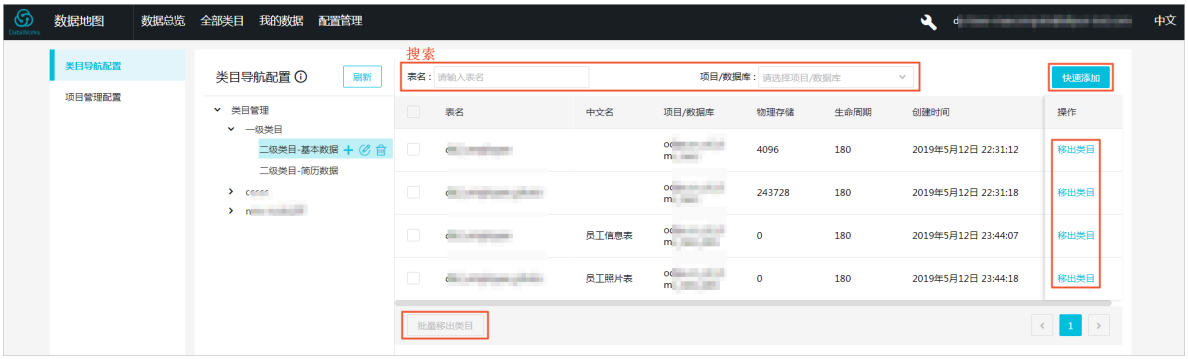
类目导航配置页面主要是可以创建类目和添加相关表到类目里。添加类目的目的是方便您更好的管理表。

1. 单击类目管理后的 ＋，添加一级类目。

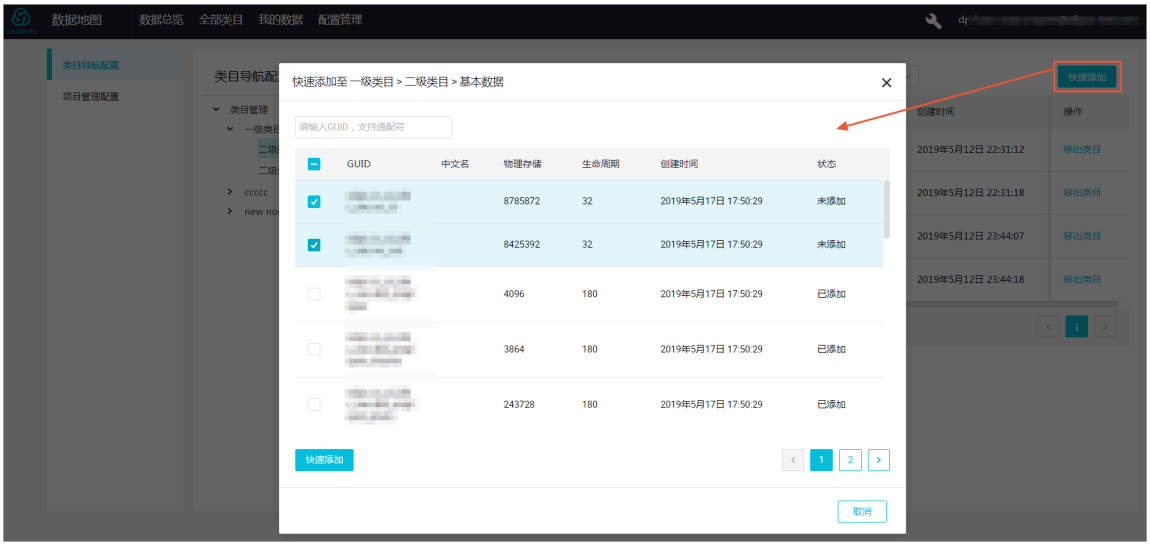
2. 单击一级类目后的 ，添加所属二级类目。

以此类推，最多可支持四级类目的设置。其中 表示编辑该类目名称， 表示删除该类目。

3. 类目配置完成后，即可进行如下操作：



- 快速添加：只能选择未添加的表。已经添加的表如果移出类目，也可以重新选择添加至类目。



- 搜索：可以通过表名和项目/数据库来搜索展示您的表信息。
- （批量）移出类目：已添加的表可以移出类目。

项目管理配置

- 线上自动化测试项目： MaxCompute表管理可以操作预览开发表的数据和预览生产表的数据两个按钮。
- Dataworks调度系统预发项目：MaxCompute表管理可以操作预览开发表的数据和预览生产表的数据两个按钮。

5 数据管理

5.1 数据管理概述

您可在数据管理模块进行组织内全局数据视图的查看、分权管理、元数据信息详情、数据生命周期管理、数据表/资源/函数权限管理审批等操作。

您可以通过点击左上方按钮切换到数据管理页面。

数据管理模块的具体功能，请参见下述文档。

数据搜索

#unique_56

新建表

收藏表修改生命周期

修改表结构

隐藏表

修改表负责人

删除表

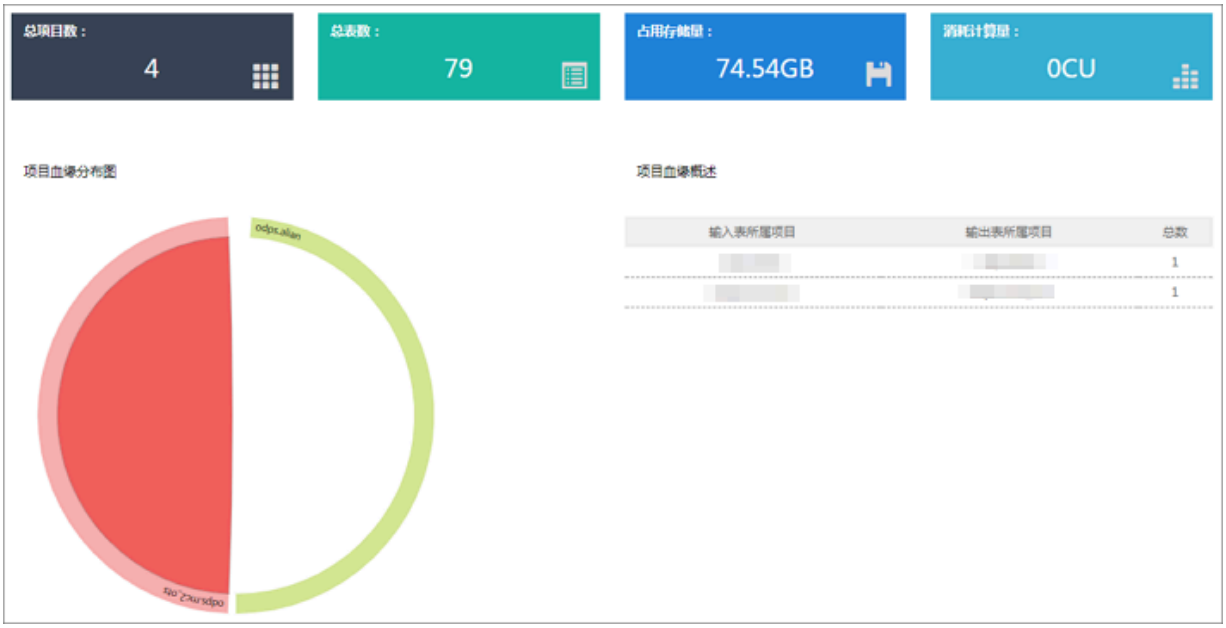
查看表详情

类目导航配置

5.2 全局概览

您可通过全局概览页面查看项目空间的整体情况。

选择数据管理 > 全局概览，进入全局概览页面，该页面中的统计都是在整个组织的前提下进行统计的，同时整个页面的数据信息都是离线产生，即页面中的数据信息为昨天的统计数据。



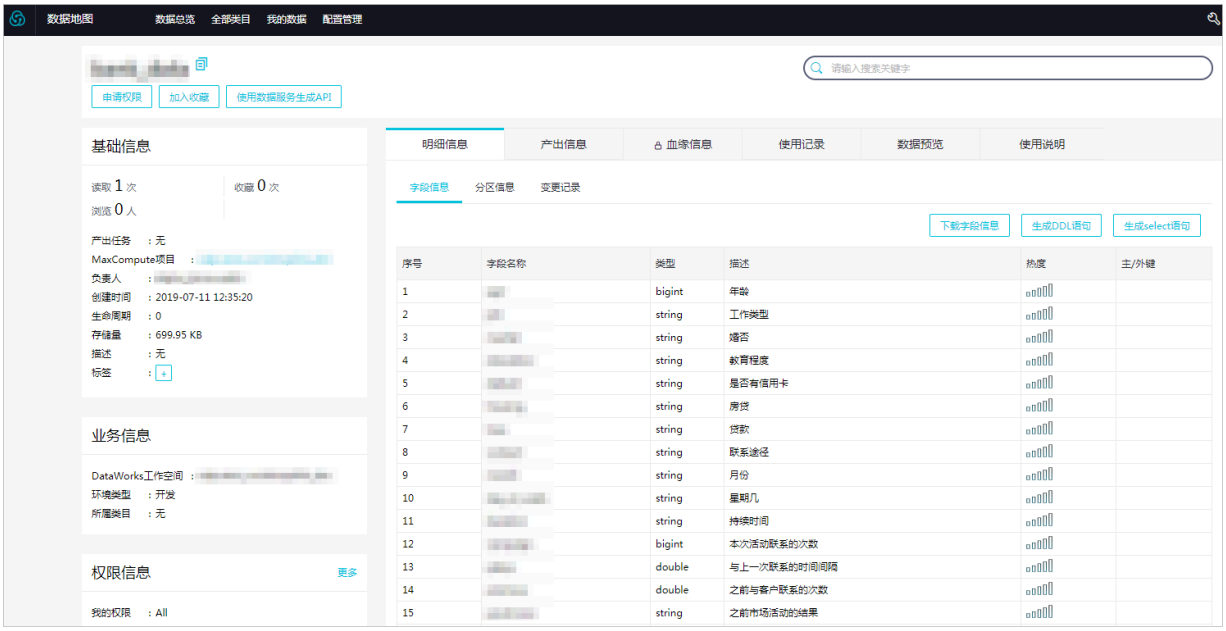
列表项	说明
总项目数、总表数、占用存储量、消耗计算量	在组织视角下，统计的项目空间个数、数据表个数、数据表占用的存储和任务运行时所消耗的计算量（CPU / minute或second等）。

列表项	说明
项目血缘分布图	在组织视角下，以网络来刻画项目空间之间的血缘关系，图中弧形代表项目空间，若存在血缘关系则两个项目空间之间进行连线。
项目血缘概述	在组织视角下，左侧为上游表所在项目空间，右侧为下游表所属项目空间，总数表示两个项目空间存在的血缘关系的数量。
项目占用存储Top	在组织视角下，各个项目空间所占用的存储排行榜，取前十名。
表占用存储Top	在组织视角下，展示数据表占用存储量前十名，并可单击具体表名跳转至表详情页。 说明: 项目存储及表占用的逻辑存储显示T+1用量，并且显示的为逻辑存储大小。项目存储量除了表存储量外，还会计算包括资源存储量、回收站存储量及其他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	在组织视角下，被引用次数最多的数据表排行，显示前十名，并可单击具体表名跳转至表详情页。

5.3 表详情页介绍

您可以通过表详情页面，查看表的具体信息。

单击数据地图模块任意列表中的数据表名称，即可跳转至表详情页面。



表详情页面为您展示表的基础信息、业务信息、权限信息、技术信息、明细信息（包括字段信息、分区信息和变更记录）、产出信息、血缘信息、使用记录、数据预览和使用说明。

申请表权限

申请表权限的详情请参见申请表权限。

收藏表

单击表名称下方的加入收藏，即可收藏该表。表被收藏后，您可以进入数据地图 > 我的数据 > 我的收藏页面进行查看。

使用数据服务生产API

单击表名称下方的使用数据服务生产API，直接跳转至数据服务页面，详情请参见#unique_49。

基础信息

表的基础信息包括表的读取次数、收藏次数、浏览人数、产出任务、MaxCompute项目、负责人、创建时间、生命周期、存储量、描述和标签等信息。

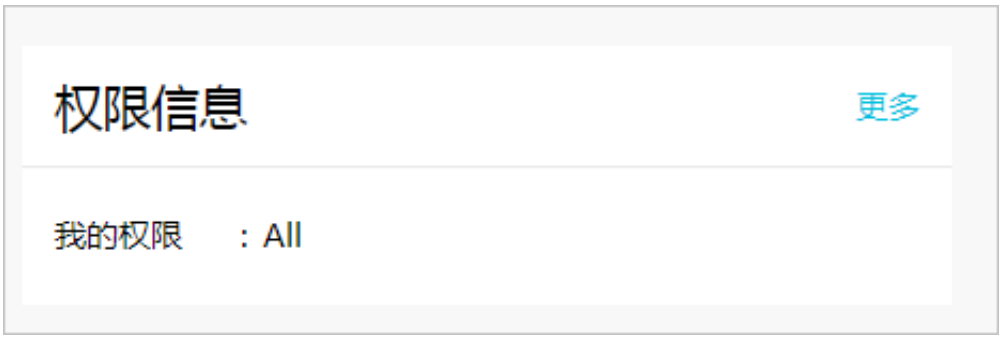
- 单击MaxCompute项目名称，即可跳转至项目详情页。
- 单击标签后面的 + ，即可为该表添加标签。

业务信息

表的业务信息主要包括DataWorks工作空间、环境类型和所属类目。

权限信息

表的权限信息为您展示您当前拥有的权限。单击右侧的更多，即可跳转至表权限申请页面。



技术信息

表的技术信息主要包括技术类型、DDL最后变更时间、最后数据变更时间、最后数据查看时间和计算引擎信息等信息。

- 时间格式默认为`yyyy-mm-dd hh:ss:mm`。
- 单击计算引擎信息后的点击查看按钮，即可查看计算引擎信息弹出框中的信息。

明细信息

表的明细信息是以表的元数据信息为主，显示字段定义、热度、等级、主外键及表结构变更情况。

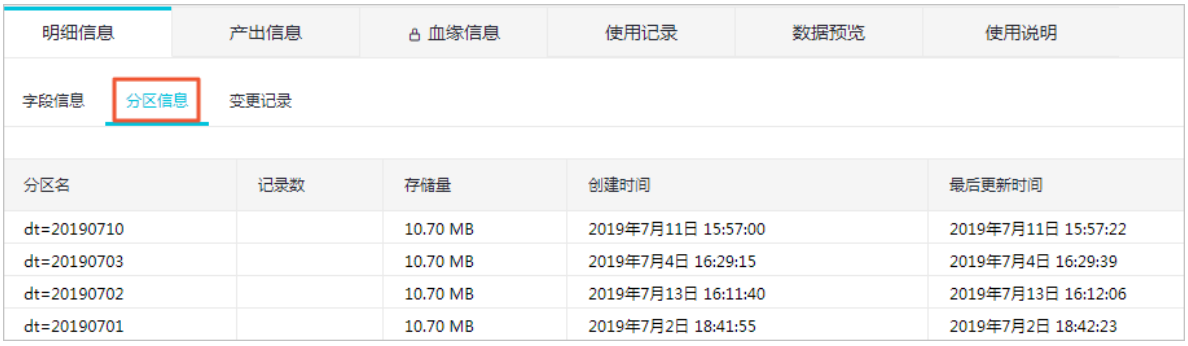
- 字段信息

  表的字段信息包括字段名称、类型、描述、热度和主/外键等信息。

操作	说明
下载字段信息	单击后，即可直接下载相应的字段信息。
生成DDL语句	单击后，即可在弹出框中显示相应的建表语句。
生成select语句	单击后，即可在弹出框中显示相应的select语句。

- 分区信息

  通过表的分区信息模块，您可以查看表当前的分区，包括分区名、记录数、存储量、创建时间和最后更新时间等信息。



- 变更记录

  表的变更记录包括分区名、变更类型、粒度、时间和操作人等信息。

  您可以在下拉框中选择不同的变更类型，来过滤、筛选变更记录。

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，您可以在该模块查看变化的情况、持续更新的数据等信息。

任务名称	操作	节点ID	执行时间(秒)	等待时间(秒)	开始时间	结束时间
[redacted]	代码 日志	700000624	8	6	2019-04-17 00:06:07	2019-04-17 00:06:15
[redacted]	代码 日志	700000624	103	3	2019-04-17 00:29:05	2019-04-17 00:30:48
[redacted]	代码 日志	700000624	10	25	2019-04-18 00:06:27	2019-04-18 00:06:37
[redacted]	代码 日志	700000624	100	8	2019-04-18 00:29:10	2019-04-18 00:30:50
[redacted]	代码 日志	700000624	11	5	2019-04-19 00:06:06	2019-04-19 00:06:17
[redacted]	代码 日志	700000624	100	3	2019-04-19 00:29:05	2019-04-19 00:30:45
[redacted]	代码 日志	700000624	10	57	2019-04-20 00:07:26	2019-04-20 00:07:36
[redacted]	代码 日志	700000624	100	7	2019-04-20 00:29:09	2019-04-20 00:30:49
[redacted]	代码 日志	700000624	12	4	2019-04-21 00:06:05	2019-04-21 00:06:17
[redacted]	代码 日志	700000624	97	6	2019-04-21 00:29:08	2019-04-21 00:30:45

血缘信息

表的血缘信息可以为您清晰地展示数据的来源和去向，并提供便捷的交互。

说明:

仅购买DataWorks标准版及以上版本的用户，可以查看表的血缘信息。

- **表血缘：您可以根据GUID搜索表的上下游表。**

- **字段血缘：您可以根据字段进行过滤。**

使用记录

- **频繁关联：频繁关联模块为您展示有多少人在使用当前的数据。**

- **访问统计：访问统计模块，以线型方式为您展示表的使用记录。**

数据预览

您可以通过表的数据预览模块，预览当前表的数据信息。

	A	B	C	D	E	F	G	H
字段	empno	firstnme	midinit	lastname	workdept	phoneno	hiredate	job
类型	STRING	STRING	STRING	STRING	STRING	STRING	DATETIME	STRING
3	000010	CH	I	HA	A00	3978	1995-01-01 00:00:00	PRES
4	000020	MI	L	TH ON	B01	3476	2003-10-10 00:00:00	MANAGER
5	000030	SA	A	KV	C01	4738	2005-04-05 00:00:00	MANAGER
6	000050	JO	B	GE	E01	6789	1979-08-17 00:00:00	MANAGER
7	000060	IR	F	ST	D11	6423	2003-09-14 00:00:00	MANAGER
8	000070	EV	D	PU	D21	7831	2005-09-30 00:00:00	MANAGER
9	000090	EI	W	HE SON	E11	5498	2000-08-15 00:00:00	MANAGER
10	000100	TH E	Q	SP	E21	0972	2000-06-19 00:00:00	MANAGER
11	000110	VI	G	LU SSI	A00	3490	1988-05-16 01:00:00	SALESREP
12	000120	SE		O' ELL	A00	2167	1993-12-05 00:00:00	CLERK
13	000130	DE	M	QU NA	C01	4578	2001-07-28 00:00:00	ANALYST
14	000140	HE	A	NI S	C01	1793	2006-12-15 00:00:00	ANALYST
15	000150	BR		AD ON	D11	4510	2002-02-12 00:00:00	DESIGNER
16	000160	EL H	R	PI	D11	3782	2006-10-11 00:00:00	DESIGNER

说明:

您需要拥有权限，方可预览生产环境的表。如果没有权限，单击去申请，前往申请页面进行申请。

使用说明

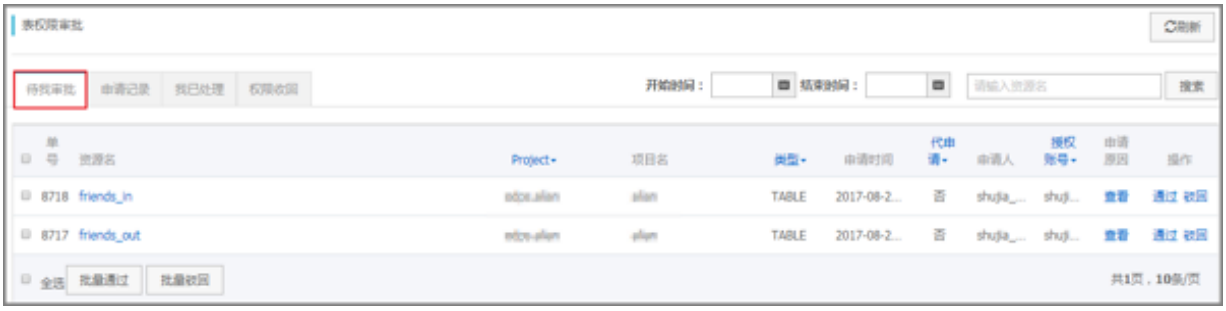
您可以编辑此界面，查看相关的历史版本和Markdown语法。您也可以根据数据的业务说明了解相关的信息。

5.4 权限管理

权限管理模块主要对待我审批、申请记录、我已处理、权限回收等相关表、资源、函数的权限申请进行管理。

待我审批

当前访问账号为管理员时，可通过待我审批模块，查看并审批其所有项目下的待审批记录，包括表、资源和函数等权限申请。



申请记录

当前访问账号可通过申请记录模块，查看其权限申请记录。

我已处理

当前账号为管理员时，可通过我已处理模块，查看其所有项目下已经处理过的权限申请记录，包括表、资源和函数等权限申请。

权限回收

当前账号为管理员时，可通过权限回收模块，查看并回收其所有项目下的已经通过的权限申请记录，包括表、资源和函数等权限申请。

5.5 数据权限申请

本文将为您介绍如何申请数据权限。

DataWorks中有以下三种数据类型。

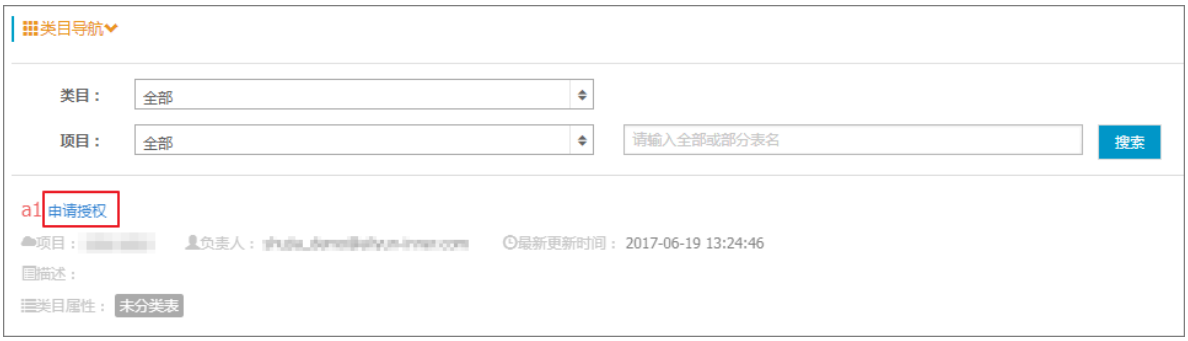
- 表：即数据表。
- 函数：即UDF，可在SQL中使用的函数。

· 资源：如文本文件、MapReduce的Jar文件等。

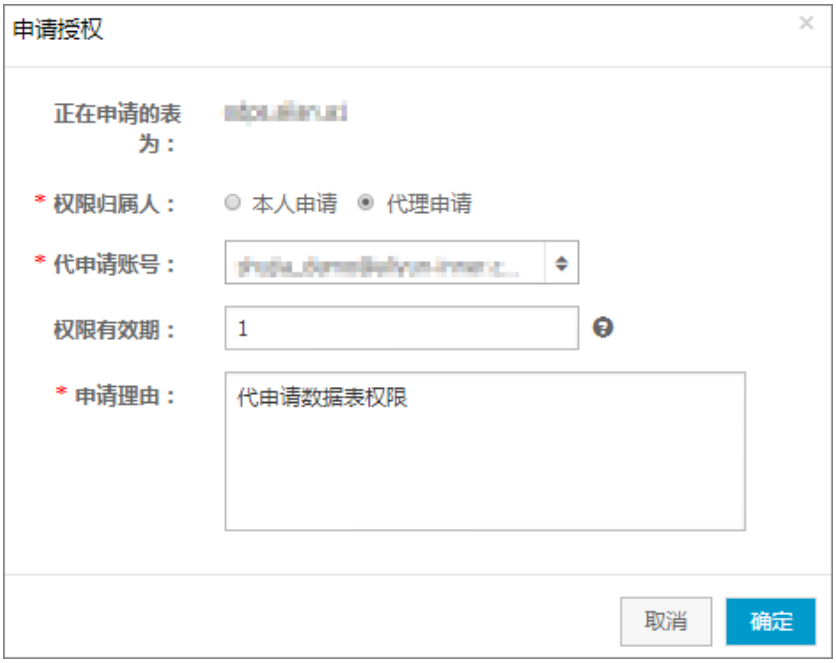
这三种数据类型有着严格的权限控制机制，您需要申请权限后才能使用。

表权限申请（只读权限）

1. 通过数据管理 > 查找数据页面找到需要申请权限的数据表。
2. 单击数据表操作栏中的申请权限。



3. 填写申请授权弹出框中的各配置项。

配置	说明
权限归属人	支持本人申请和代理申请。 · 本人申请：选择该项，审批通过后权限归属于当前用户。 · 代理申请：选择代理申请，需填写代申请账号（系统右上角显示的登录名）。审批通过后权限归属于被代理人。 
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，该权限将被系统自动回收。

配置	说明
申请理由	请简要填写申请理由以便更快地通过审批。

4. 单击确定后提交，然后等待审批。您可在权限管理 > 申请记录中查看申请状态。

函数和资源权限申请

1. 进入数据管理 > 查找数据页面。
2. 单击列表右上方的申请数据权限。

3. 填写申请授权对话框中的配置。

配置	说明
申请类型	支持函数和Resource资源两种类型。

配置	说明
权限归属人	支持本人申请和代理申请。 · 本人申请：选择该项，审批通过后权限归属于当期用户。 · 代理申请：选择代理申请，需填写代申请账号。审批通过后权限归属于被代理人。
项目名	选择所需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。资源名称请填写完整，包括文件后缀，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，该权限将被系统自动回收。
申请理由	请简要填写申请理由以便更快地通过审批。

4. 单击确定后提交，然后等待审批。您可通过权限管理 > 申请记录查看申请状态。

5.6 管理配置

当前用户（需组织管理员权限）可在类目导航配置页面中进行新建表所属类目的配置。

操作步骤

1. 以开发者身份进入DataWorks 管理控制台，单击项目列表下对应项目后的进入数据开发。

2. 单击顶部菜单栏中的数据管理。

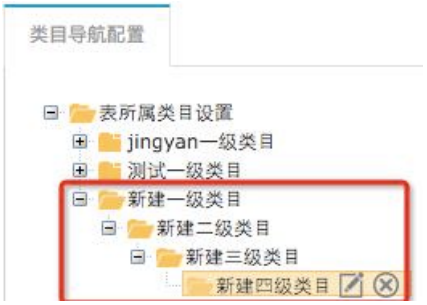
3. 在数据管理页面，单击左侧导航栏中的管理配置。

4. 单击表所属类目设置后的⊕，添加一级类目。

5. 单击一级类目后的⊕，添加所属二级类目。

以此类推，最高可支持四级类目的设置。其中✎表示编辑该类目名称，⊗表示删除该类目。

类目配置完成后，即可在新建表页面中选择已配置类目。

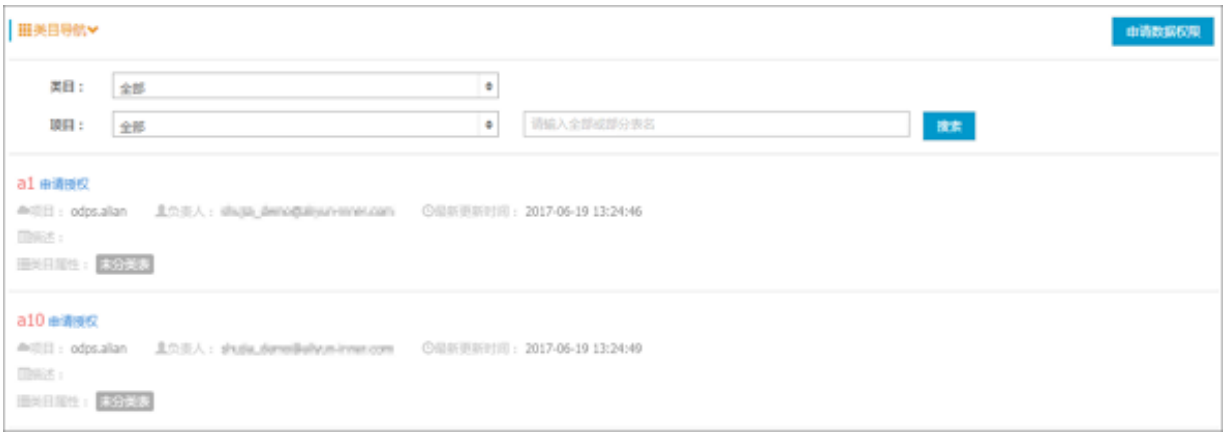


新建表的所属类目，如下图所示：

5.7 查找数据

本文将为您介绍如何查找数据。

您可进入数据管理 > 查找数据 页面，对本组织范围内（多项目空间）的数据表进行搜索。在全部数据页面中，通过选择数据类目导航 + 搜索框中输入表名进行模糊匹配的方式快速查找需要的表。



您可通过以下三种方式对数据表进行搜索。

- 类目导航选择：查看当前类目下所有表。
- 项目名称选择：查看当前项目下所有表，可与类目筛选条件一起使用。
- 搜索条件：在搜索框里输入表名（支持表名模糊搜索），同时支持备注搜索。

5.8 数据表管理

数据表管理模块对数据表进行分类，并为各分类提供不同的表信息以及表操作管理功能，以便您管理自己的数据表。

在数据表管理中，您可对表进行生命周期设置、表管理（包括修改表的类目、描述、字段、分区等）、表隐藏/取消隐藏和表删除等操作。

表的分类

- 我收藏的表

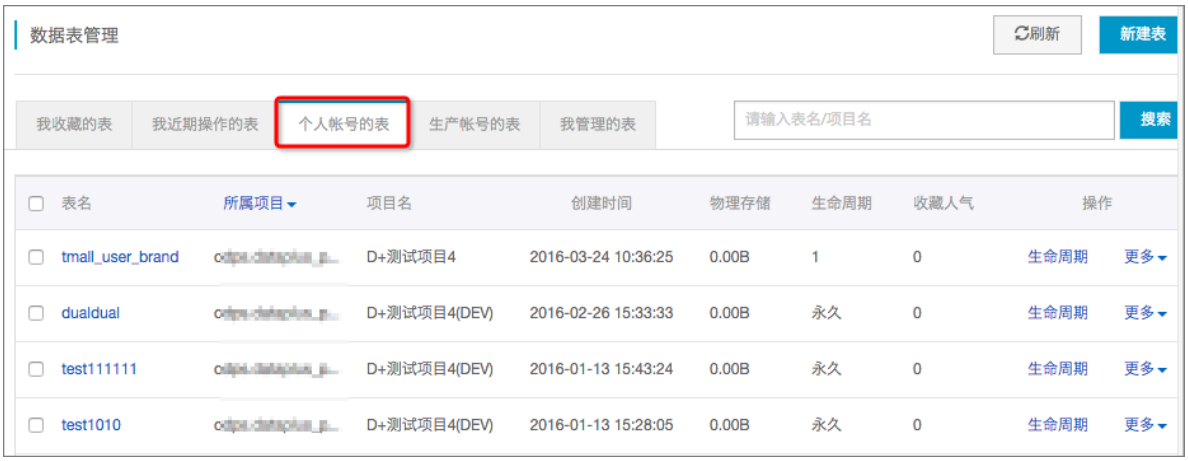
  展示了您所收藏的数据表列表，在此您也可以进行取消收藏操作。

- 我近期操作的表

  展示了您近期操作过的表，在此您可以进行表生命周期设置、表管理（包括修改表的类目、描述、字段、分区等）、表隐藏/取消隐藏、表删除等操作。

- 个人账号的表

  展示了您在组织内所创建的数据表列表，即Owner为当前用户的表。



本模块支持通过表名称模糊匹配查找，同时也可通过所属项目筛选查找表。您可在此对表进行的操作同我近期操作的表。

- 生产账号的表

  展示了表Owner为MaxCompute访问身份配置为计算引擎指定账号（即生产账号）的表。您可在此对表进行的操作同我近期操作的表。

- 我管理的表

  若您为项目管理员，将在此页面显示其所管理的项目空间中的所有数据表，管理员可在此可对表进行各类操作，包括修改表Owner。

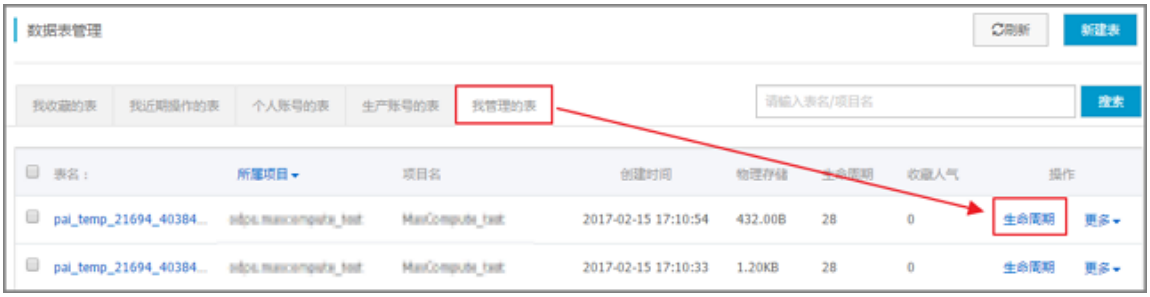
表的管理操作

- 收藏表

  数据管理模块提供收藏表功能，您在表详情页中单击收藏即可，也可在我收藏的表页面中取消收藏。

- 修改生命周期

  1. 单击列表的操作栏中的生命周期。

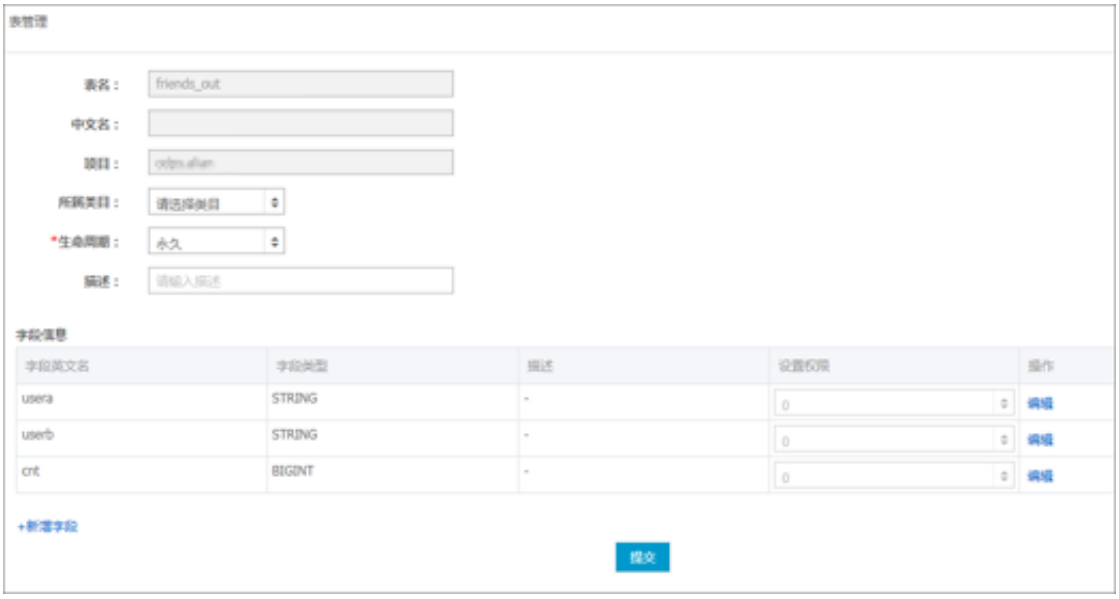


  2. 修改生命周期弹出框中表的生命周期。

· 修改表结构

1. 单击列表操作栏中的更多选项，选择表管理进行表结构修改。

2. 在打开的表管理页面中修改相关信息。



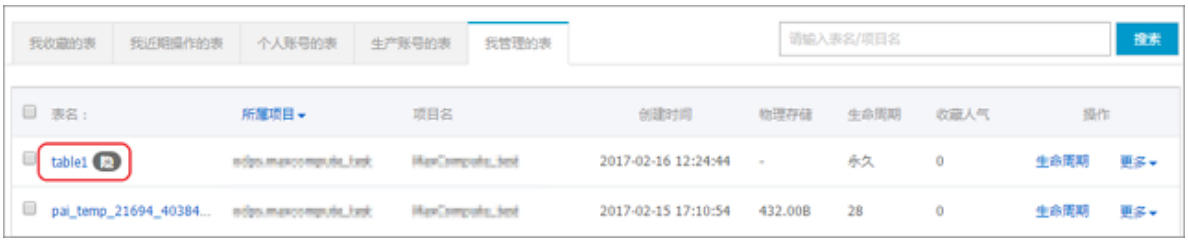
3. 修改完成后，单击提交。

- 隐藏表

  隐藏表功能即表负责人或项目管理员可对表进行隐藏，让其他成员不可见。

  单击列表操作栏中 更多选项，选择隐藏即可隐藏表。隐藏后的表，也可在此单击取消隐藏。

  被隐藏的表名后将有隐藏标识。



  说明:

  主账号隐藏的表子账号不能查看隐藏表内容，单击会报相应的提示：表被隐藏，请联系管理员或owner，子账号隐藏表主账号可以查询表内容。

- 修改表负责人（表Owner）

  项目管理员可修改表负责人，具体操作如下：

  1. 进入我管理的表模块，单击列表操作栏中的更多，选择修改owner。

  2. 在修改表owner弹出框中输入新Owner的云账号名称，该Owner必须为本项目的成员。
  3. 修改完成后，单击提交。

- 删除表

1. 单击列表操作栏中的更多，选择删除。

2. 单击确定，确认删除操作。

说明:

数据表一旦删除，该表的结构信息及表的所有数据均不可恢复，请谨慎操作。

5.9 创建表

通常数据开发过程中需要创建表来存储数据同步、数据加工的结果数据，本文将为您介绍可视化建表、DDL建表两种创建表的方式。

说明:

通过数据地图（原数据管理）模块创建的表可以进行业务类目划分。当组织内业务很多时，给数据表划分类目可以方便元数据的管理。

前提条件

- 实名认证云账号，生成AccessId和Accesskey。

  建表所用的云账号都是当前登录人账号，必须有AccessId和Accesskey才能发请求到MaxCompute进行建表，所以该云账号必须要实名认证生成AccessId和Accesskey。详情请参见#unique_69。

- 给云账号赋建表权限。

  如果您需要创建表，必须先给建表的云账号授权。MaxCompute项目的owner直接运行授权语句进行授权。

```
use projectname; --打开项目空间
add user aliyun$云账号;  --添加用户
grant CreateInstance,CreateTable,List ON PROJECT projectname TO
aliyun$云账号; --给用户赋权
```

> **说明:**
> 因为此处建表都是用当前登录的云账号创建，因此表的owner即为当前登录账号。

可视化建表

1. 以开发者身份登录DataWorks控制台，单击相应工作空间后的进入数据开发。
2. 单击左上角的图标，选择全部产品 > 数据地图（数据管理），即可进入数据地图页面。
3. 单击左侧菜单栏中的数据表管理，然后单击右上角的新建表。

4. 填写新建表页面的基础信息，单击下一步。

配置	说明
项目名	列表中显示当前用户已加入的MaxCompute项目空间。
表名	以字母、数字、下划线组成。
别名	表的中文名称。
所属类目	当前表所处的类目，最多支持四级类目导航，详情请参见#unique_60。
描述	当前表的简要说明。
生命周期	即MaxCompute的生命周期功能。填写一个数字表示天数，那么该表（或分区）超过一定天数未更新的数据会被清除。 支持1天、7天、32天、永久和自定义5种选项。

5. 填写新建表页面的字段和分区信息。

   - 添加字段信息设置。
   - 设置分区。

配置	说明
字段英文名	字段英文名，由字母、数字、下划线组成。
字段类型	MaxCompute数据类型（STRING、BIGINT、DOUBLE、DATETIME和BOOLEAN）。
描述	字段详细描述。
操作	上移、下移和删除。
是否设置分区	如果选择设置分区，需配置分区键的具体信息，支持STRING和BIGINT类型。

6. 单击提交。

   新建表提交成功后，系统将自动跳转至数据表管理页面，单击我管理的表，即可查看新建的表。

DDL建表

1. 以开发者身份登录DataWorks控制台，单击相应工作空间后的进入数据开发。
2. 单击左上角的图标，选择全部产品 > 数据地图（数据管理），即可进入数据地图页面。
3. 单击左侧菜单栏中的数据表管理，然后单击右上角的新建表。
4. 单击新建表对话框中的DDL建表。

5. 填写MaxCompute SQL建表语句。

```
create table if not exists table2
(
 id string comment'用户ID',
 name string comment'用户名称'
) partitioned by(dt string)
```

```
LIFECYCLE 7;
```

6. 单击提交，出现如下页面。

基础信息页面除表别名、所属类目和生命周期配置项外，都会自动填充。而字段和分区信息页面中字段的中文名称以及字段的安全等级等，需要您进行编辑添加。

7. 补充新建表基础信息页面中的配置项，单击下一步。

8. 单击提交。

新建表提交成功后，系统将自动跳转至数据表管理页面，单击我管理的表，即可查看新建的表。

6 资源优化

6.1 资源优化概述

资源优化从数据存储、数据计算和数据采集3个领域进行扫描，帮您扫描出可以优化的表和节点，从而合理、高效地运行DataWorks上的任务。

说明:

目前仅华东2（上海）地域支持资源优化，正在内部邀测中。如果您有相关需求，请提交工单进行申请。

资源优化以列表的形式为您展示不规范的使用问题，您可以根据对应优化项的解读，来进行优化。

资源优化页面主要面向个人和管理员2种用户群体：

- 个人资产优化页面主要供个人使用，为您展示个人名下的总任务数、总表数、优化趋势和个人资产优化模块。
- 工作空间资产优化页面主要供管理员使用，为管理员展示当前工作空间下的总任务数、总表数、优化趋势、可优化计算管理个人、可优化存储管理个人和工作空间资产优化模块。

  管理员可以通过可优化计算管理个人和可优化存储管理个人的排行信息，联系对应的资产责任人进行优化。

说明:

个人资产和工作空间管理这两个查看维度，需要添加一下数据的更新时间，目前个人资产优化和工作空间资产优化页面的总实例数、总表数的数据是离线（T+1）更新的，会存在数据延迟的情况。

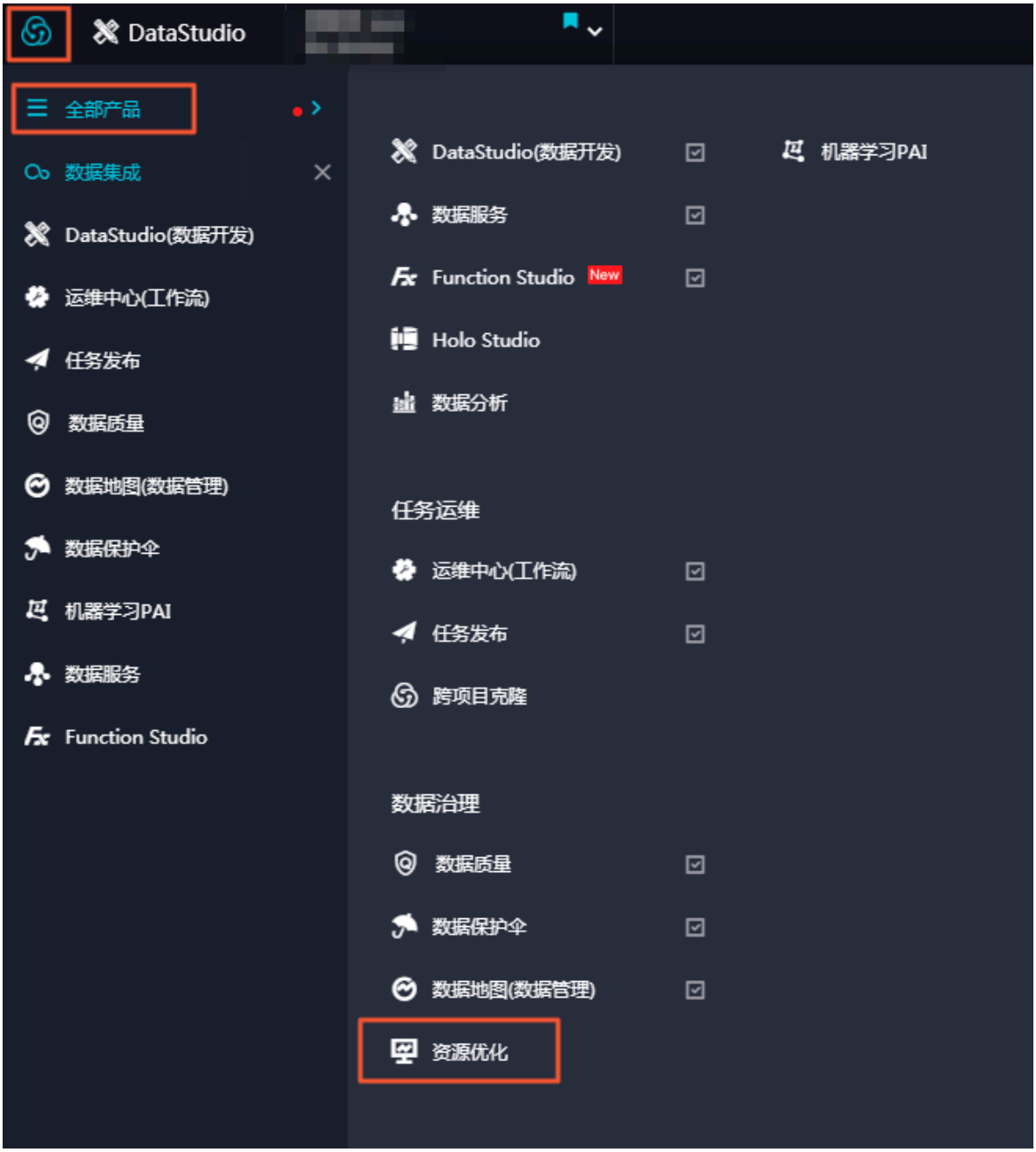
6.2 个人资产优化

个人资产优化页面主要供个人使用，为您展示个人名下可优化的任务和表。

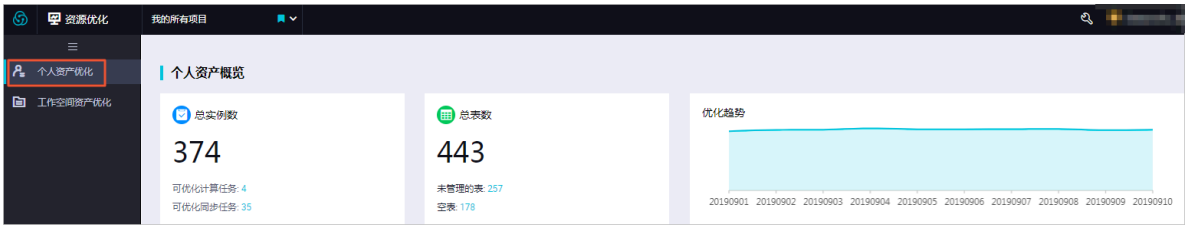
进入个人资产优化页面

1. 登录DataWorks控制台，单击相应工作空间后的进入数据开发。

2. 单击左上角的图标，选择全部产品 > 资源优化。



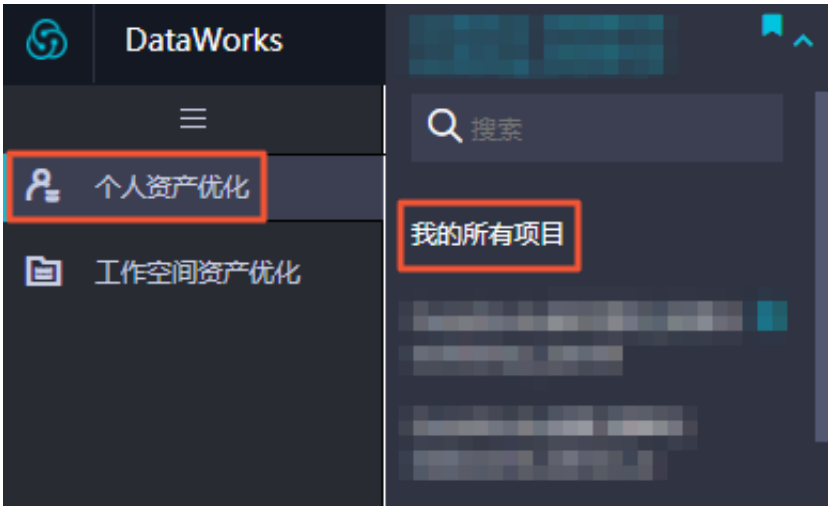
3. 单击左侧菜单栏中的个人资产优化，即可进入个人资产优化页面，查看个人名下的总实例数和总表数。



优化趋势为您展示最近10天内可优化项的变化趋势，您可以在此查看最近完成可优化项的数量。

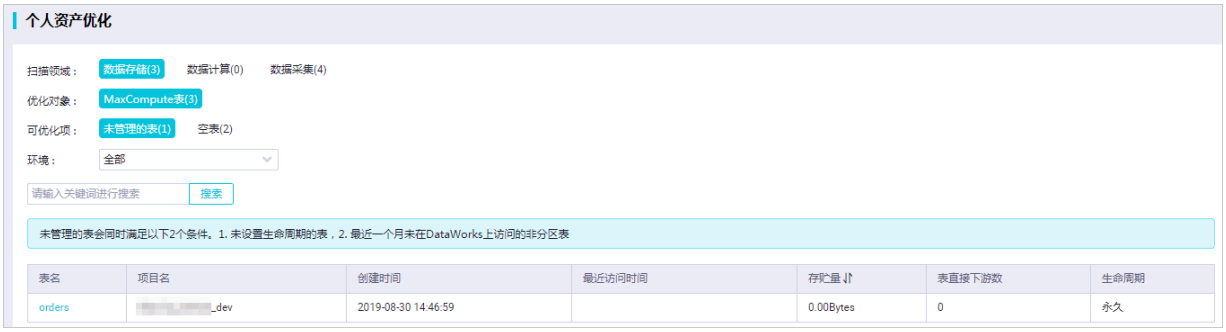
> **说明:**
> 优化趋势中的数据是离线计算生成的，您可以查看最新的日期，以获取数据的最近更新时间。

您可以单击顶部工作空间下拉框，选择相应的工作空间，也可以选择我的所有项目。



查看个人资产优化页面

个人资产优化从数据存储、数据计算和数据采集3个领域进行扫描。目前，数据存储和数据计算扫描的是MaxCompute的表和SQL任务。数据采集扫描的是写入至MaxCompute中的同步任务。



> **说明:**

DataWorks支持开发环境和生产环境隔离的标准工作空间模式，即1个DataWorks工作空间支持底层有2个MaxCompute项目，此时您可以通过环境进行筛选。

扫描领域	优化对象	可优化项	说明
数据存储	MaxCompute表	未管理的表	未管理的表需要满足以下2个校验条件： · 未设置生命周期的表。 · 最近一个月未在DataWorks上访问的非分区表。 同时满足上述条件的表，会被扫描出来。针对上述扫描条件，您可以通过设置表的生命周期，解决上述扫描问题。表的生命周期详情请参见#unique_73。 说明: 表的生命周期到期后，会回收表数据，请谨慎操作。
		空表	存储量为0的表即为空表。不建议您直接删除表，推荐您根据表的创建时间，对早期创建的表进行审计。
数据计算	MaxCompute任务	冲突任务	与其它任务写入同一张表，可能导致非预期结果。在数据开发过程中，不建议多个任务操作同一张表的同一个分区，避免其中一个任务报错，补数据时出现数据质量问题。 开发任务时，需要考虑数据幂等性问题。建议根据任务的下游数量，暂停其中一个任务，调整暂停任务下游节点的依赖关系。
		数据倾斜	产生数据倾斜的根本原因是：有少数fuxi实例处理的数据量超过其它实例处理的数据量，导致少数实例的运行时长超过其它实例的平均运行时长，从而导致整个任务的运行时间较长，造成任务延迟。 数据倾斜的解决方案请参见#unique_74。

扫描领域	优化对象	可优化项	说明
数据采集	同步任务	持续导入一致	扫描出连续15天导入数据量持续一致的数据同步任务，请关注源数据是否不再更新。 此时需要考虑业务是否已暂停。如果业务暂时停止，建议暂停任务的调度，以节省调度、计算和存储等资源。
		导入为空	扫描出导入数据量持续为0的数据同步任务，您可以暂停或下线该任务。
		同源导入	扫描出有相同数据源的数据同步任务，该类任务会存在重复导入MaxCompute的情况，导致存储及调度资源的浪费。您可以合并作业。
		OSS同步优化	扫描出通过公网进行OSS数据传输的数据同步任务，该类任务会导致额外的网络费用。 您可以进入工作空间管理 > 数据源管理页面，修改数据源的Endpoint地址为内网地址，即可减少公网流量费用、提升传输速度。

6.3 工作空间资产优化

工作空间资产优化页面主要以项目管理员为维度，为项目管理员展示指定工作空间下的可优化项信息，以及可优化计算排行和可优化存储排行。

说明:

可优化计算排行和可优化存储排行最多展示10行信息。

进入工作空间资产优化页面

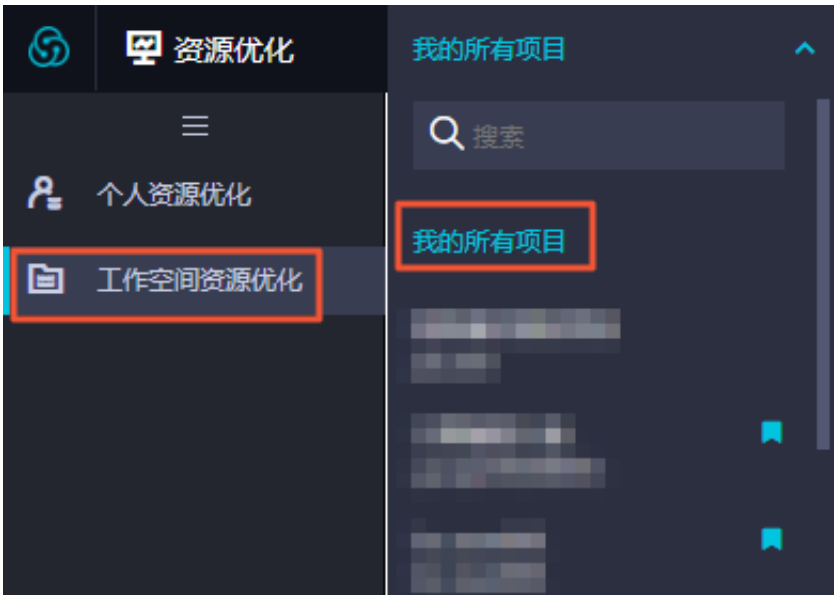
1. 登录DataWorks控制台，单击相应工作空间后的进入数据开发。
2. 单击左上角的图标，选择全部产品 > 资源优化。

3. 单击左侧菜单栏中的工作空间资产优化，即可进入工作空间资产优化页面，查看工作空间下的总实例数总表数。

优化趋势为您展示最近10天内可优化项的变化趋势，您可以在此查看最近完成可优化项的数量。

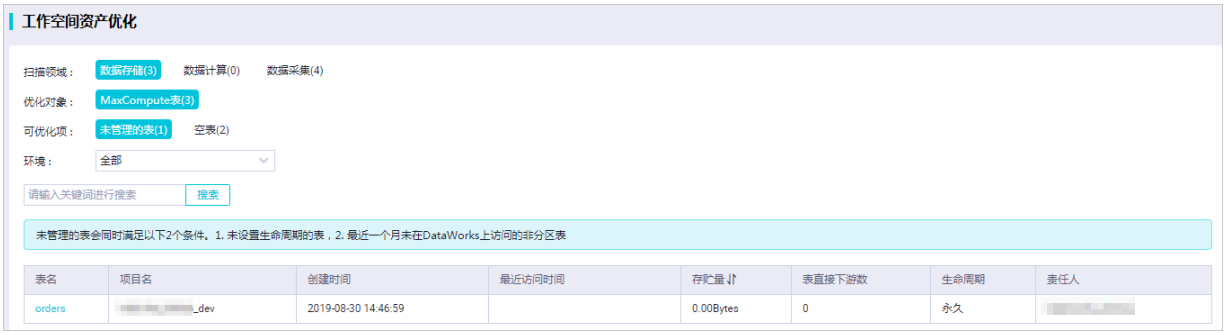
> **说明:**
> 优化趋势中的数据是离线计算生成的，您可以查看最新的日期，以获取数据的最近更新时间。

您可以单击顶部工作空间下拉框，选择相应的工作空间，也可以选择我的所有项目。



查看工作空间资产优化页面

工作空间资产优化和个人资产优化的可优化项基本一致，只是查看的视角不同。



扫描领域	优化对象	可优化项	说明
数据存储	MaxCompute表	未管理的表	未管理的表需要满足以下2个校验条件： · 未设置生命周期的表。 · 最近一个月未在DataWorks上访问的非分区表。 同时满足上述条件的表，会被扫描出来。针对上述扫描条件，您可以通过设置表的生命周期，解决上述扫描问题。表的生命周期详情请参见#unique_73。 说明: 表的生命周期到期后，会回收表数据，请谨慎操作。
		空表	存储量为0的表即为空表。不建议您直接删除表，推荐您根据表的创建时间，对早期创建的表进行审计。
数据计算	MaxCompute任务	冲突任务	扫描出与其它任务写入同一张表，会导致非预期结果的任务。
		数据倾斜	扫描出存在部分实例的处理数据量，及时间超过其它实例的情况，会导致整体任务执行时间变长的任务。
数据采集	同步任务	持续导入一致	扫描出连续15天导入数据量持续一致的数据同步任务，请关注源数据是否不再更新。 此时需要考虑业务是否已暂停。如果业务暂时停止，建议暂停任务的调度，以节省调度、计算和存储等资源。
		导入为空	扫描出导入数据量持续为0的数据同步任务，您可以暂停或下线该任务。
		同源导入	扫描出有相同数据源的数据同步任务，该类任务会存在重复导入MaxCompute的情况，导致存储及调度资源的浪费。您可以合并作业。

扫描领域	优化对象	可优化项	说明
		OSS同步优化	扫描出通过公网进行OSS数据传输的数据同步任务，该类任务会导致额外的网络费用。 您可以进入工作空间管理 > 数据源管理页面，修改数据源的Endpoint地址为内网地址，即可减少公网流量费用、提升传输速度。