

政务数据资源管理平台

V1.0

用户使用手册

2021.10

目 录

一、文档概述.....	3
二、功能介绍.....	3
1 平台管理.....	3
1.1 节点管理.....	3
1.2 许可证管理.....	4
1.3 引擎服务.....	5
1.4 主题设置.....	12
2 项目管理.....	14
2.1 用户列表.....	15
2.2 项目列表.....	17
3 项目切换.....	18
4 项目设置.....	19
4.1 项目信息.....	19
4.2 用户权限管理.....	19
4.3 日志管理.....	27
4.4 页面管理.....	29
5 元数据管理.....	32
5.1 数据源.....	32
5.2 元数据开发.....	37
5.3 规范中心.....	47
6 数据采集.....	55
6.1 库表采集.....	55
6.2 API 接口方式.....	73
7 数据治理.....	74
7.1 表输入.....	77
7.2 质量探查.....	78
7.3 标准化.....	79
7.4 基础加工.....	82
7.5 SQL 加工.....	82
7.6 表输出.....	84
8 数据资源管理.....	89
8.1 基础信息库.....	89
8.2 专题信息库.....	91
9 任务运维.....	93
9.1 作业总览.....	93
9.2 实时运维.....	95
9.3 周期运维.....	96
9.4 告警管理.....	102
10 数据资产.....	104
10.1 资产概览.....	104
10.2 资产目录.....	105

10.3 问题数据.....	110
11 数据共享.....	111
11.1 专题管理.....	111
11.2 发布管理.....	120
12 数据门户.....	124
12.1 首页.....	124
12.2 新闻政策.....	125
12.3 资源目录.....	126
12.4 应用成效.....	130
12.5 个人中心.....	137
三、关于德拓.....	138



一、文档概述

该文档主要内容为政务数据资源平台的快速入门和使用指导，旨在帮助用户快速掌握政务数据资源平台的使用方法。

二、功能介绍

1 平台管理

平台默认存在一个平台管理员 admin，初始密码 123456。



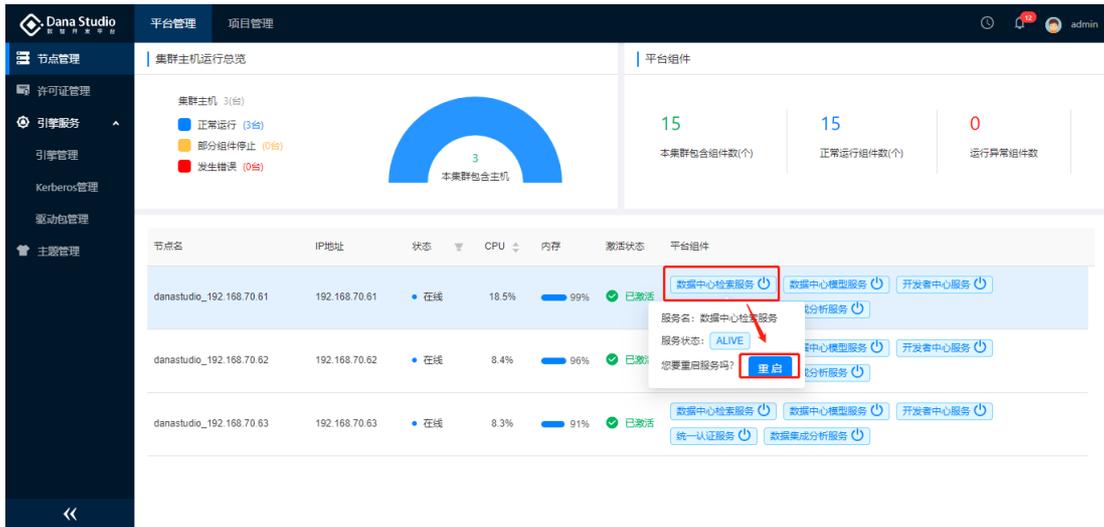
注：使用了非 Chrome 浏览器和低版本浏览器登录政务数据资源平台，登录页面上方会给出提示且通过提示信息会给到谷歌浏览器下载的官网。

1.1 节点管理

提供了整个数据资源平台的状态总览，包括集群的整体运行状况、节点的运行压力、服务组件的状态运维。

重启服务：

平台管理员登录——平台管理——节点管理——点击服务名——重启服务

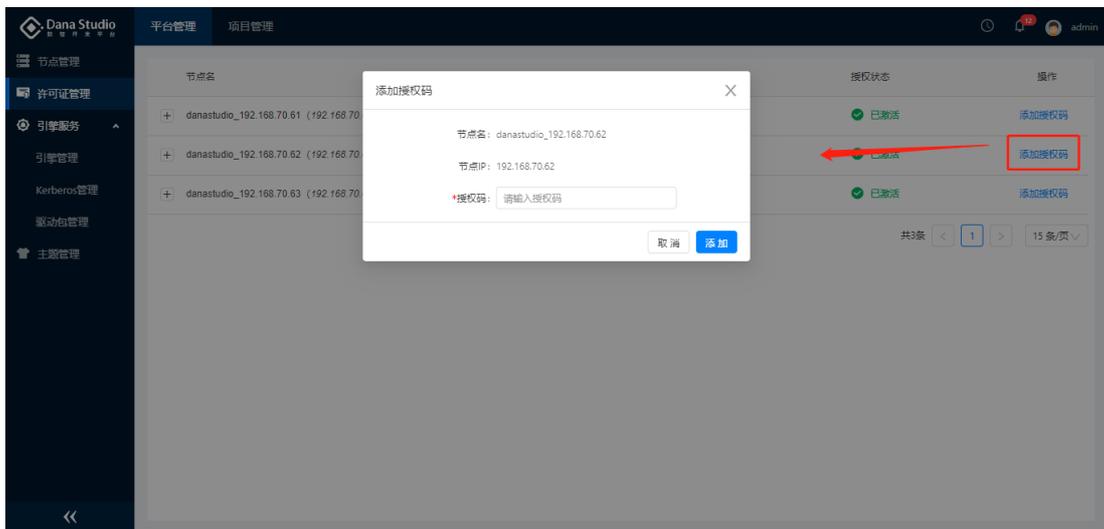


1.2 许可证管理

用于注册和激活许可证，管理员向 datatom 申请激活码和注册码后激活平台以使用完整功能。

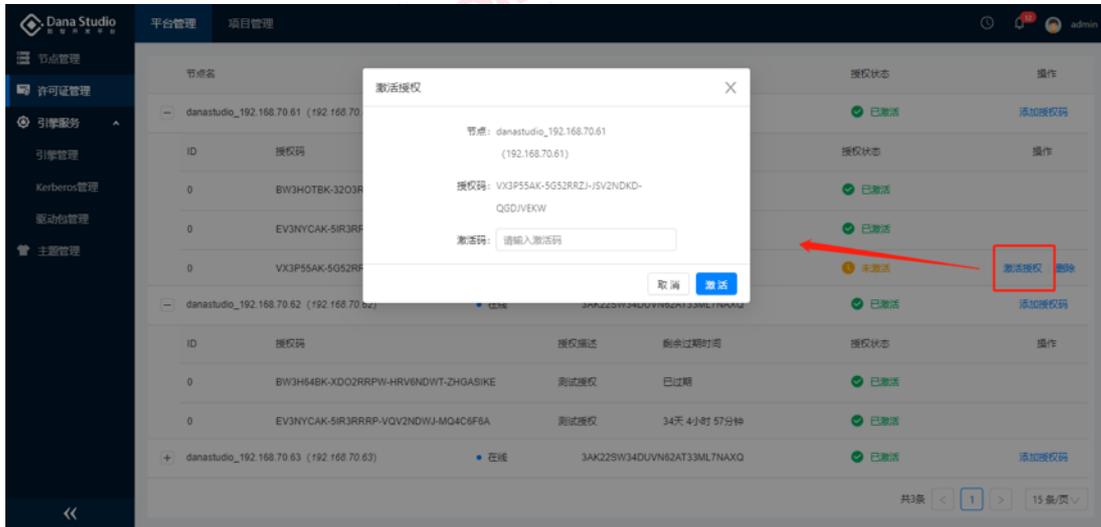
添加授权码：

平台管理员登录——平台管理——许可证管理——添加授权码——填写授权码——确定



添加激活码：

点击“激活授权”按钮——输入激活码——确定激活



1.3 引擎服务

引擎服务中决定了平台所有可用的引擎资源。注册成功的引擎，能够被其他模块用。在引擎服务可以注册各类的引擎资源，支持 datatom 自研的引擎：如 eagles, stork, teryx, dodox, spark 之外，还可以对接所有通用协议的 Hadoop 大数据平台的 hive（主要支持的版本有：DDP4.5、CDH6.2、TDH5.2、TDC2.1、华为 1.2.1 FusionInsight）、presto、hdfs、达梦，优炫以及 GaussDB 引擎等。同时通过对引擎的驱动管理，系统能同时适配多平台、多版本 hive 的对接。

大数据运维人员在 DM 上安装 DS 平台以及 stork, datax 和 dodox 等引擎，其他引擎的注册需用户或专业的大数据运维人员在 DS 平台进行安装，下面对这些引擎注册过程进行说明。

1.3.1 引擎管理

不同引擎注册需配置字段说明表

引擎类型	配置字段	备注
Stork Teryx GaussDB	IP 地址, 端口号, 初始数据库, 用户名, 密码	跟连接 数据库 相同的方式
Hive	驱动包, URL, 用户名, 密码, 配置文 件, Kerberos 认证, 配置参数, 配置项	注册详情见下表
Presto	主节点 IP, 备节点 IP, 端口号, catalog 路径, 存储路径, 执行文件路径	后续版本会去除这两个引擎注 册, 因为统一使用 MG 网关
spark	IP 地址, root 密码, 命令所在位置, kerberos 账户, keytab 路径	
达梦	IP 地址, 端口号, 用户名, 密码	跟连接 数据库 相同的方式
优炫	IP 地址, 端口号, 初始数据库, 用户名, 密码	跟连接 数据库 相同的方式

在 DanaManager 上为 DS 安装需要的组件

使用 DanaManager 的管理员账号 admin/admin 登录, 进入主界面--服务---添
加服务---勾选服务---点击“下一步”根据页面提示进行安装, 详细步骤可参考[安装部
署手册](#)

各组件作用说明如下:

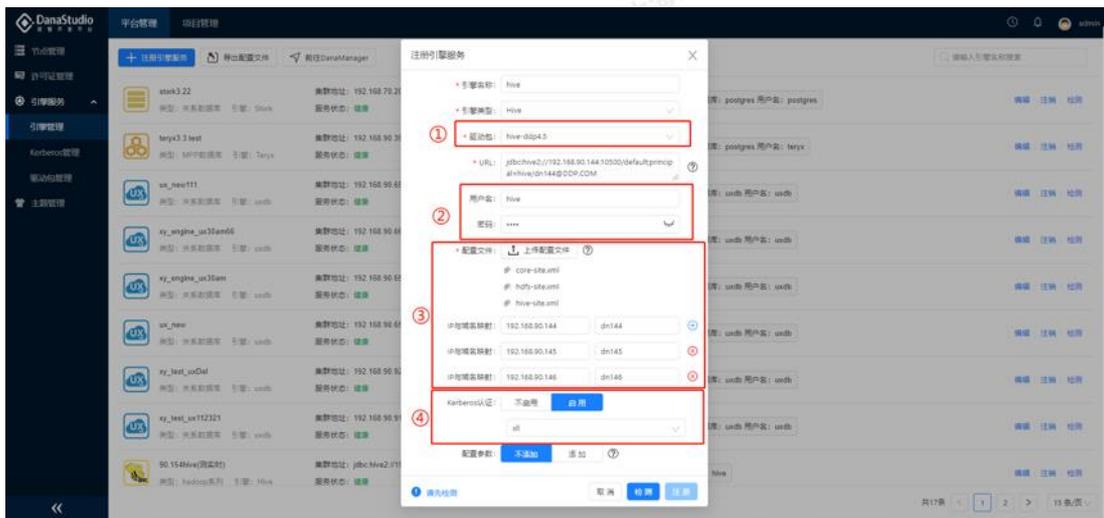
序号	产品	章节	组件	组件说明	是否必装	备注
①	DS	1.1	Eagles	存放 DS 数据	✓	
		1.2	Dodox	任务调度服务	✓	强依赖 Eagles
		1.3	Datax	离线采集服务	✓	
			DanaStu dio	数智开发平台	✓	强依赖 Eagles
			DanaGua rd	DS 中的告警服 务	✓	
			Metadata Gateway	DS 中的 MG 服 务	✓	
		1.4	Stork	数据库	✓	依赖 Eagles
		1.5	Kafka+Z	实时服务	✗	

			ookeeper			
			QuickSilver			强依赖于 kafka 和 zookeeper
②	HD	2	HDFS YARN+MapReduce2 TEZ Hive ZooKeeper Spark2	大数据基础引擎平台	✓	若 Zookeeper 已安装, 此处无需安装 Zookeeper
③	Teryx	3	teryx	数据库	✓	
DM 安装 DanaStudio: ① DM 安装 DanaStudio+DDP-HD: ①+② DM 安装 DanaStudio+Teryx: ①+③						

注: 通过 DM 安装的 dodox、datax、kafka、quicksilver 引擎, 不会在 ds 页面中展示

手动连接引擎:

平台管理员登录——平台管理——引擎服务——注册引擎服务——输入连接信息——检测——引擎注册

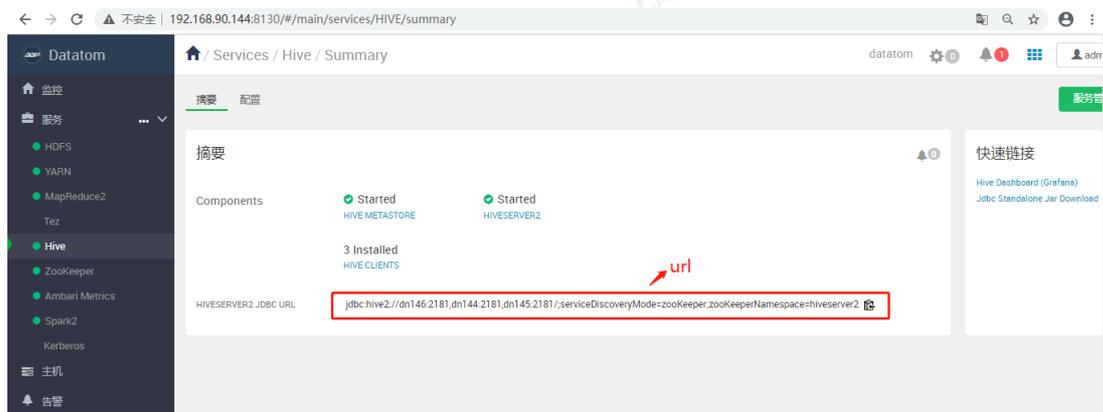


Hive 配置字段说明表如下:

字段名	字段说明
驱动包	选择对应驱动包，平台内置 hive-ddp4.5、hive-cdh6.2.0 等驱动包，也可在平台管理-驱动包管理中自己上传驱动包
URL	hive 元数据 IP，从 DanaManager 平台或大数据运维人员处直接获取 注：如果有多个节点需将所有节点的 ip 和端口号都写明，如： jdbc:hive2://dn146:2181,dn144:2181,dn145:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=hiveserver2
用户名 密码	若 hive 引擎开启密码认证，则需要填写 hive 的用户名和密码；未开启则无需填写
配置文件	要求文件名称与格式保持一致，core-site.xml、hdfs-site.xml、hive-site.xml
IP 与域名映射	如果配置文件中使用了域名，需要配置 IP 与域名映射，IP 若是集群需要填写完整
Kerberos 认证	hive 引擎若开启 kerberos 认证，则需要配置 kerberos。可以在平台管理-kerberos 管理模块配置
配置项	设置大数据平台支持的参数。若不配置，系统以配置文件中的参数访问 Hive 平台 支持对配置文件中已有参数重新设值，系统会优先以用户设定的参数值访问 Hive 平台 支持增加配置文件之外的参数，系统以配置文件及新增的参数访问 Hive 平台

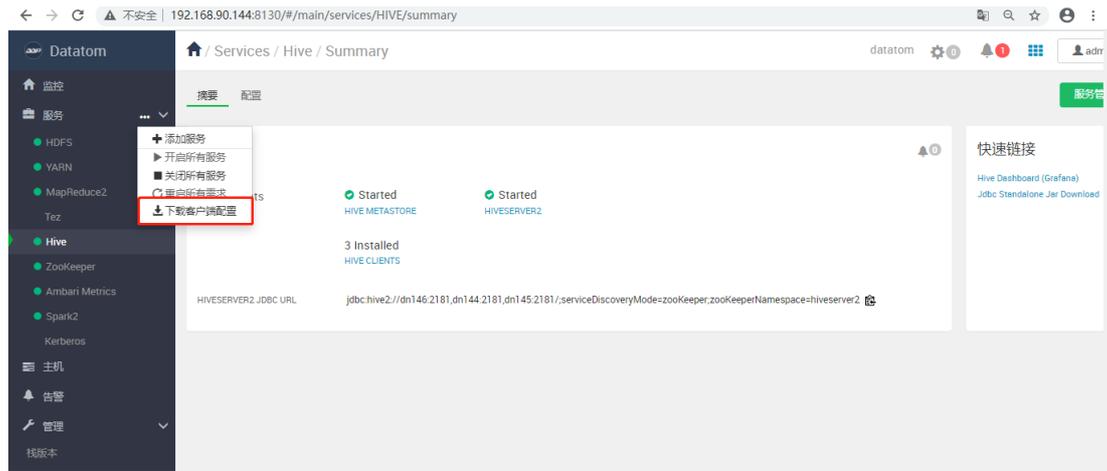
具体连接信息可在以下位置获取：

url:



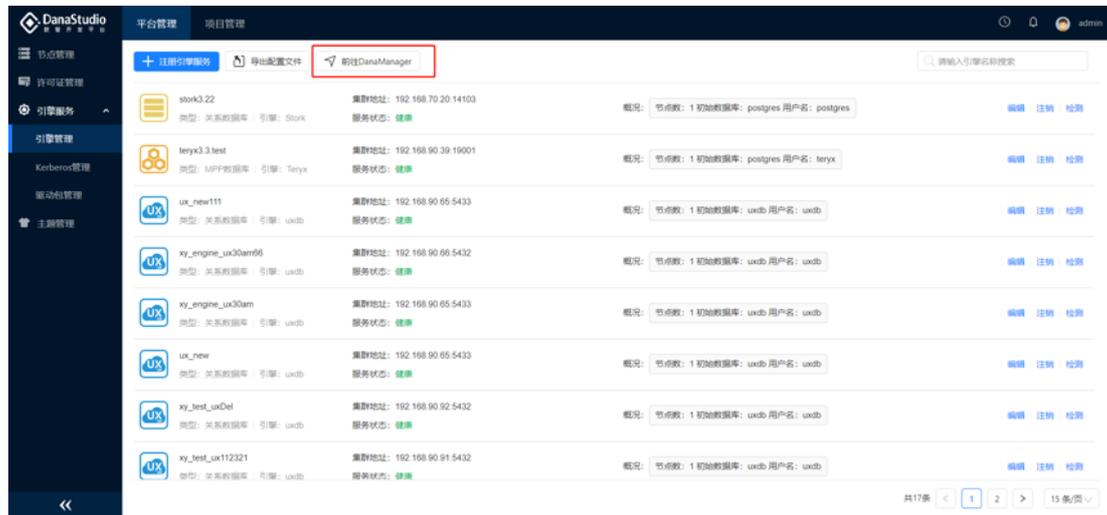
配置文件:

找到 core-site.xml、hdfs-site.xml、hive-site.xml 三个文件并在 DS 平台上上传



前往 DanaManager:

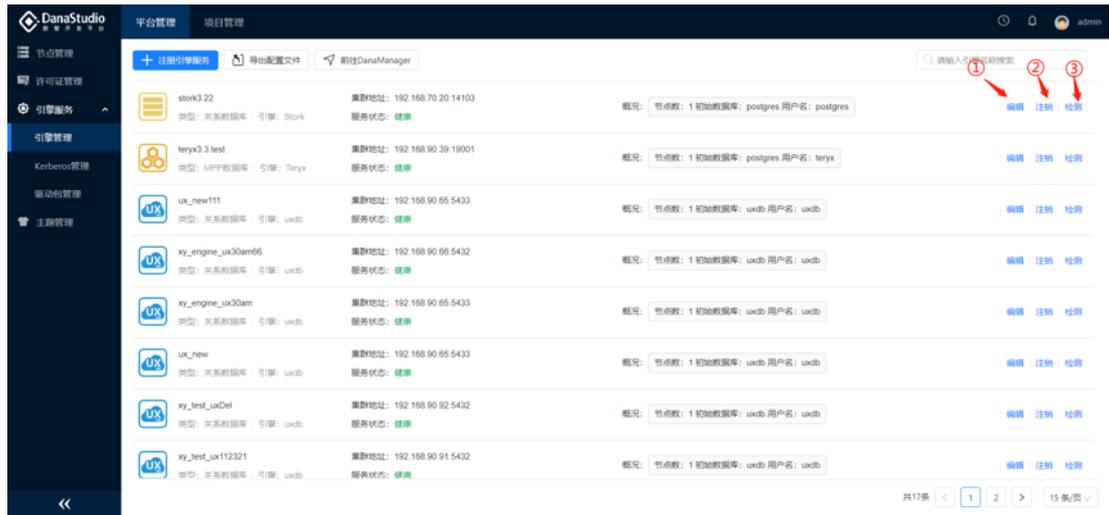
平台管理——引擎服务——前往 DanaManager



设置引擎:

对添加的引擎服务可以修改、注销和检测的操作。

平台管理——引擎服务——选择某一服务，可选择修改、注销、检测操作



- ①点击“编辑”按钮，可以对该引擎的信息进行编辑修改
- ②点击“注销”按钮，类似删除操作，将该引擎服务在本节点进行删除
- ③点击“检测”按钮，可以查询服务状态

1.3.2 Kerberos 管理

Kerberos 是秘密密钥网络认证协议，使用数据加密标准加密算法进行加密和认证。Kerberos 是为对网络资源的请求进行认证而设计的。与其他秘密密钥系统一样，Kerberos 基于可信任的第三方概念，这个第三方对用户和服务执行安全认证。

如果用户需要开启 Kerberos 认证，则需使用平台管理员账号新增 Kerberos 认证信息。

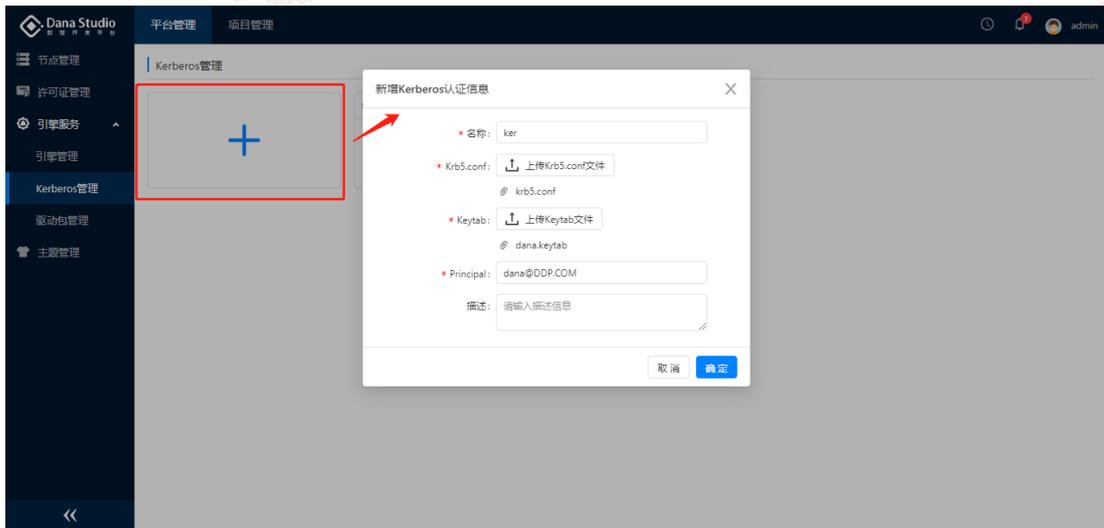
Kerberos 认证信息必须上传 Krb5.conf 配置文件和 Keytab 密钥表文件，及填写 Kerberos 认证名称和 Principal 认证账户名称。您为不同的引擎服务开启 Kerberos 认证时，需在该引擎的配置文件下找到 Krb5.conf、Keytab 和 Principal 三个字段的的信息，一般会存放在-etc 的文件目录下，但不同的引擎其存放的目录可能不一致，您可以咨询专业的大数据运维人员直接获取。

添加 kerberos:

平台管理员登录——平台管理——引擎服务——Kerberos 管理——上传

kerberos——配置 krb5.conf、keytab 等信息——确定

注: 已添加的 kerberos 可在注册 hive 引擎时被选择



1.3.3 驱动包管理

为了应对不同版本的 Hive 引擎（CDH、TDC 和 DDP 等）注册服务的需求，需先将不同版本的 Hive 引擎注册所需的驱动包上传至 DanaStudio 平台。平台内置 hive-v1、hive-v2、hive-v3、hive-ddp4、hive-ddp4.0、hive-ddp4.5、hive-cdh6.2.0 的驱动包，支持手动上传驱动包。

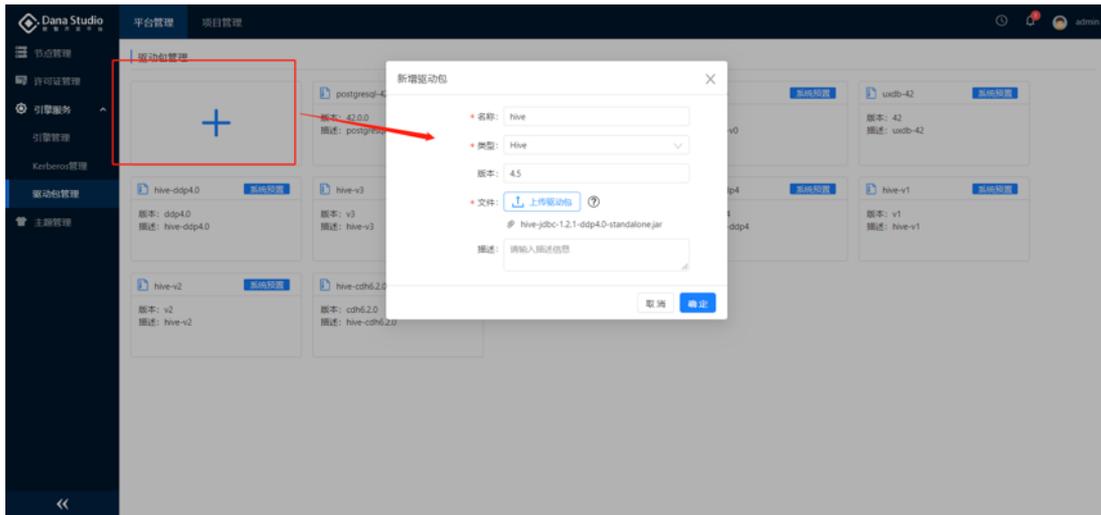
您可在不同平台下分别获取驱动包或咨询大数据运维人员直接获取。

添加驱动包:

平台管理员登录——平台管理——引擎服务——Kerberos 管理——新建驱动包

——上传驱动包——确定

注: 已添加的驱动包可在注册 hive 引擎时被选

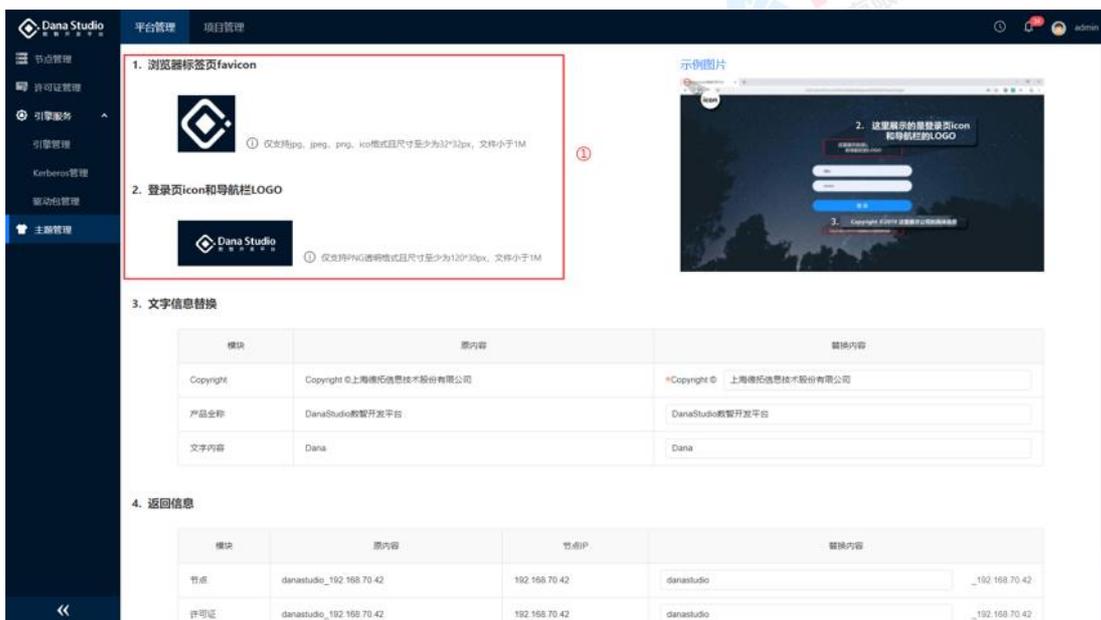


1.4 主题设置

管理员可以在该页面进行更换产品品牌信息相关图片、LOGO、文字和返回信息。

设置主题：

平台管理员登录——平台管理——主题设置——修改产品相关信息——保存



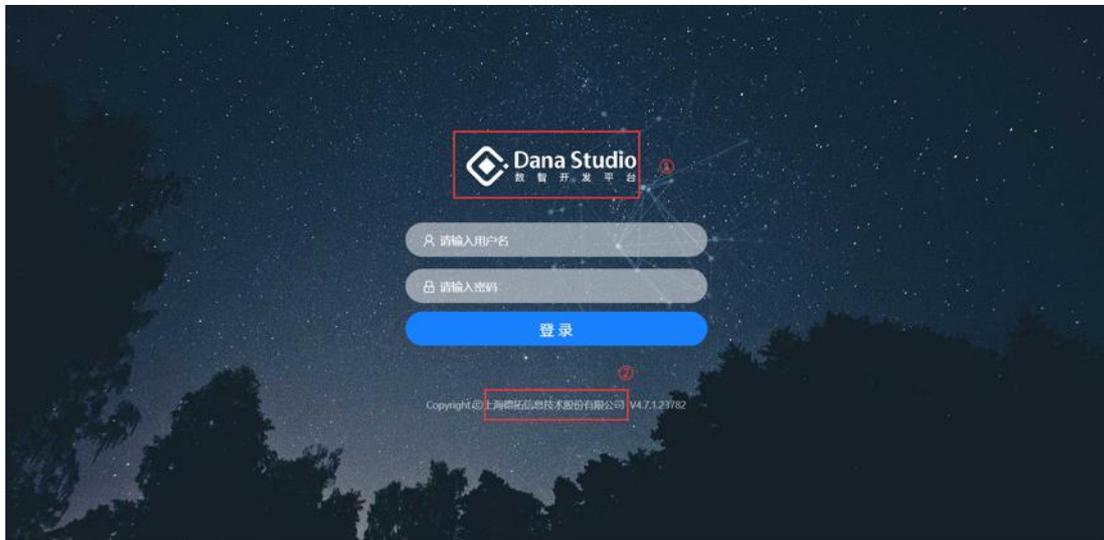
①支持鼠标悬浮于图片上，点击出现的上传按钮进行图片上传或通过直接拖拽图片进行上传

②在集群各节点的/opt/dana/web/host/oem/1 路径下会保存上传成功的浏览器标签页 favicon，且自动将图片转化为 ico 格式

注：保存修改信息成功后，3 秒后自动登出平台

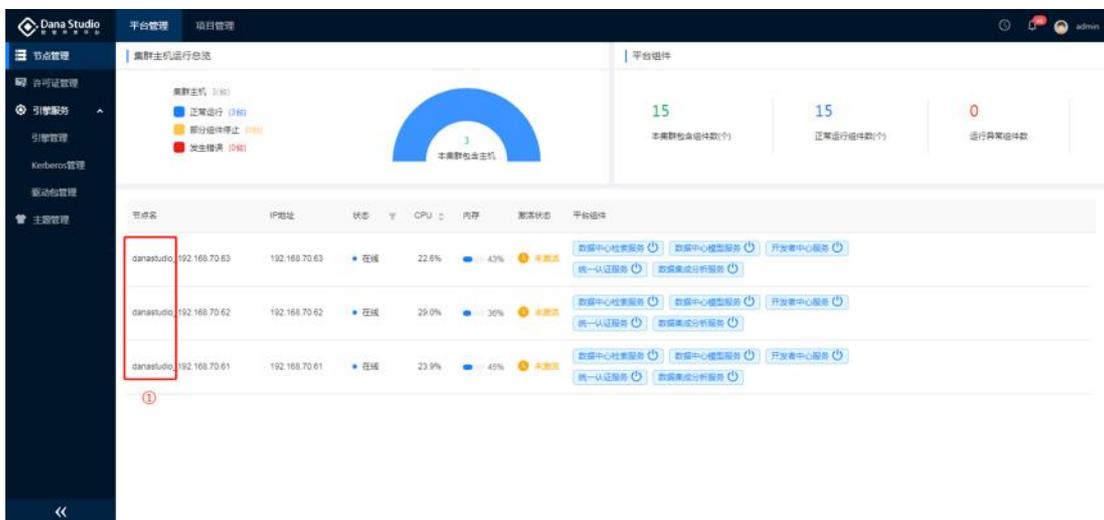
查看修改信息：

信息修改并保存后，可以在对应页面查看到修改后的产品信息



①展示在主题设置页面中修改的登录页图标

②展示在主题设置页面中修改的产品名称



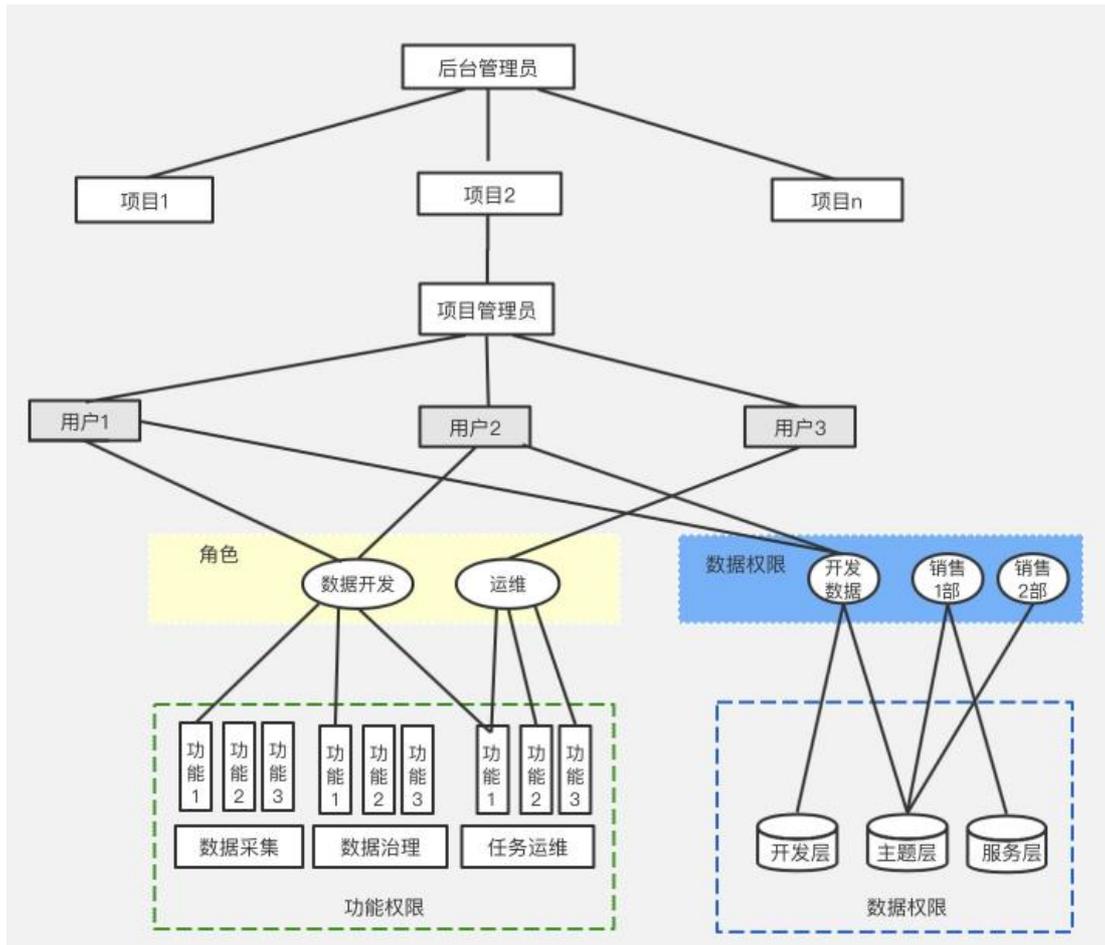
①展示在主题设置页面中修改的节点信息

节点名	状态	机器码	授权状态	操作
danastudio_192.168.70.61 (192.168.70.61)	在线	3AK22SW34DUVN62AT33ML7NAXQ	已激活	添加授权码
danastudio_192.168.70.63 (192.168.70.63)	在线	3AK22SW34DUVN62AT33ML7NAXQ	已激活	添加授权码
danastudio_192.168.70.62 (192.168.70.62)	在线	3AK22SW34DUVN62AT33ML7NAXQ	已激活	添加授权码

①展示在主题设置页面中修改的许可证替换信息

2 项目管理

功能、数据权限都是以项目组织并隔离的。不同项目之间的功能、数据都是相互隔离的。用户只有加入该项目并被分配权限后，才能使用功能产生的数据。项目由平台管理员创建，项目内的权限则由项目管理员管理。

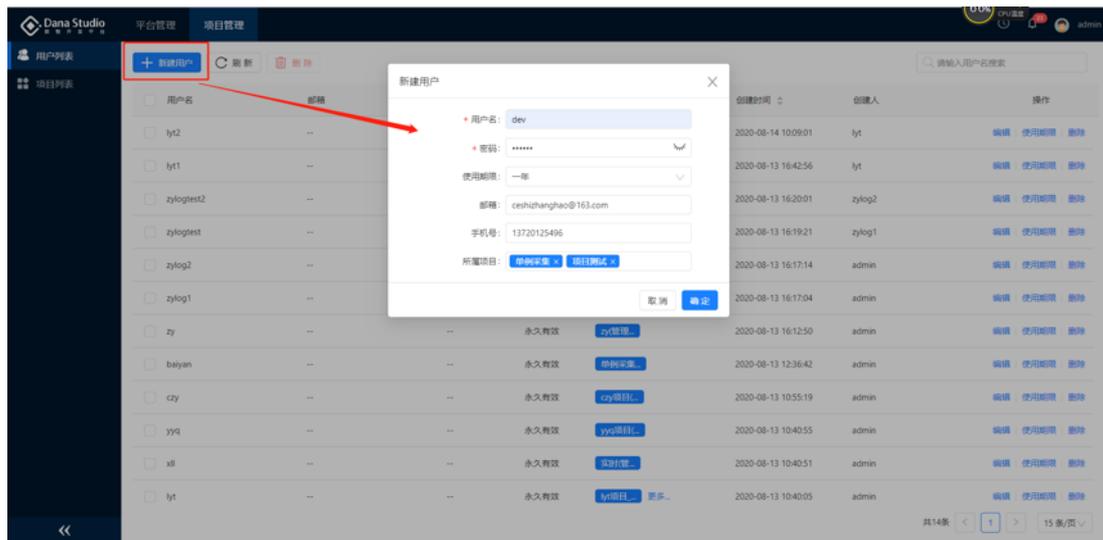


2.1 用户列表

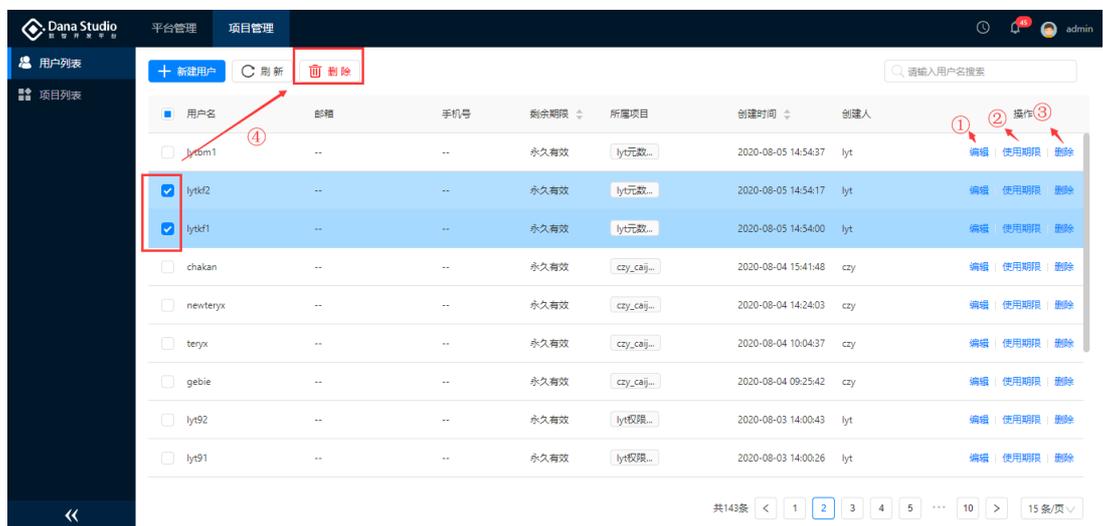
该模块可以对平台中所有用户进行管理，包括管理用户所属项目和使用期限等信息。

新建用户：

平台管理员登录——项目管理——用户列表——新建用户



查看用户列表：



- ①点击操作下的“编辑”按钮，可以编辑用户的基本信息：用户名、密码、邮箱、手机号和所属项目
- ②点击操作下的“使用期限”按钮，编辑用户使用期限，用户到期后，不可再登录平台
- ③点击操作下的“删除”按钮，删除用户
- ④勾选多个用户，点击上方“删除”按钮，批量删除用户

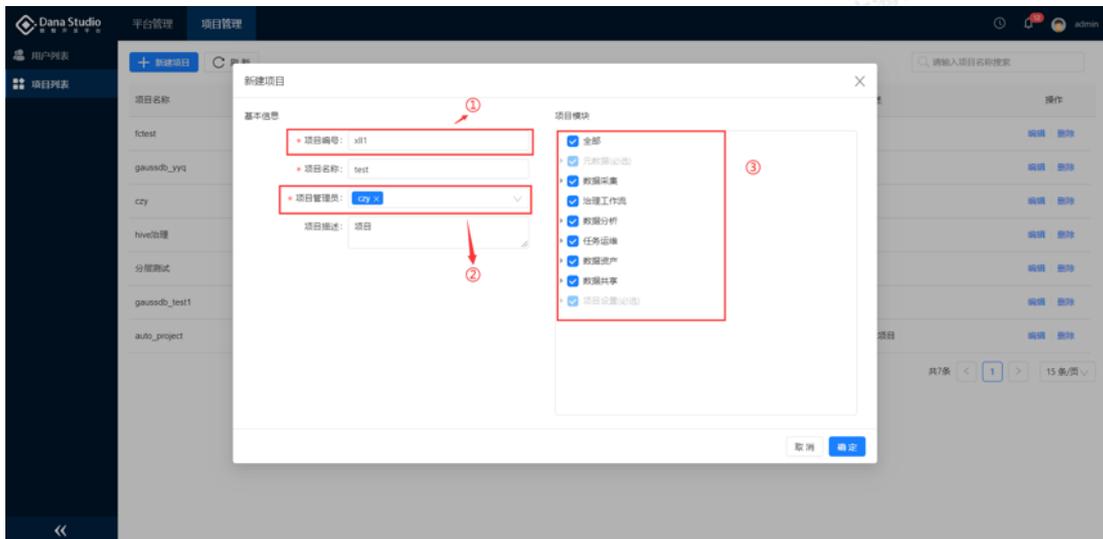
注：当删除的用户为某项目管理员，而该项目仅有这一个项目管理员，则不允许删除该用户

2.2 项目列表

该模块对项目进行创建，并分配项目管理员。在创建项目时可以自由配置项目使用的平台功能模块。后续对项目的管理，包括用户、权限、数据都由项目管理员负责管理。

新建项目：

平台管理员登录——项目管理——项目列表——新建项目

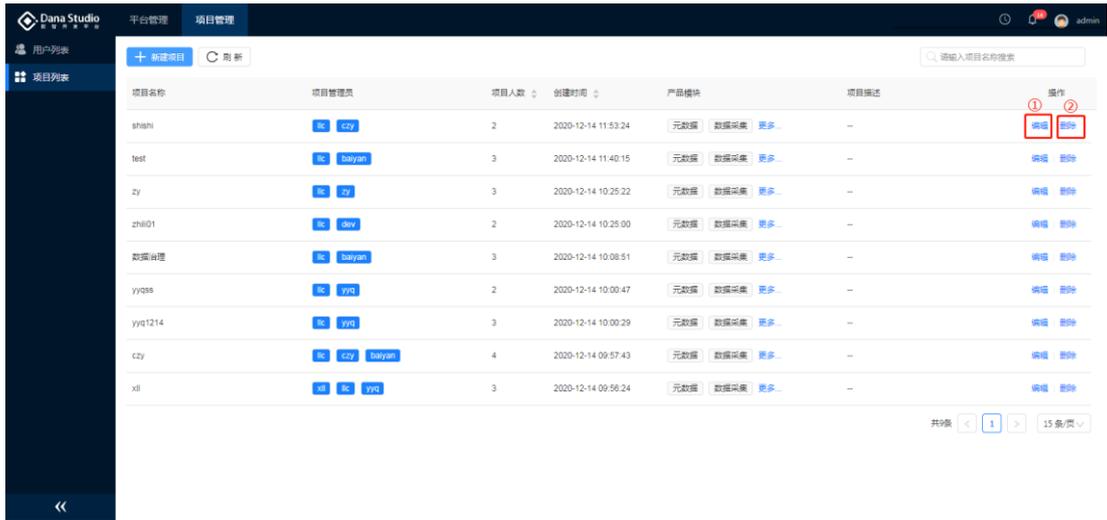


①项目编号：作为项目的唯一标识，后续为项目建库的时候需要自动加项目编号作为前缀，如 demo_odsdB (demo 为项目标识)，以作为各项目的数据权限隔离。

②项目管理员：用户被指定为项目管理员后，默认该用户具有系统预制的项目管理员的角色。项目管理员能操作该项目的所有功能、资源及数据，并能进行用户创建及权限管理。

③产品模块：指定项目的功能模块后，该项目只能使用设定的功能模块。其中元数据和项目设置模块为项目管理员默认能使用的功能模块。

查看项目列表



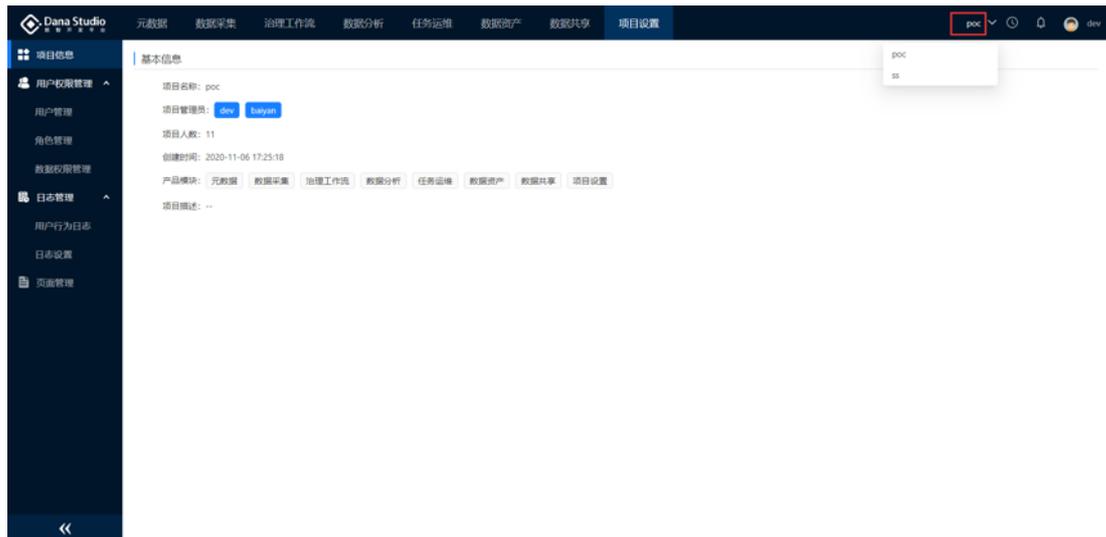
①点击操作下的“编辑”按钮，可以编辑项目信息：项目名称、项目管理员、产品模块和项目描述，其中项目编号不支持修改

②点击操作下的“删除”按钮，删除项目

注：删除项目后，该项目所有的资源、设置的用户权限都自动被删除，底层数据保留。

3 项目切换

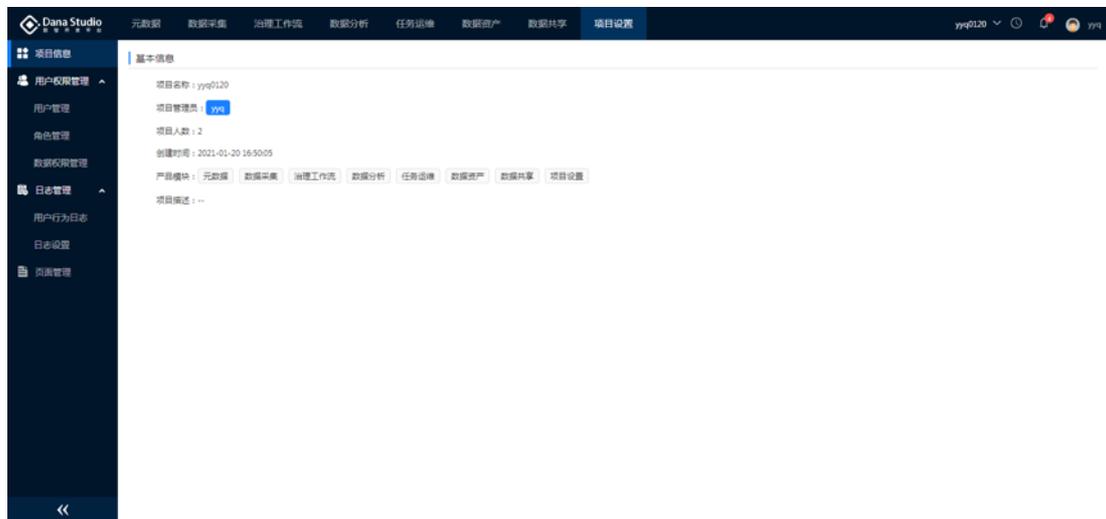
用户可以参与多个项目，平台提供自由切换项目功能，点击右上角项目名称，方便用户在不同项目间的自由切换。不同项目间，用户的能访问的功能、资源、数据都是彼此隔离的。



4 项目设置

4.1 项目信息

项目管理员可查看项目的基本信息。

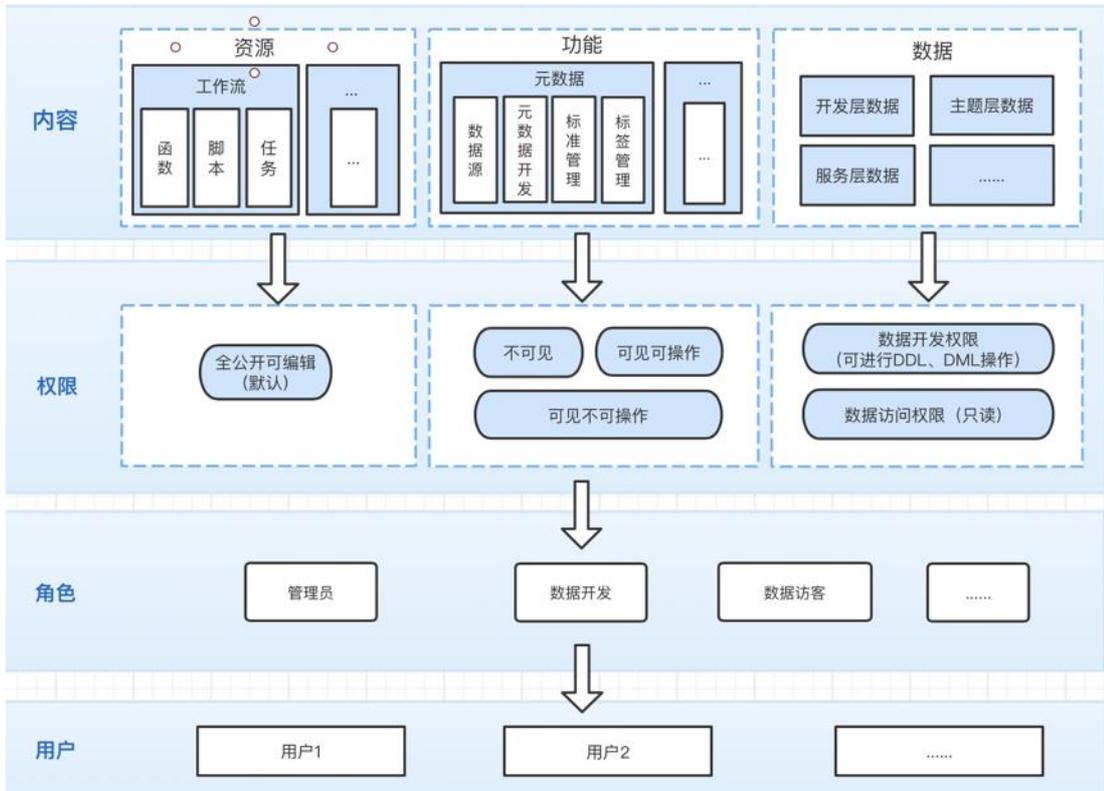


4.2 用户权限管理

平台提供灵活而细致的用户权限管理机制，既可配置系统功能权限，也可配置数据采集、元数据、 workflow、数据分析各模块的数据权限。通过角色管理实现对功能权限的控制，通过数据权限管理实现对数据权限的控制。

功能权限：对功能做权限控制，系统功能主要包括：平台管理、项目管理、项目设置、元数据、数据资产、数据采集、 workflow、数据分析、任务运维。

数据权限：对开发层、主题层、服务层、其他自定义层的数据权限做设置，能对这些数据做到库、表级别的权限管控。



4.2.1 用户管理

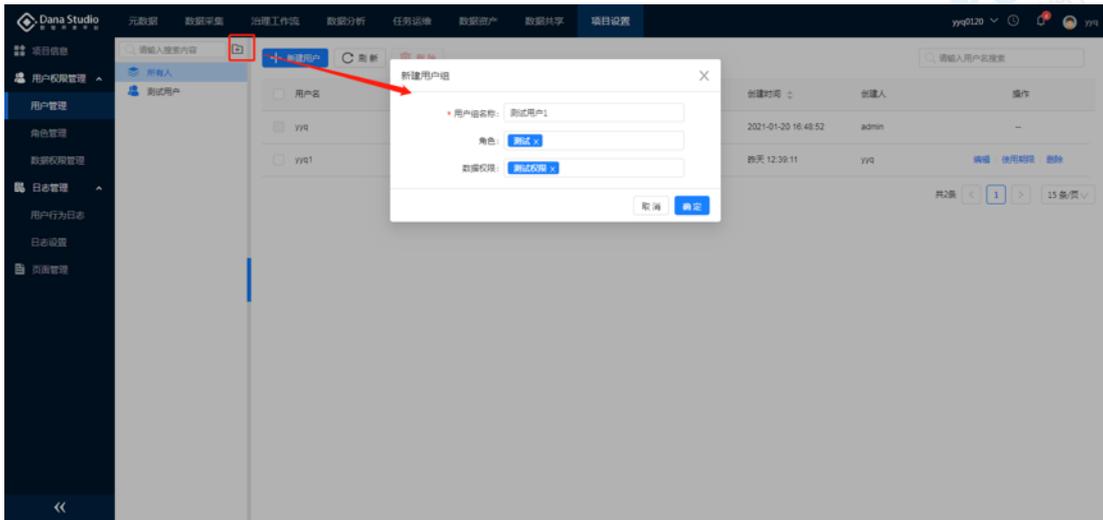
项目管理员可通过建立用户组实现对该项目下的所有用户的分组管理，也可在此模块新建用户并赋予对应权限。

新建用户组：

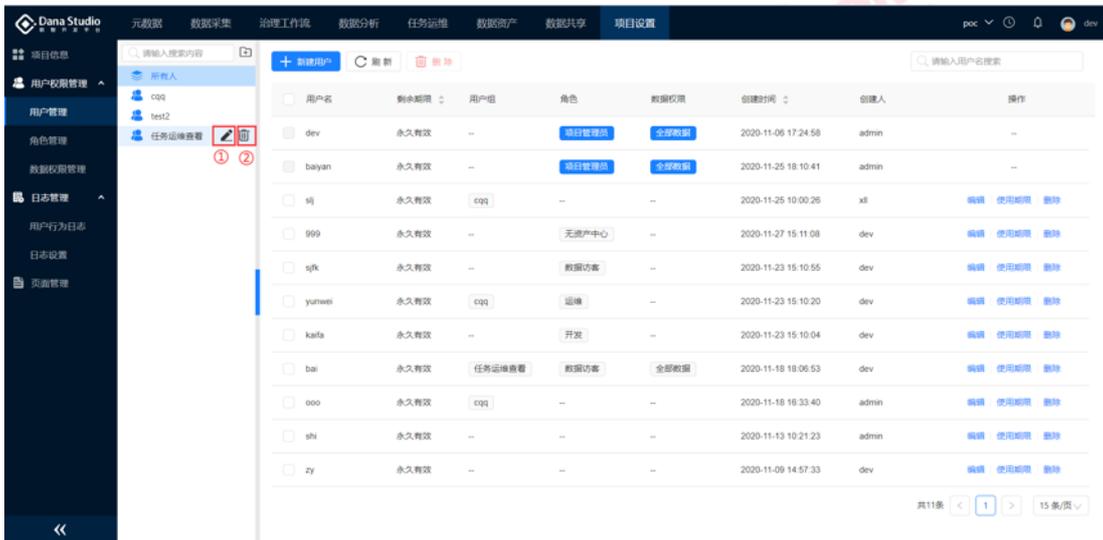
项目管理员——项目设置——用户管理——点击“新建用户组”图标，配置用户组的功能权限和数据权限

用户组设置多个角色、数据权限，该用户组中的所有用户将拥有该用户组所设置的权限。

注：角色和数据权限可分别在角色管理和数据权限管理模块中新增



修改、删除用户组：

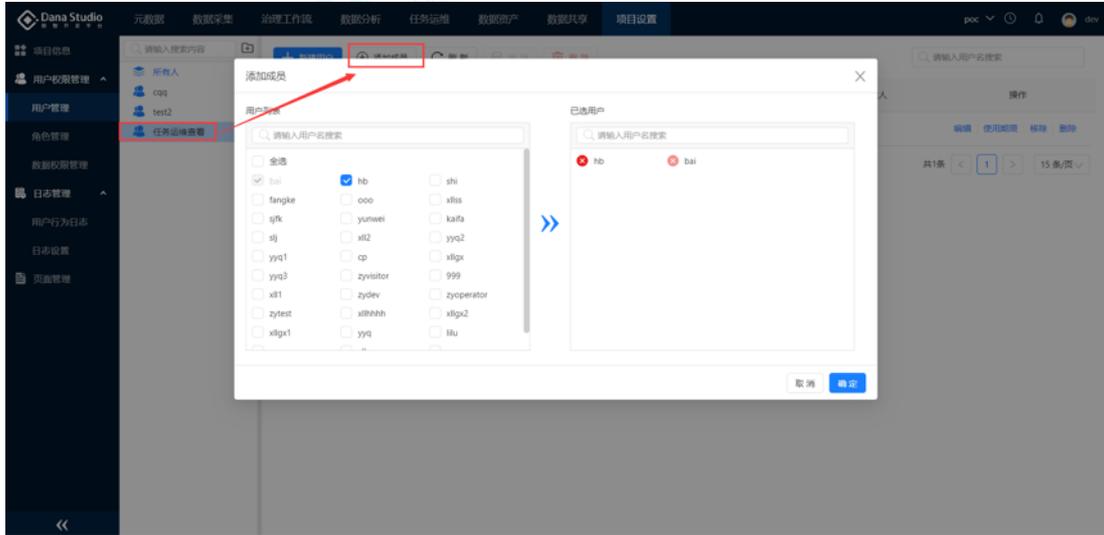


①点击用户组右侧的“编辑”按钮，可修改用户组的名称、角色和数据权限。

②点击用户组右侧的“删除”按钮，用户组被删除，基于用户组授权的权限都会被回收。原本属于该用户组下的用户，其该用户组的权限被收回。

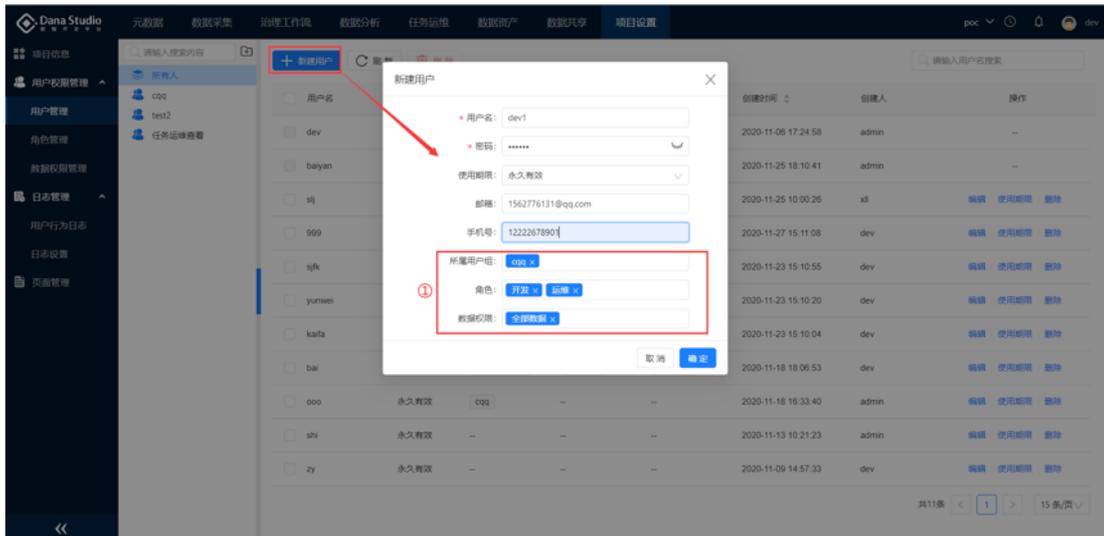
添加成员：

项目管理员——项目设置——用户管理——选中对应用户组——点击上方的“添加成员”，可批量将平台中其他项目的用户添加至该用户组，拥有该用户组的权限。



新建用户：

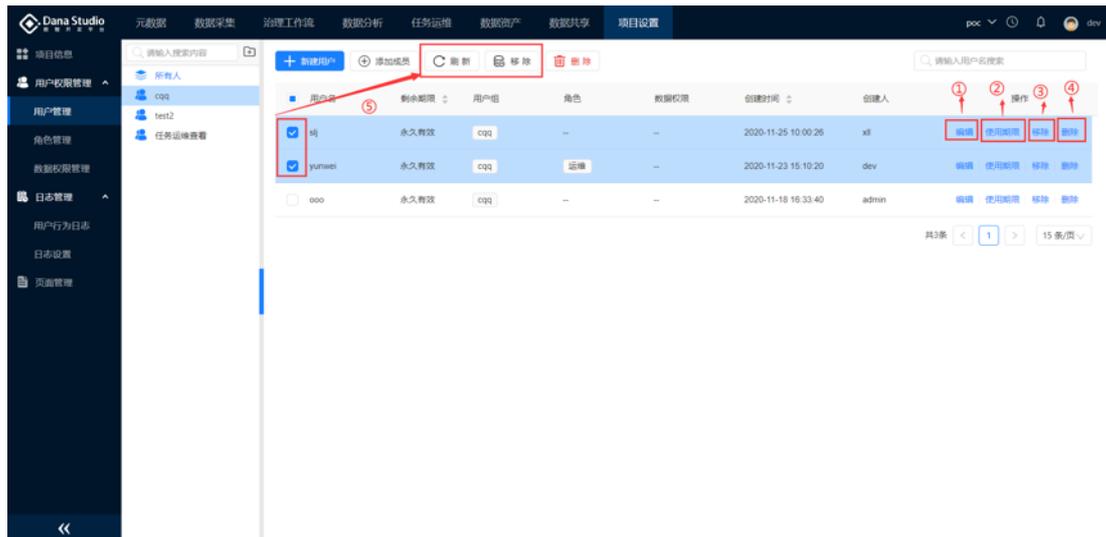
项目管理员——项目设置——用户管理——点击“新建用户”



①支持设置多个所属用户组、角色、数据权限，该用户所操作的功能模块权限、数据权限为所属用户组权限与用户本身的角色权限、数据权限的并集

注：角色和数据权限可分别在角色管理和数据权限管理模块中新增

查看用户列表：



①点击操作下的“编辑”按钮，编辑用户基本信息和权限信息

②点击操作下的“使用期限”按钮，编辑用户使用期限

③点击操作下的“移除”按钮，将该用户从此用户组中移除，但该用户仍在该项目中

④点击操作下的“删除”按钮，将该用户从平台中删除，若该用户加入其他项目，则无法删除

⑤支持批量移除和批量删除操作

注：项目管理员无法被删除和移除

4.2.2 角色管理

通过角色来实现对功能权限的管控，平台预制了丰富的角色以满足数据开发、数据管理过程中的不同人员分工需求，同时支持自定义角色，能灵活按需进行功能权限设置。

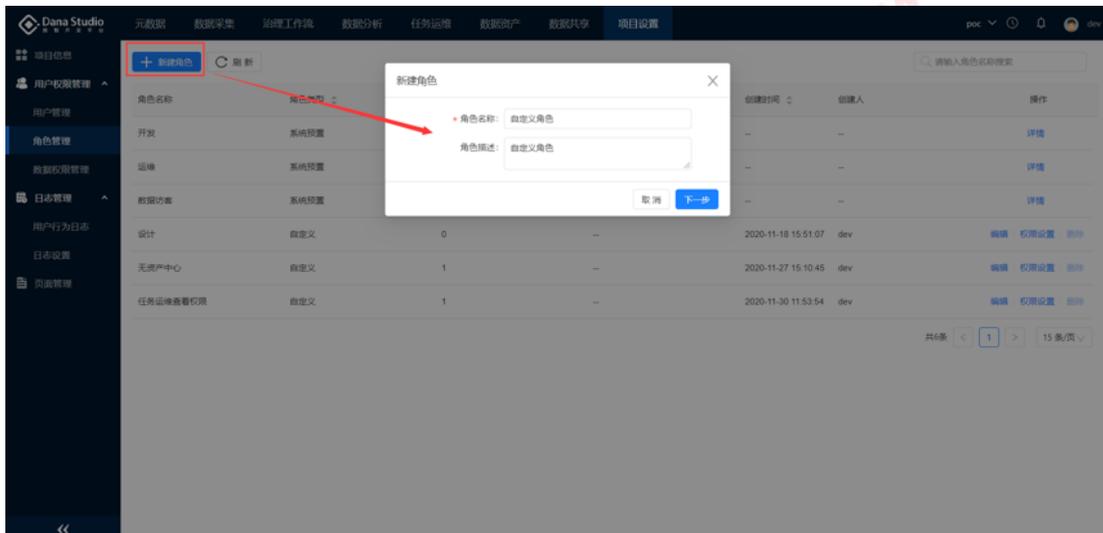
预制角色：预制的角色是系统级别的，所有项目均可使用，具体角色如下：

用户角色	角色描述
平台管理员	负责平台管理，平台管理员对平台的基础资源进行监管，包括平台组

	件、节点授权、引擎服务等全局资源的调度与分配；平台管理员还负责新建用户和项目，其中新建项目时需为项目指定项目管理员，设置项目拥有的功能权限等。
项目管理员	负责项目管理，项目管理员可基于项目创建用户（组），角色，数据权限，并进行用户授权，不同项目的管理员仅可使用所属项目内的功能、资源、数据。
开发	负责项目的数据开发工作。例如企业的治理专家、数据工程师及使用产品的用户。开发者在平台上对数据治理的全流程进行开发与监控，包括对数据采集任务、治理任务的配置与管理、任务运维信息等。
运维	负责对数据采集及治理任务进行运维管理。运维者主要具有查看详细信息、运维调度、资源监控等权限，对数据采集、数据治理和工作流等任务进行监控，帮助运维人员实现端到端的可视化作业调度。
数据访客	数据访客仅用于访问数据资产。
自定义角色	用户可根据业务需求新建自定义角色，并为角色赋予平台各功能模块的读写权限。

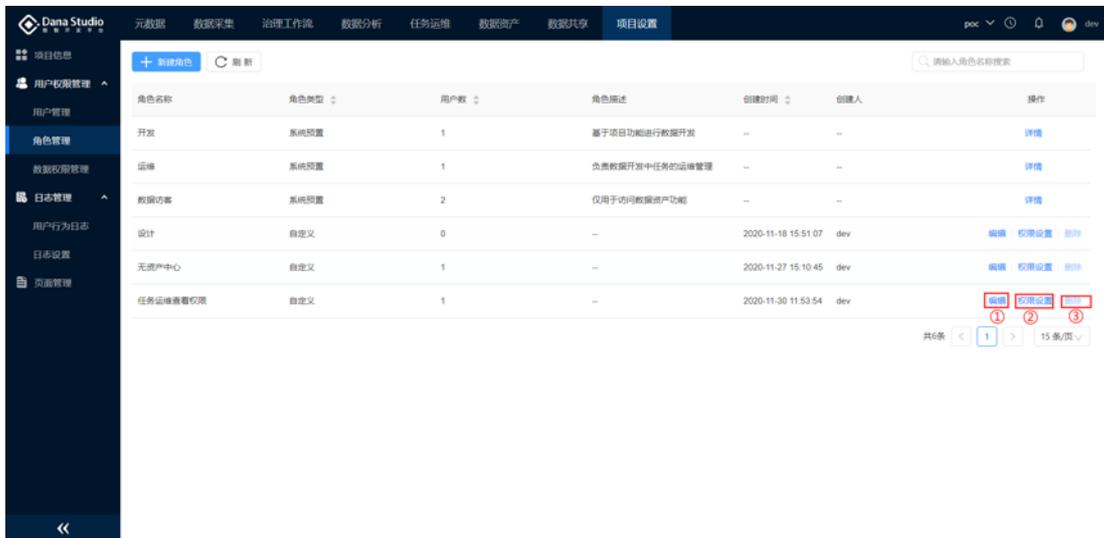
新建角色：

项目管理员——项目设置——角色管理——点击“新建角色”，填写角色名后，点击“下一步”配置角色的模块权限





查看角色列表：



- ①点击操作下的“编辑”，修改角色的名称和描述
- ②点击操作下的“权限设置”界面，编辑该角色的权限
- ③点击操作下的“删除”按钮，删除角色

注：系统预制的角色及被使用的角色不支持删除

4.2.3 数据权限管理

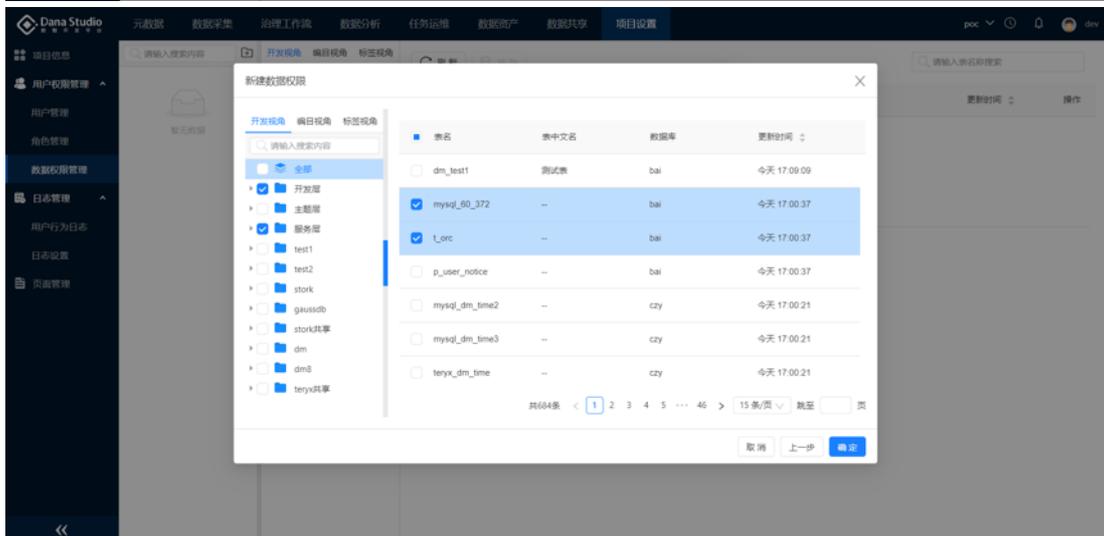
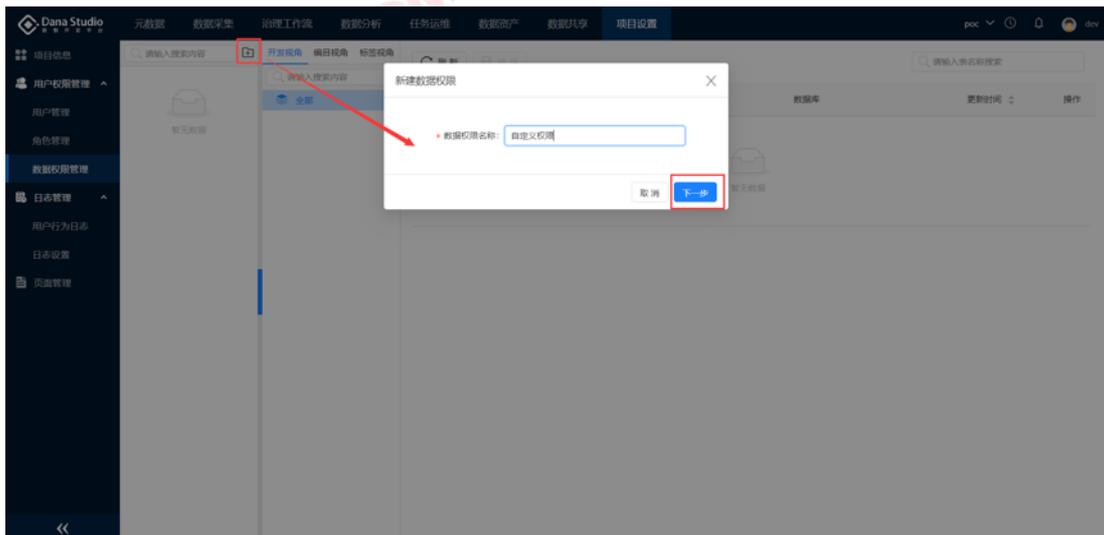
支持对用户设置数据访问权限。通过数据权限来控制对各层级数据的操作及访问权限。

平台提供三种视角来进行数据权限设置，开发视角、编目视角、标签视角，更方便的满足不同场景下的授权需求，如对于开发者而言，可以使用开发视角，对于数据资产查看时使用编目视角或标签视角更方便。当然，授权的都是同一份数据，无论哪种视角的授权都会同步到其他视角。

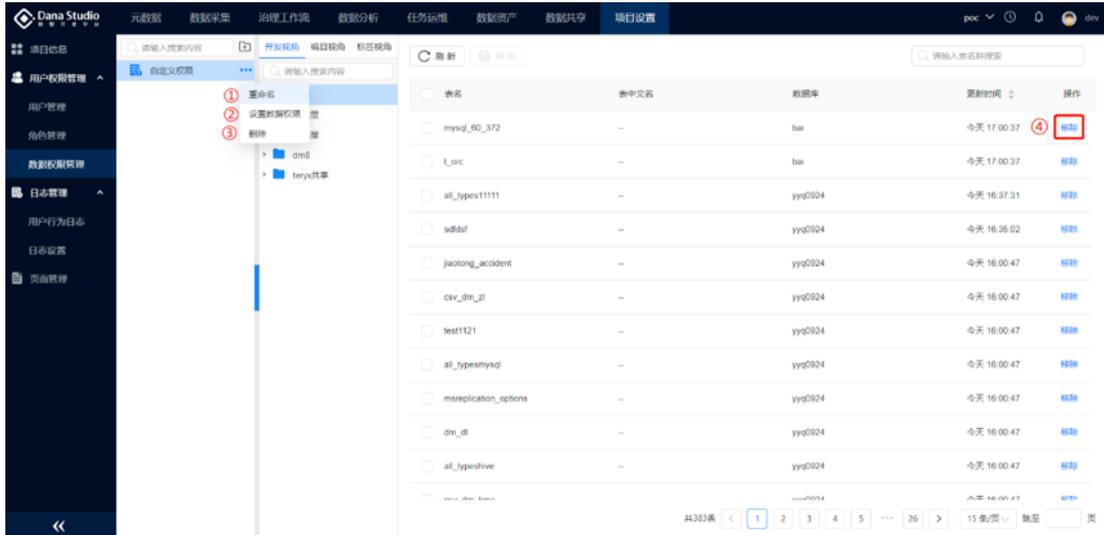
系统预制的“全部数据”数据权限，能操作所有数据。同时也提供自定义数据权限进行灵活设置。

新建数据权限：

项目管理员——项目设置——用户管理——点击“新建数据权限”图标，输入权限名称，点击“下一步”，配置该数据权限可操作的库和表。



查看权限列表：



①点击权限的“重命名”按钮，可修改数据权限名称

②点击权限的“设置数据权限”按钮，可修改权限的可操作的表

③点击权限的“删除”按钮，可删除权限。只有该数据权限没有与任何用户组或用户绑定时，才允许删除。

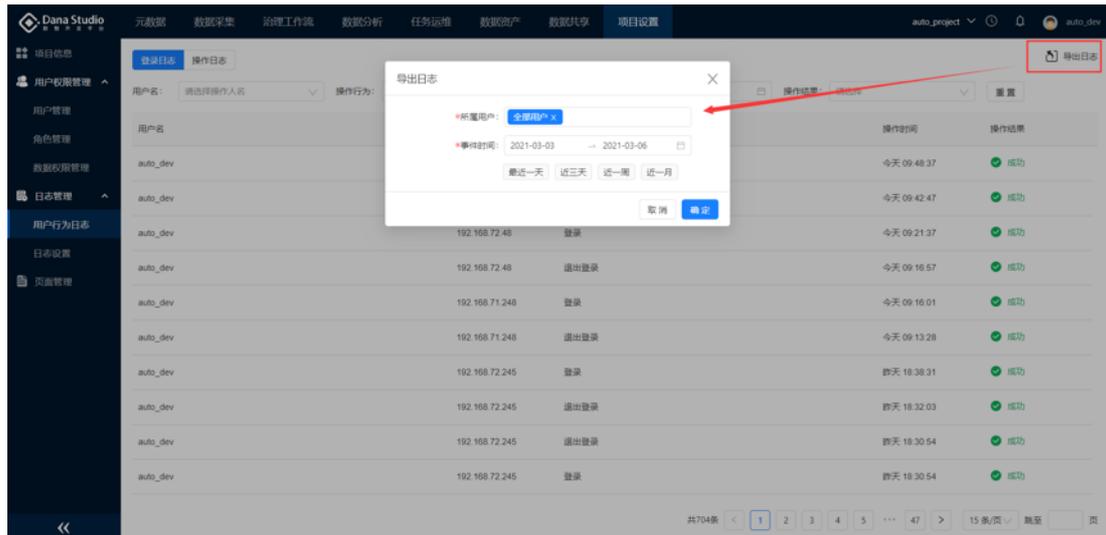
④点击权限下的某个表的“移除”按钮，拥有该权限的用户不再拥有该表的操作权限

4.3 日志管理

日志管理提供对各类日志进行归档、清理的机制，能提供日志的精确查询，并能根据设置的条件将日志导出到本地。日志管理不仅能解决由于大规模日志存储带来的平台运行问题，也提高了平台的运维能力。

4.3.1 用户行为日志

支持查看、导出用户的登录和操作日志。



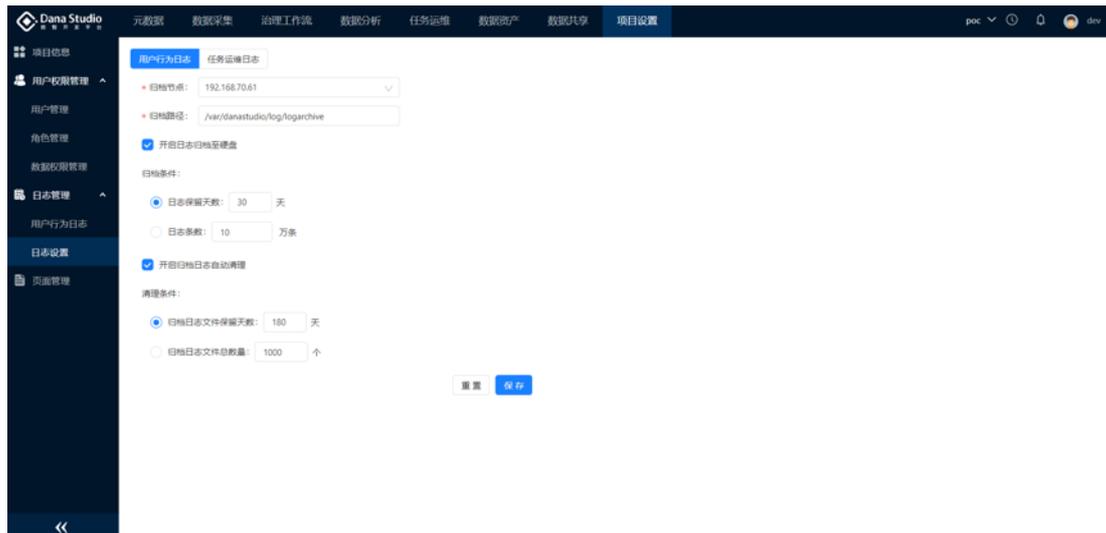
注：若日志未归档，则将筛选后展示的页面日志导出成 csv 文件，若日志已归档，则将已归档日志文件与页面导出的 csv 文件打成 zip 包导出。

4.3.2 日志设置

日志设置分为用户行为日志和任务运维日志，可为不同日志配置不同的归档方案。

日志归档：开启日志归档开关后，根据设置的归档目录、条件将在线日志进行离线归档。可选择根据日志保留天数或日志条数进行归档。

日志清理：开启归档日志自动清理开关后，根据设定的清理条件，对归档的历史日志记录进行清理。可选择根据归档日志文件保留天数或归档日志文件总数量进行清理。



注：针对项目生成归档日志文件，日志归档后不再在页面展示，每周一零点进行日志归档和清理操作。

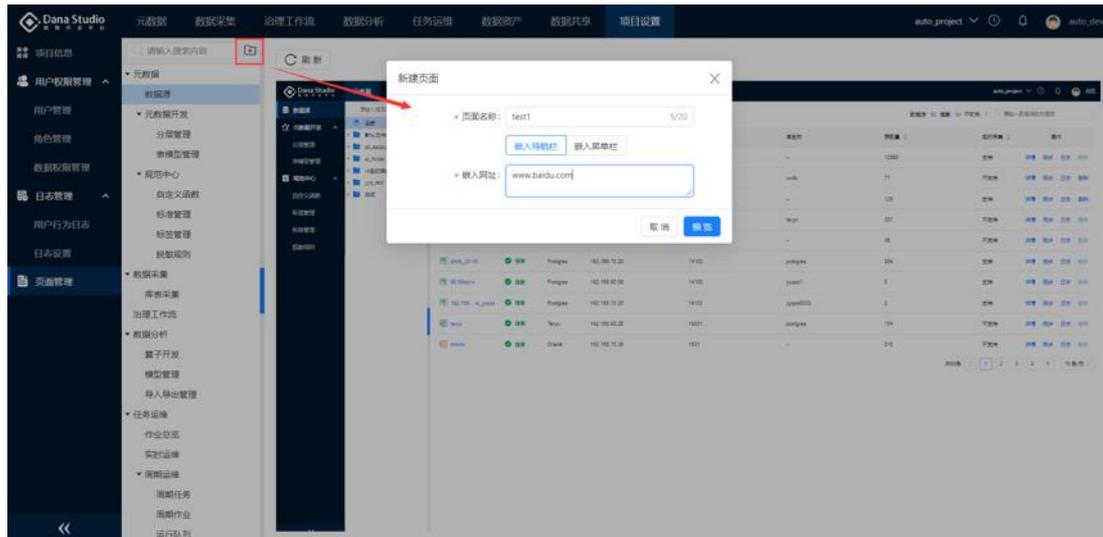
4.4 页面管理

用户可以在此页面嵌入自己的页面到平台中，可以对页面进行重命名、移动顺序，并可以在平台管理——日志审计中查看相关的操作记录。

新增页面：

项目设置——页面管理——新建页面——配置网页信息——预览

注：不支持嵌入同节点的 URL 地址



设置页面：

对新增的页面可以进行发布、编辑和刷新操作。

项目设置——页面管理——选择左侧新增的导航栏或菜单栏——进行发布、编辑、刷新操作



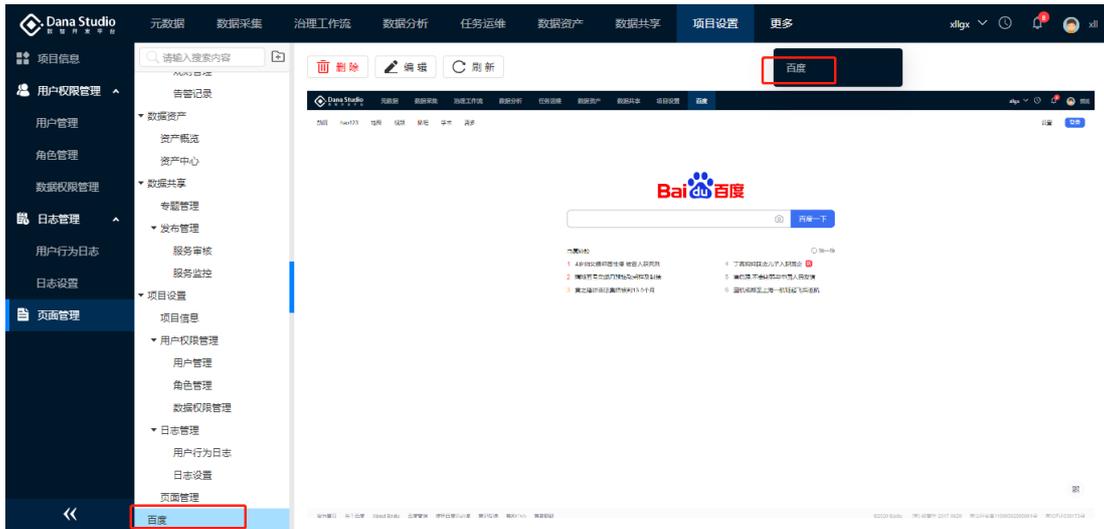
①点击“发布”按钮发布页面后，可以在对应的模块查看到该页面

②点击“编辑”按钮，可以对此页面进行重新编辑

③点击“刷新”按钮，对此页面进行刷新

查看页面：

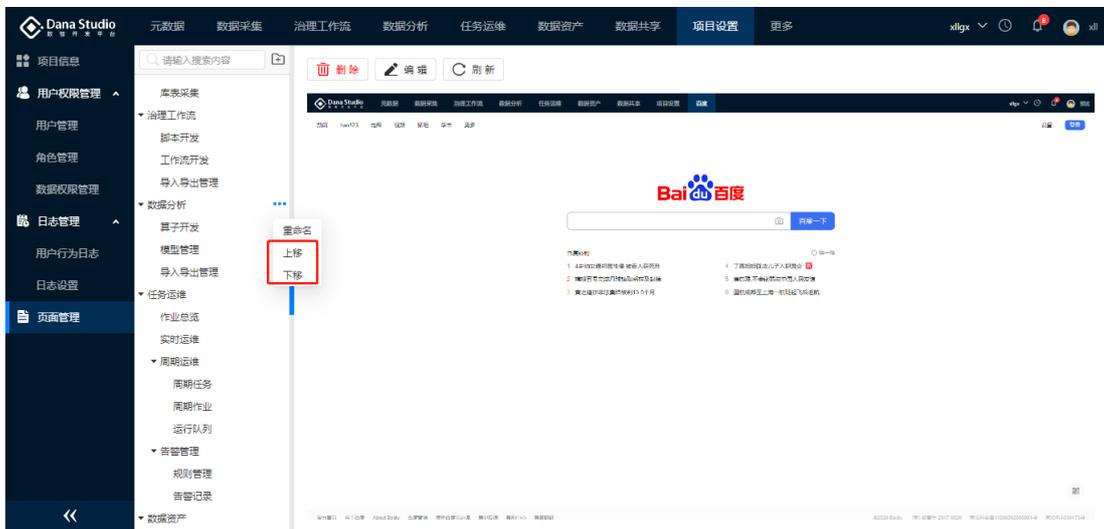
页面发布后，在对应的模块可以查看到该页面。



页面移动:

项目设置——页面管理——鼠标悬浮于导航栏或菜单栏上——选择“...”——选择上移或下移

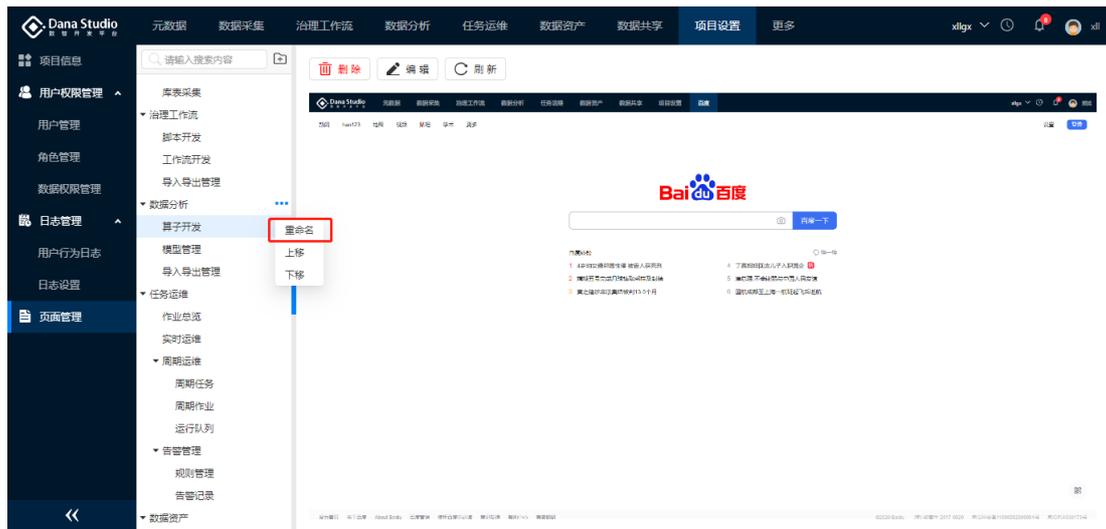
注: 移动只能在本层级移动，处在当前层级顶部时不能上移，底部时不能下移



页面重命名:

项目设置——页面管理——鼠标悬浮于导航栏或菜单栏上——选择“...”——选择重命名

注: 所有导航栏、菜单、目录平级内容均不能出现重名，不同级内容允许重名



注：1) 新建的页面仅同一项目组下用户可见

2) 同一项目下，其余用户新建的页面，需进入角色管理页面重新配置才可见

5 元数据管理

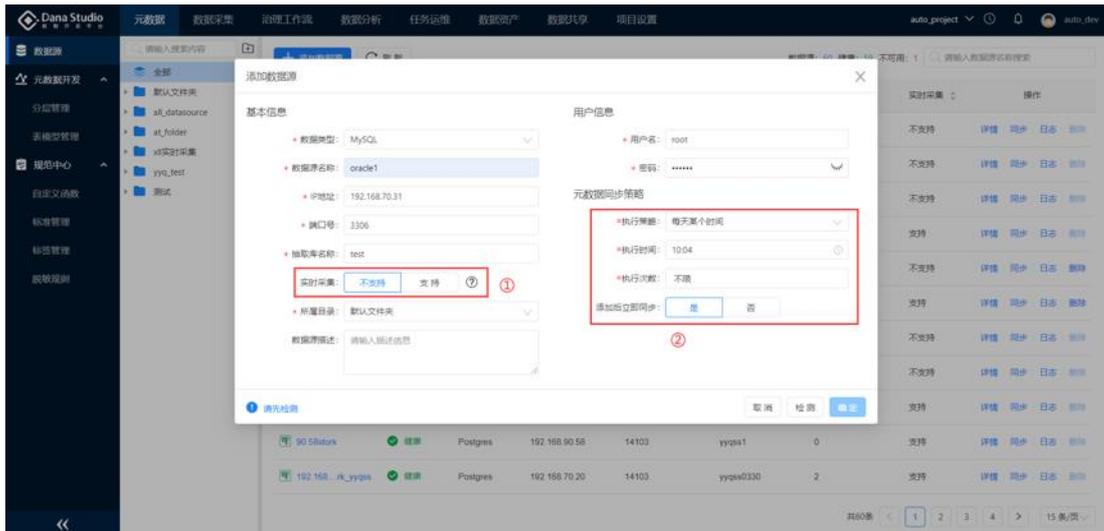
5.1 数据源

数据源管理是数据同步的前提。数据资源平台提供统一数据源管理功能，以解决不同类型的数据源差异性所带来的问题。

支持连接十余种数据库，包括关系型数据库、分布式数据库，涵盖 Mysql (5.7、8.0)、Oracle (10g、11g、12c)、DB2 (11.5)、SQL server (2005、2008)、Postgres (3.0-3.2)、Teryx (3.0、3.2、3.3)、GaussDB200/300、达梦 8、Hive (3.1)、优炫 (2.1)、CSV 多个版本的数据库类型。

添加数据源：

元数据——数据源——点击“添加数据源”——配置信息——检测成功后，点击“注册”



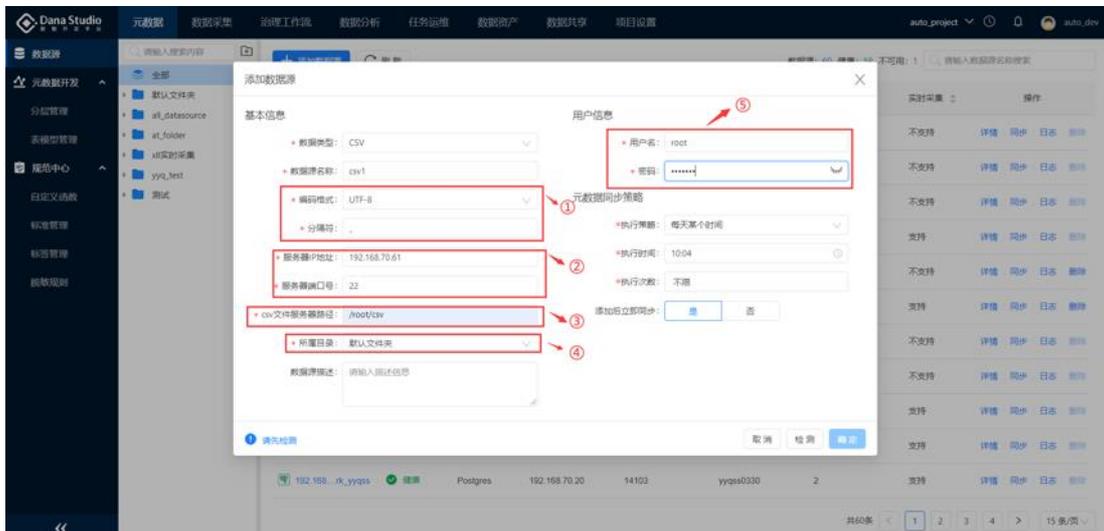
①当前仅 mysql5.7/8.0.x、oracle10/11g 数据库支持实时，且数据源配置选择支持实时采集，才可在库表采集——实时采集中使用该数据源。

②元数据的同步策略，数据源根据策略将库下的表的同步在平台中，当数据源中表较多时，数据不会立即完成同步

注：1) 注册优炫、达梦分布式数据库时，使用主节点 IP

2) 适用于整库迁移，数据源端注册信息发生变更后，与其关联的采集、库表推送任务不受影响，能正常运行

注册 csv 数据源：

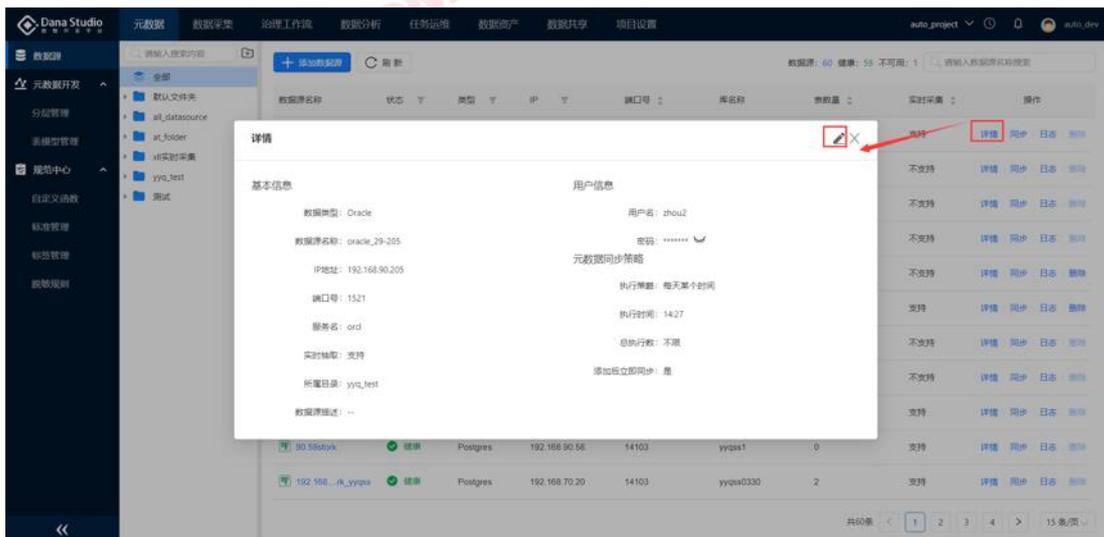


- ①编码格式、分隔符：填写 csv 文件对应的编码格式和分隔符，否则会检测失败
- ②服务器 IP 地址、端口号：填写 csv 文件所在的 ip 地址和端口
- ③csv 文件服务器路径：填写 csv 文件所在服务器中对应的路径
- ④所属目录：系统预置一个默认文件夹，可新建自定义文件夹
- ⑤用户名、密码：填写 csv 文件在服务器登录的用户名、密码

注： csv 文件路径的操作权限需要给到 hive 用户

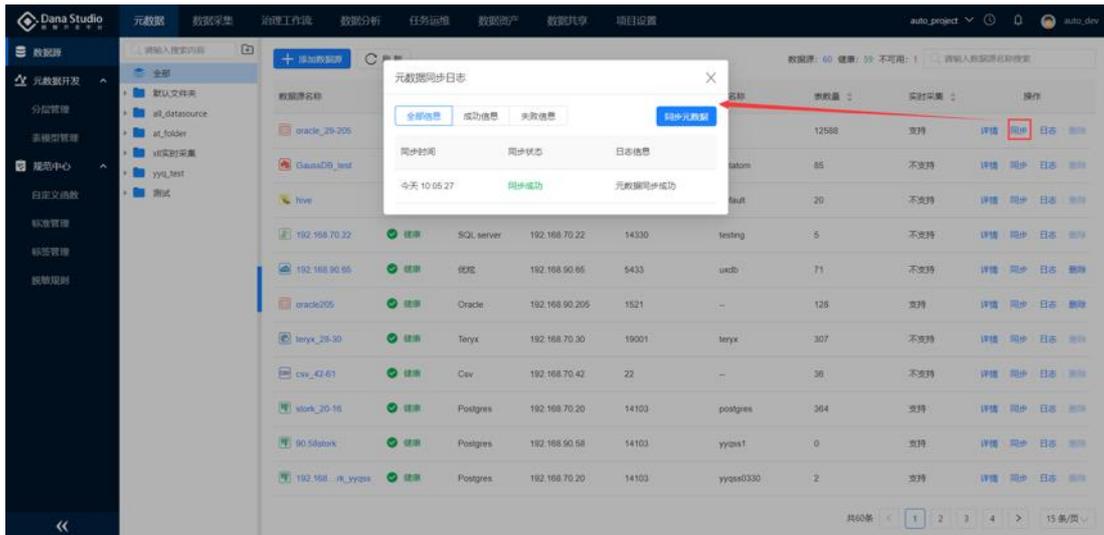
编辑数据源：

元数据——数据源——点击操作下的“详情”，进入数据源详情弹窗——点击“编辑”



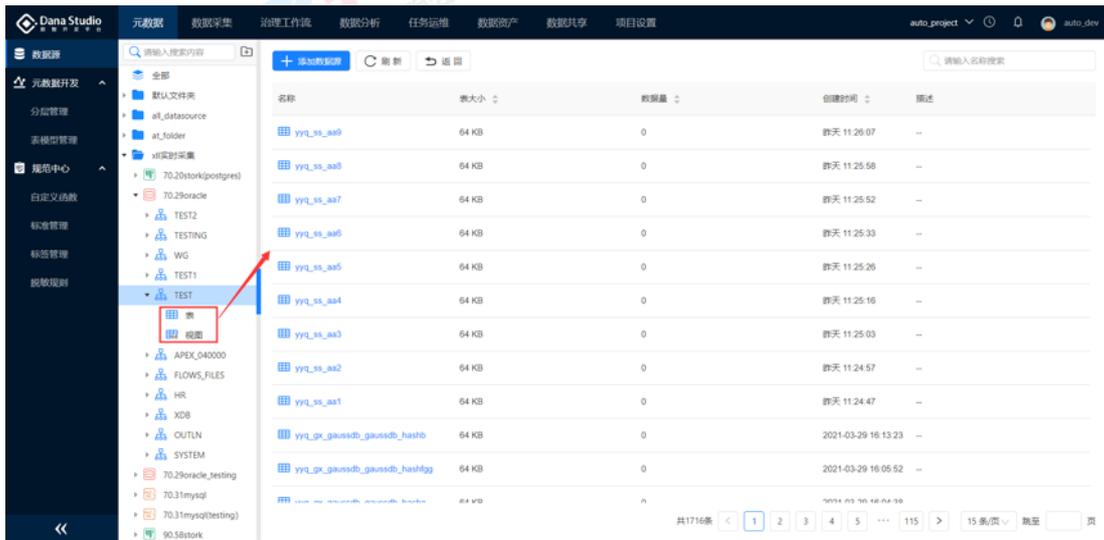
立即同步：

元数据——数据源——点击操作下的“同步”弹窗中的“同步”按钮，数据源会进行一次立即同步



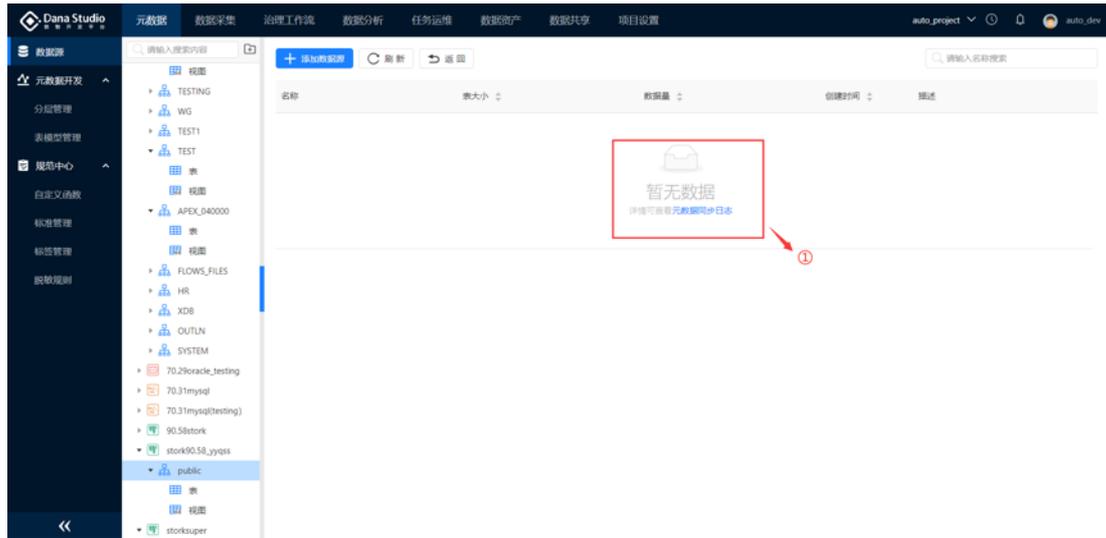
查看数据表：

元数据——数据源——点击列表或目录树中数据源名称，可查看该数据源下的表和视图



注：如数据库中表的表描述、字段注释带有单引号 (')、分号 (;)、双短横线 (--) 特殊字符，页面上展示时会以空格代替

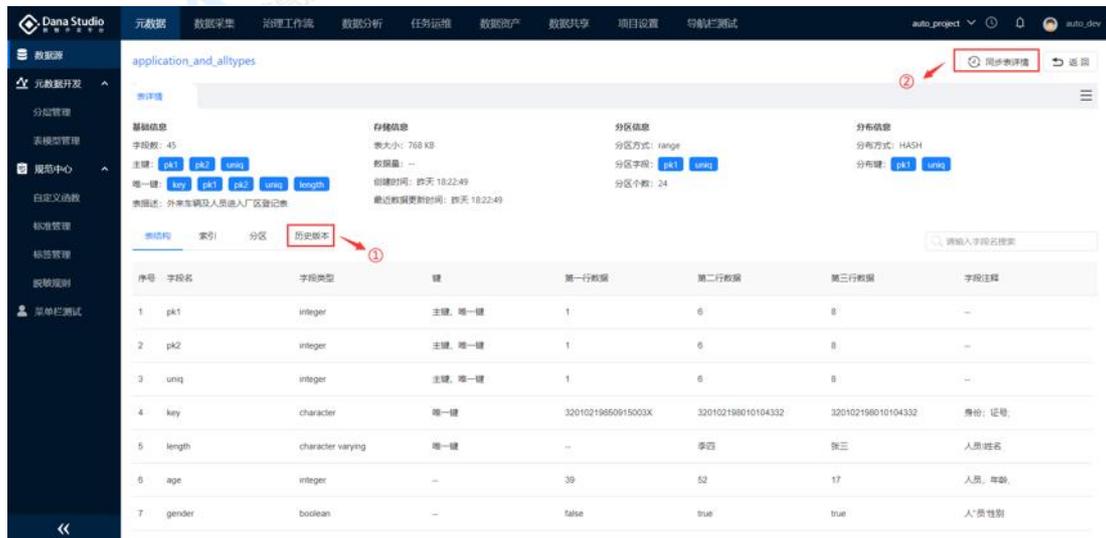
数据未同步完成时，可点击页面中的“元数据同步日志”，查看日志



注：弹窗中的日志以打开时的状态为准，状态有变化时，不会自动刷新

查看数据表详情：

元数据——数据源——点击列表或目录树中数据源名称，进入数据表列表界面——
 点击表名称可查看该表的详情信息



①若该表的表结构发生过变更，数据源同步后会生成一个历史版本，可查看该表的历史版本信息

②点击“同步表详情”按钮，同步源表中所有信息

注：视图表不展示分区、分布、索引信息

5.2 元数据开发

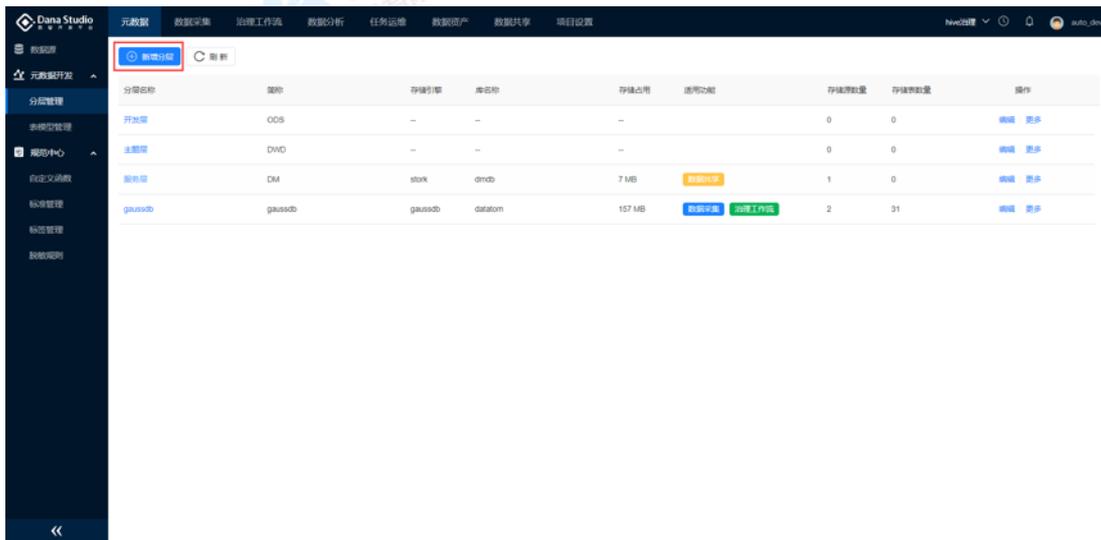
通过元数据开发可以进行可视化建模，为数据采集、数据分析治理和资产管理打下基础。通过元数据开发可构建数据仓库分层管理模型，数据表、资产管理的编目。

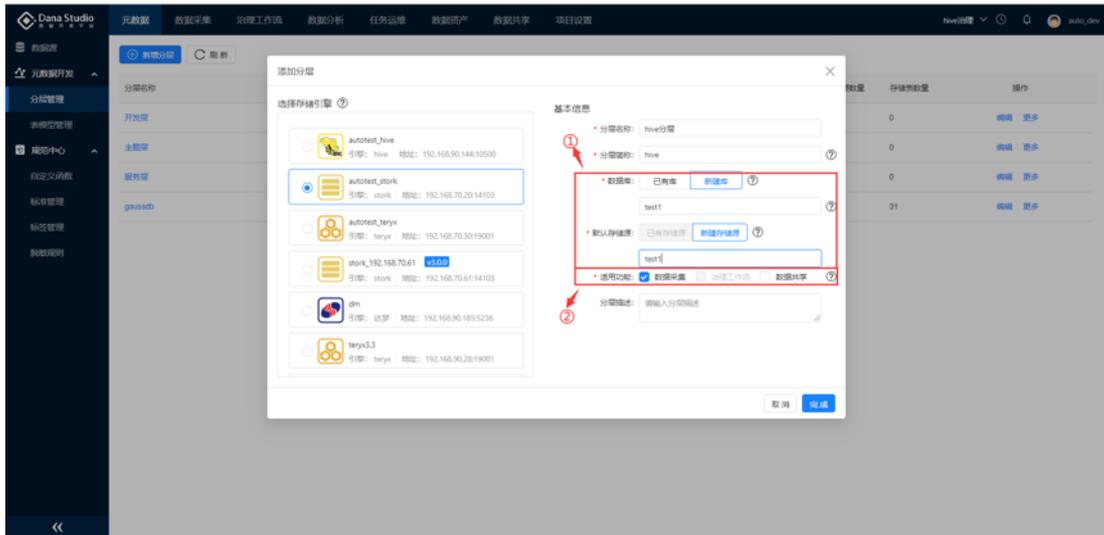
5.2.1 分层管理

新增分层：

元数据——元数据开发——分层管理——在分层管理 tab 页中——点击“新增分层”

注：建议所有的项目下新建的分层合计不超过十个





①引擎选择 stork、teryx、GaussDB、优炫存储源是以 schema 为基础来添加；

引擎选择 hive、达梦存储源是以 database 为基础来添加；

②选定的功能可将本层的存储源作为数据输出源，输入源无此限制。

注：1) 支持配置 hive3.1、teryx3.2 和 3.3、stork3.2、GaussDB200/300、达梦 8、优炫 2.1 六种存储引擎，其中 teryx3.2 版本不支持治理 workflow 功能

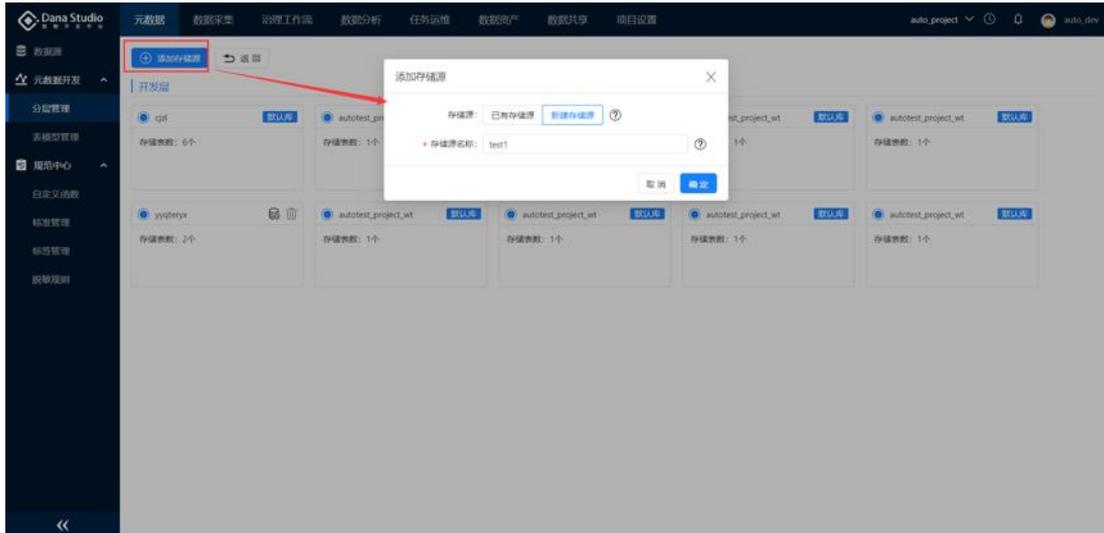
2) 平台预置开发层简称 ODS 层、主题层简称 DWD 层、服务层简称 DM 层。默认的三个分层初始不设置存储引擎，需要用户自己通过“编辑”去添加引擎

3) 普通用户若没有全部数据权限，则无法新建分层

4) 分层管理中适用于采集时支持添加多个不同的 hive 引擎，若有 hive 分层适用于治理 workflow 时，其他 hive 分层若是适用治理 workflow 必须为一个 hive 引擎

添加存储源：

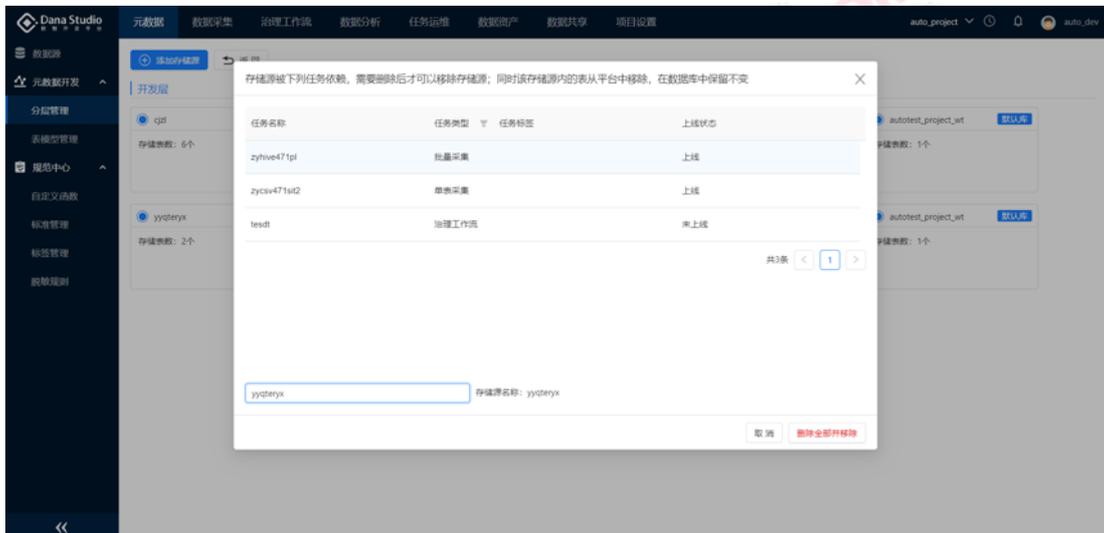
元数据——元数据开发——分层管理——点击分层名称，进入分层页面——点击“添加存储源”



- 注：**
- 1) 支持添加该分层引擎中的已有存储源或新建存储源
 - 2) 普通用户若没有该分层的全部权限，则无法新建存储源
 - 3) 第一个适用于治理工作流的分层中会存在一个问题库（项目简称_wt）

移除存储源：

元数据——元数据开发——分层管理——点击分层名称，进入分层页面——点击存储源右上方的“移除”按钮



- ①当存储源没有被任务依赖，输入存储源名称确认后可直接移除
- ②当存储源有任务依赖，弹出被依赖的任务弹窗提醒，确认移除后，被依赖的任务将删除，同时该存储源内的表从平台移除，数据在数据库中保留不变

查看分层列表：

分层名称	别名	存储引擎	库名称	存储占用	适用功能	存储库数量	存储表数量	操作
开发层	OOS	--	--	--		0	0	编辑 更多
主题层	DWD	--	--	--		0	0	编辑 更多
数据层	DM	--	--	--		0	0	编辑 更多
teryx	teryx	teryx	postgres	135 MB	数据仓库 治理工作流	3	66	编辑 更多
gaussdb	gaussdncj	gaussdb	datatom	157 MB	数据仓库	4	55	编辑 更多
gauss共享	gaussgx	gaussdb	datatom	157 MB	数据共享	1	1	编辑 更多

①点击操作下的“编辑”按钮，编辑分层信息

②点击操作下“更多”的“移除”按钮，移除分层，所有分层均支持移除

1) 其中当分层没有被任务依赖，可以直接移除分层

2) 当分层有任务依赖，存在弹窗列举分层存在的依赖的任务，确认移除分层后，所依赖的任务及产生的表（包括上线的任务）将一并移除，底层数据库中产生的数据不变

3) 包含问题库的层被移除掉，如果有两个及两个以上的层用于治理工作流，则问题库转移到第二个层里；若只有一个层用于治理工作流，则问题库及表也随之移除，底层数据库中产生的数据不变

③点击操作下“更多”的“上移、下移”按钮，对分层进行移动

注：1) 项目管理员默认具有分层管理的操作权限

2) 其他内置的角色中开发和运维只具有分层管理的查看权限

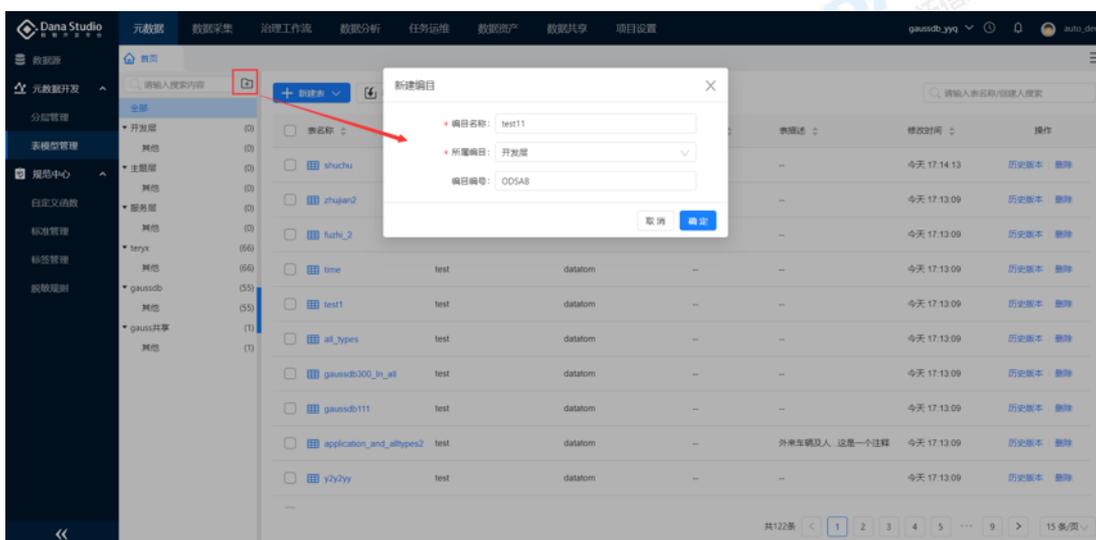
5.2.2 编目管理

编目信息是表的数据目录，编目信息可以按照不同数据的来源、主题、用途等进行设置，帮助整理，归类表，方便后边对于表的查询和管理。平台支持多至 4 级的编目体系，可以根据具体业务场景自由对编目的级别进行调整。

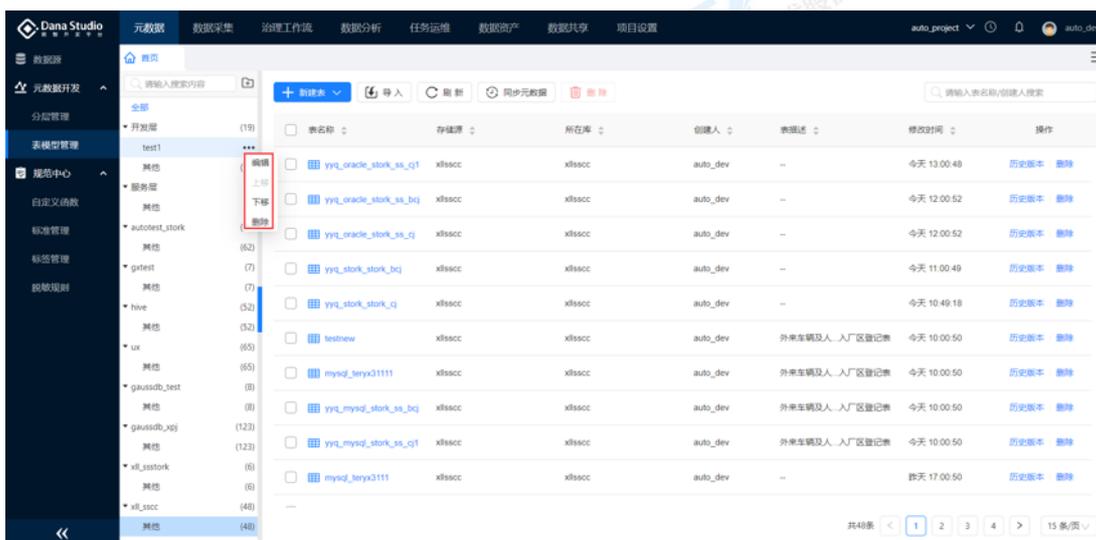
新建编目：

元数据——元数据开发界面——点击目录树中的“新建编目”图标

注：每个分层下默认存在一个“其他”编目，用于存放未选择编目的表



支持对于编目进行位置顺序的移动，重命名、新增、删除等操作。



5.2.3 表模型管理

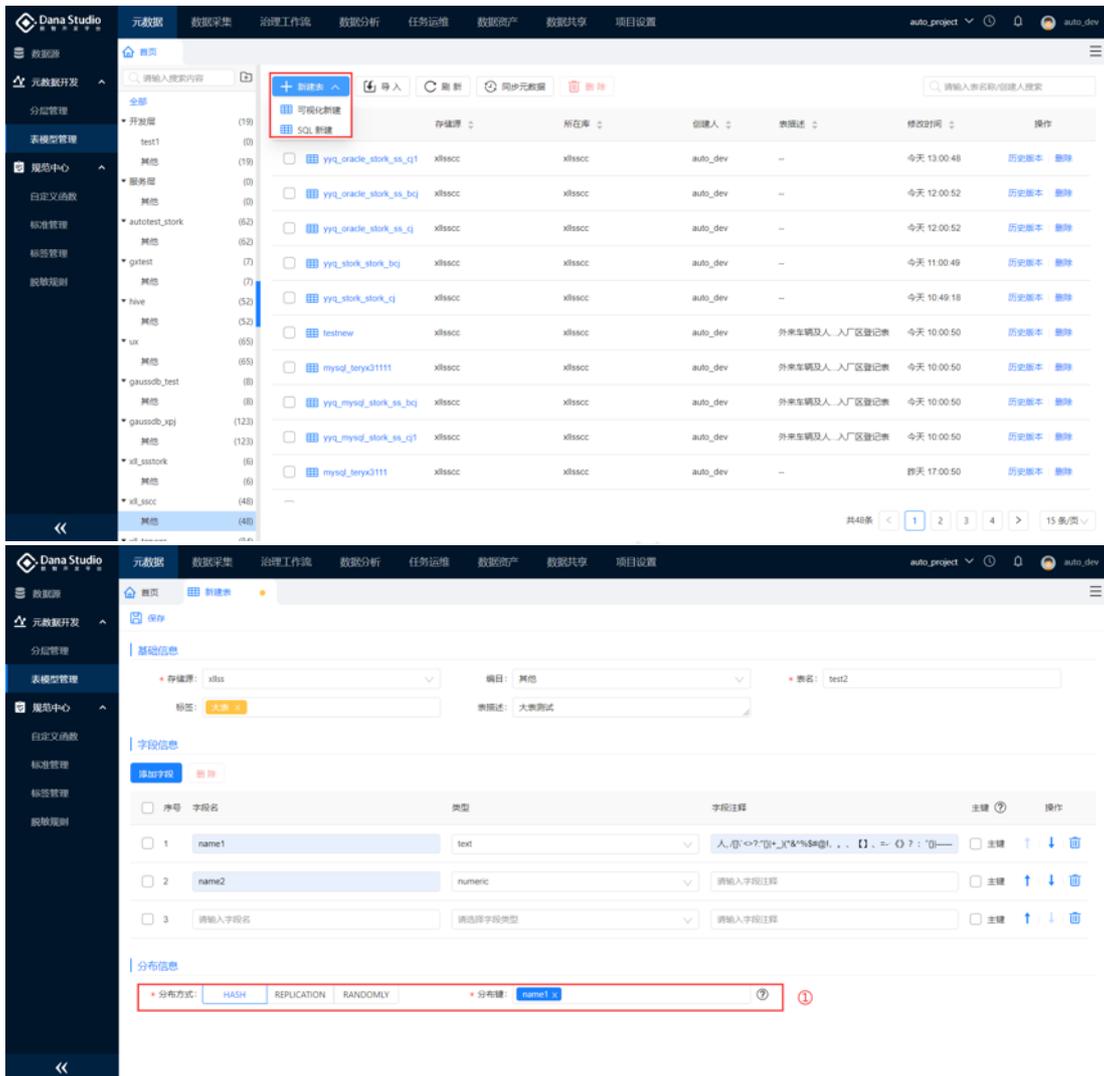
元数据开发列表展示当前项目中存储源中的所有表。

可视化新建表：

元数据——元数据开发——表模型管理——新建表——点击“可视化新建表”按钮

注：1) hive 实时表会默认建成拉链表，增加两个字段
_op(string),_source_ts_ms(timestamp)

2) 普通用户若没有整库权限，则无法往该库中新建表



①teryx、GaussDB、达梦、优炫作为存储源时，存在分布方式、分布键的设置

1) teryx、达梦分布方式：有三种类型 HASH、REPLICATION、RANDOMLY

2) 优炫分布方式：仅有一种 HASH

3) GaussDB 分布方式：有两种类型 HASH、REPLICATION

a.HASH：选择 HASH 分布方式时，会出现分布键；当表中存在主键时，分布键只可以在主键中选择，如果表中不存在主键时，分布键可以任意设置；

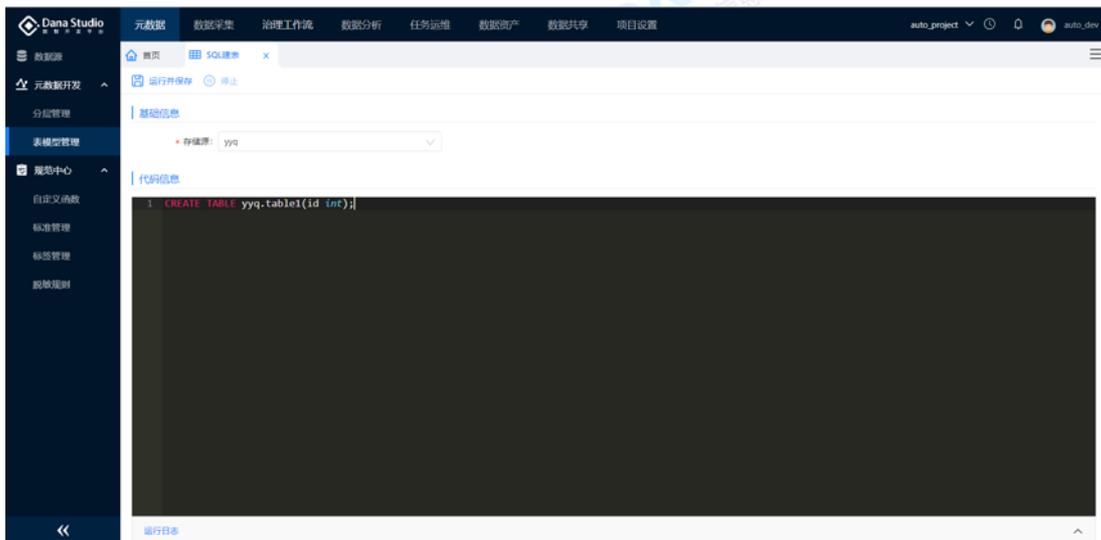
b.REPLICATION：选择复制分布时，不会出现分布键选择；

c.RANDOMLY：选择随机分布时，不会出现分布键选择；当表存在主键时，RANDOMLY 分布选项禁用；如果已经选择了 RANDOMLY 分布，后设置表主键，则默认选择到 HASH 分布

注：字段注释、表描述不支持输入单引号（'）、分号（;）、双短横线（--）等特殊字符

SQL 新建表：

元数据——元数据开发——表模型管理——新建表——点击“SQL 新建表”按钮——编写 sql 语句——运行并保存



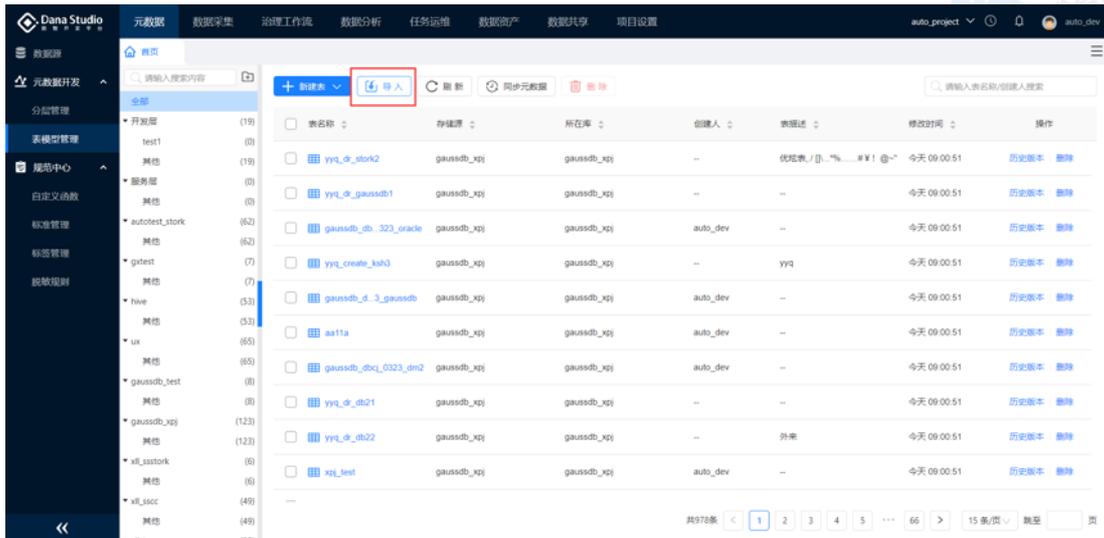
注：1)如字段注释、表描述语句中有单引号（'）、分号（;）、双短横线（--）等特殊字符时，运行会报错

2)建表语句中表名称前需要加上所属存储源名称

导入表:

元数据——元数据开发——表模型管理——点击“导入”按钮，将数据源中的表直接导入至存储源

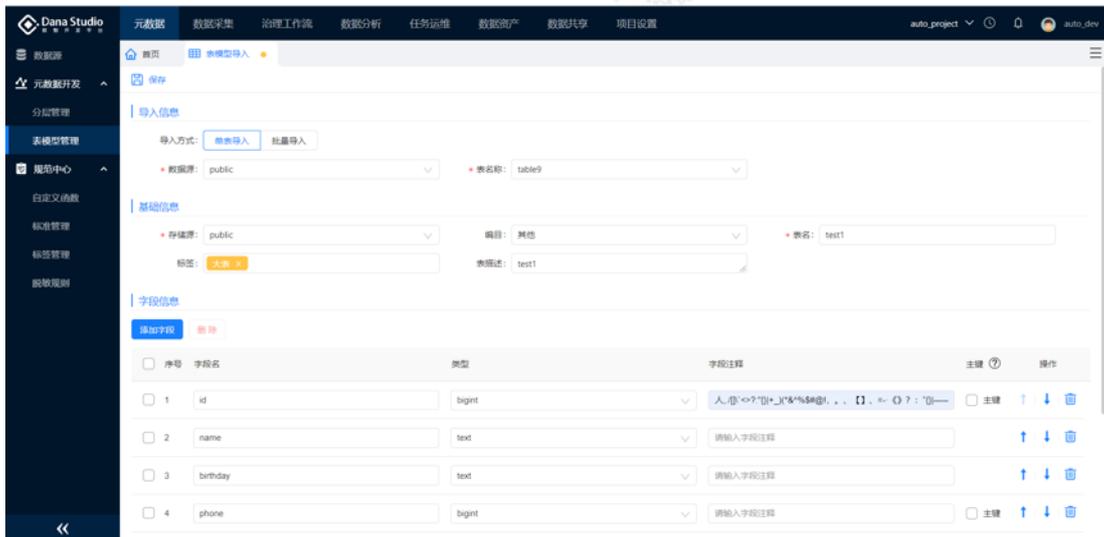
注: 支持单表导入和批量导入



单表导入:

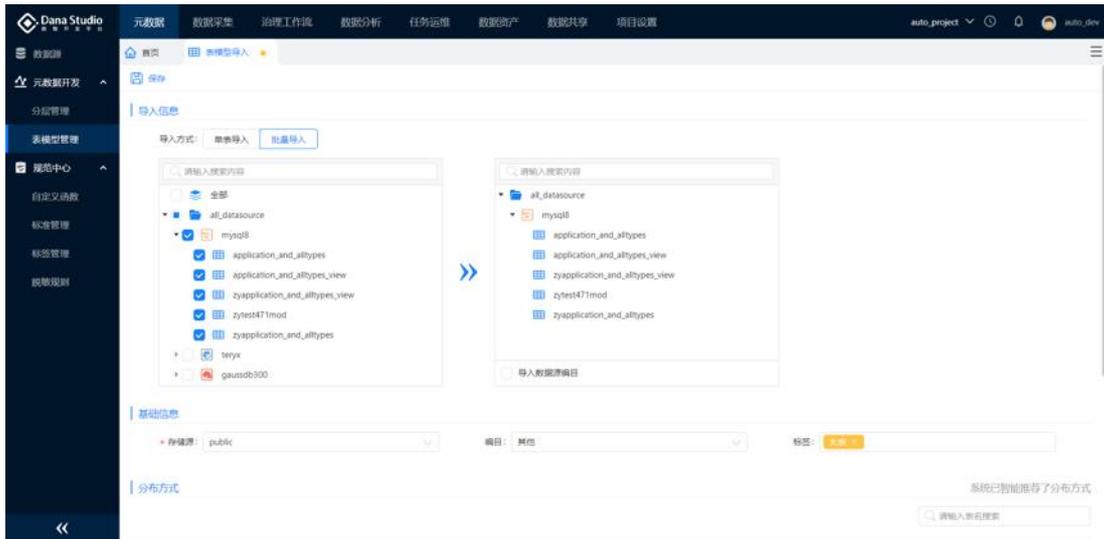
导入——单表导入——选择数据源、存储源——配置基础信息、字段信息等——

保存

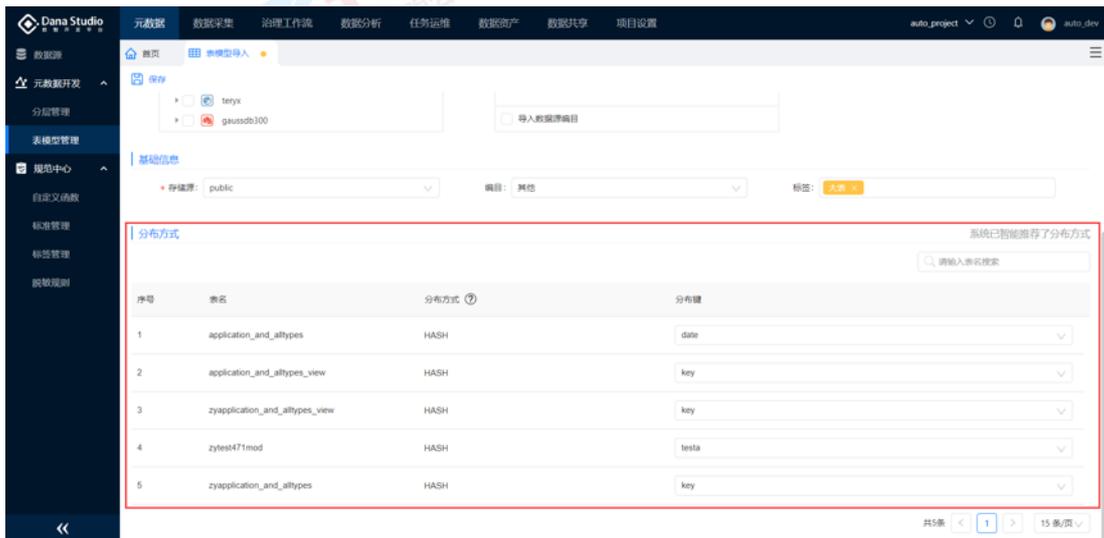


批量导入:

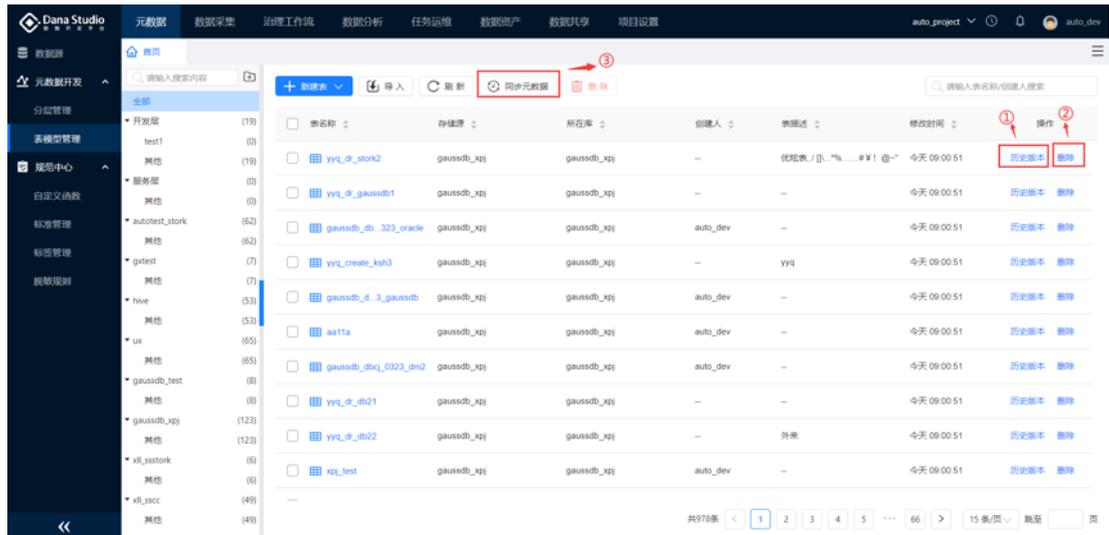
导入——批量导入——选中多张表——配置基础信息——保存



注：导入表到 teryx、GaussDB、达梦、优炫存储源时可以配置分布方式、分布键，其他数据库不展示此选项



表列表：



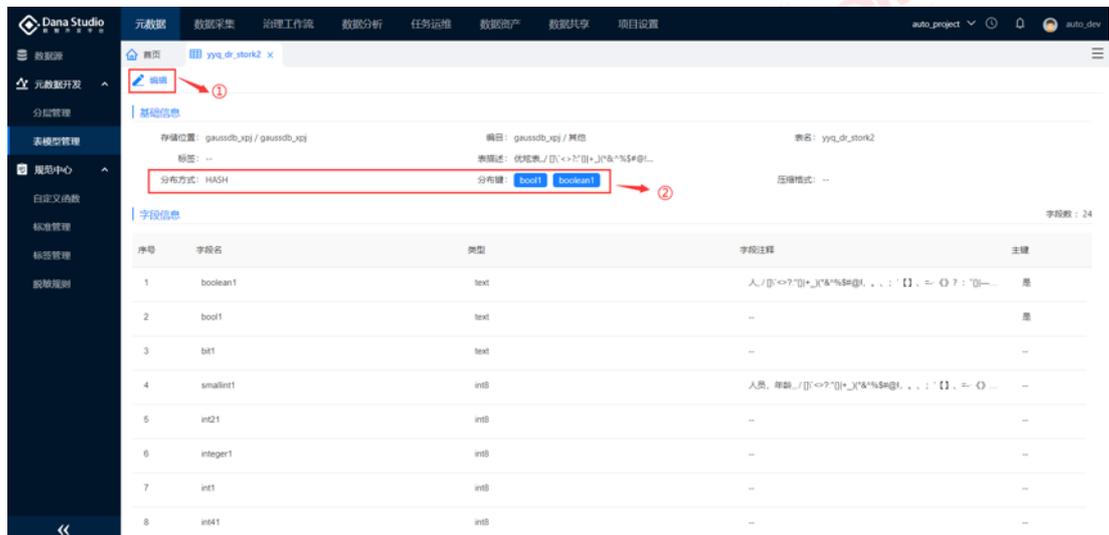
①若表结构有所变更，原表会被存为历史版本，可点击操作下的“历史版本”查看历史表信息

②点击操作下的“删除”按钮，删除表

③点击“同步元数据”，将存储源中的所有表同步至平台上

表详情：

元数据——元数据开发界面——点击表名称，查看表详情信息，支持编辑表信息



①点击“编辑”按钮，可以编辑表信息

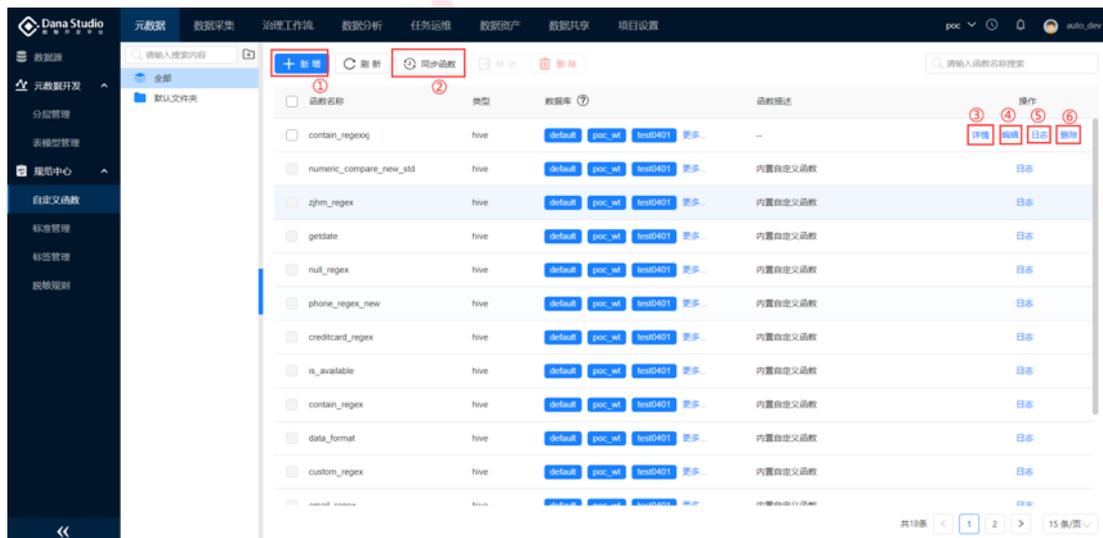
②teryx、GaussDB、达梦、优炫中的表展示分布方式、分布键，其他存储源不会展示

注：元数据开发处导入的表支持修改表的全部信息，被任务依赖、有数据的表仅支持编目、标签和表描述和字段注释的修改

5.3 规范中心

5.3.1 自定义函数

开发者可以借助自定义函数功能导入事先写好的函数包，在平台上一键注册并使用。支持上传并管理自定义 JAR 包函数，自定义函数兼容 Presto SQL 语法的 JAR 包和 spark SQL 语法的 JAR 包格式。



①自定义函数支持新增功能，点击“新增”，打开新增自定义函数的弹窗；

②点击“同步函数”，更新函数状态；

③仅用户上传的自定义函数有“详情”按钮，点击“详情”，可以查看自定义函数的详情；

④仅用户上传的自定义函数有“编辑”按钮，点击“编辑”，可以编辑自定义函数的信息；

⑤点击“日志”，可以查看自定义函数同步的日志；

⑥仅用户新建的自定义函数有“删除”按钮，点击“删除”，可以删除自定义函数；

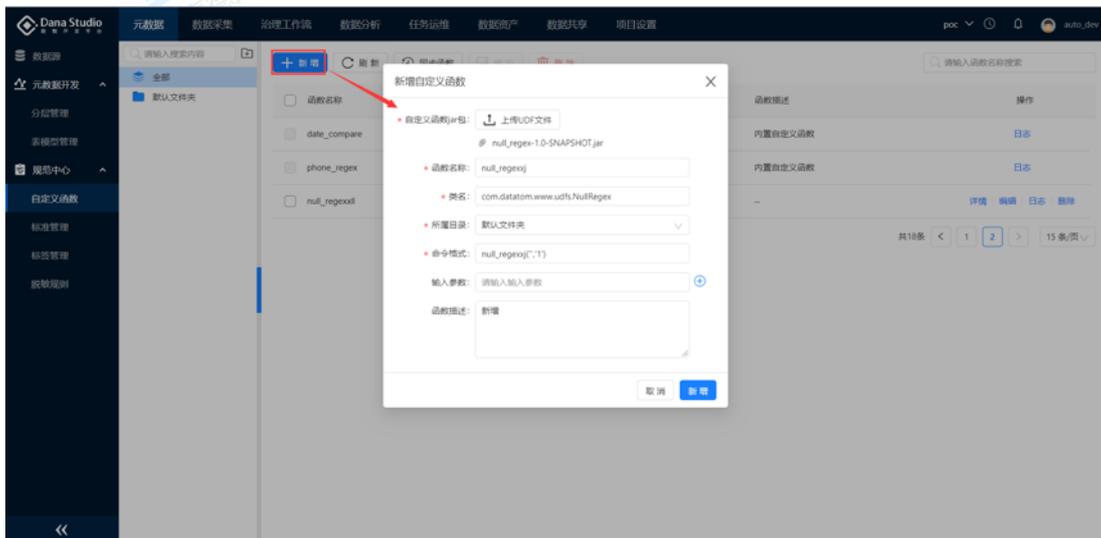
注：1) 更换治理引擎时，平台内置自定义函数类型做相应变更（需要用户手动点击一次同步函数），旧引擎下用户自定义上传的 UDF 保留最后一次状态

2) 内置自定义函数只有在分层有适用于治理工作流的时候展示出来，否则不展示，且不支持操作

3) 第一次进入自定义函数页面时，需要点击同步函数，才可以同步内置自定义函数

新建自定义函数：

元数据——自定义函数——新建——选择自定义函数 jar 包，并填写相关必填项——确定

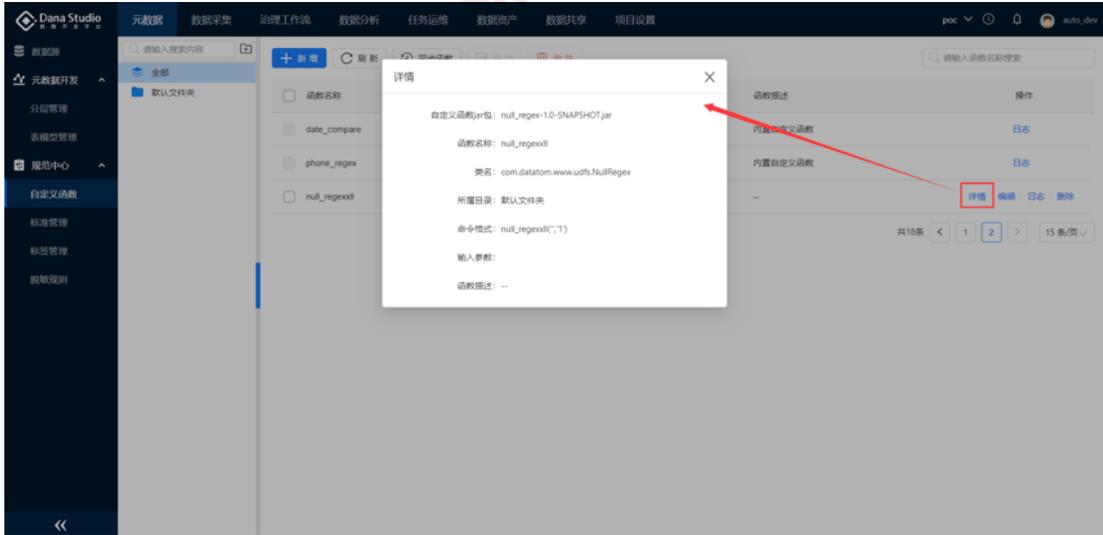


注：1) 自定义函数新建完成后不支持修改名称
2) 目前 Hive 支持展示平台自用和用户上传，Teryx/GuassDB200/300 只支持展示平台自用，不支持界面上上传

3) 如果函数注册成功，数据库展示适用于治理工作流的分层的所有存储源，显示为蓝色，如果注册失败，则显示为红色

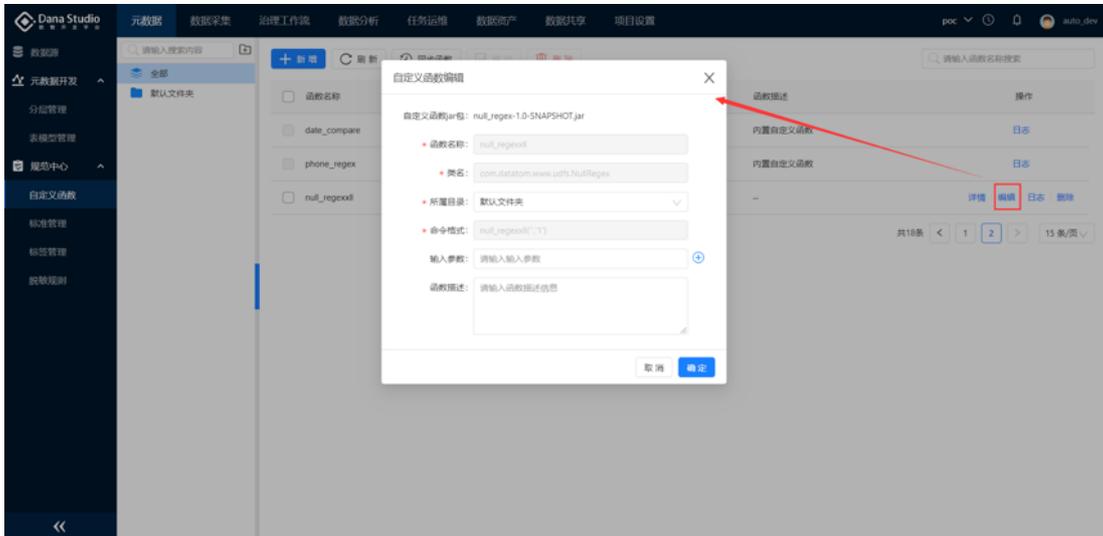
查看自定义函数详情:

元数据——自定义函数——用户已经添加的自定义函数后的“详情”按钮



编辑自定义函数:

元数据——自定义函数——用户已经添加的自定义函数后的“编辑”按钮——输入修改的信息——确定



注: 仅可以修改所属目录、输入参数、函数描述

5.3.2 标准管理

标准管理提供了创建自定义标准、导入代码表以及标准审核功能，用于规范用户数据结构与所用代码表。其目的是为了规范和统一字段名称、代码表映射，从而达到消除歧义快速治理。

数据标准分为三大类：

1、文本标准：

- (1) 基础规则，支持字母、数字、中文、特殊字符
- (2) 自定义规则，支持正则表达式

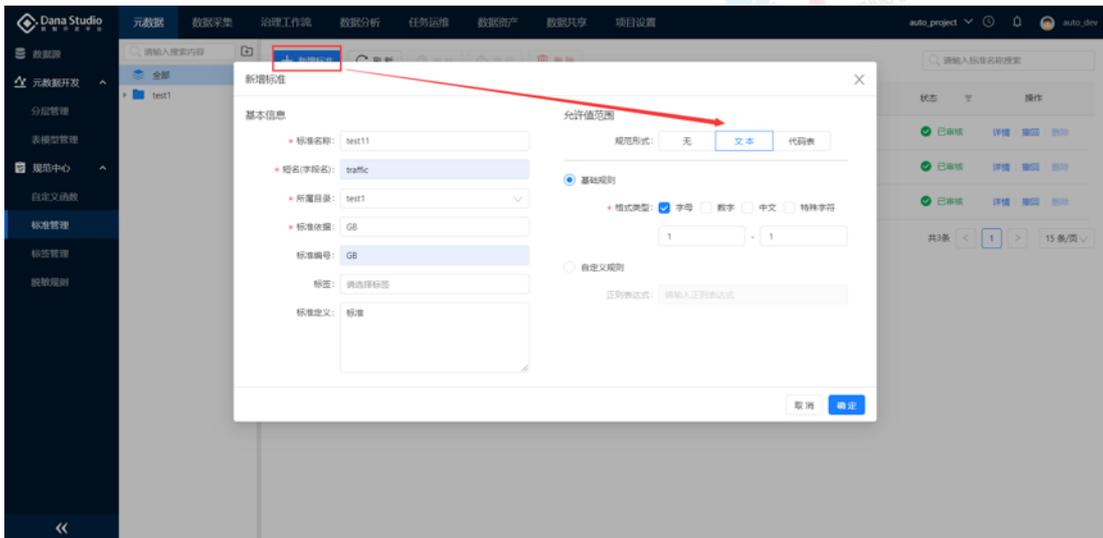
2、代码表标准，通过已有的代码表加载数据标准

3、无标准

新增标准：

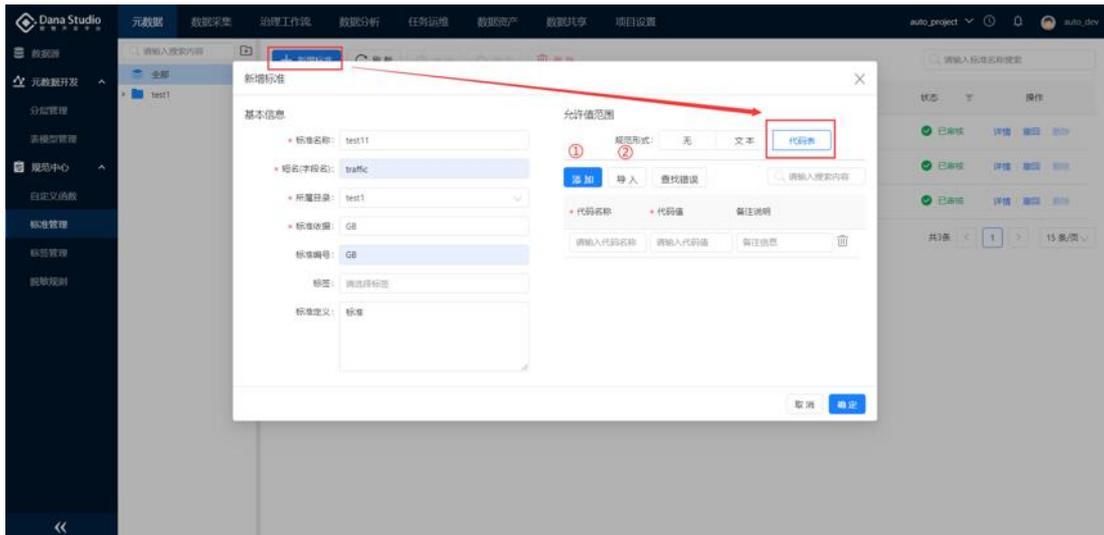
1、文本类标准

元数据——标准管理——点击“新增”按钮——填写信息——选择“文本”规则——填写规则——确定



2、代码表标准

元数据——标准管理——点击“新增”按钮——填写信息——选择“代码表”规则——手动添加或导入代码表



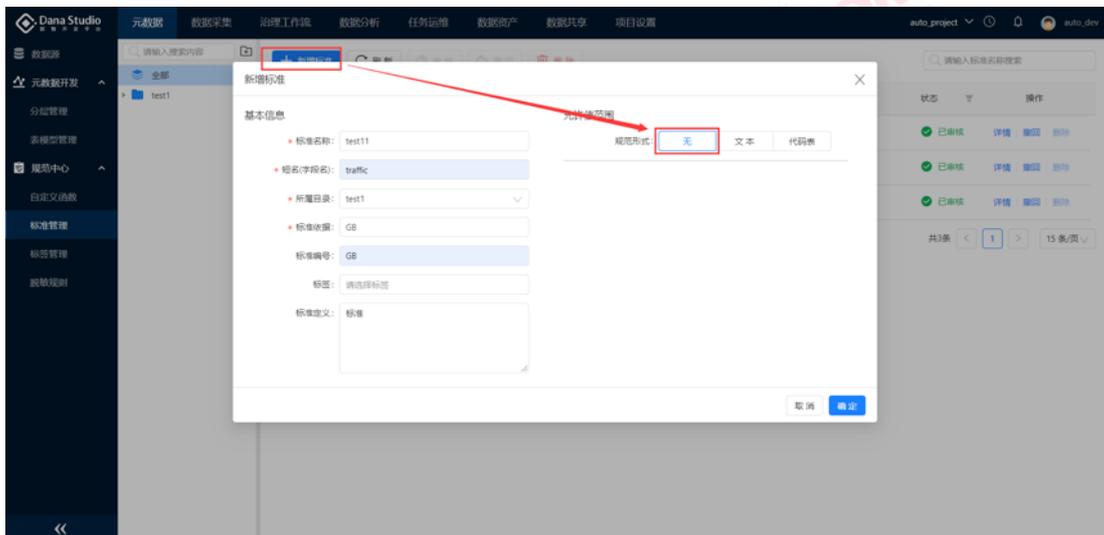
①支持手动添加代码表

②支持导入代码表

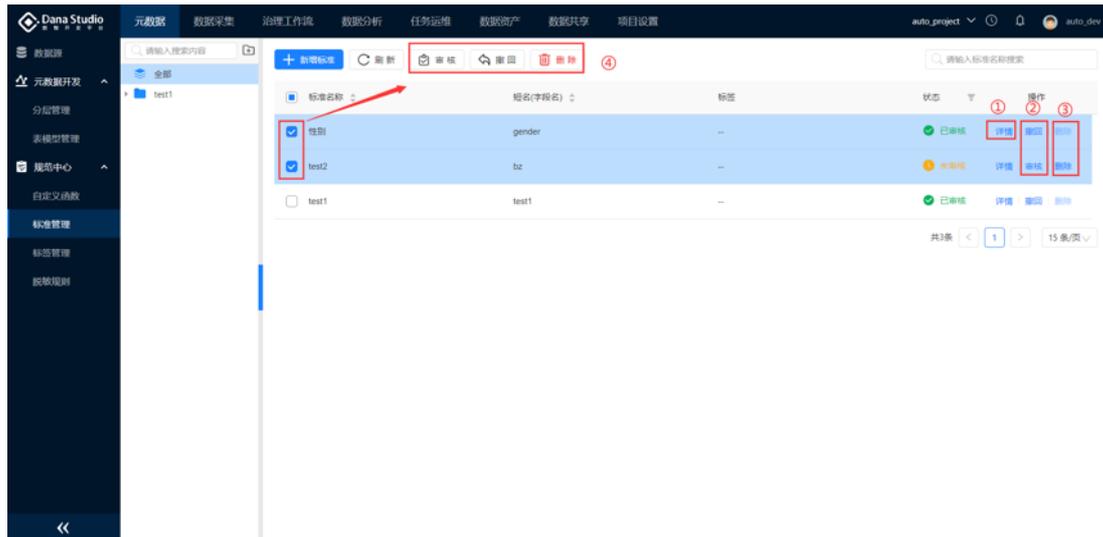
注：支持导入的文件：csv、excel，内容格式三列：代码名称、代码值、备注说明(必须有表头，前两列必须有值)

3、无标准

元数据——标准管理——点击“新增”按钮——填写信息——选择“无”



查看标准列表：



- ①查看数据标准的详情以及修改（未审核的标准才可被修改）
- ②对数据标准的审核、撤回（已审核的标准才可在后续治理任务中被使用）
- ③删除数据标准（未审核的标准才可被删除）
- ④支持批量审核、撤回、删除操作

5.3.3 标签管理

为达到数据表、任务、数据标准、脚本的精准筛选的效果，可以使用给数据打标签的方式来实现，然而如果没有良好的标签管理则会杂乱无序，失去标签原本用于标记整理的意义，因此规范化的标签管理是很有必要的。

标签支持按照表、任务、标准、脚本进行分类。支持每个类型下建立 4 级独立标签体系。

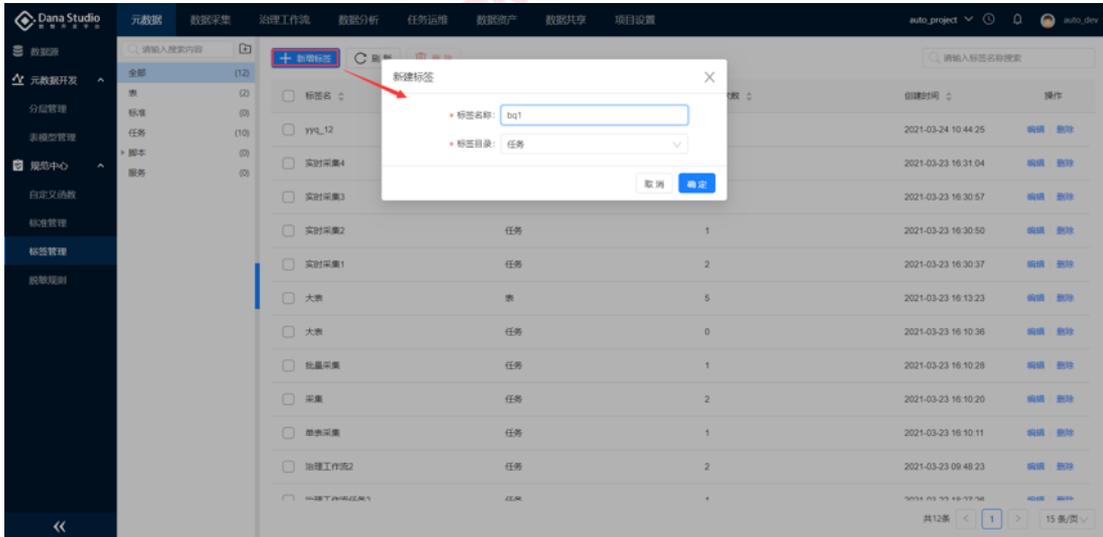
新建标签：

元数据——标签管理——点击“新增标签”按钮

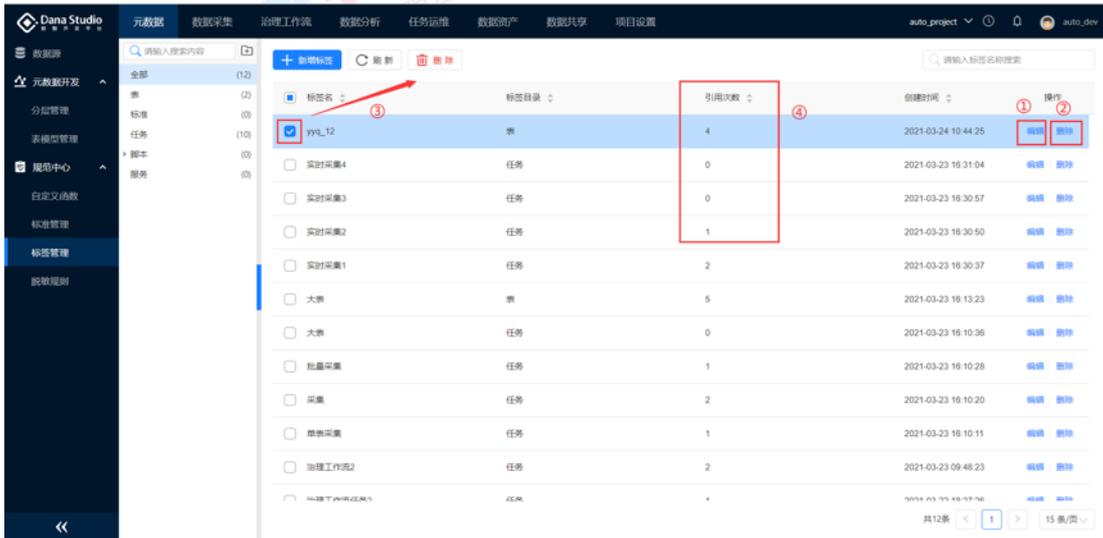
注：1) 每个分类下的标签只可在对应的功能模块选择使用

2) 标签目录可以选择最底层目录或者“其他”

3) 同一个目录下的标签不允许重名



查看标签列表：



①点击操作下的“编辑”，修改标签信息，包括标签名称、标签目录（标签目录的修改限于该标签本身所在的分类下的其他最底层目录）

②点击操作下的“删除”按钮，删除标签，引用该标签的对象中此标签也删除

③支持批量删除操作

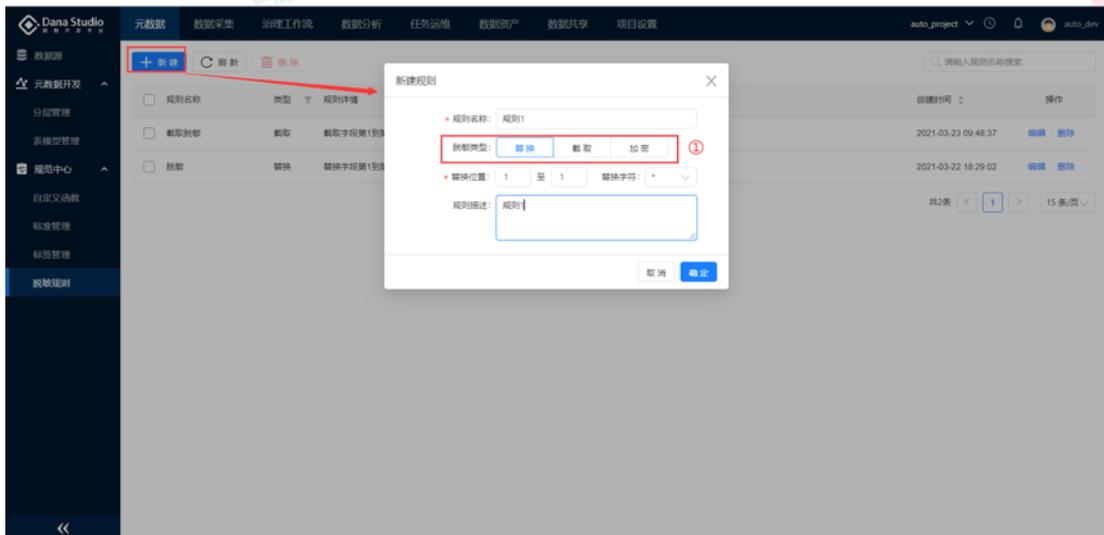
④显示当前标签被使用的次数，表、标准、任务、脚本被删除时，引用次数对应减少

5.3.4 脱敏规则

设置脱敏规则，用于数据共享中字段脱敏。

新建脱敏规则：

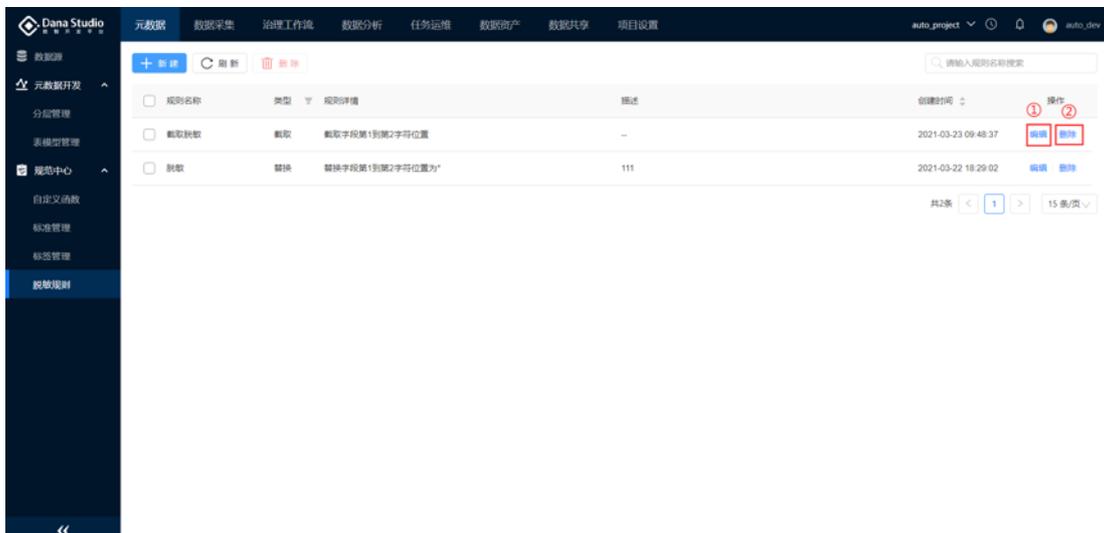
元数据——规范中心——脱敏规则——新建——填入必填项——确定



①支持的脱敏类型如下：

- i. 替换：设置区间内的内容替换为 * - .;
- ii. 截取：可截取指定区间内的值；
- iii. 加密：包含 BASE64、AES 加密

脱敏规则列表：



①点击操作下的“编辑”按钮，可修改脱敏规则

②点击操作下的“删除”按钮，可删除脱敏规则

6 数据采集

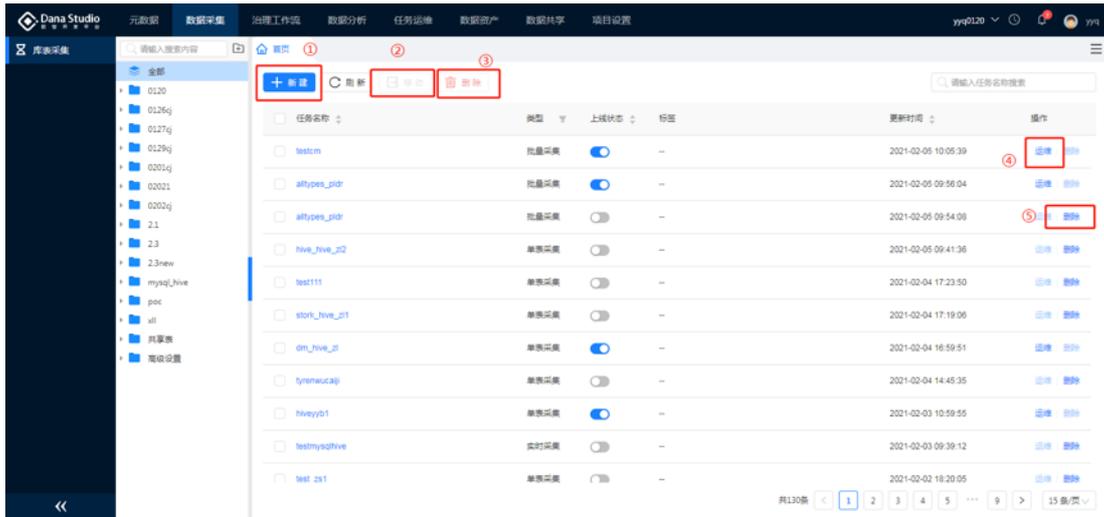
6.1 库表采集

数据采集提供了多源异构的数据接入，并可对接所用通用协议的 Hadoop 大数据平台的能力，灵活的将数据存储到大数据平台中。

支持关系型数据库、MPP 数据库、CSV 文件的多版本数据采集，其中关系型数据库包括 Mysql、Oracle、DB2、SQL server、Postgres，MPP 数据库包括 Teryx、GaussDB、Hive、达梦、优炫。数据采集支持离线、实时两种同步方式，实现数据的全量、增量同步，单表、批量支持覆盖、全量追加、增量追加三种入库策略。同时提供批量创建采集任务的快捷模式，能快速实现多表甚至整库的数据迁移，节省大量时间和人力成本。

数据库	hive (DDP 4.5)				stork (3.2)				teryx (3.2/3.3)				GaussDB				达梦				优炫																			
	分布方式：无								分布方式：hash, replication, randomly								分布方式：hash, replication								分布方式：hash, replication, randomly, append, list								分布方式：hash							
	单例	批量	增量	实时	单例	批量	增量	实时	单例	批量	增量	实时	单例	批量	增量	实时	单例	批量	增量	实时	单例	批量	增量	实时	单例	批量	增量	实时												
mysql 5.7	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	不支持	√	√	√	不支持	√	√	√	√	√	√	√	不支持	√	√	√								
mysql 8	√	√	√	未测试	√	√	√	未测试	√	√	√	未测试	√	√	√	未测试	不支持	√	√	√	不支持	√	√	√	√	√	√	√	不支持	√	√	√								
stork 3.0/3.1/3.2 (对应 postgresql 10.1/10.7/12.X)	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	√	不支持	√	√	√	√	√	√	√	不支持	√	√	√								
oracle 10g/11g	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	不支持	√	√	√	不支持	√	√	√	√	√	√	√	不支持	√	√	√								
oracle 12c	√	√	√	未测试	√	√	√	未测试	√	√	√	未测试	√	√	√	未测试	不支持	√	√	√	不支持	√	√	√	√	√	√	√	不支持	√	√	√								
teryx 3.0/3.1/3.2 (对应 greenplum9.2.X/--9.4.X)	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	√	不支持	√	√	√	√	√	√	√	不支持	√	√	√								
DB2 11.1/11.5	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	√	不支持	√	√	√	√	√	√	√	不支持	√	√	√								
DB2 10.5	√	√	√	未测试	√	√	√	未测试	√	√	√	未测试	√	√	√	未测试	不支持	√	√	√	不支持	√	√	√	√	√	√	√	不支持	√	√	√								
sqlserver 2005/2008	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	√	不支持	√	√	√	√	√	√	√	不支持	√	√	√								
hive 3.1	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	√	不支持	√	√	√	√	√	√	√	不支持	√	√	√								
csv	√	不支持	不支持	不支持	√	不支持	不支持	不支持	√	不支持	不支持	不支持	√	不支持	不支持	不支持	√	不支持	不支持	不支持	√	不支持	不支持	不支持	√	不支持	不支持	不支持	√	不支持	不支持	不支持								
GaussDB200/300	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	√	不支持	√	√	√	√	√	√	√	不支持	√	√	√								
达梦	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	√	不支持	√	√	√	√	√	√	√	不支持	√	√	√								
优炫	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	不支持	√	√	√	√	不支持	√	√	√	√	√	√	√	不支持	√	√	√								

查看采集任务列表：



- ①点击“新建”，新建采集任务，可选：单表、批量、实时、自定义
- ②点击“移动”，可以批量移动采集任务，目前重名规则为同一目录下不重名
- ③点击“删除”，可以批量删除采集任务
- ④点击操作下的“运维”按钮，可以前往该任务对应的周期任务页面
- ⑤点击操作下的“删除”按钮，删除采集任务

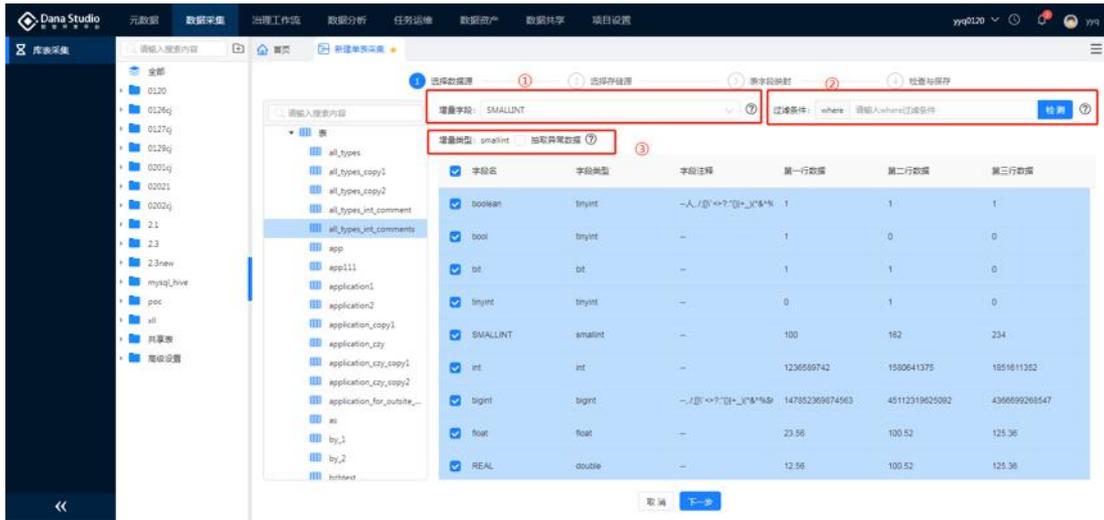
6.1.1 离线采集（单表、批量、自定义）

新建单表采集：

数据采集——库表采集——单表采集——选择数据源——选择存储源——表字段

映射——检查与保存

选择数据源：



①增量字段可以根据设置的时间类型、string 类型、整型字段进行增量采集，设置的时间字段根据当天采集到数据的最大时间(采集数据是开始时间到当日执行时间之间数据)作为下次采集的开始时间，设置的整型字段根据当前采集到的最大值作为下次采集的依据。

目前支持数据库：oracle、mysql、DB2、postgres、teryx、sqlserver、GaussDB、达梦、优炫、hive

string 类型时间格式支持：

yyyy-MM-dd HH:mm:ss yyyy-MM-dd

yyyy/MM/dd HH:mm:ss yyyy/MM/dd

yyyyMMddHHmmss yyyMMdd

时间类型支持：date、datetime、timestamp

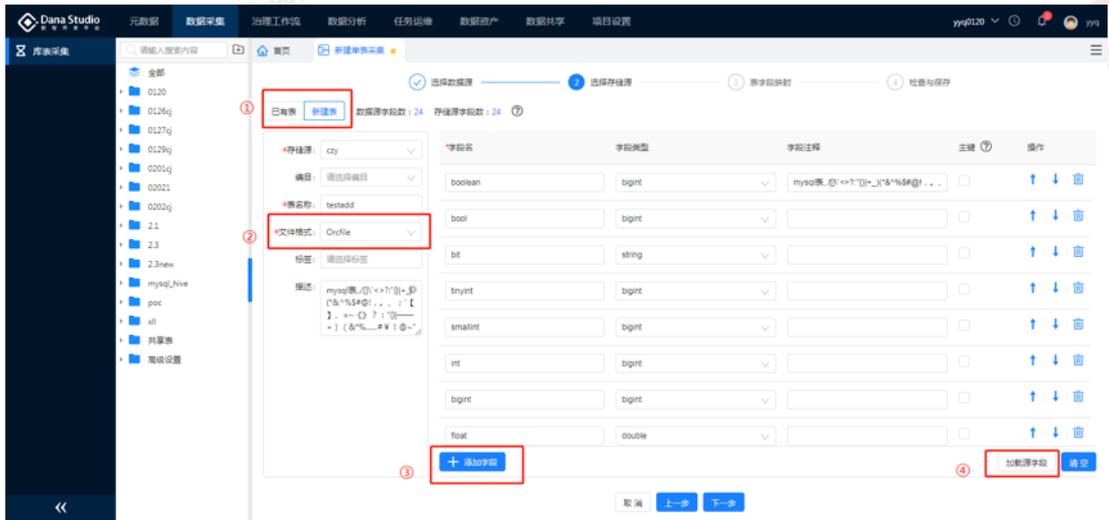
②设置增量字段后，支持选择是否抽取异常数据

勾选“抽取异常数据”时，第一次运行抽取任务时将所有非业务型异常数据(时间字段为空或者为乱码时)抽取到存储源，后续增量抽取还是按照增量条件只抽取符合增量规则的数据；

不勾选“抽取异常数据”时，增量抽取将不抽取异常数据，只抽取上次执行到本次执行时间范围内的数据。

③支持输入 SQL 的 WHERE 筛选字段（不包括 WHERE）进行字段筛选

选择存储源：



①支持选择新建表或选择存储源中的已有表

②存储引擎为 hive 时，可抽取 textfile 和 orcfile 文件格式；

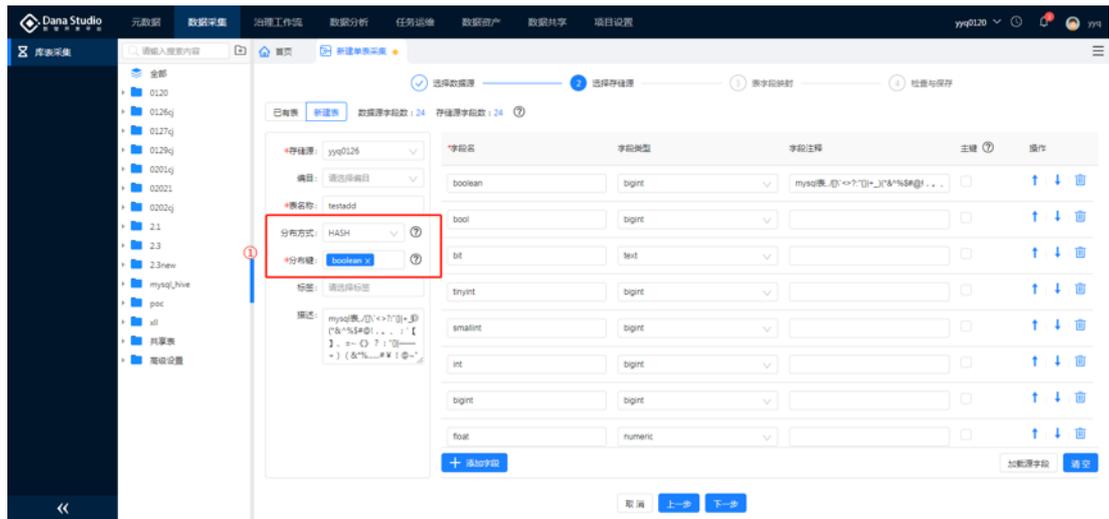
存储引擎为 stork、teryx、GaussDB、达梦、优炫时，不支持选择文件格式

③点击“添加字段”，可对表进行添加字段的操作

④点击“加载源字段”，可自动加载出源表的所有字段

注：1) 选择新建表时，表描述、字段注释不支持单引号 (')、分号 (;)、双短横线 (--) 等特殊字符

2) 选择已有表时，如果已有表的表描述、字段注释带有单引号 (')、分号 (;)、双短横线 (--) 等特殊字符时，展示时会以空格替代



①存储引擎为 teryx、GaussDB、达梦时，可设置分布方式、分布键，存储引擎为优炫时，可设置分布键

teryx、达梦分布方式：有三种类型 HASH、REPLICATION、RANDOMLY

GaussDB：有两种类型 HASH、REPLICATION

优炫分布方式：仅有一种类型 HASH

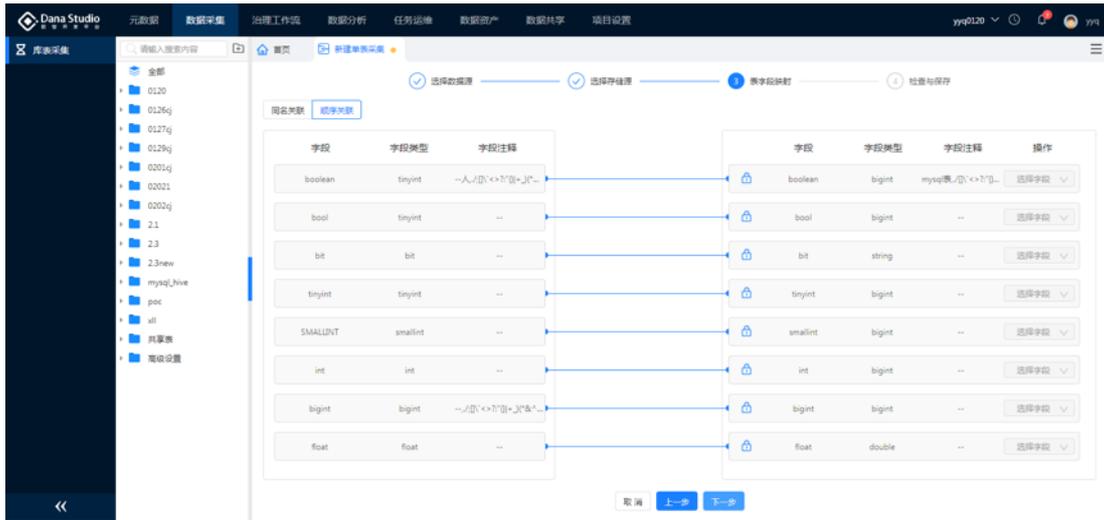
a.HASH：选择 HASH 分布方式时，会出现分布键；当表中存在主键时，分布键只可以在主键中选择，如果表中不存在主键时，分布键可以任意设置；

b.REPLICATION：选择复制分布时，不会出现分布键选择；

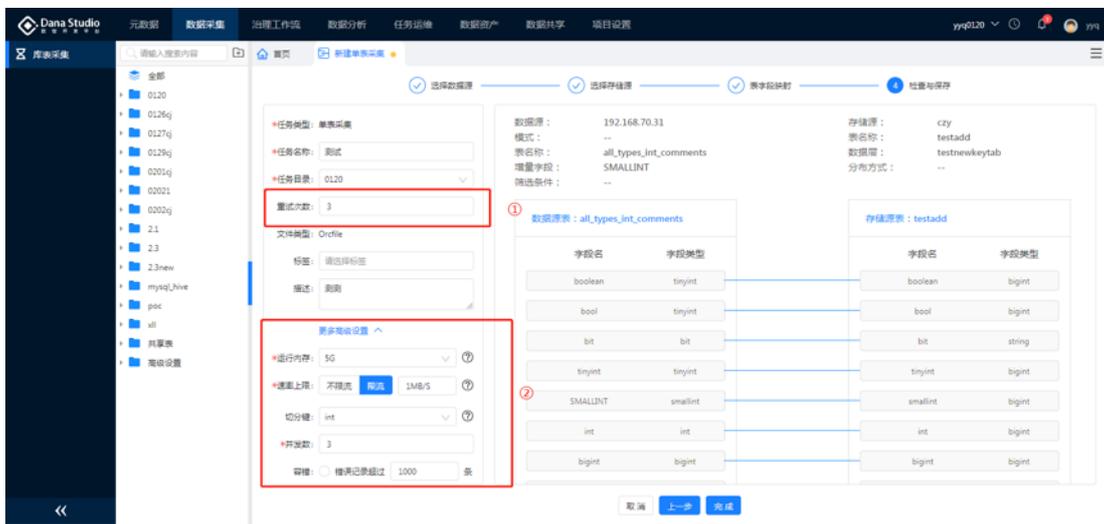
c.RANDOMLY：选择随机分布时，不会出现分布键选择；当表存在主键时，RANDOMLY 分布选项禁用；如果已经选择了 RANDOMLY 分布，后设置表主键，则默认选择到 HASH 分布；

注：普通用户若没有整库权限，则无法往该库中新建表

表字段映射：



检查与保存:

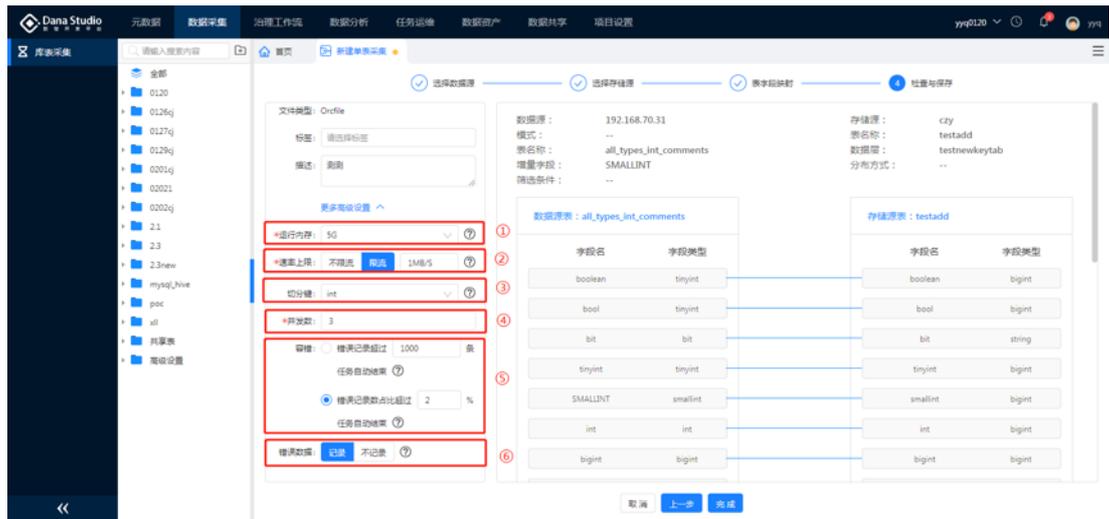


①支持设置重试次数：当重试次数大于 0，任务运行失败后会进行重试

②支持高级设置：根据设置运行采集任务，提升离线采集性能

注： csv 不支持高级设置，其余数据源类型均支持

检查与保存—高级设置:



高级设置支持配置的参数：运行内存、速率上限、切分键、并发数、容错、错误数据记录；

①运行内存设置范围：1-32G，如果超出当前环境运行的内存，任务会运行失败；

②速率上限：通过限流可以保护数据来源端或者数据去向端的读写压力，不限流的情况下则会提供现有硬件环境下最大的传输性能；

③切分键：根据配置的字段进行数据分片采集，与作业并发数配合使用实现并发采集。可以将源数据表中某一列作为切分键，建议使用主键或有索引的列作为切分键，目前仅整型和字符串类型支持设置切分键，整型类型字段值包含中文时不可以设置为切分键，任务会运行失败；

④并发数：配置并行抽取的任务数，只有配置切分键，采集可以配置并发数，如果未选择切分键时，并发数默认为 1，如果选择了切分键字段，并发数默认 1，可配置取值 ≥ 1 的正整数；

⑤容错：支持单选，有两个选项：错误记录超过设置值，任务自动结束和错误记录数占比超过设置值，任务自动结束；

a.错误记录超过设置值，任务自动结束：当该周期内采集的错误数据记录数超过设置的错误阈值时，任务会报错；若不超过设置的错误记录数，任务可以执行成功，但错误数据会被记录到 datax 本地；

b.错误记录数占比超过设置数，任务自动结束：当该周期内采集的错误数据超过占比超过设置的值时，任务会报错；若不超过设置的错误百分比，任务可以执行成功。但错误数据会被记录到 datax 本地；

⑥错误数据：可以选择记录、不记录，选择记录时，Datax 能记录错误数据到运行节点本地，存储路径为：设置的 datax 错误数据存储目录/errorData/任务名/任务名_时间戳.csv，时间戳格式如：20200101080000；

单表采集其他操作：



①点击“转化为自定义任务”，可将此单例抽取任务转化为自定义采集任务（此操作不可逆）

②点击“查看代码”，可查看该采集任务的脚本代码

注：为保证“转化为自定义任务”和自定义采集的模板格式一致，可根据如下两步操作直接在自定义采集任务中运行

1) 拷贝 content 中 job 的内容（注意包括 job 前的大括号）

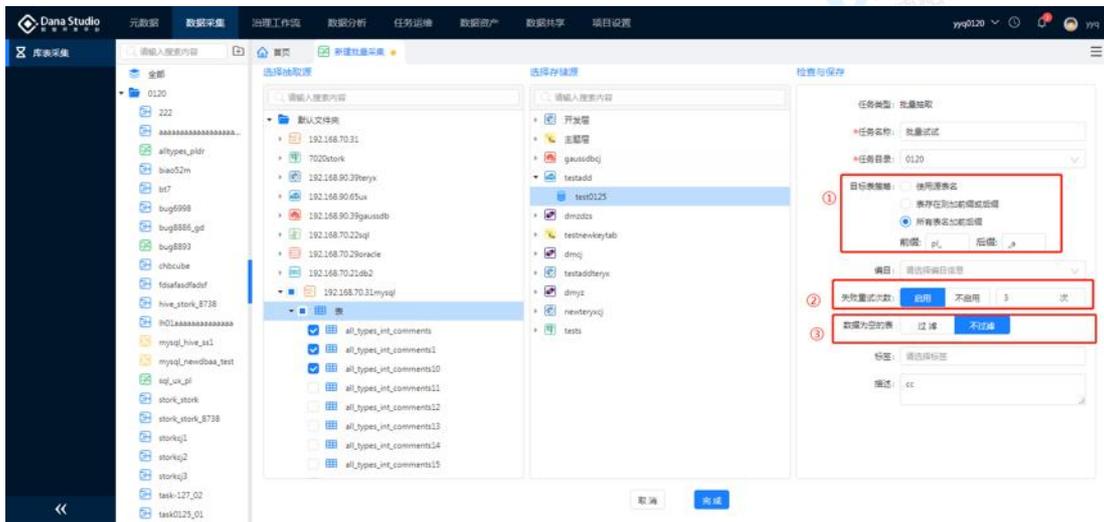
2) 将内容中的“\\”全部替换为空

新建批量采集：

数据采集——库表采集——批量采集——选择数据源——选择存储源——检查与保存

注：1) 当前最多支持 2000 张表的批量采集

2) 普通用户若没有整库权限，则无法往该库中新建表

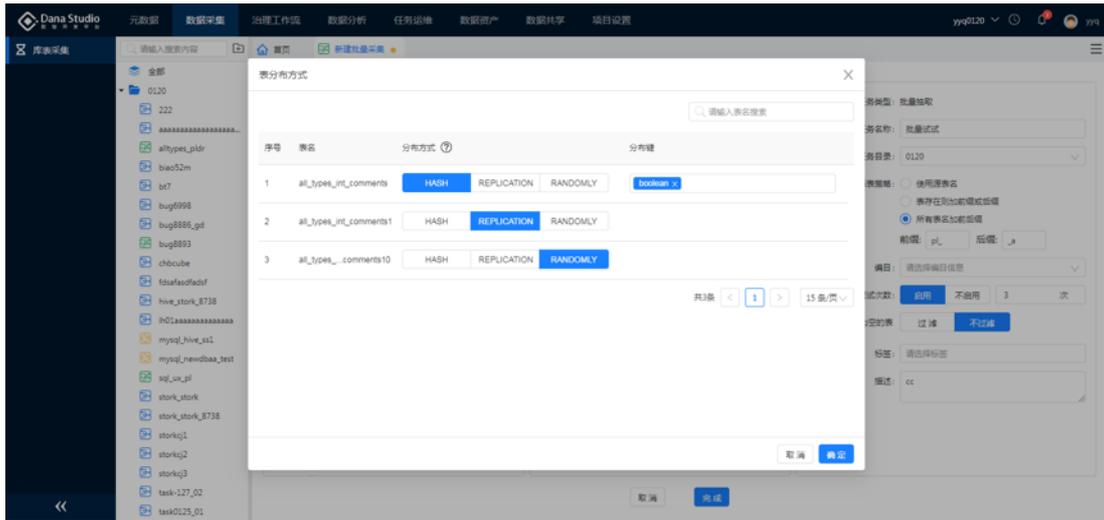


①目标表策略：使用源表名、表存在则加前缀或后缀、所有表名加前缀或后缀三种策略

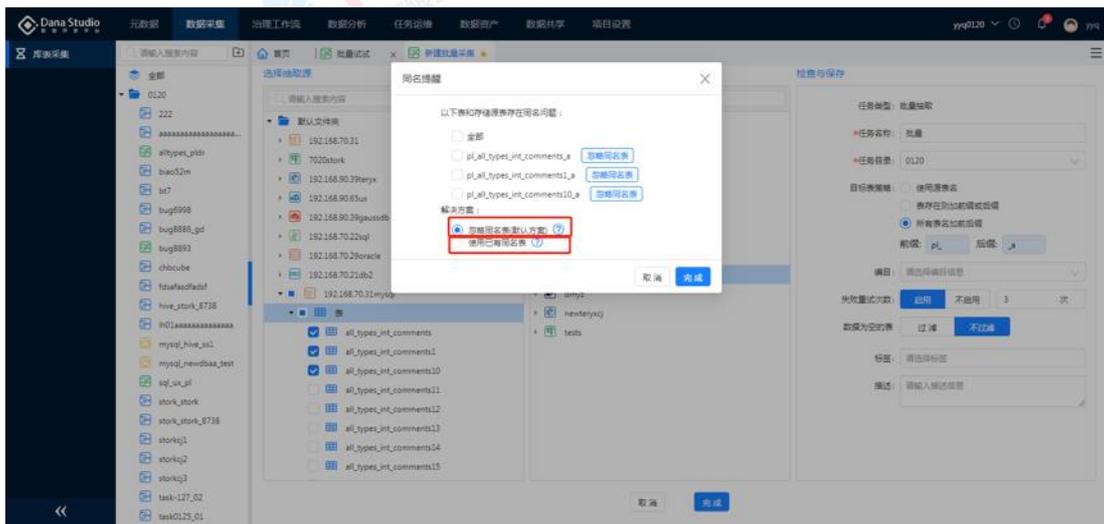
②失败重试次数：可以选择启用、不启用；选择启用时，可以配置重试次数，当重试次数大于 0，任务运行失败后会进行重试

③数据为空的表：可以选择过滤、不过滤，选择过滤，任务运行后，所有空表不会被抽取

点击“完成”后，如选择存储源为 teryx、GaussDB、达梦引擎，会弹出表分布方式弹窗，设置表的分布方式、分布键，如选择存储源为优炫引擎，会弹出表分布方式弹窗，设置表的分布键；



其他存储源点击“完成”或 teryx、gaussdb、达梦、优炫存储源点击“确定”后，若出现同名表的情况，会弹出表同名提醒弹窗，给每个同名表设置解决方案（选择需要操作的表，再选择方案）

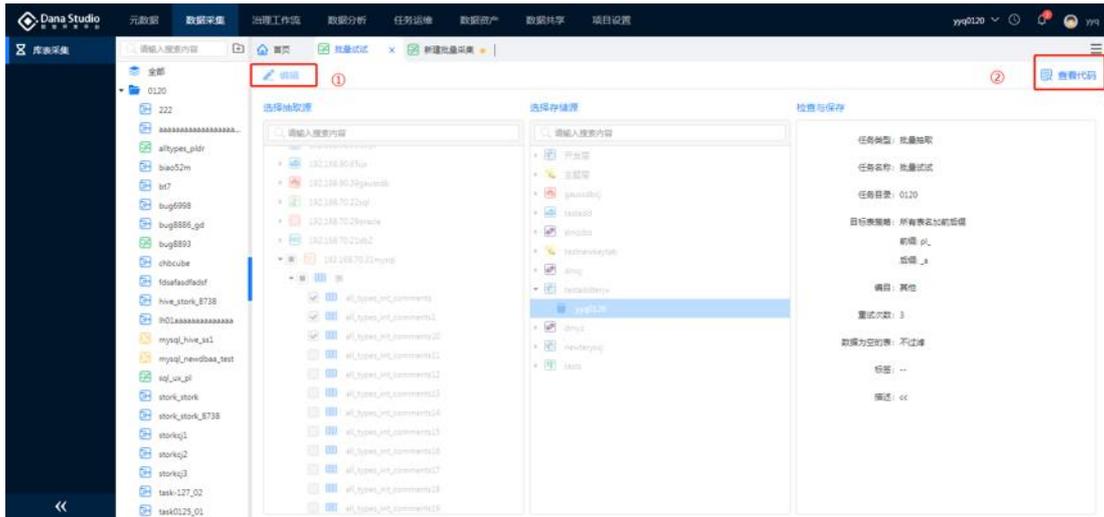


①忽略同名表（默认方案）：同名表不进行抽取

②使用已有同名表：目标表存在则不创建直接使用

注：使用已有同名表需保证源表字段名与同名表表字段一致（忽略大小写），否则任务运行失败

批量抽取的其他操作：

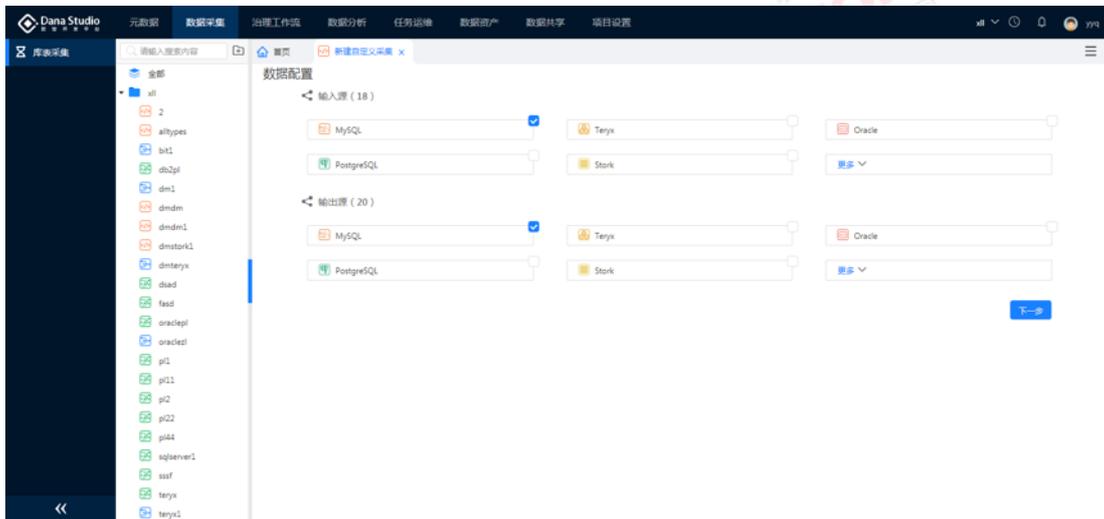


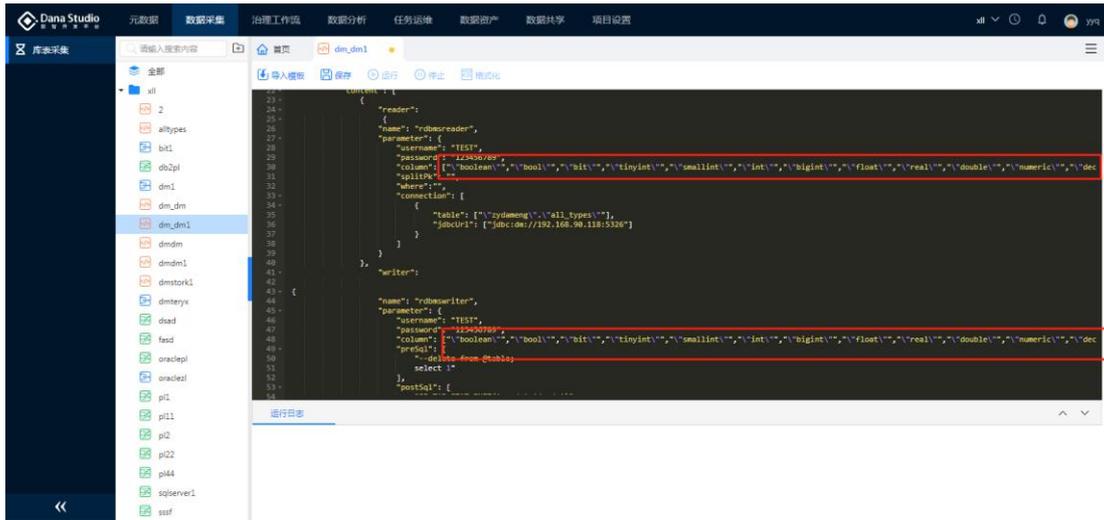
①可以编辑批量采集任务；

②查看批量采集任务的代码；

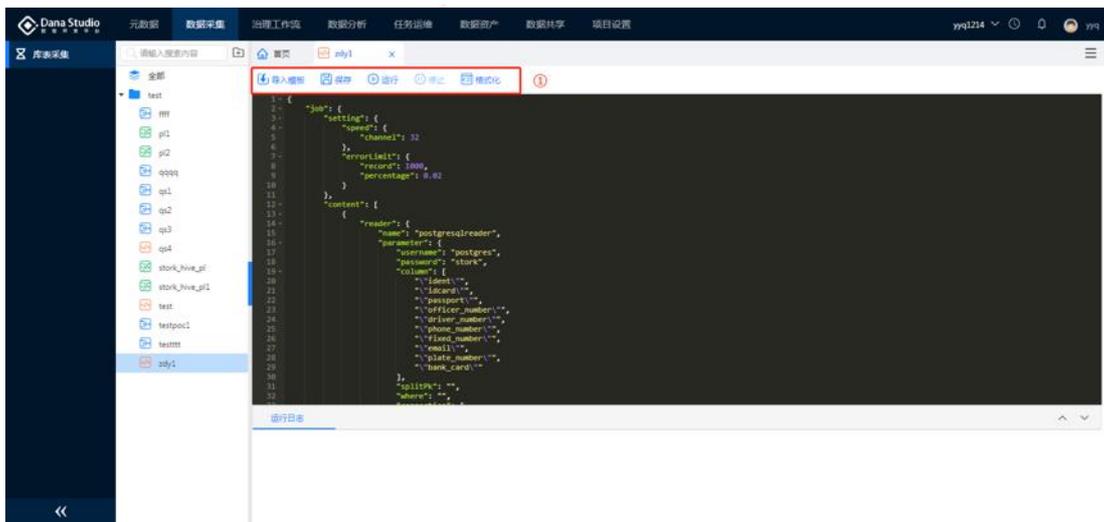
新建自定义采集：

数据采集——库表采集——新建自定义采集——选择输入源、输出源——新建——配置相应的参数，如 ip、port、dbname、cols 等——保存——选择保存脚本的脚本名称、所属目录——确定——上线——配置时间策略，提交任务





注：存储源为达梦（RDBMS）、stork、teryx、gaussdb、优炫，字段名为关键字时，字段前后需要加上 `和 `进行转义



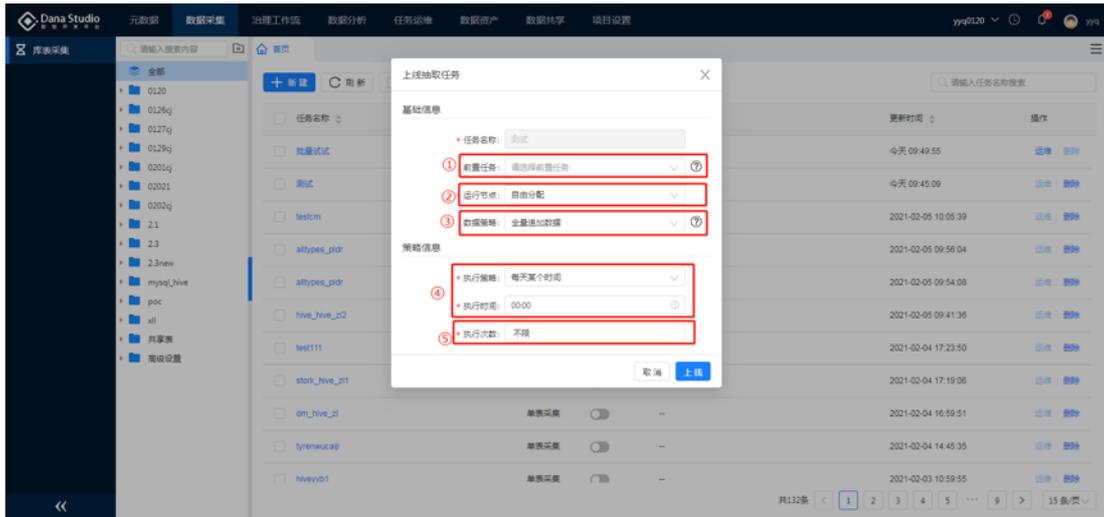
①支持导入、保存、运行、停止等操作

注：自定义采集任务上线后，在任务运维-作业管理中每运行一次，数据全量追加一次。其中，存储源选择 hive 时，可通过修改脚本的方式实现数据的覆盖。

上线：

离线同步成功添加后，可进行“上线”“删除”等操作

点击“上线”按钮，配置时间策略、数据策略，成功后在“任务运维——周期作业”页面查看执行记录



①可设置前置任务，支持在下拉框中搜索，方便找寻任务

若配置了前置任务，后置任务在前置任务执行成功后，才根据执行策略开始执行；

前置任务一旦执行失败，后置任务不再执行。

②自由分配为负载均衡，若需要任务在单一节点运行可手动选择

③数据策略：

a：未设置增量字段的任务可以选择覆盖存储表数据、全量追加数据：

1) 全量追加数据（默认）：每次执行将抽取源所有数据追加至存储源表中

2) 覆盖存储表数据：任务运行成功，数据覆盖原有表；任务运行失败，数据不进行覆盖，原有数据也不清空

b：设置了增量字段的任务可以额外选择增量追加数据：

1) 全量追加数据（默认）：上线第一次执行，全量采集，并追加到已有数据后。第二次及以后执行，向已有数据后边追加增量。

2) 覆盖存储表数据：上线第一次执行，全量采集，并覆盖已有数据。第二次及以后执行，向已有数据后边追加增量。

3) 增量追加数据：上线第一次执行①能获取到存储源表中时间字段或整型字段时，获取存储源表最大时间到本次执行时间之间的抽取源数据或获取存储源表最大值到本次抽取源数据最大值之间的数据追加到存储源表中。②不能获取存储源表中的时间字段或整型字段时，则全量采集并追加到存储源表中。第二次及以后执行，向已有数据后边追加增量。

④可选择按天、周、月某个时间段执行，也可以手动或者使用 cron 命令执行

cron 表达式规则：

1) 默认无第二个参数

2) 不支持 second year 字段，只支持五位字段 min hour day month

weekday

3) 各字段只支持* - , /四种特殊字符，day、weekday 字段不支持? L M C # 字符

month weekday 字段支持缩写、大小写

4) 不支持表达式收尾空格，各字段间不支持多空格

⑤可设置执行总数，执行相应次数后停止策略执行

注：如果设置前置任务，任务上线设置时间策略时建议和前置任务的时间策略设置至少一分钟的时间差

6.1.2 实时采集

前提：

1、数据源已开启归档日志

1) oracle10/11g 开通实时需确保拥有以下权限： create session 连接数据库的权限、select any transaction 支持任意事务权限、select any dictionary 支持任何数据字典权限

2) mysql5.7/8.0.x 开通实时需检测权限： SELECT 查询权限, RELOAD 加载权限, SHOW DATABASES 数据库展示权限, REPLICATION SLAVE 日志读取权限, REPLICATION CLIENT 服务启动权限

2、DanaManager 上已安装 QuickSilver 服务。

实时采集支持场景如下：

hive

是否开启 Kerberos 认证		备注
hive	Kafka	
x	x	
✓	✓	Kafka 必须要安装在 Hive 的节点上
✓	x	

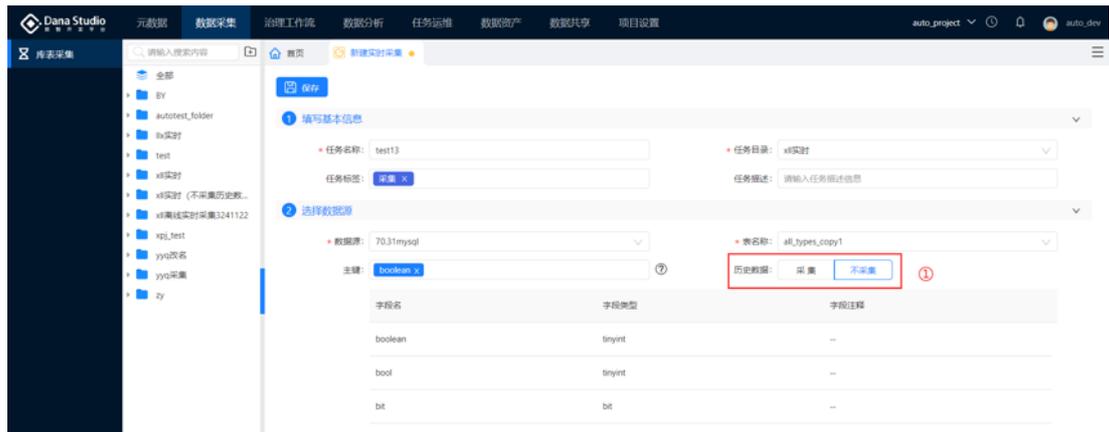
stork、teryx

Teryx	kafka (是否开启 kerberos 认证)	备注
✓	x	
✓	✓	

新建实时采集：

数据采集——库表采集——新建实时采集——分别填写基本信息、选择数据源、选择存储源、字段映射步骤栏——保存

填写基本信息、选择数据源：



①如勾选此项，在创建实时抽取的同时新建一个对应的离线任务，将实时抽取开始之前的数据先行抽取到目标表中（清除表中原数据），离线任务执行完毕后才将实时任务抽取的数据存放到目标表中。该离线任务仅在实时抽取开始前执行一次。该离线任务的名字为：“实时任务名字_同步历史数据”。

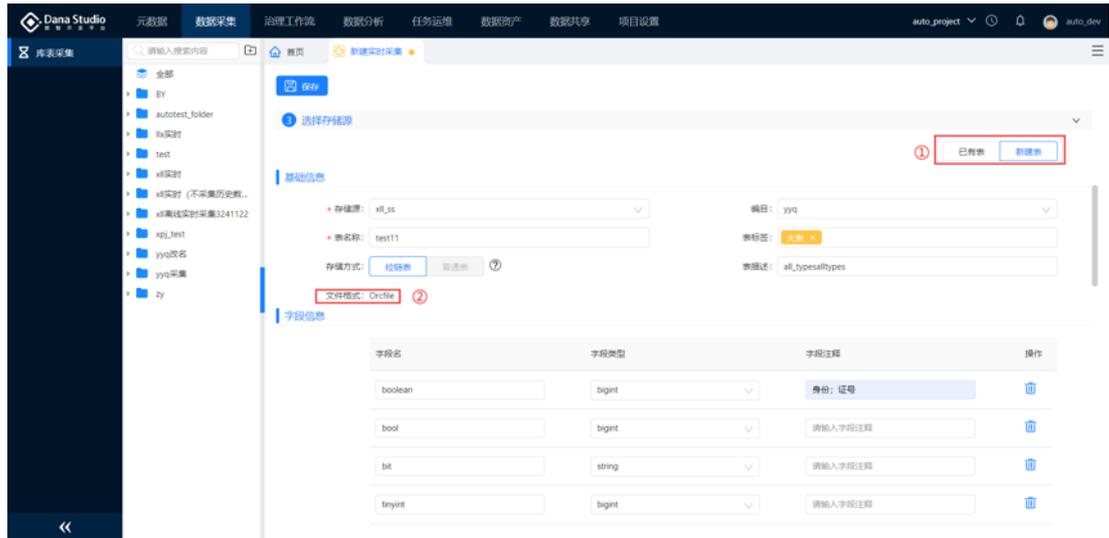
注：1) 只展示支持实时采集的数据源

2) 若源表有主键信息，则自动同步过来，且支持修改，若源表没有主键信息，则支持手动设置

3) 若选择没有主键的表或未设置主键，无法采集 teryx 存储源

4) 输入源的同一张表仅支持建一个实时任务。若表已经建过实时任务，则该表无法选择。

选择存储源：



①支持选择新建表或选择存储源中的已有表

②存储引擎为 hive 时，可抽取 orc 文件类型；存储引擎为 stork、teryx 时，不支持选择文件类型

注：1) 暂不支持实时采集到 GaussDB200/300、达梦

2) hive 实时表会默认建成拉链表，增加两个字段

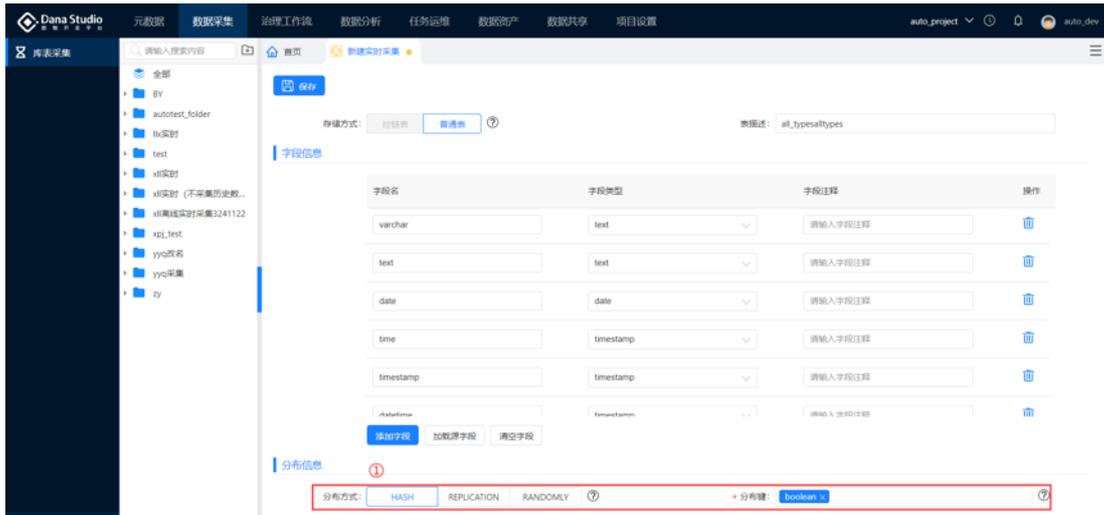
`_op(string),_source_ts_ms(timestamp)`

`_op` 标识记录的变化状态：增、删、改

`_source_ts_ms` 标识记录的修改时间

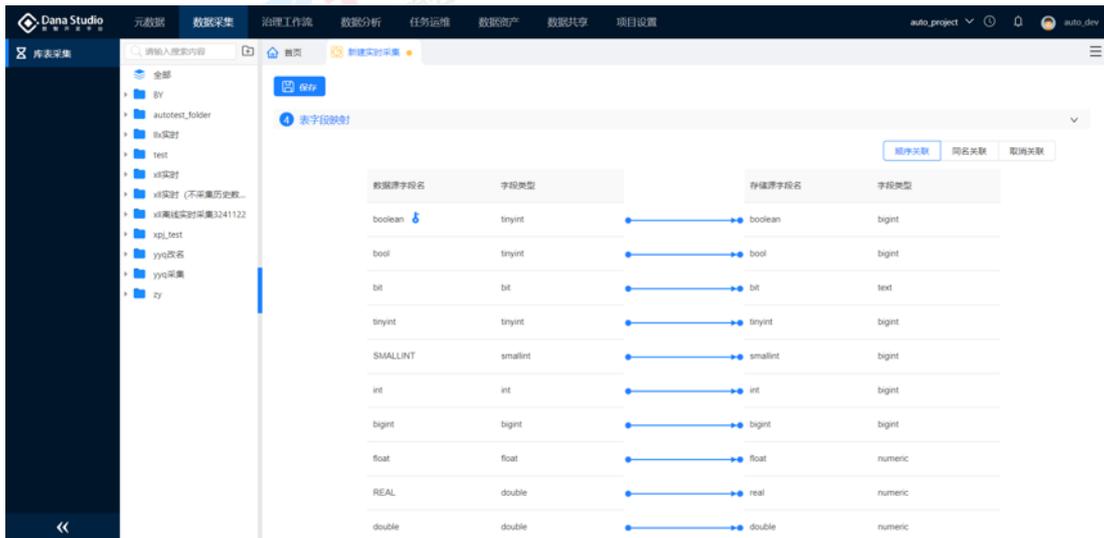
3) 选择已有表时，如果已有表的表描述、字段注释带有单引号 (')、分号

(;)、双短横线 (--) 等特殊字符时，展示时会以空格替代



①存储引擎为 teryx 时，支持设置分布方式；存储引擎为 hive 和 stork 时，不支持设置分布方式

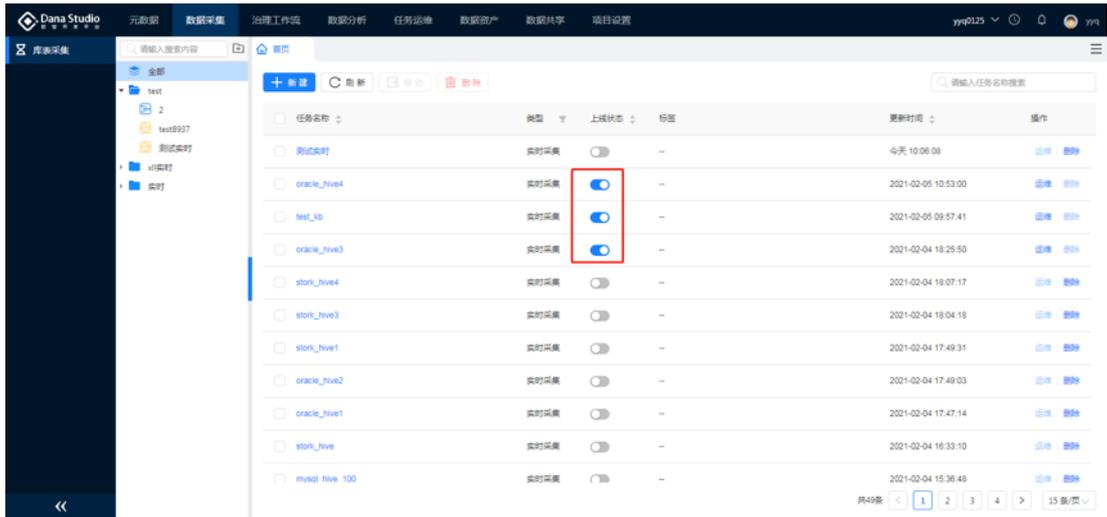
字段映射:



上线实时采集任务:

在任务列表界面，点击实时采集任务的上线按钮——在“任务运维——实时作业”中运行

实时抽取点击上线按钮，直接上线成功，离线抽取开始抽取数据，实时抽取展示为未运行，此时不可以下线、暂停。离线抽取成功后，实时抽取展示成功，失败则展示失败，离线抽取任务日志展示相关信息。



注：1) 原始表无时间戳，抽取后的表数据顺序可能混乱

2) 实时任务中的离线任务未运行结束时，无法下线实时任务

3) DS V4.7.1 版本的实时模块不兼容 DS V4.6.4 版本之前的发布版本，建议清空历史数据使用

6.2 API 接口方式

通过提供的 API 接口服务，资源需求方获取接口服务使用授权后，可以对服务接口进行调用。

接口服务模式是属于一种较高层次的共享模式，各业务部门的应用系统建立平台的接口，通过系统与系统之间的接口交互模式，实现数据的共享与调用。支持常见的接口服务有 WebService、API 等多种接口类型，满足平台的接口服务采集需求。

6.2.1 数据共享 API 接口

通过数据共享接口方式，实现平台上所有数据的访问、分类、索引、检索、关联、分享、流转等功能。对于结构化数据支持通过标准 SQL 进行数据查询，对于视

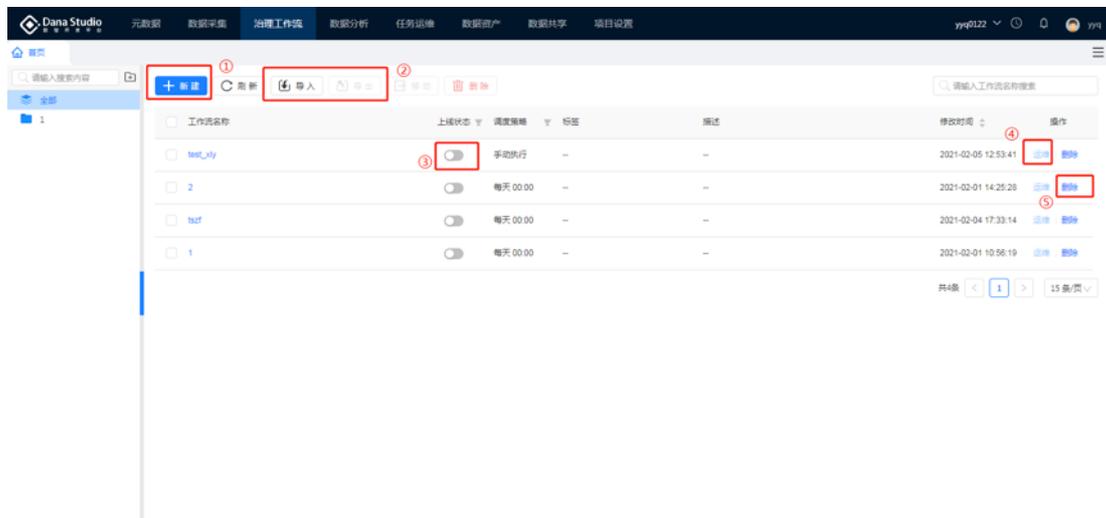
频、图像、文本等非结构化数据，支持对文件的预览、转码、截取、全文检索等功能。

6.2.2 应用开发 API 接口

应用开发接口支持完备的上层应用开发所需的各种功能接口，包括存储服务、文件系统管理、用户及权限、硬件管理、消息、流程、地图、报表、传输、应用、流媒体、数据可视化、检索、目录管理、虚拟化等多种类型开发接口，并且提供多种开发语言的支持。

7 数据治理

治理 workflow 页面



①点击“新建”按钮，新建 workflow

②点击“导入”、“导出”按钮，可以导入、导出 workflow

③点击上/下线按钮，可以上下线 workflow

④点击“运维”按钮，打开新的 tab，定位展示该 workflow 的周期任务页面

⑤点击“删除”按钮，可以删除 workflow

注：①如果设置前置任务，任务上线设置时间策略时建议和前置任务的时间策略设置至少一分钟的时间差；

②导入规则：

1) 导入文件内的工作流与已有部分重名，不勾选重名时覆盖，则提示导入异常，重名 workflow 导入失败，其余 workflow 导入成功（脚本导入同此规则）

2) 导入文件内的工作流与已有重名，不勾选重名时覆盖，workflow 导入失败（脚本导入同此规则）

3) 导入文件内的工作流与已有重名，勾选重名时覆盖，workflow 导入成功（脚本导入同此规则）

4) workflow 包含的脚本按照原目录同步到现有脚本目录中

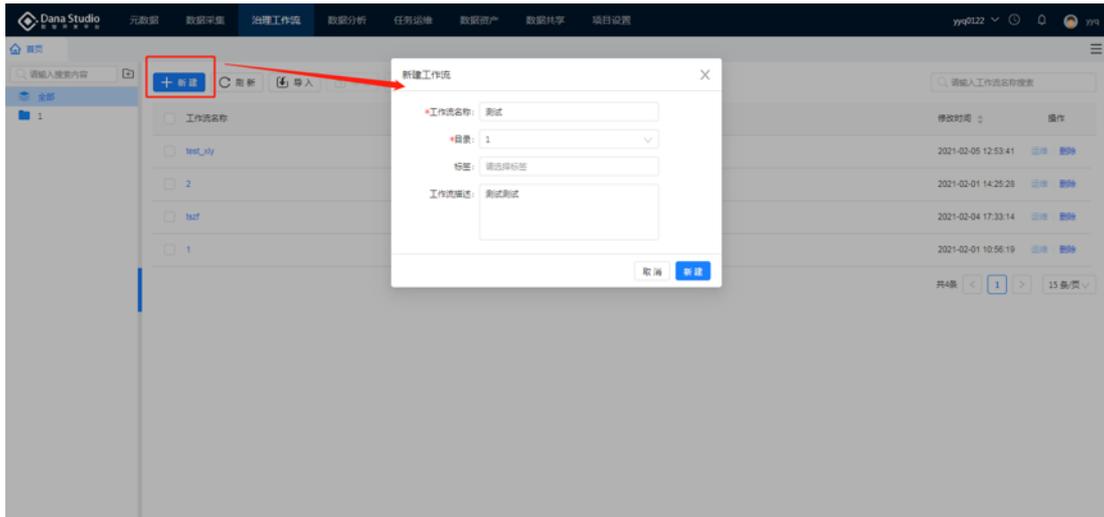
导出规则：

1) 已上线的工作流、数据同步的工作流、包含数据治理节点的工作流不允许导出

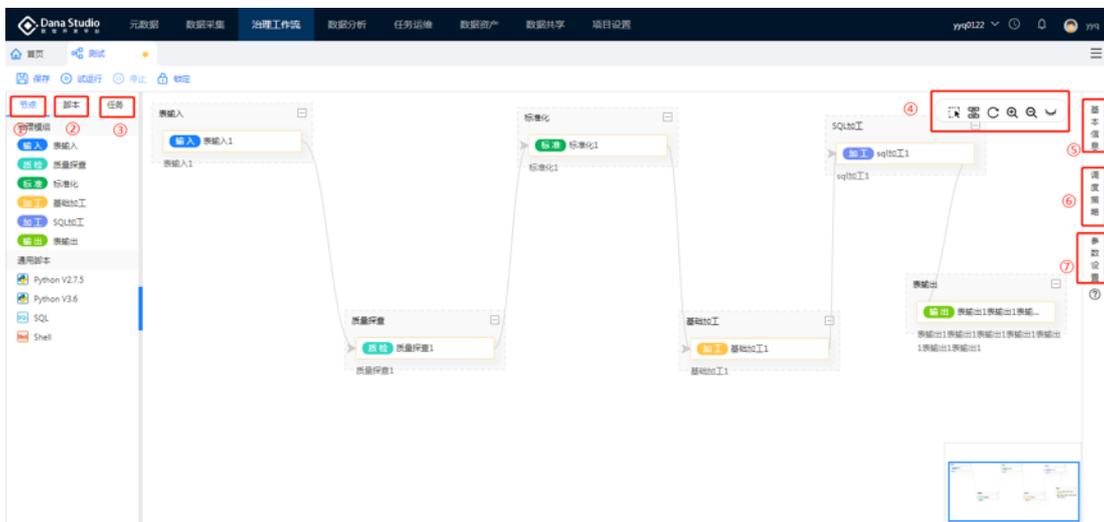
2) 一次导出不可超 10 条 workflow

新建治理 workflow

治理 workflow——点击“新建”按钮——填入相关必填项——新建



治理 workflow 编辑页面



①节点下包括治理模组、通用脚本，可拖拽新建模组、脚本

治理模组包括：表输入、质量探查、标准化、基础加工、SQL 加工、表输出

通用脚本包括：Python V2.7.5、Python V3.6、SQL、Shell

②脚本下包含本项目的脚本，支持新建、导入

注：如果模组名称过长时，可以 hover 在模组上，气泡展示全部名称

节点：

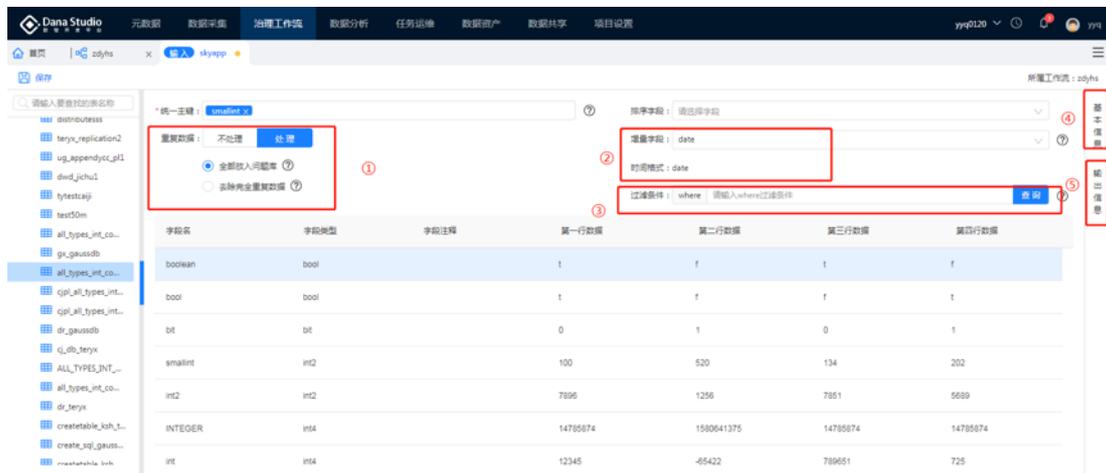
治理流程：治理 workflow —— workflow 任务 —— 可通过拖拽选择新建治理模组或者脚本 —— 保存 —— 试运行

7.1 表输入

表输入中可以选择配置过的可操作表资源中的表作为处理对象，通过设置增量字段以实现增量输入、排序规则进行排序、过滤条件进行过滤，设置统一主键以及两种去重方式配置项。

配置表输入：

双击进入表输入模组——选择输入源——填写配置信息——查看输出信息——保存



①重复内容处理：

- 1) 全部放入问题库：所有重复数据统一放入问题库，不进入后续治理
- 2) 去除完全重复数据：对于所有字段信息完全重复的数据，系统去除全部重复

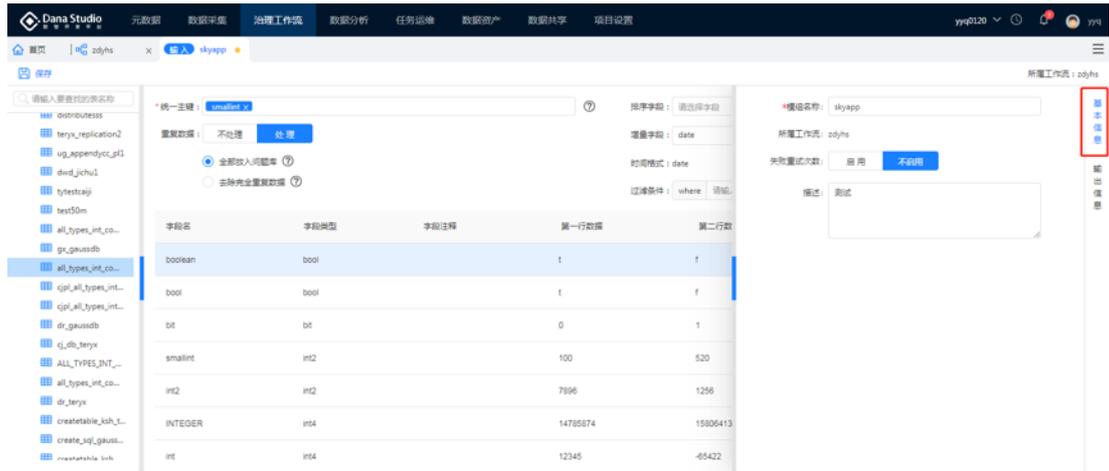
数据只保留一条，其他重复数据被完全删除，不会放入问题库

②增量字段：支持选择时间类型或时间格式的 string 类型字段作为增量字段

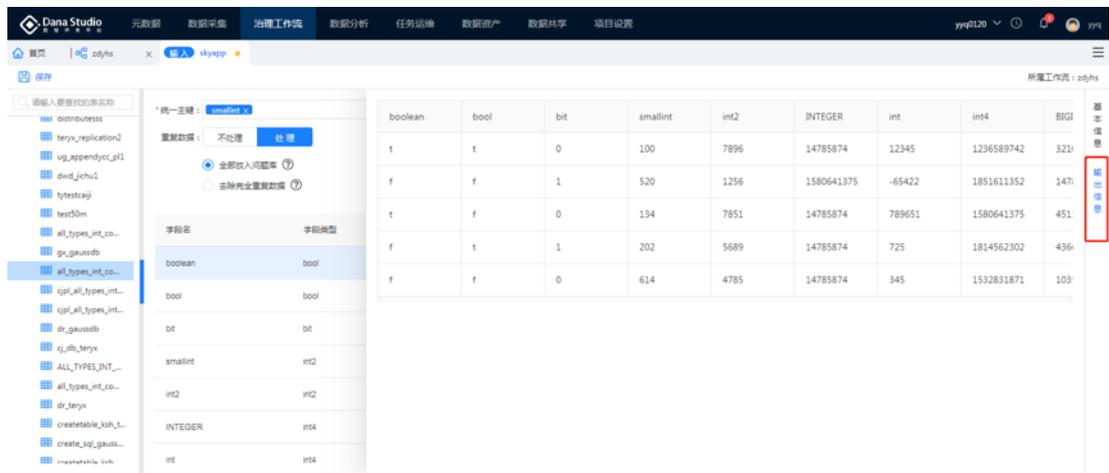
③过滤条件：根据所填写的 where 语句进行条件查询后预览输出

④点击“基础信息”按钮，打开基础信息抽屉，可以修改模组名称、配置失败重

试次数



⑤点击“输出信息”按钮，展示按照设置的条件查询出的前 1000 条记录



7.2 质量探查

质量探查是对输入数据的质量检测，是开发者对数据的质量（比如脏数据比例）进行解和判断的工具，该模组对数据不产生直接处理操作，但有过滤作用。

配置质量探查：

双击质量探查模组——选中某个字段，配置探查规则——配置完所有字段的探查规则后——查看输出信息——保存



探查规则分为空值探查、基础探查和深度探查，不符合探查规则的数据会被识别为问题数据：

①空值探查：勾选后，空值会被识别为问题数据并放入问题库

②基础探查有：

- i. 字段内容：字母、数字、中文、特殊字符（勾选后出现内容长度），int 类型仅有内容长度、数值范围
- ii. 字段内容包含
- iii. 字段内容不包含

③深度探查有：标准规则、业务规则、正则表达式

7.3 标准化

标准化是对数据中不规范的字段和数据进行修改，使其符合治理标准。

配置标准化：

双击标准化模組——设置标准字段名以及关联标准代码表——查看输出信息——

保存

字段名	字段类型	第一行数据	第二行数据	第三行数据	标准字段名	标准代码表	字段注释
boolean	bool	1	f	1	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息
bool	bool	1	f	f	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息
bit	bit	0	1	0	+ 添加标准字段名	sex标准代码 X	请输入注释信息
smallint	int2	100	520	134	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息
int2	int2	7896	1256	7851	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息
INTEGER	int4	14785874	1580641375	14785874	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息
int	int4	12345	-65422	789651	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息
int4	int4	1236589742	1851611352	1580641375	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息
BIGINT	int8	321084190005201400	147852369674563	451123196204275100	cd X	+ 添加标准代码表	请输入注释信息
int8	int8	147852369674563	147852369674563	451123196204275100	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息

①添加标准字段名：将字段更改为数据标准中的别名。

②添加标准代码：将字段中的数据进行聚类，并替换为映射的标准代码值。支持AI 推荐代码表，支持自动映射代码。

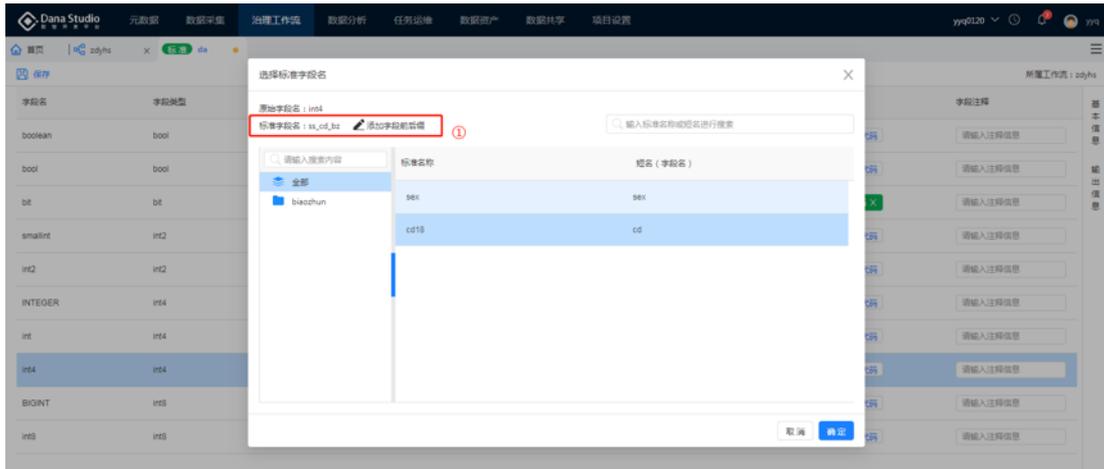
添加标准字段名：

标准化模組——点击“添加标准字段名”——进入选择数据标准字段名弹窗——

确定

字段名	字段类型	第一行数据	第二行数据	第三行数据	标准字段名	标准代码表	字段注释
boolean	bool	1	f	1	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息
bool	bool	1	f	f	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息
bit	bit	0	1	0	+ 添加标准字段名	sex标准代码 X	请输入注释信息
smallint	int2	100	520	134	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息
int2	int2	7896	1256	7851	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息
INTEGER	int4	14785874	1580641375	14785874	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息
int	int4	12345	-65422	789651	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息
int4	int4	1236589742	1851611352	1580641375	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息
BIGINT	int8	321084190005201400	147852369674563	451123196204275100	cd X	+ 添加标准代码表	请输入注释信息
int8	int8	147852369674563	147852369674563	451123196204275100	+ 添加标准字段名	+ 添加标准代码表	请输入注释信息

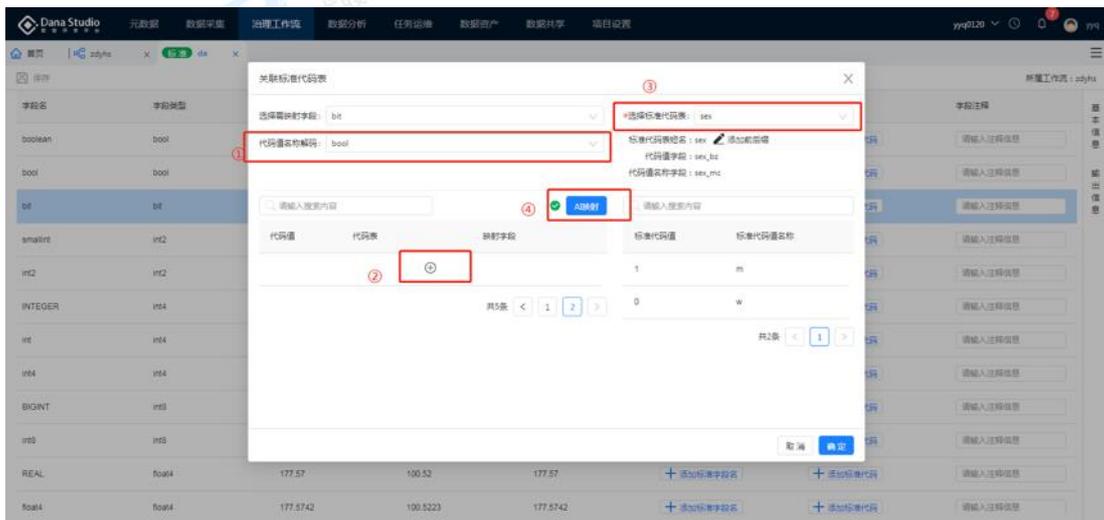
①支持对标准字段名重新选择、删除操作



①支持通过添加字段前后缀，避免与其他字段重名

关联标准代码表：

点击“添加标准代码”——选择代码值名称解码——选择标准代码表——AI 映射——确定



①代码表名称解码：用户手动选择其他字段进行解码

②支持手动添加代码表

③支持对标准代码表名称添加前后缀，避免与其他字段重名

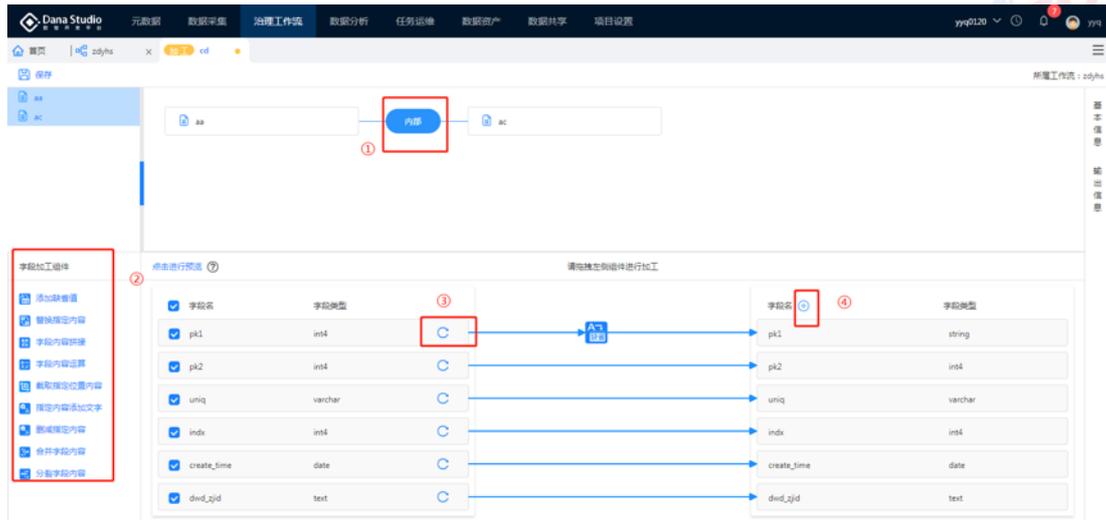
④点击“AI 映射”，能够根据代码值和代码值名称自动去匹配映射，也可手动通过拖动的方式进行映射

注：添加映射后，该表会增加两个字段（代码值字段、代码值名称字段）

7.4 基础加工

配置基础加工：

双击进入基础加工模组——拖入表（支持多表关联）——点击进行预览——拖入
 字段加工组件——查看输出信息——保存



①支持配置多种关联方式（内部、左侧、右侧、完全外侧）

注：多表关联时，需将次表拖入主表上才可关联

②支持以下九种基础加工组件：

- | | | |
|----------|------------|------------|
| 1.添加缺省值 | 2.替换指定内容 | 3.字段内容拼接 |
| 4.字段内容运算 | 5.截取指定位置内容 | 6.指定内容添加文字 |
| 7.删减指定内容 | 8.合并字段内容 | 9.分裂字段内容 |

③点击字段右侧的刷新图标，可清空该字段所有加工组件

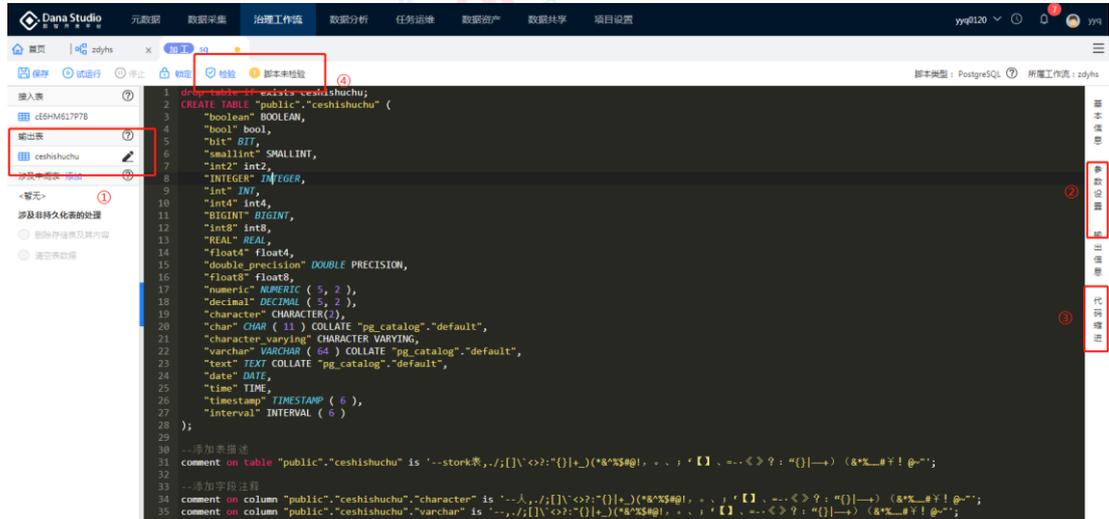
④支持添加输出字段

7.5 SQL 加工

配置 SQL 加工：

双击进入 SQL 加工模组——编写脚本——配置输出表——配置涉及中间表——

试运行——检验——保存



①支持设置涉及表，可对涉及非持久化表进行删除存储表及其内容或清空表数据

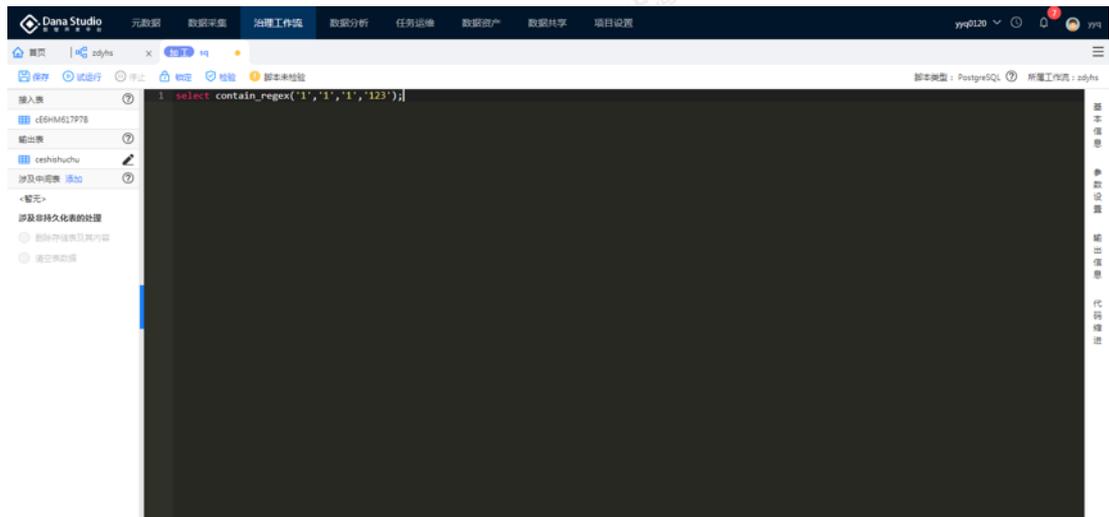
操作

②支持使用时间参数：时间参数与外部工作流时间参数同步，此处只支持查看

③支持代码缩进：空格 2 和 空格 4（默认值）

④点击“检验”按钮，检验脚本是否正确

注：涉及表、输出表需在治理层 basedb 库中存在



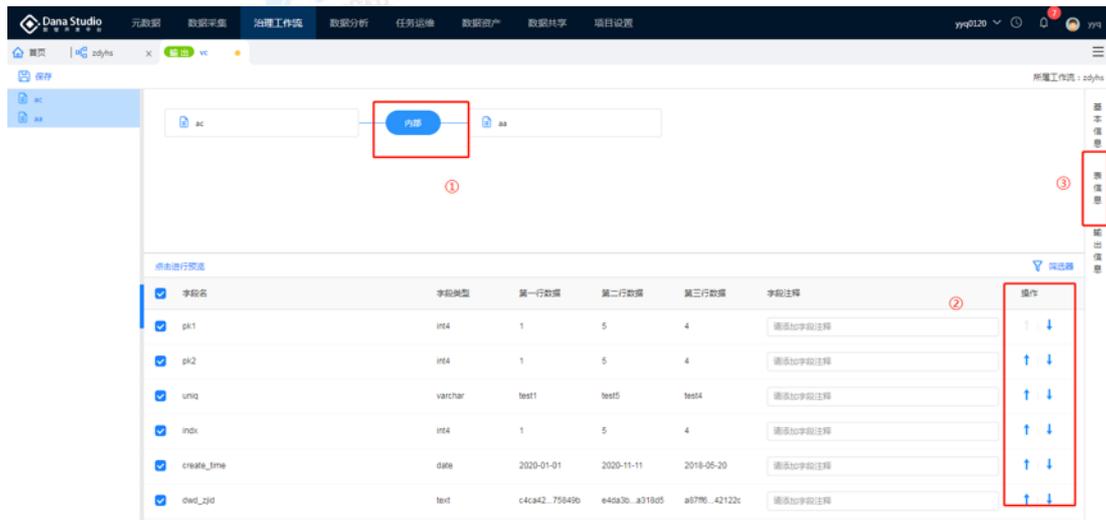
SQL 加工模组可以使用已有的自定义函数，teryx 作为治理 workflow 分层时可以使用内置的自定义函数，hive 作为治理 workflow 分层时可以使用内置或者用户添加的自定义函数，函数调用需和此函数的命令格式一致

7.6 表输出

表输出将通过数据治理后的表输出到主题层进行存储。此模块支持联表查询，设置输出信息和字段。

配置表输出：

双击进入表输出模组——拖入表（支持多表关联）——点击进行预览——勾选需要输出的字段——配置表信息——查看输出信息——保存



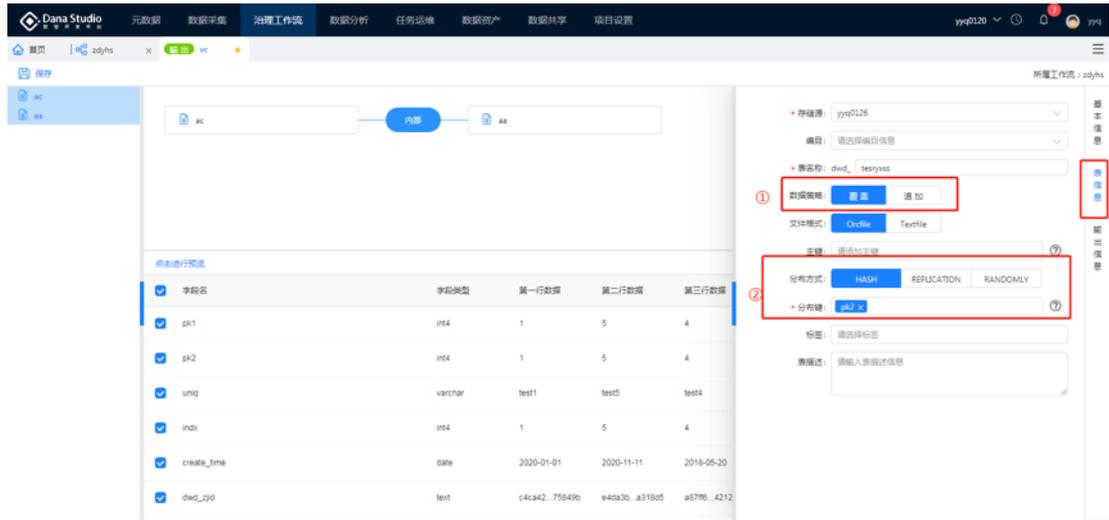
①支持配置多种关联方式（内部、左侧、右侧、完全外侧）

②支持对字段顺序进行调整

③配置表信息，配置输出表表名等信息

注：多表关联时，需将次表拖入主表上才可关联

配置表信息：



① 数据策略支持：覆盖、追加。选择覆盖，实现每次运行治理任务，输出表的数据进行全量覆盖；选择追加，每次运行治理任务，都追加数据到同一张输出表。

② 存储源为 teryx/gaussdb 时，支持分布方式的设置

注：用户进行表输出保存时，自动检测输出表是否同名，若同则检测表结构是否一致，当表结构一致时能够保存成功，否则提示表结构不一致，会再次给出询问是否保存

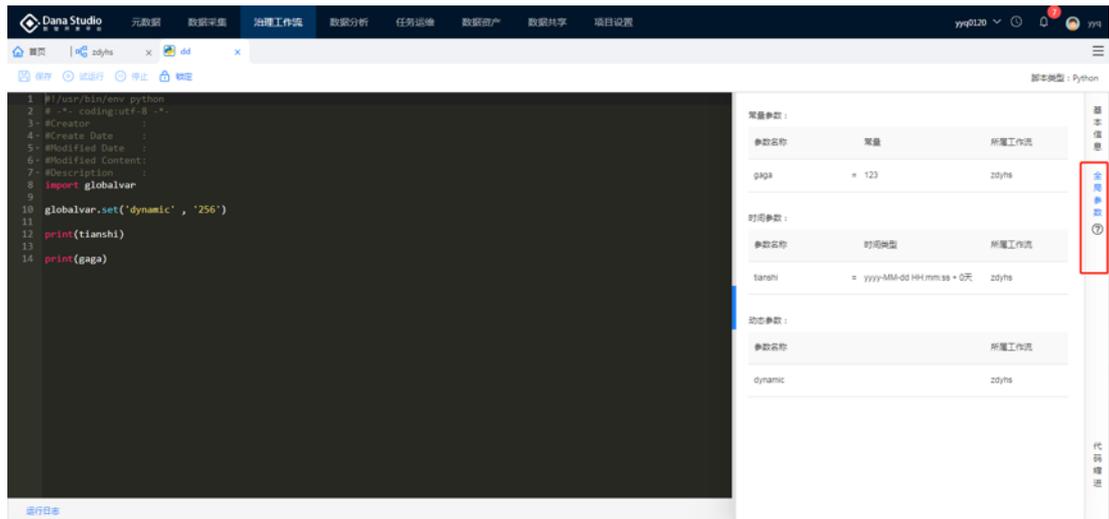
脚本：

脚本支持新建、导入，可在节点下通用脚本处拖拽脚本新建，或在脚本下点击新建按钮新建脚本、导入脚本

1、新建脚本

脚本——新建——新建脚本——配置脚本信息——确定

现支持 python、php、shell、sql 脚本中可设置常量参数、时间参数和动态参数



注：SQL 脚本可以使用已有的自定义函数，teryx 作为治理 workflow 分层时可以使用内置的自定义函数，hive 作为治理 workflow 分层时可以使用内置或者用户添加的自定义函数，函数调用需和此函数的命令格式一致，具体使用方式和 SQL 加工模组一致；

动态参数支持脚本关系如下：

1) Python -> Python

- ①在统一声明变量处声明全局变量
- ②通过系统的 set 和 get 方式设置和获取变量（优先 import 指定头文件）

2) Python -> SQL

- ①在统一声明变量处声明全局变量
- ②在 Python 脚本中通过系统的 set 和 get 方式设置和获取变量（优先 import 指定头文件）
- ③在 SQL 中获取变量，使用方式同常量参数，直接脚本中使用即可

3) 调用参照：dynamic 为动态参数

A.py :

```
import globalvar

globalvar.set('dynamic', '256')
```

B.py :

```
import globalvar

numrice = globalvar.get('dynamic')

print(numrice) # 输出 256

globalvar.set('dynamic' , '2020-09-01 12:00:00')
```

C.sql:

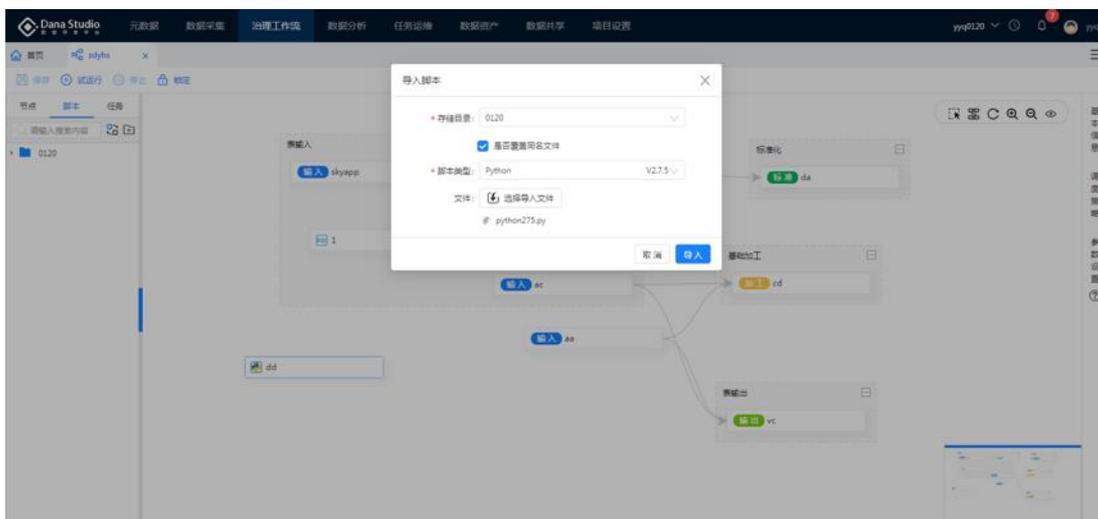
```
select * from table where t = 'dynamic'

# 等同于 select * from table where t = '2020-09-01 12:00:00'
```

- 注:** 1) 常量参数、时间参数和动态参数参数名称都支持 \$ 符号
 2) 使用 python 脚本时, 注意不要将 python 解释器和格式声明去掉

2、导入脚本

脚本——导入——选择存储目录、脚本类型, 上传文件——导入
 支持 Shell、Python、Python3、SQL、Kettle 脚本导入



③包含本项目下自定义采集任务，仅支持新建的自定义采集任务，不支持单表采集转换的自定义采集

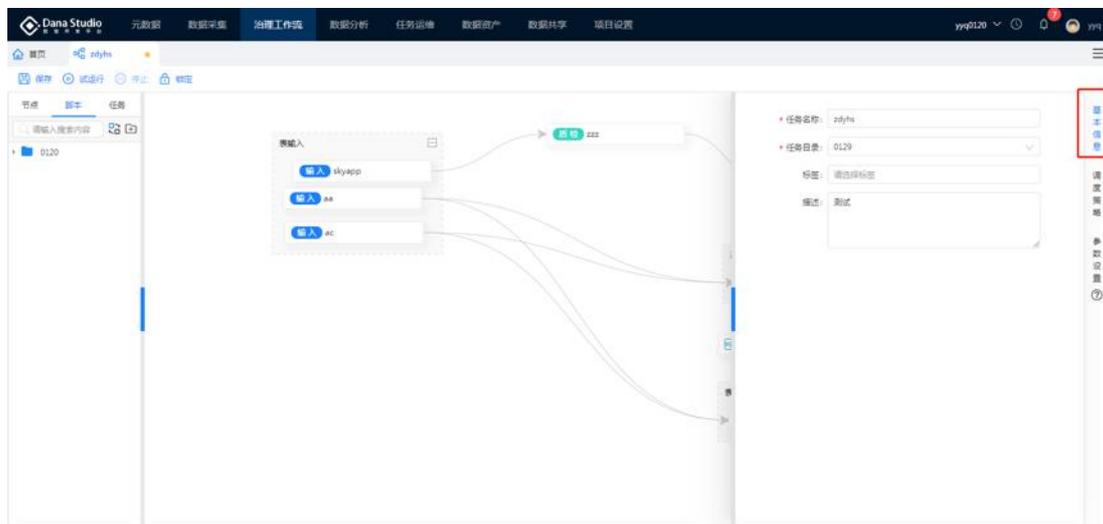
④画布统一操作，包括框选、自动布局、刷新、放大、缩小、展示

框选：框选多个模组，且可以进行分组命名

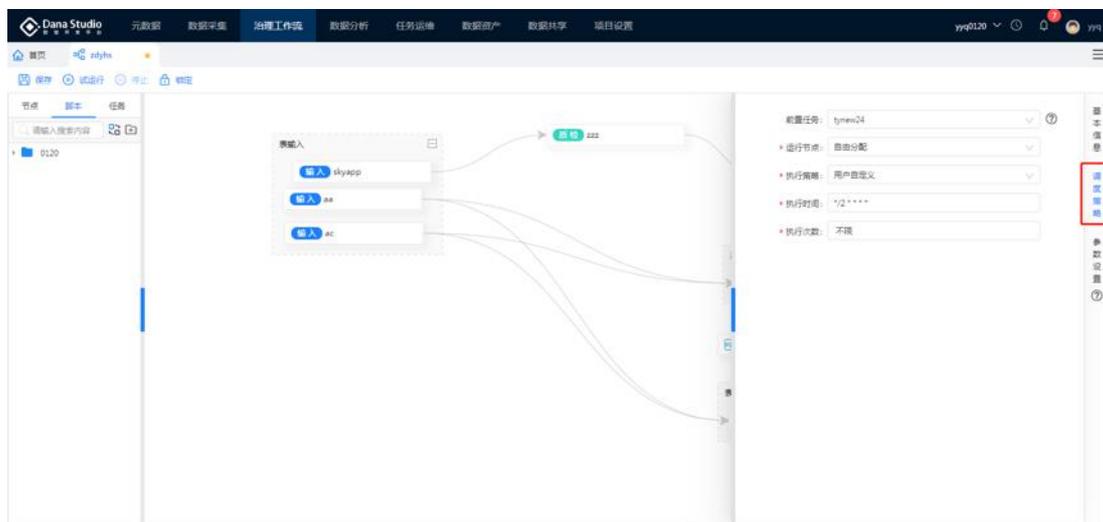
自动布局：自动排列模组，从左到右排列

展示：可在模组下方展示其全部名称详情

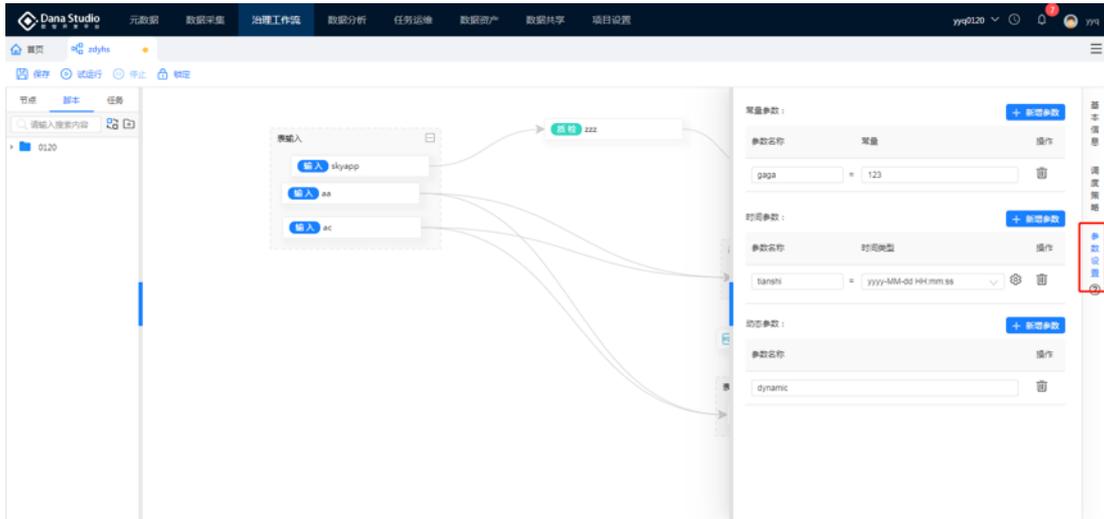
⑤基本信息处可配置：任务名称、任务目录、标签、描述



⑥调度策略处可配置前置任务、运行节点等，如选择前置任务，策略默认同前置任务策略，可修改



⑦参数设置处可配置脚本使用的参数



8 数据资源管理

通过数据融合、标准化、元数据、数据质量等专业数据治理服务，归集治理政府各部门、社会各领域数据。通过数据归集、数据治理服务形成高质量的基础数据资源，完成建设全市统一的各类基础库、主题库、专题库。通过数据共享交换、开放等大数据能力中心建设，为各职能部门提供基础性、战略性的信息服务和决策支持，支撑城市“智慧城市”建设。

8.1 基础信息库

将各委办的相关业务员信息数据进行统一的整合汇集，形成统一、完善、标准的数据信息仓库，为各职能部门提供基础性、战略性的信息服务和决策支持。

8.1.1 人口基础库

依托于数据资源管理平台，对接现有公安局、人社局、民政局、住建、组织部、教育局、卫健等单位人口信息资源，建设统一规范的人口基础数据库，实现全市人口基础信息的集中、统一与规范管理，对接人口库，后续将身份证号作为自然人基础信息的唯一标识，并以此标识作为数据交换与共享的关键字段。人口基础数据库包括人从出生、教育、就业、人口流动、婚姻、住房、医疗、社保、司法、养老到死亡的全生命周期数据信息。通过从各个部门获取到的数据进行清洗、整理、关联、比对、加载、编目、入库等工作，进行核查比对后，形成人口基础、人口专项、人口变动等专题应用数据。

8.1.2 法人基础库

建成以组织机构代码为统一标识的标准统一、信息规范、查询快捷、动态管理、安全可靠的法人单位基础信息数据库及相关应用服务系统。该数据库将集聚整合质监、工商、民政、编办、税务和统计等参建部门管理系统中涉及法人单位的基础信息，实现各参建部门业务管理系统的互联互通，信息资源共享，最终目标实现集法人单位基础信息采集、处理、更新、交换、服务为一体的规范化的政务信息运行管理共享体系。

8.1.3 空间地理库

基于已建时空大数据平台进行升级改造，统筹推动建设城市各类地下管线、地面建（构）筑物等各类设施、土地资源、河网水系、道路交通、园林绿化等相关信息空

间化改造，形成城市统一的地理空间框架、时空云平台、主城区三维模型，实现“一张图”的共享与综合应用。

8.1.4 电子证照库

依据《电子证照-总体技术架构》、《电子证照-目录信息规范》、《电子证照-元数据规范》、《电子证照-标识规范》、《电子证照-文件技术要求》、《电子证照-共享服务接口规范》等6项国家标准，建设城市电子证照库。

8.1.5 信用库

建设个人信用信息数据库和企业信用信息数据库，可以压缩失信者和违法分子的生存与发展空间，促进社会诚信度，对提高党的执政能力有深远的意义。

8.2 专题信息库

专题信息库设计面向管理常态与非常态业务需求，通过将数据湖和基础库数据进行二次抽取装载的方法重新组织数据，并按照不同领域专题应用的需求重新整合形成专题库。按照需求牵引的方式建设，建设包含党建、城管、交通、应急等行业专题和评估考核等综合专题。

8.2.1 “数字治超”专题库

交通局现有“数字治超”大应用建设需求，要求建立交通部门、公安局、市监局、文旅局、生态局、应急局、城管局部门间的信息共享渠道，由城市数据发展中心搭建专题库进行数据共享、数据互认、数据管理、数据服务等服务。

8.2.2 “大数据评估考核”专题库

大数据评估考核专题库针对汇聚的评估指标数据进行统一管理及评估考核数据。根据国家评估指标的要求，结合实际情况，汇聚指标评估所需的相关业务数据，并实时对业务数据进行动态更新。调研各地市数据使用需求情况，按照业务逻辑建立评估模型，关联各部门数据进行建模。通过提供数据抽取、数据快照功能，对评估数据进行对接、采集，在采集过程中根据评估需求，动态化对不同数量、不同类型的指标进行按需组合，配置特定的数据采集流程，形成具有针对性的采集任务。系统提供根据对应的指标体系，自动化计算各区划的指标数据，可视化展示体系指标，对考核进行数据支撑。

8.2.3 车辆专题库

车辆在城市管理中是一种非常重要的实体信息，对车辆建立全方位的专题库是非常有必要的。车辆专题库基于物专题库生成，信息一般包括：车辆现有与历史拥有者及拥有者信息、车辆基本属性（如品牌、出产地、购买地、型号、颜色、排量、发动机等）车辆历史使用者及使用者信息、车辆乘坐者及乘坐者信息。

8.2.4 产业发展专题库

构建产业发展专题库，主要涵盖产业升级、营商环境和民族经济等专题数据指标，包含产业经济、工业升级、新兴产业、大数据产业、服务业升级、人才引进、专利申报、高新技术引进等指标内容，为形成产业发展指数云图及辅助决策建议提供支撑。

8.2.5 经济运行专题库

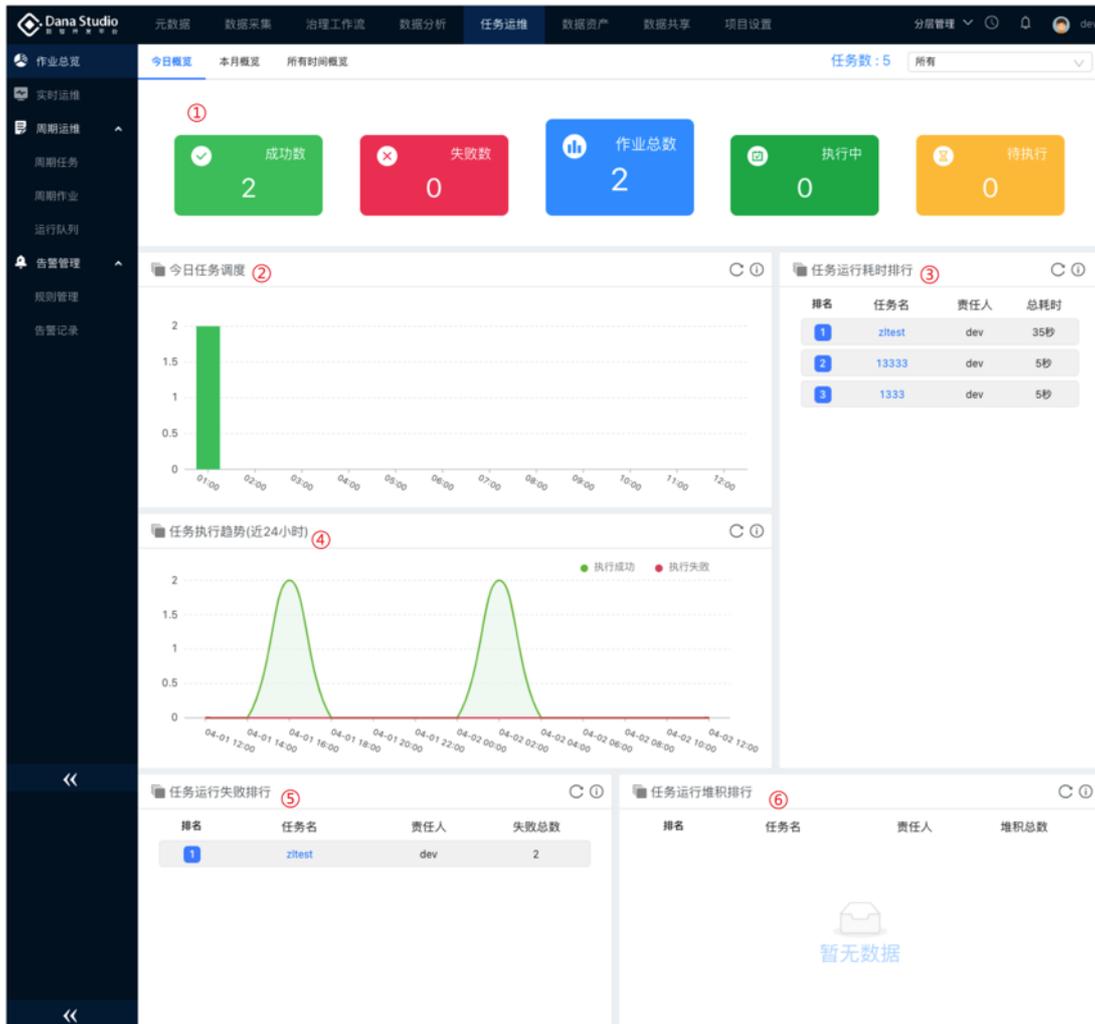
构建经济运行专题库，主要涵盖扩大投资、区域经济、民营经济和消费监测等专题数据指标，包含固定资产投资、产业投资、投资项目、创投情况、生产总值、民营经济概况、民营企业等指标内容，为形成经济运行指数云图及辅助决策建议提供支撑。

9 任务运维

任务运维模块承载平台调度和任务监控功能，保障工作流有序、可控地执行。满足业务批处理要求以及任务异常监控。从数据采集、工作流、数据共享等模块提交上来的作业，将会在这里被调度执行。

9.1 作业总览

作业总览对整个作业集群的重要指标提供了可视化监控,通过图形化方式展示了运维人员常用的集群性能指标以及作业的执行统计，可以切换显示所有、采集、工作流、数据共享任务信息。



①切换显示今日、本月、所有时间下的任务执行的成功数、失败数、总数等信息

②显示每个时间段的成功和失败个数柱状图

1) 图例只保留成功和失败，横轴只展示历史的执行结果，纵轴指标为每个小时内作业的成功和失败总数

2) 统计内容为作业次数，作业的含义是任务的实例化（一个任务执行一次算一个作业）

例：假如当前是 16:18，那么在 17:00 的柱状图是要显示 16 点——17 点之间的作业数的。

③分页展示作业运行耗时排行

④统计过去 24 小时的作业执行情况，纵轴指标为每 2 个小时内作业的成功或失败总数

⑤展示作业运行失败排行

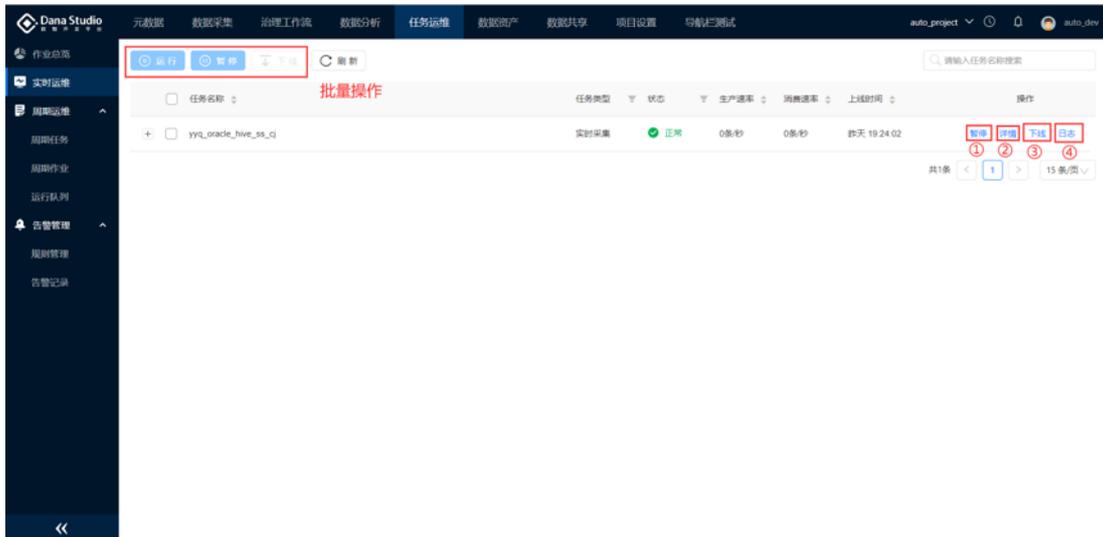
⑥展示队列详情内作业待执行的堆积总数排行

9.2 实时运维

实时作业运维提供实时任务的监控运维总览，针对进行中的实时采集任务展示正在运行任务的详情、生产端和消费端速率。

针对实时作业管理中的实时作业可进行单个/批量运行、暂停、下线，支持查看任务日志。

任务运维——实时运维



支持批量运行、暂停、下线任务；

①点击“暂停”按钮，暂停任务执行。暂停后再启动同步暂停期间的数据。

②点击“详情”按钮，查看作业详情

③点击“下线”按钮，下线该任务。下线后再上线，带着离线抽取的实时抽取，下线之后上线，清除数据离线抽取；未带有离线抽取的实时抽取，下线之后上线，追加上线之后变更的数据

④点击“日志”按钮，查看作业运行日志

注：1) 实时作业中的离线任务仅支持查看日志

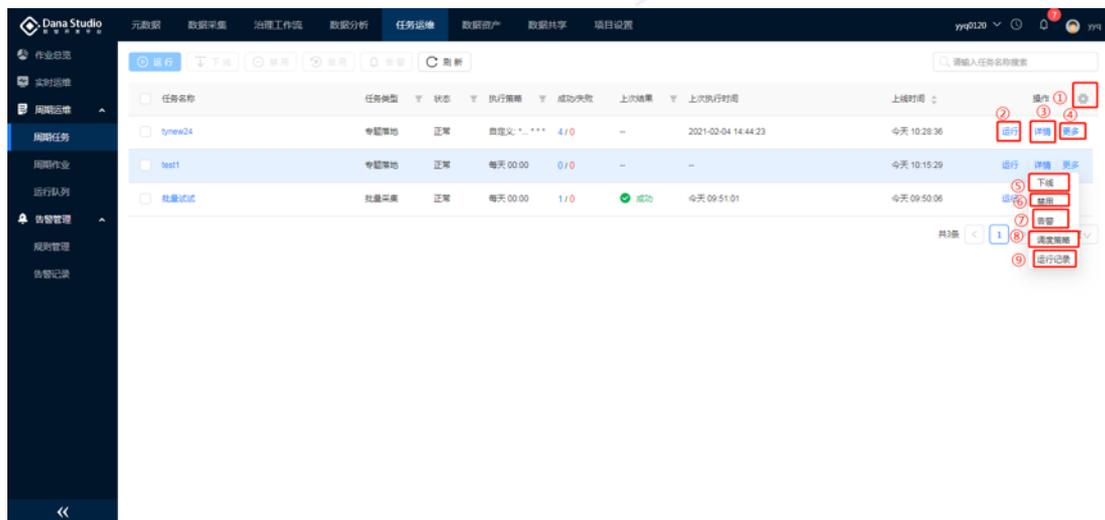
2) 目标端为 Stork、teryx 引擎时，当源表数据发生变化增、删、改时，目标端进行同样的操作

3) 目标端为 Hive 引擎时，系统会默认比源表多建两个字段；当源表数据发生变化增、删、改时，目标端不能进行同样的操作，仅会增加数据，但会将发生变化的数据记录同步到目标表，根据_op 标识记录的变化状态：增-c、删-u、改-d

4) 实时任务失败后不支持暂停

9.3 周期运维

9.3.1 周期任务



支持批量运行任务、下线任务、禁用任务、启用任务、添加告警规则；

- ①点击“设置”按钮，可以设置列表字段展示
- ②点击“运行”按钮，可运行该任务
- ③点击“详情”按钮，查看任务详情
- ④点击“更多”按钮，显示上/下线、启/禁用、告警、运行记录四个按钮
- ⑤点击“上/下线”按钮，上线任务可下线，下线任务可上线
- ⑥点击“启/禁用”按钮，启用任务可禁用，禁用任务可启用
- ⑦点击“告警”按钮，可添加告警规则，触发告警规则后在告警记录处相应增加

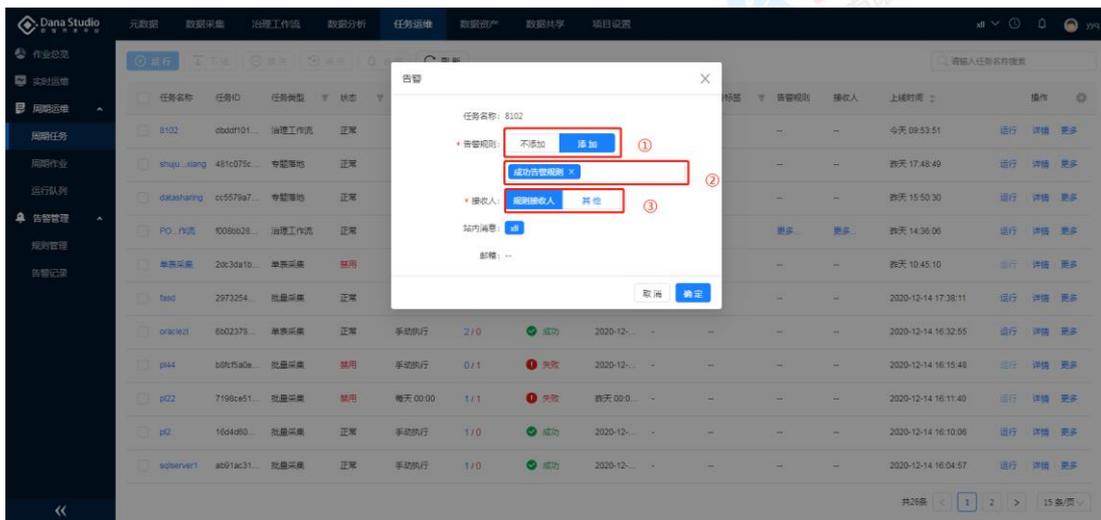
记录

- ⑧点击“调度策略”按钮，可查看、修改任务的调度策略
- ⑨点击“运行记录”按钮，可跳转周期作业页面，展示该任务的作业运行记录

告警：

点击任务后的“更多”——点击“告警”——选择告警规则——添加接收人——

确定



- ①可选择是否添加告警规则
- ②告警规则选择框，可选择告警规则，告警规则在告警管理——规则管理（详情

见 9.4.1 规则管理) 处新建

③可配置告警消息接收人、接收方式

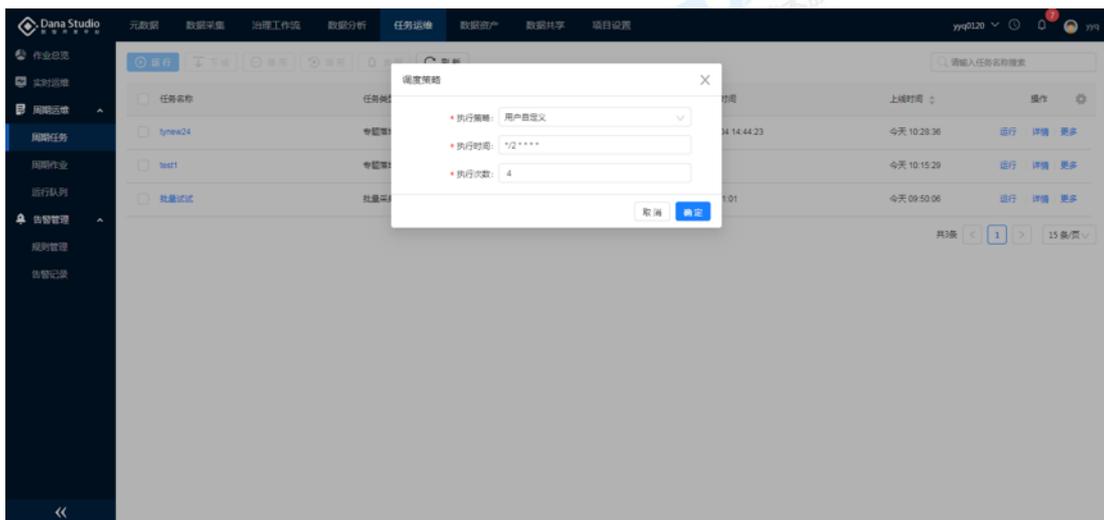


触发告警规则会向配置的接收人、通过接收方式进行通知

例：接收人：当前用户、接收方式：站内消息，触发告警规则后，点击页面右上角的通知按钮可以查看告警消息

调度策略：

点击任务后的“更多”——点击“调度策略”——选择执行策略——修改执行时间、执行次数——确定



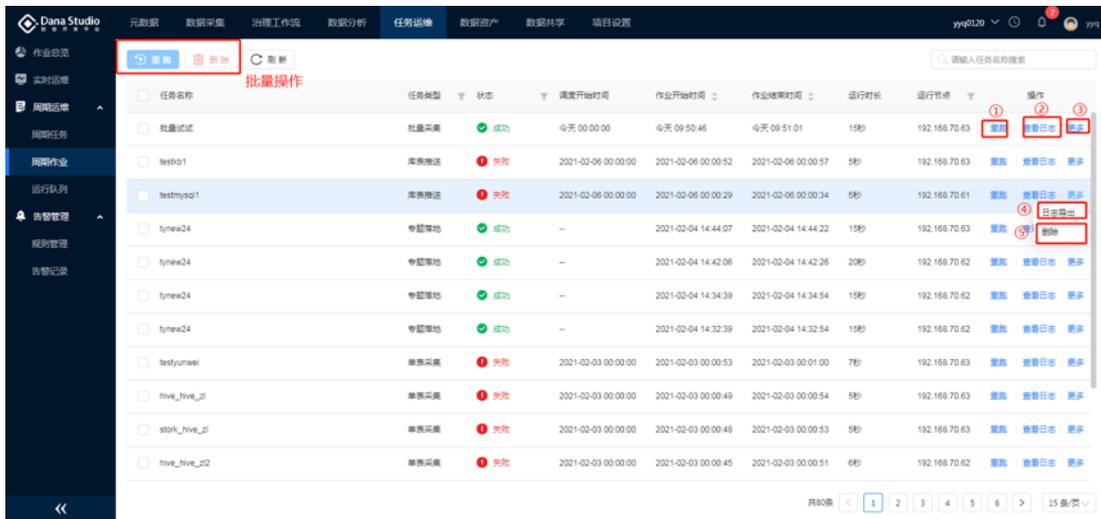
支持任务上线后修改调度策略，修改后的调度策略，各模块任务处调度策略也同步，任务将按照修改后的调度策略执行；

注：1) 正在运行的任务不支持修改调度策略，运行中调度策略按钮置灰，运行结束调度策略按钮恢复

2) 执行次数设置的数值需要大于任务已经运行过的数值，否则任务不会运行

9.3.2 周期作业

周期作业帮助运维人员管理线上作业。对于已上线的作业可以修改预设的调度周期策略、执行节点。



支持批量重跑作业、删除作业；

①点击“重跑”按钮，可针对这条作业重新执行

②点击“查看日志”按钮，可查看该作业的日志，

③点击“更多”按钮，展示“日志导出”、“删除”按钮

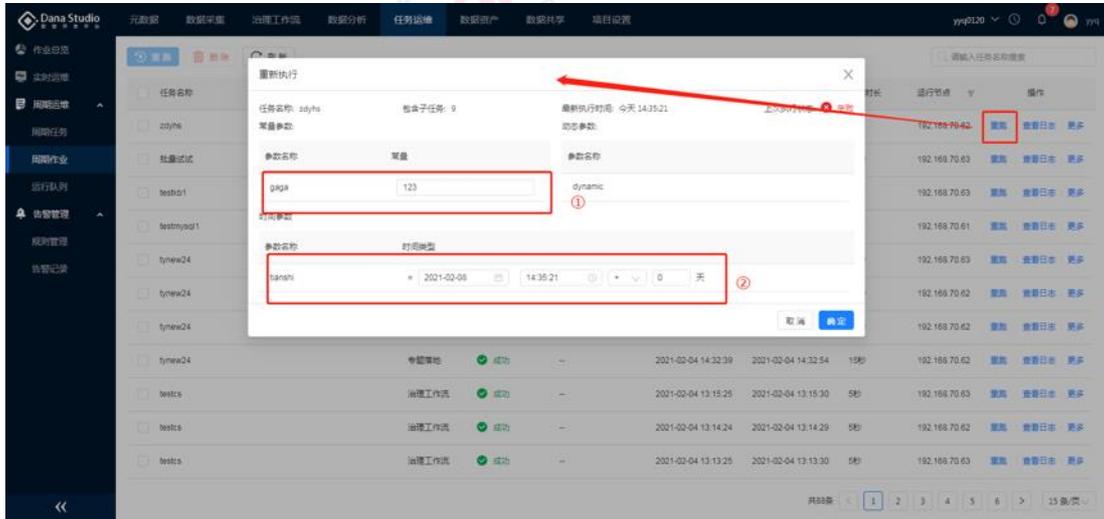
④点击“日志导出”按钮，可导出该作业运行的日志，如该作业重跑多次，则导出该作业最近一次重跑的日志

⑤点击“删除”按钮，可删除该条作业

注：重跑后不会产生作业数和新的作业记录，可从日志中查看重跑日志

重跑:

周期作业——重跑——配置参数（仅针对于包含参数的任务）——确定



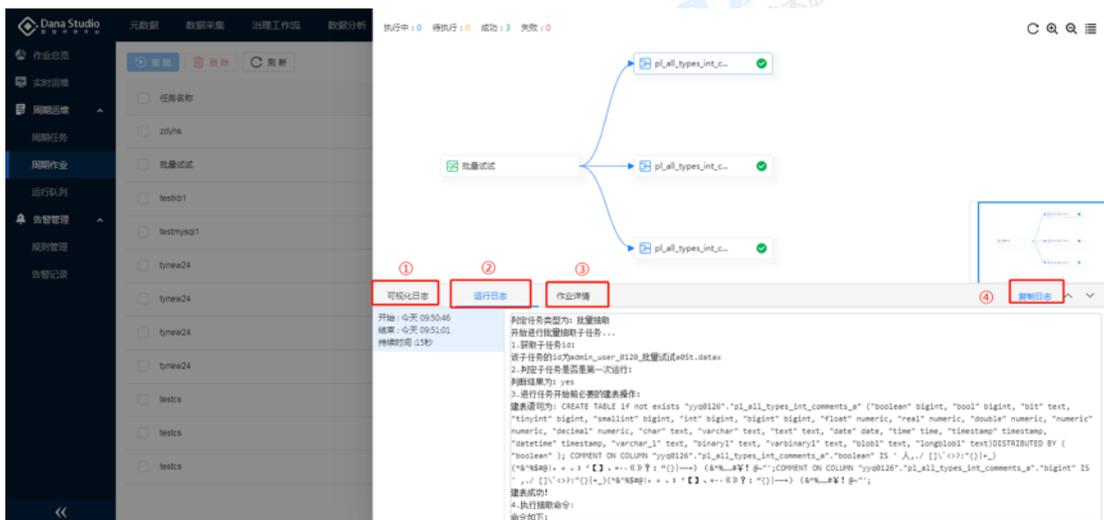
①可配置常量参数进行重跑

②可配置时间参数进行重跑

查看日志:

任务运维——周期作业——点击采集任务的作业记录里后“查看日志”——打开

日志页面



①点击“可视化日志”按钮，可查看作业运行记录及日志重跑记录

②点击“运行日志”按钮，可查看作业运行日志详情及日志重跑记录

③点击“作业详情”按钮，可查看详情信息

④复制日志，仅切换到“运行日志” Tab 时展示，点击“复制日志”按钮，可复制当前作业的运行日志

注：仅采集任务有可视化日志

9.3.3 运行队列

堵塞队列是当作业中心作业调度的并发度超过预设的最大并发时，后续的作业会暂时加入堵塞队列，等作业中心压力释放后，会继续执行。



①点击“中断”按钮，可对执行中任务作业进行中断执行操作

②点击“优先”按钮，可对任务作业进行优先运行操作

③点击“降级”按钮，可对任务作业进行降级运行操作

④点击“移除”按钮，可将任务作业移除

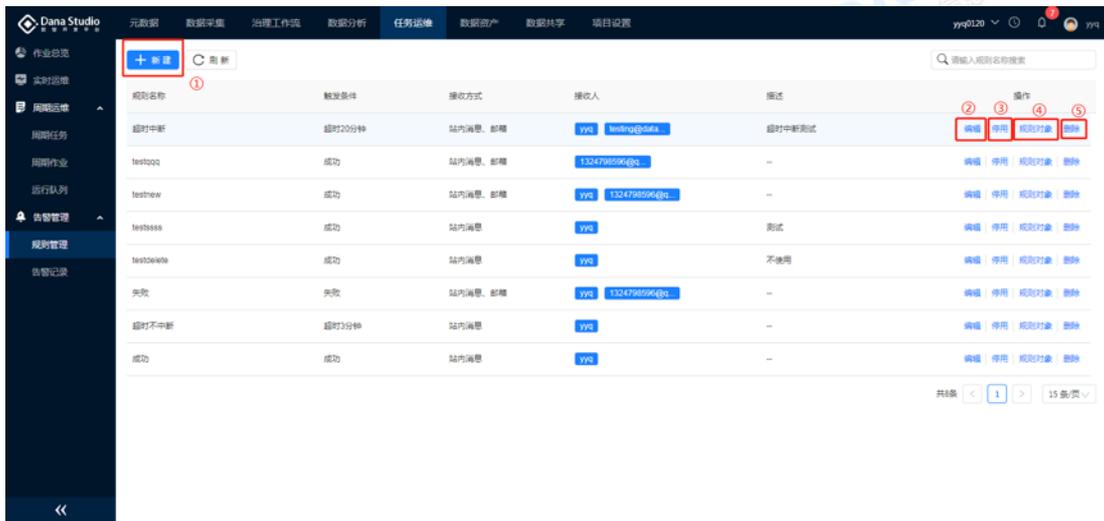
注：任务阻塞多次时，如资源足够作业会顺序执行；如已阻塞了多次作业，第一次作业执行失败后任务将被禁用，被阻塞的作业不会正常执行，会被直接清空。

9.4 告警管理

前提：

DanaManager 上已安装 DANAGUARD 服务

9.4.1 规则管理



①点击“新建”按钮，可新建告警规则，其中邮箱格式支持：163 网易邮箱、公司邮箱 (datatom.com)、qq 邮箱

②点击“编辑”按钮，打开告警规则详情页面，可编辑告警规则

③点击“启/停用”按钮，启用时，启用所有关联任务的告警规则，停用规则时，所有添加该规则的任务都停止该触发条件

④点击“规则对象”按钮，列举出添加该规则的任务列表

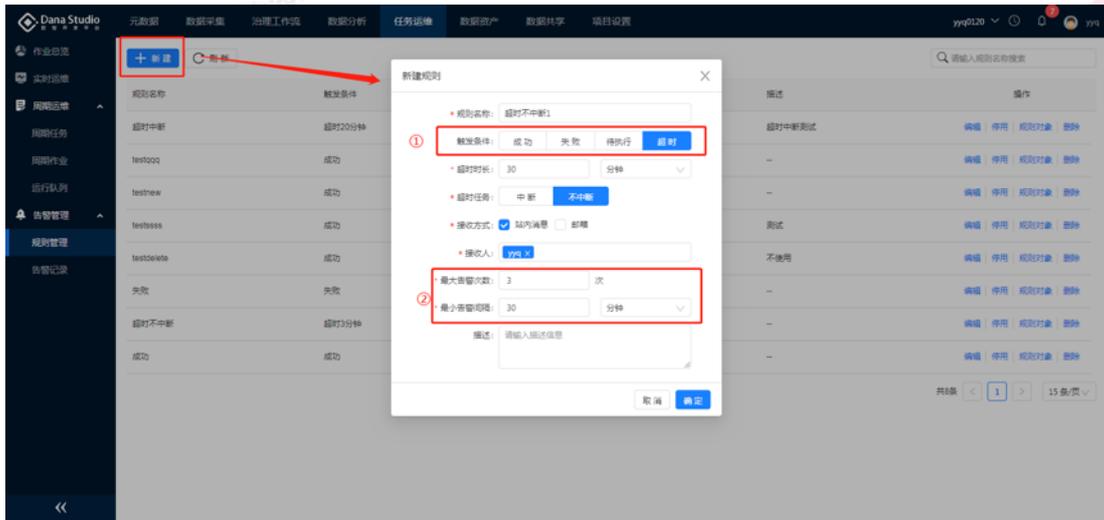
⑤点击“删除”按钮，可删除该告警规则

注：1) 删除规则后，与该规则绑定的任务全部解绑且不产生告警记录，原告警记录保留

2) 编辑规则后，任务按照新的规则告警，之前的告警规则产生的告警记录信息不变

新建：

规则管理——新建——填入必填项——确定



①拥有四种触发条件：成功、失败、待执行、超时，超时包含超时中断和超时不中断

②告警次数：任务作业触发条件后，最多会产生告警的次数，最大告警次数默认3次

告警间隔：任务作业触发条件后，两次告警消息间隔的时间，最小告警间隔默认30分钟

9.4.2 告警记录

任务运维——告警管理——告警记录——查看任务触发告警规则的记录

告警时间	规则名称	触发条件	接收人	任务名称	任务类型	操作
2021-03-31 00:02:58	成功告警1	成功	test_yys	autotest_single_e_ct_task_byexistb	单表采集	查看作业 删除
2021-03-31 00:02:47	成功告警1	成功	test_yys	脚本工作流1	治理工作流	查看作业 删除
2021-03-30 00:02:59	成功告警1	成功	test_yys	autotest_single_e_ct_task_byexistb	单表采集	查看作业 删除
2021-03-30 00:02:46	成功告警1	成功	test_yys	脚本工作流1	治理工作流	查看作业 删除
2021-03-29 16:50:49	成功告警1	成功	test_yys	脚本工作流1	治理工作流	查看作业 删除
2021-03-29 16:49:39	成功告警1	成功	test_yys	脚本工作流1	治理工作流	查看作业 删除
2021-03-29 00:02:45	成功告警1	成功	test_yys	脚本工作流1	治理工作流	查看作业 删除
2021-03-29 00:02:38	成功告警1	成功	test_yys	autotest_single_e_ct_task_byexistb	单表采集	查看作业 删除
2021-03-28 00:02:46	成功告警1	成功	test_yys	脚本工作流1	治理工作流	查看作业 删除
2021-03-28 00:02:41	成功告警1	成功	test_yys	autotest_single_e_ct_task_byexistb	单表采集	查看作业 删除
2021-03-27 00:02:46	成功告警1	成功	test_yys	脚本工作流1	治理工作流	查看作业 删除
2021-03-27 00:02:39	成功告警1	成功	test_yys	autotest single e_ct task byexistb	单表采集	查看作业 删除

①支持对告警时间、任务名称、接收人进行查询

②支持对任务类型进行筛选

③点击“查看作业”按钮，跳转到周期作业处，并针对任务名称筛选展示

④点击“删除”按钮，对该条告警记录进行删除

10 数据资产

数据资产以项目中所涉及到的数据为对象，以统计和查看为核心功能，为用户提供完整丰富的资产视图。用户可以在平台上总览数据的可视化统计分析图表，也可以按照编目、搜索、排序等方式查找数据，查看数据的详情。

10.1 资产概览

资产概览模块主要汇总了整体的数据资产视图，可以看到相应的分层信息以及各层数据的一些维度统计信息。

注：资产概览中每个图表每个整点刷新一次



10.1.1 数据统计

此页面综合展示数据总量、数据总项数（健康数据量、中断数据量）、敏感数据占比等情况；

点击【中断数据量】，页面展示中断数据列表以及中断数据详情。

点击【敏感数据占比】，页面按敏感数据类型进行分类展示敏感数据。

10.1.2 数据趋势

左上方按月展示数据源增长曲线；

上方中间按月展示原始数据项增长曲线；

右上方按月展示原始数据条目增长曲线；

左下方按月展示治理数据项增长趋势；

右下方按月展示治理数据生成条目增长趋势；

10.2 资产目录

资产目录提供数据源编目、原始数据编目、主题/治理数据编目等视角的表目录展示，点击表名称可进入查看表的详细信息。

数据资产——资源目录——搜索表——点击某表名，进入表详情页面



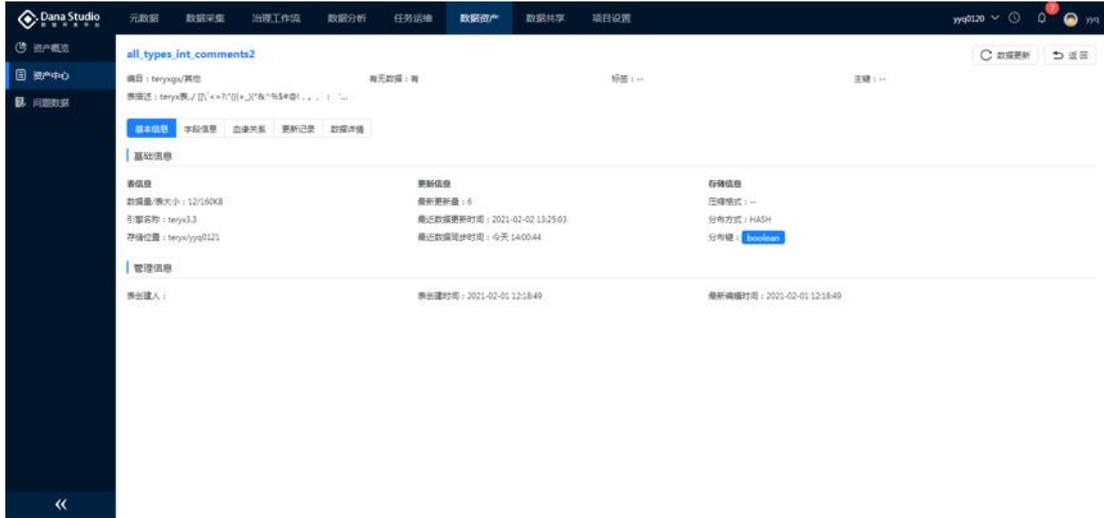
①点击列表右上角的“刷新”按钮，可以立即同步元数据，数据立即更新

10.2.1 基本信息

支持查看表名、数据来源单位、标签、表描述等表概要信息，表存储量、更新情况和分布方式等基础信息，表创建人、创建时间和表元数据更新时间等管理信息，其中只有 teryx、gaussdb、达梦、优炫展示分布方式，其余数据源不展示。

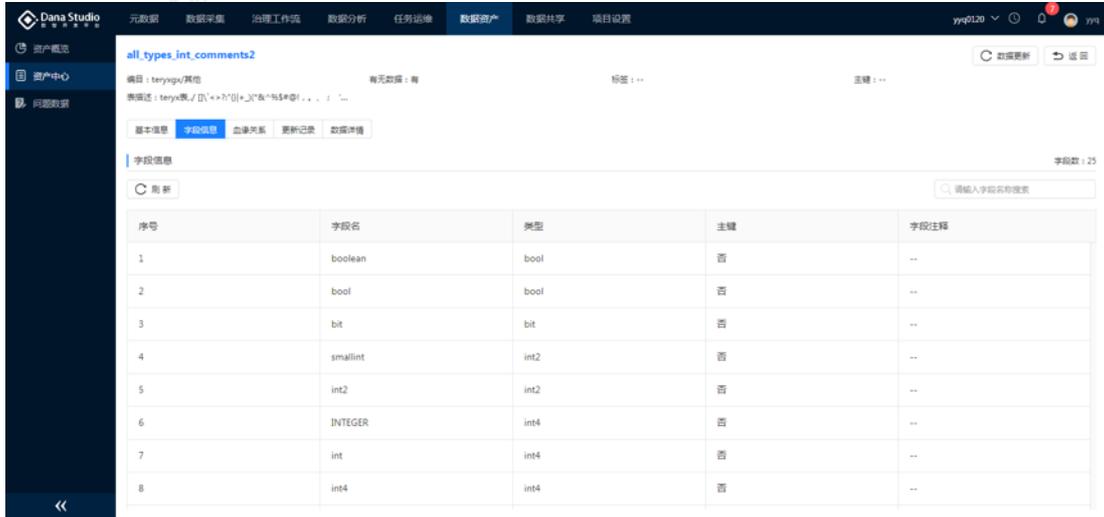
注：1) 基本信息每个整点更新一次

2) 元数据-元数据开发或资产中心页面，操作同步元数据后，资产中心处数据立即更新



10.2.2 字段信息

支持查看表结构，包括字段名、字段类型、主键信息、字段描述、数据敏感度标志、是否加密字段标志的查看，也支持对字段进行快速搜索以及字段个数的查看。



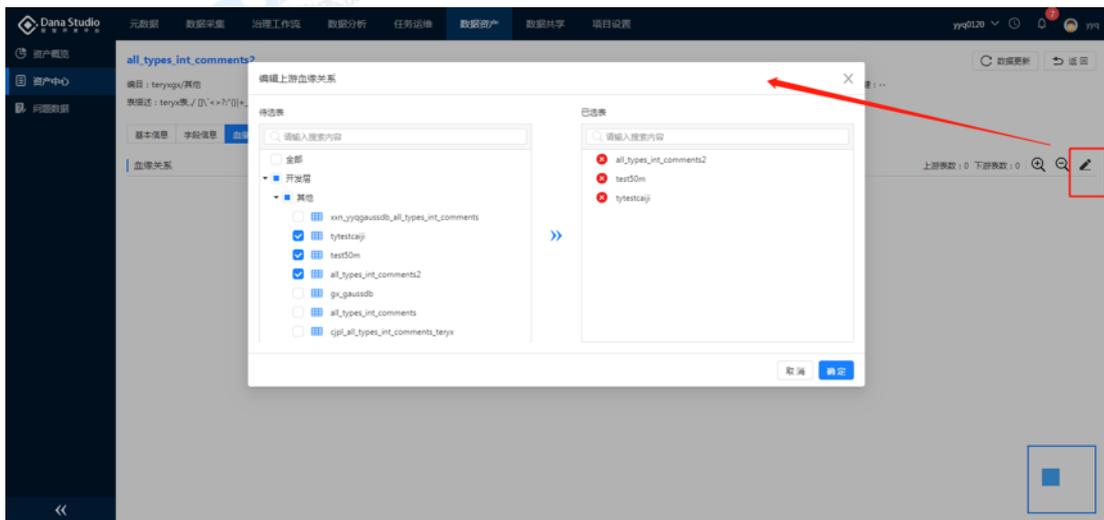
10.2.3 血缘关系

支持查看表的血缘关系获知数据表的父级血缘和子级血缘。以方便定位表数据的来源和去向，快速锁定数据变动的范围。



10.2.4 编辑上游血缘关系

支持对血缘关系中的父级关系进行编辑以修正血缘关系。修正后的血缘关系将自动扩散影响到全平台内相关的表的血缘关系。



10.2.5 更新记录

支持按照近 7 天、近 15 天、近 30 天查看表的数据变化趋势图，使用户便于掌握数据变化情况，发现异常任务。



10.2.6 数据查询

本数据查询模块提供编目查询和搜索查询两种方式。

编目查询支持按行政单位编目、主题编目、涉密编目以及政务编目进行逐级展开查看或者输入关键字进行数据资源定位。

通过搜索查询框输入关键字，可实现精准查询或模糊查询。

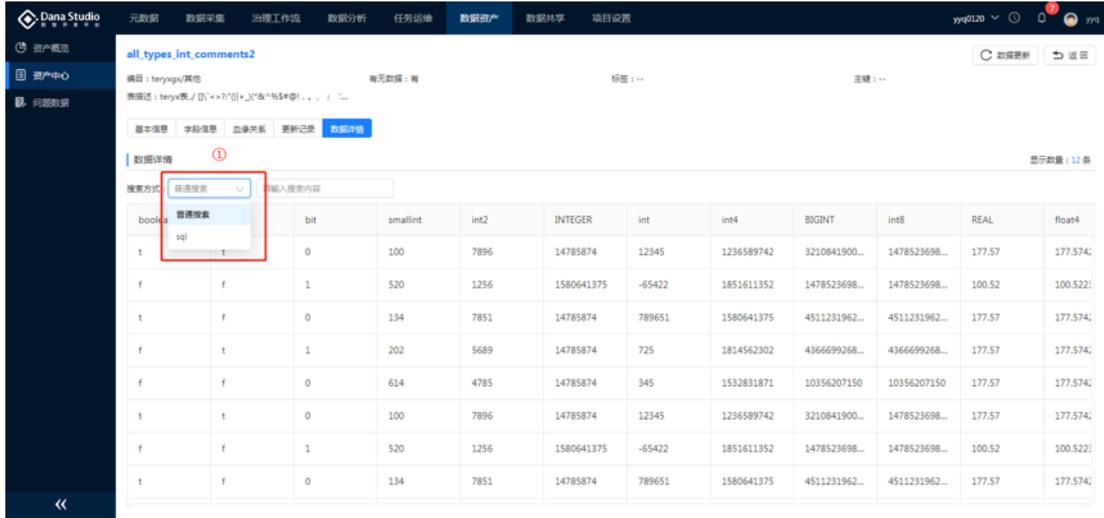
当输入的是数据项唯一编码，即可实现数据资源的精准定位。

模糊查询支持的筛选条件包括标签、行政单位、起止时间、数据量、数据状态、主题等；

10.2.7 数据详情

支持查看数据详情，预览表数据。

注：只展示该表的前 1000 条数据



①支持普通搜索、sql 语句搜索，搜索对应表数据

10.3 问题数据

问题数据是数据治理过程中，针对不符合数据质量、数据不规范等治理加工过程中筛选出来的不合格数据进行汇总展示，可为开发者提供具体的数据的问题原因，方便开发者对后期问题数据进行二次治理或形成相对于的数据质量报告，便于数据的质量提升。



①点击“同步更新数据”按钮，同步生成的问题数据

②点击表名进入到问题表的问题数据详情页面

问题数据详情页面：

nano1	nano2	cf_问题
c	c	2020-12-14 17:47:56
c	c	2020-12-14 17:47:56
b	b	2020-12-14 17:47:56
b	b	2020-12-14 17:47:56
a	a	2020-12-14 17:47:56
a	a	2020-12-14 17:47:56

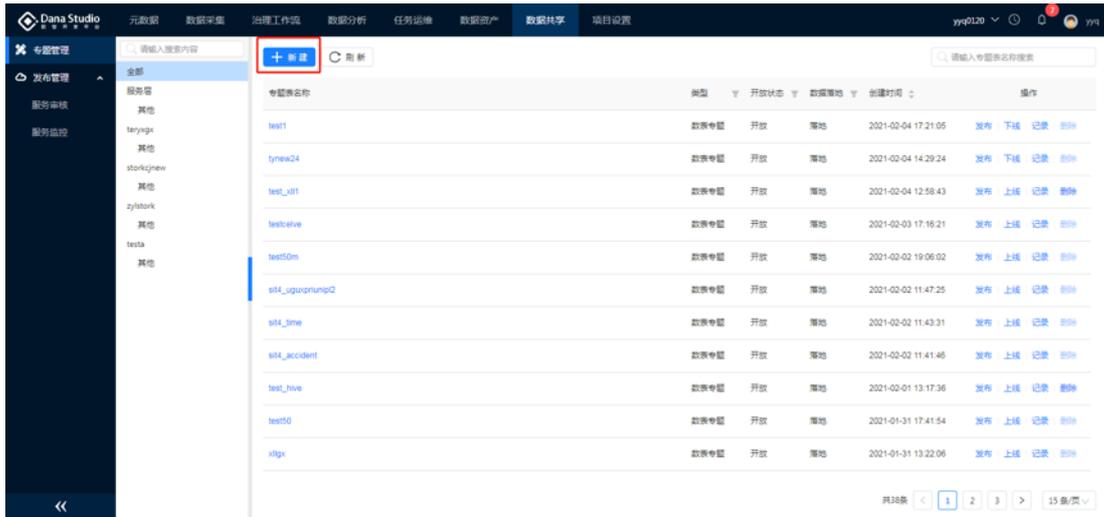
11 数据共享

数据资源管理平台对各种数据进行组织和存储，这时需要进行数据的提取、转换和映射，为了实现对各类社会公共信息和专业业务信息的共享及应用系统之间的互联互通，对系统间包括跨行业系统间和同行业不同系统间的异构数据库数据交换及应用系统之间的业务流程自动化，平台可以通过各种协议使数据和各市县、机构实现互联。通过对外交换数据，使得大数据平台的数据能够为更多的部门服务，使现有数据产生更大的价值。

11.1 专题管理

11.1.1 新建专题

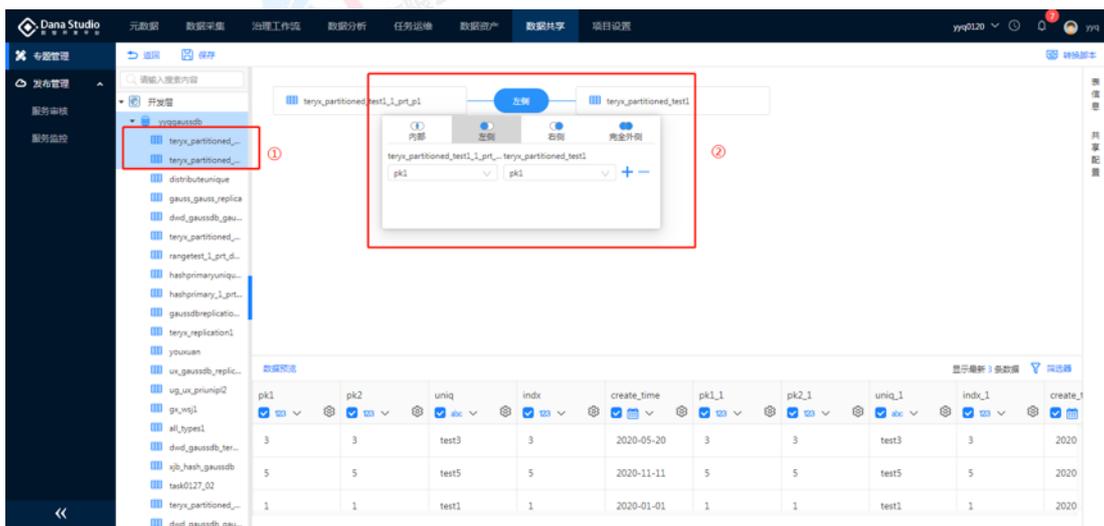
数据共享——专题管理——点击“新建”——选择源表，关联——数据编辑——配置表信息——配置共享配置——保存



(1) 选择源表, 关联

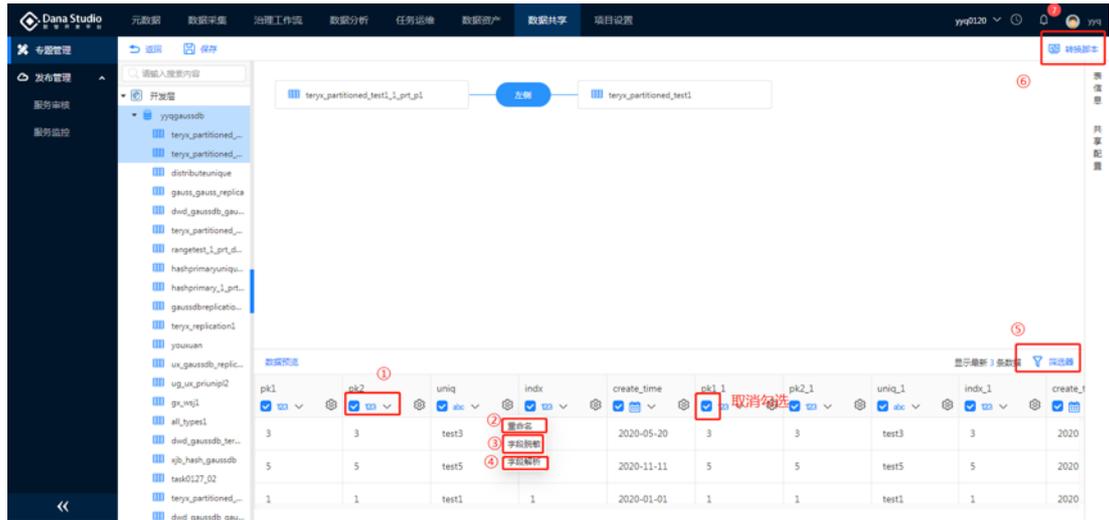
①将需要关联的表拖入编辑区——②配置关联关系——下方展示关联后的数据预览

览



(2) 数据编辑

支持修改字段类型、重命名字段、字段脱敏、字段解析、筛选数据、取消字段, 同时支持转换为 sql 脚本编辑数据



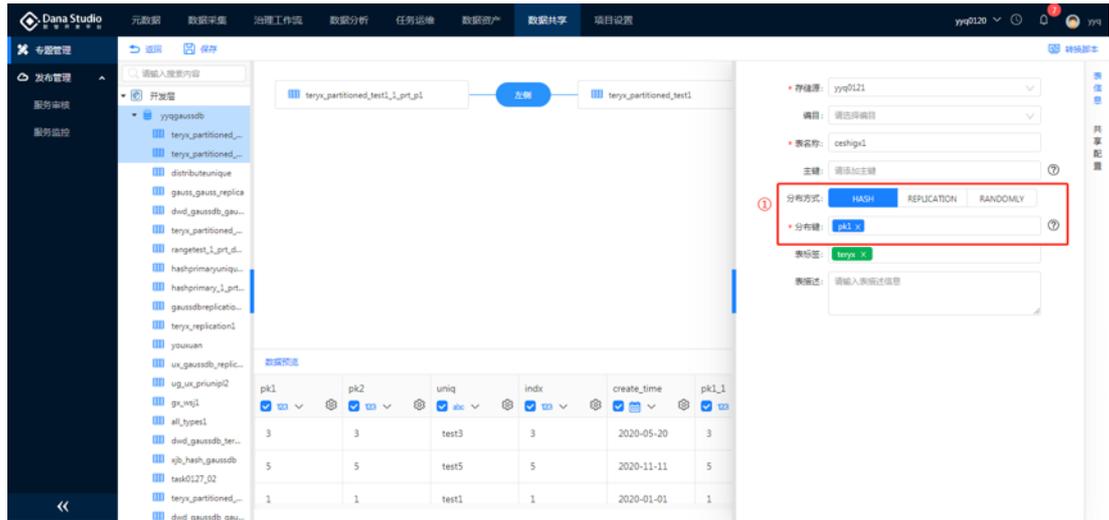
- ①点击字段下方的字段类型可以变更为其他类型：整数、小数、字符串、日期、日期和时间
- ②重命名：将字段名重新命名
- ③字段脱敏：数据根据所选择的脱敏规则脱敏，脱敏规则可在元数据-规范中心-脱敏规则界面设置
- ④字段解析：数据根据所选择的代码表进行解析，代码包可在元数据-标准管理界面设置
- ⑤筛选器：支持设置筛选条件筛选数据
- ⑥转换脚本：支持转化为 sql 脚本，通过 sql 脚本处理数据，不支持切回可视化页面编辑

(3) 配置表信息

当前支持 stork、hive 源中的表 生成专题表 存至 stork 存储源中

teryx 源中的表 生成专题表 存至 teryx 存储源中

gaussdb 源中的表 生成专题表 存至 gaussdb 存储源中



①分布方式、分布键设置仅在 teryx-teryx、gaussdb-gaussdb 时展示

teryx 分布方式：有三种类型 HASH、REPLICATION、RANDOMLY

gaussdb 分布方式：有两种类型 HASH、REPLICATION

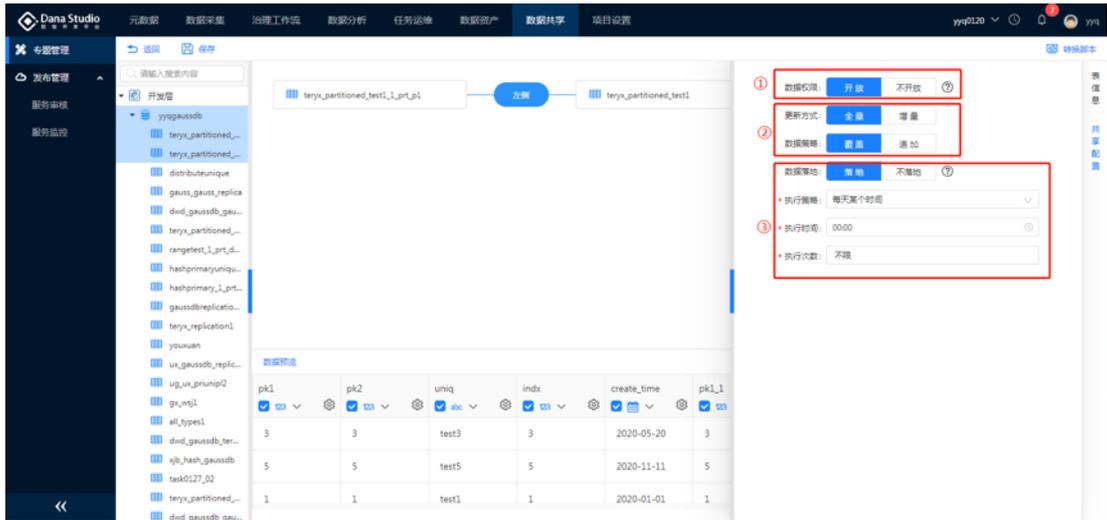
a.HASH：选择 HASH 分布方式时，会出现分布键；当表中存在主键时，分布键只可以在主键中选择，如果表中不存在主键时，分布键可以任意设置；

b.REPLICATION：选择复制分布时，不会出现分布键选择；

c.RANDOMLY：选择随机分布时，不会出现分布键选择；当表存在主键时，RANDOMLY 分布选项禁用；如果已经选择了 RANDOMLY 分布，后设置表主键，则默认选择到 HASH 分布

注：不支持新建专题的数据源，在新建专题表时不会展示

(4) 共享配置



①数据开放：数据开放后可发布数据服务，不开放的专题数据无法发布

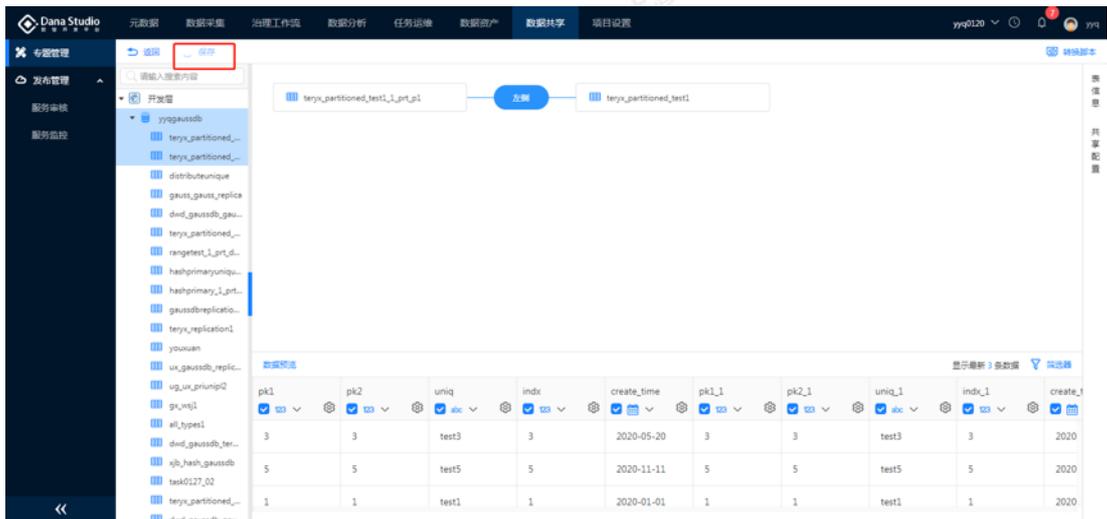
②更新方式、数据策略：

a：更新方式选择“增量”，数据策略选择“覆盖”，每次运行会将表中的数据覆盖；数据策略选择“追加”，每次运行数据追加在表中；

b：更新方式选择“增量”，配置增量字段后，每次运行数据增量追加在表中。

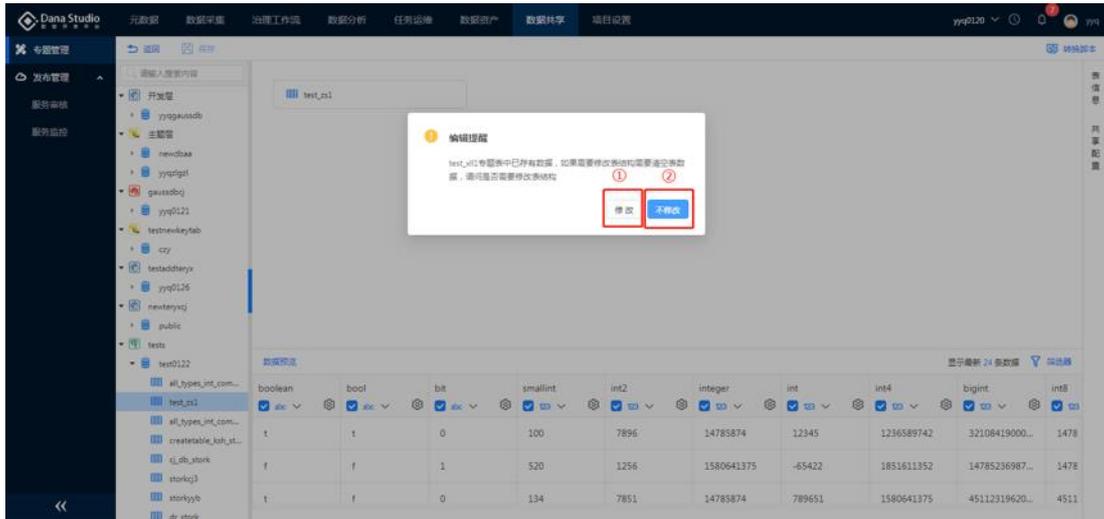
③数据落地：选择落地，系统会生成数据落地任务，根据执行策略，将专题数据从数据仓库抽取到服务层库中；选择数据不落地，数据将不会生成到专题表中。

(5) 保存



11.1.2 编辑专题

已上线且数据采集成功的专题表——下线并编辑专题表——操作编辑提醒弹窗



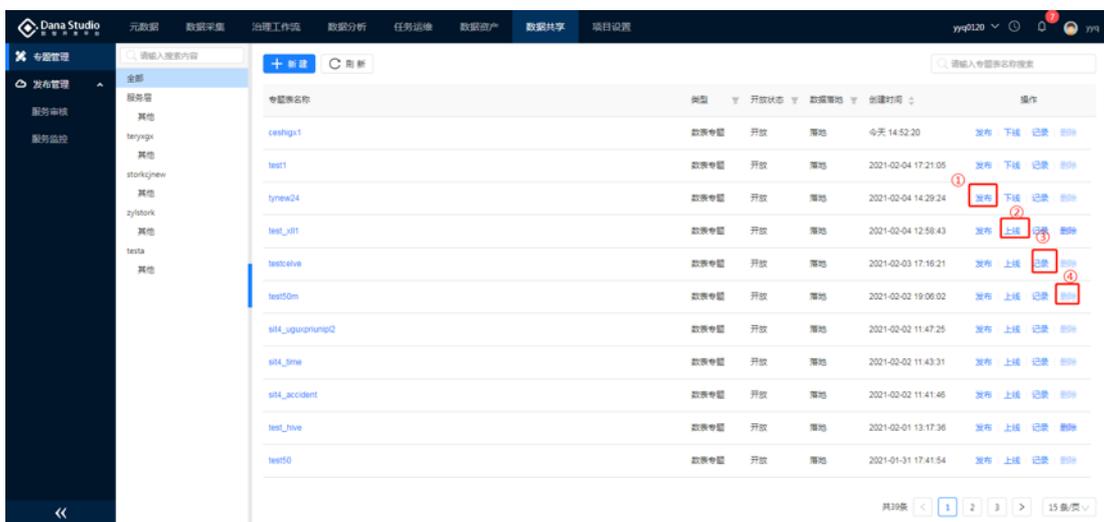
①选择修改，支持更改专题表字段、存储源、表名称等信息

②选择不修改，不支持更改专题表字段、存储源、表名称等信息，可修改标签、筛选器条件、共享配置

注：①专题已上线或已发布数据服务，则进入编辑界面只能查看，无法修改信息

②若专题表中不存在数据，可以直接修改所有内容，无编辑提醒弹窗，修改表结构后将会重建表

11.1.3 专题列表



①点击操作下的“发布”按钮，可将该专题服务发布，进入服务审核页面待审核。未开放的表无法发布。

②点击操作下的“上线”按钮，可将改任务上线运行，任务运维处运行。未落地的任务无法上线。

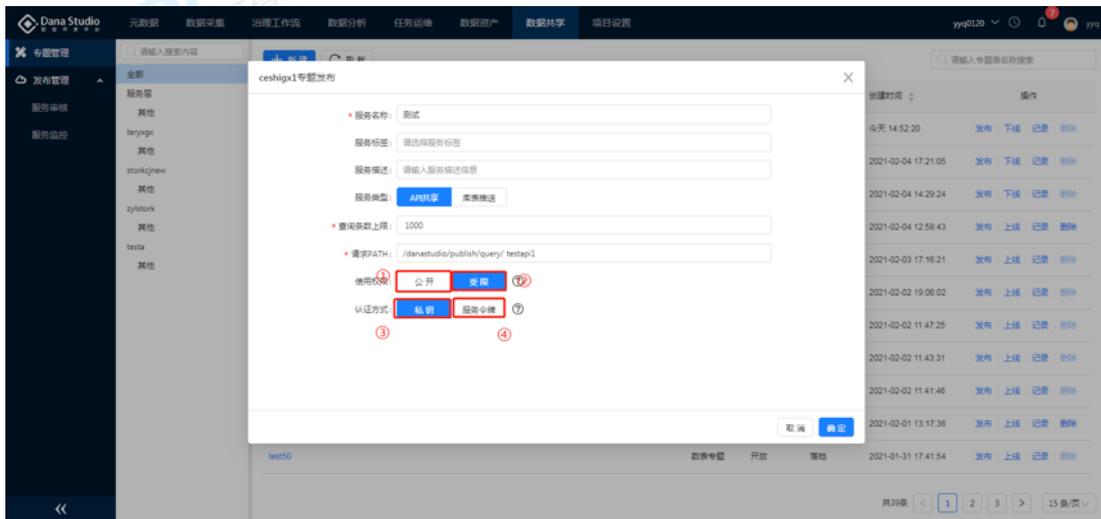
③点击操作下的“记录”按钮，可以查看到该专题发布的服务发布记录

④点击操作下的“删除”按钮，可以将该专题删除

注：落地任务已上线、专题存在发布的数据服务，专题无法删除

11.1.4 发布 API 共享服务

数据共享——专题管理——点击专题表后的“发布”——服务类型选择“API 共享”——填入必填项——确定——服务发布后可进入服务审核页面查看



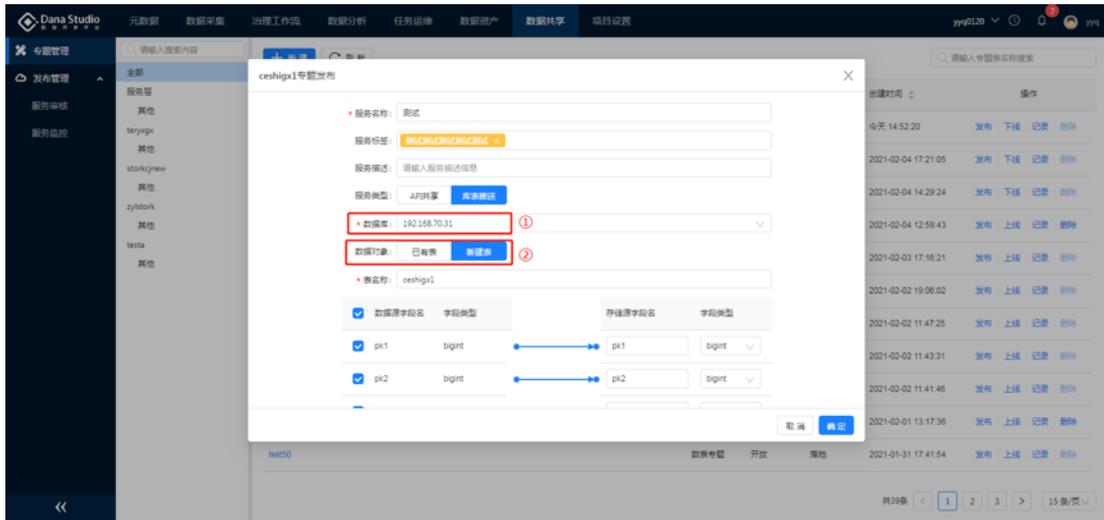
①使用权限为公开时，无认证方式可选，API 调用时，不需要授权认证

②使用权限为受限时，认证方式可选私钥或 Token，API 调用时，需要先获取认证码

③私钥：用户 API 调用后，生成一个只能由申请人使用的私人秘钥

④token：用户 API 调用后，发送一个 token 给申请人使用

11.1.5 发布库表推送服务



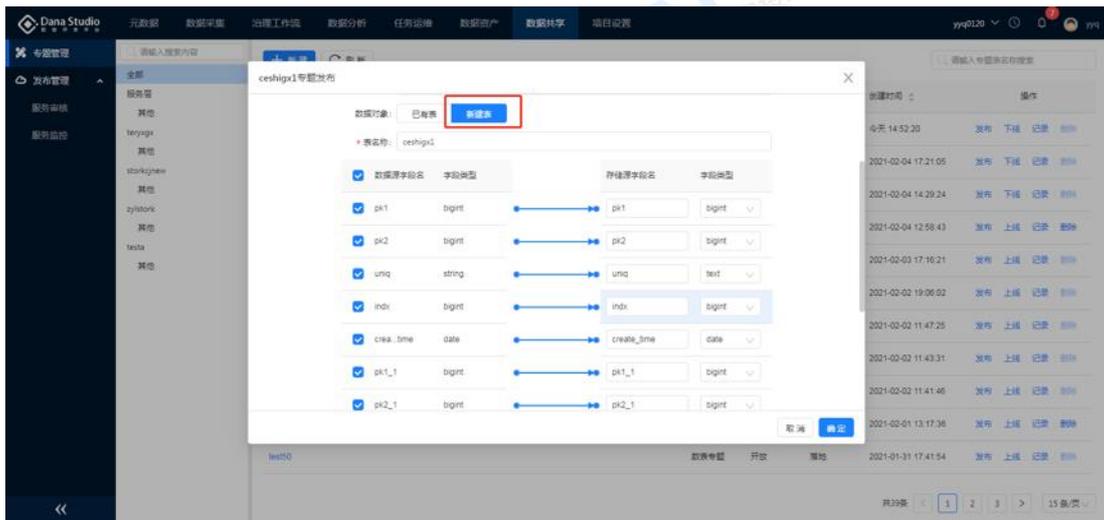
①数据库展示所有含写入数据权限的 stork、mysql、oracle 类型数据源

②数据对象包含已有表和新建表两种，选择的数据库没有建表权限时，新建表置灰不可选择

注：1) 库表推送不支持推送至外部数据源中

2) 选择的数据库必须是 utf-8 编码格式

数据对象——新建表：

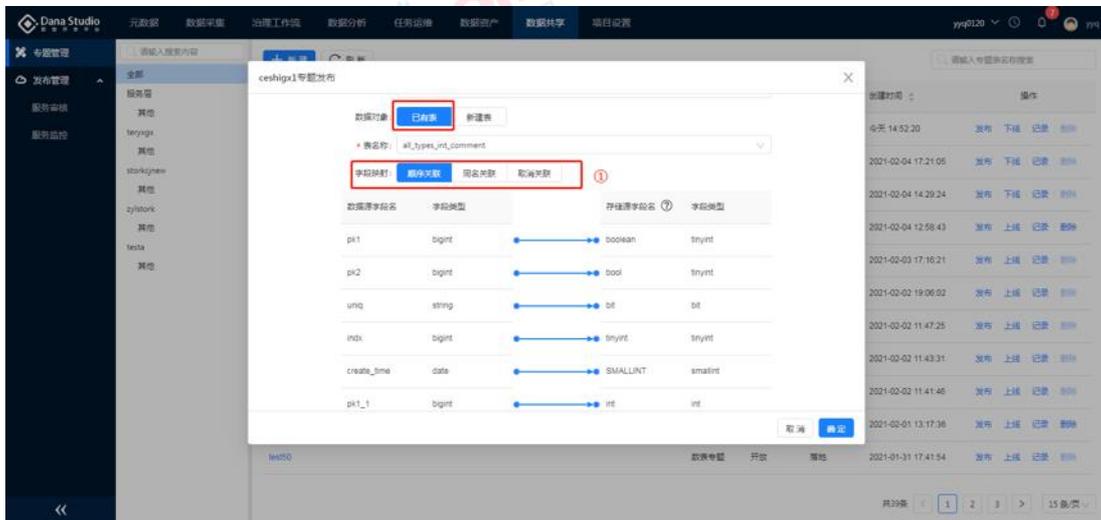


注：1) 选择新建表时，支持编辑新建表字段名称和字段类型，不可改变映射关系

2) 默认勾选全部字段，可取消勾选，但至少勾选一个字段

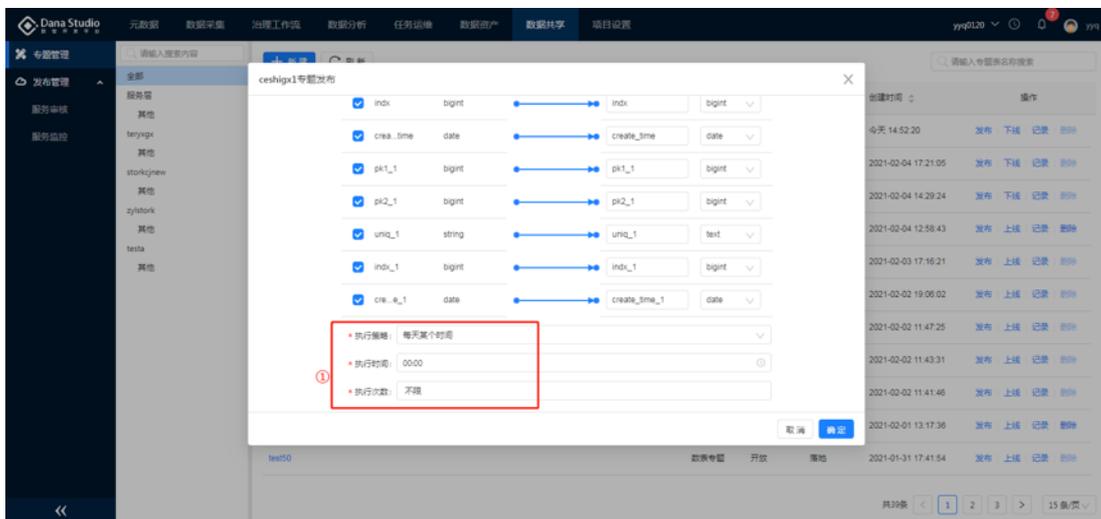
3) 新建表时 Oracle 数据库表名字段名称只支持 30 个字符

数据对象——已有表：



①选择已有表时，需要设置字段映射，字段映射包括：同名关联、顺序关联、取消映射，至少存在一条字段映射否则执行策略等不会展开

执行策略：



①库表推送需要设置执行策略：每天/周/月某个时间、用户自定义、手动执行

注：库表推送的数据策略：

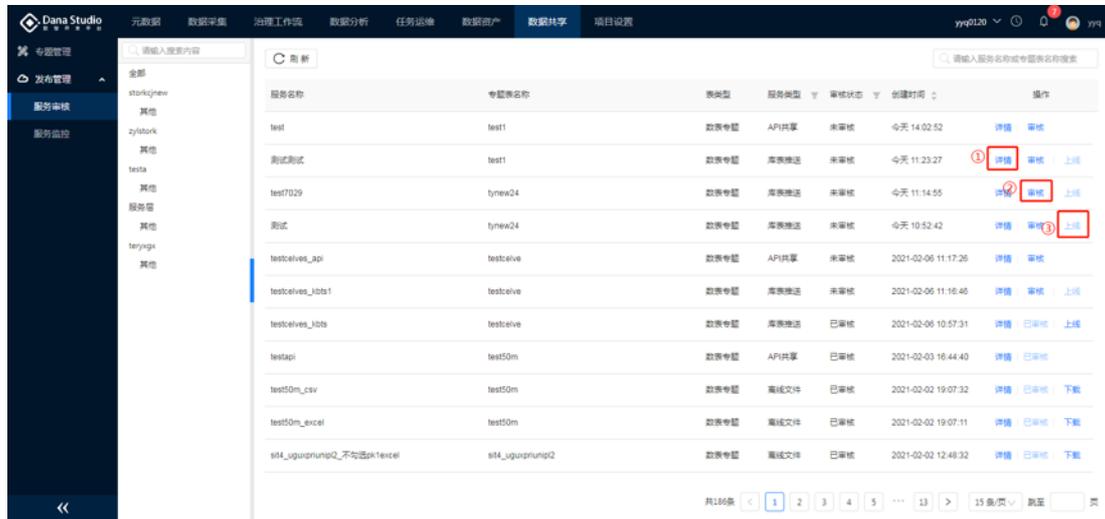
①如果该库表推送服务的专题共享信息中选择了全量更新的方式，库表推送默认为全量追加的方式推送数据到用户库中；

②如果该库表推送服务的专题共享信息中选择了增量更新的方式，库表推送默认为增量追加的方式，增量字段使用专题数据中设置的时间字段；

11.2 发布管理

11.2.1 服务审核

数据共享——发布管理——服务审核——查看已发布的服务



①点击操作下的“详情”按钮，可将该专题服务发布的详细信息

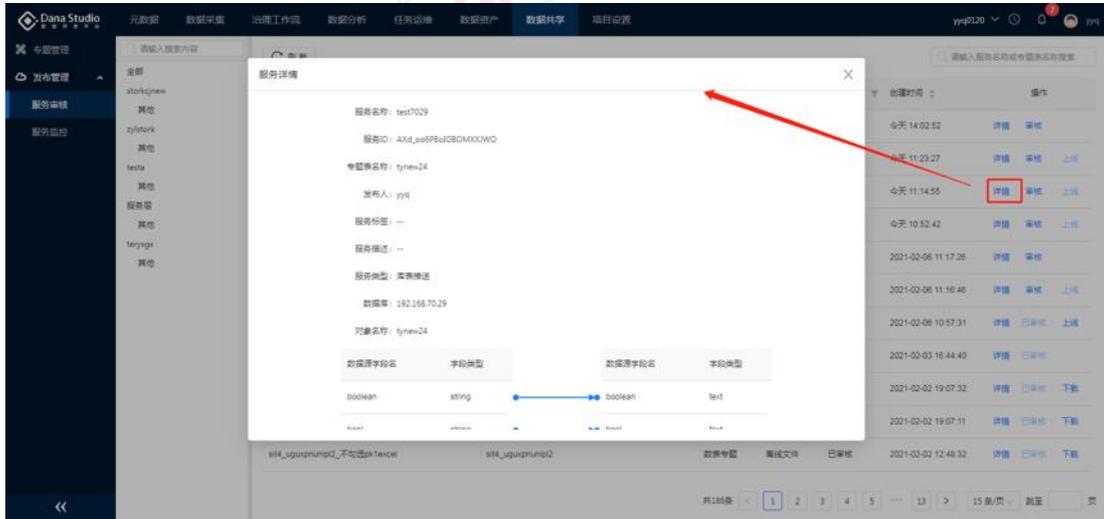
②点击操作下的“审核”按钮，可通过服务或打回服务，打回的服务在服务审核页面不展示

③仅库表推送的服务有上下线按钮，审核通过的库表推送服务，才可以进行上线，点击操作下的“上线”按钮，可将该专题服务上线，上线执行策略为审核该专题时的策略

注：审核通过的服务，API 共享后才可查看到数据

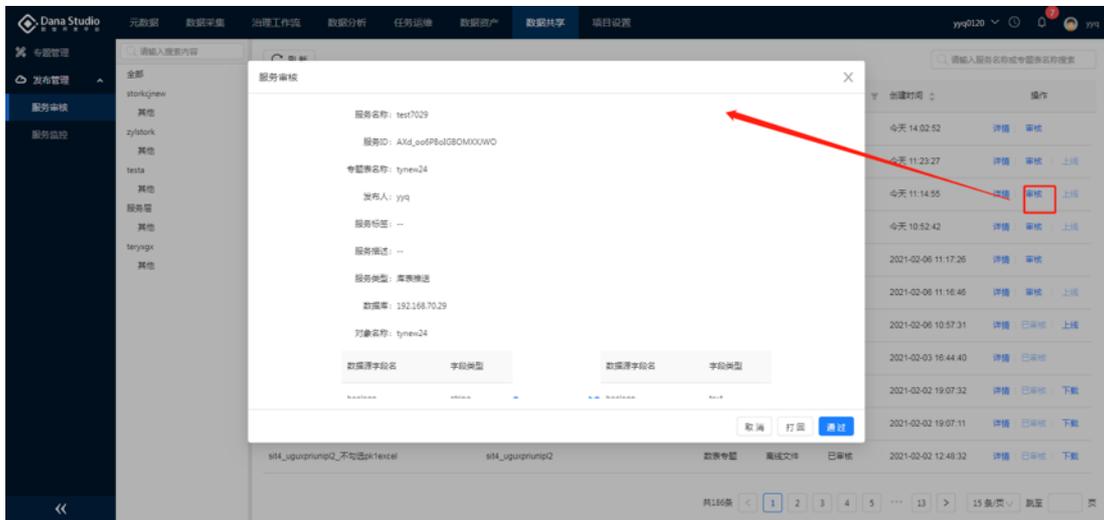
服务详情：

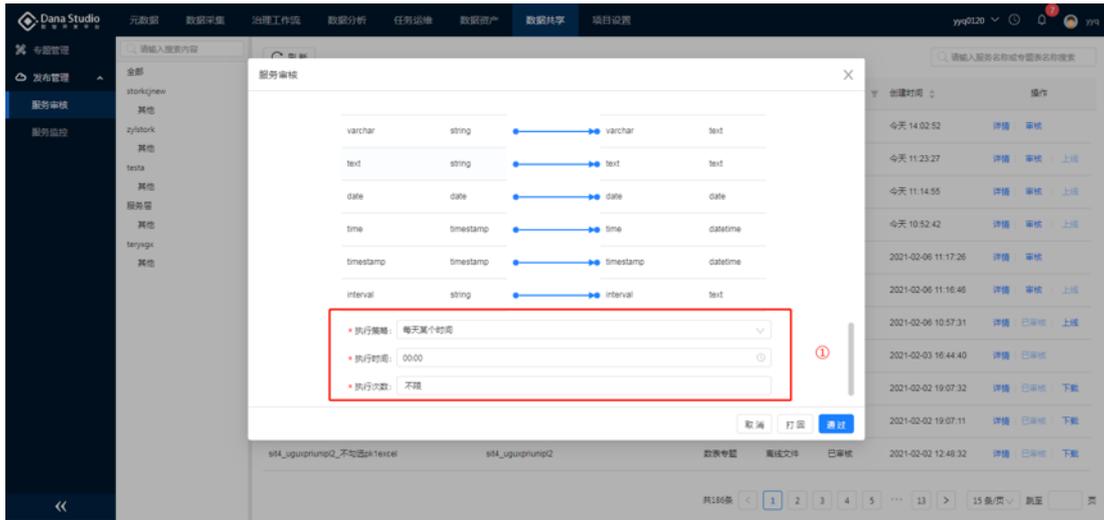
数据共享——发布管理——服务审核——点击服务后的“详情”按钮——查看发布服务的详情



库表推送服务审核:

数据共享——发布管理——服务审核——点击库表推送服务后的“审核”按钮——审核或打回服务

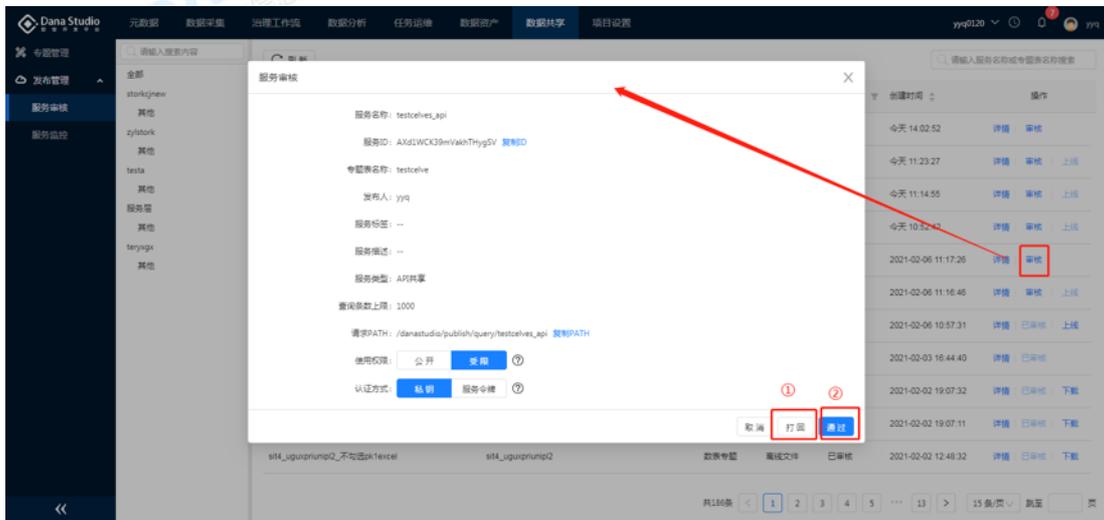




①审核时可以修改执行策略

API 共享服务审核:

数据共享——发布管理——服务审核——点击 API 共享服务后的“审核”按钮——审核或打回服务



①点击服务审核页面的“打回”，服务审核不通过，服务打回至专题管理页面，重新修改后可再次审核

②点击服务审核页面的“通过”，服务审核通过，可进行 API 分享

11.2.2 服务监控

数据共享——发布管理——服务监控——可查看审核通过的服务使用情况



①点击上方卡片，可以切换下方展示的服务列表

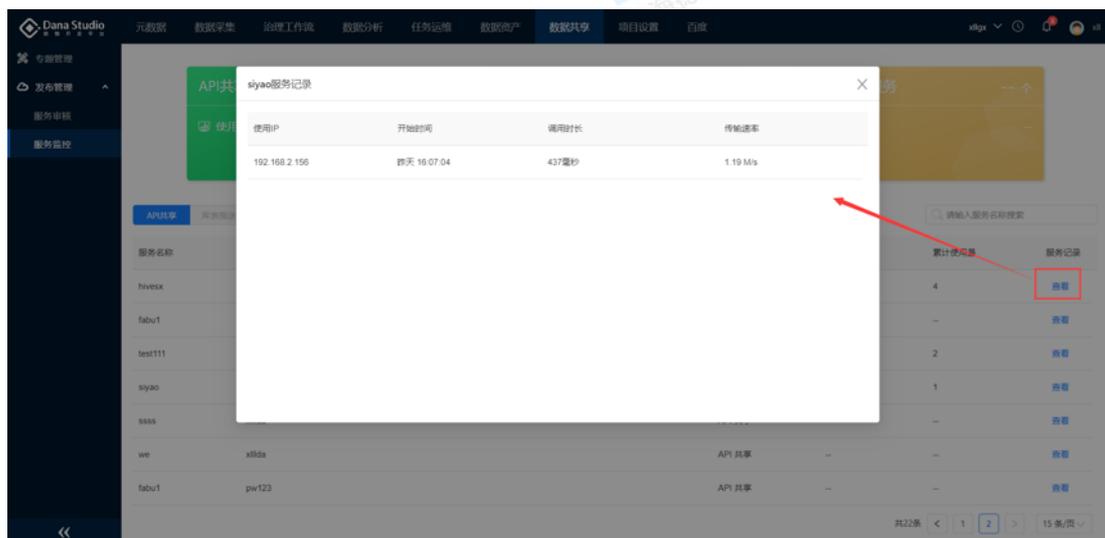
②点击中间 Tab，可以切换下方展示的服务列表

③点击“查看”按钮，可以查看服务使用记录

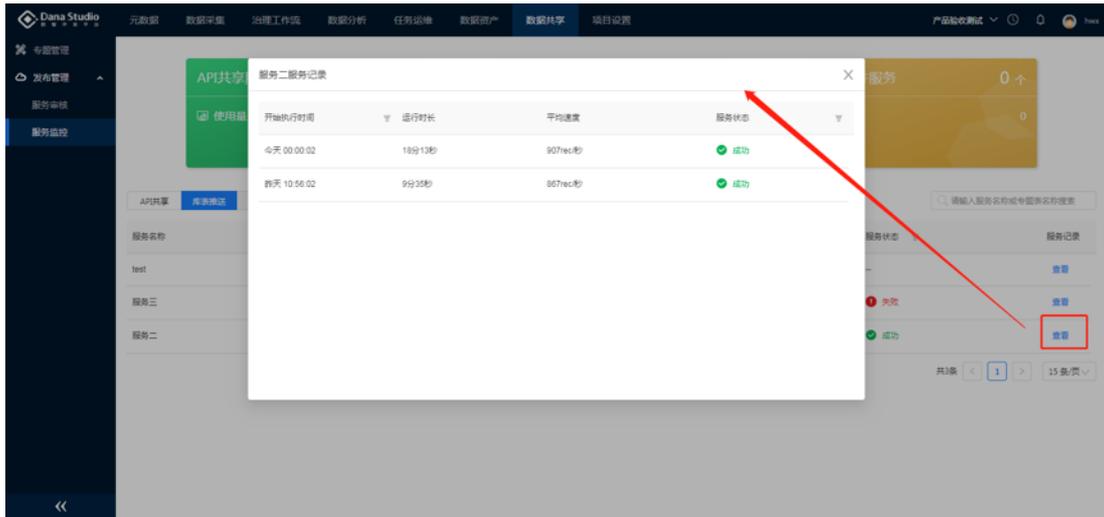
服务查看：

数据共享——发布管理——服务监控——点击服务后的“查看”按钮——查看服务的使用记录

API 共享服务查看：



库表推送服务查看：



12 数据门户

通过对不同中心平台能力的整合，实现数据全链路管理，包括跨平台统一目录管理、统一运维管理、归集能力集成、治理能力集成、资产管理能力集成、共享交换能力集成。同时，对于各委办局，提供统一的信息门户界面，并根据用户角色及权限的区分，展示不同的信息展示及服务交互内容，构建委办局提数据需求、大数据中心审核数据需求、专业厂商响应数据需求的共建模式。

12.1 首页

对于所有委办局租户平台用户的访问，建设统一风格、统一功能、统一认证、统一推送、统一的信息门户界面及门户管理功能，能根据用户角色及权限的区分，展示不同的信息展示及服务交互内容，门户中聚合数据采集、归集、治理、融合、共享、开放、分析、应用、联动和协同等多方面数据内容。并以接口的形式与省级数据服务系统无缝对接。

12.1.1 实时看板

实时看板能实时提供今日各项基础数据（如新增、活跃、累计用户数），即时监测数据运营趋势和健康表现。

12.1.2 用户概览

用户概览包含昨日新增用户数、昨日访问人数、昨日总访问次数、昨日打开次数、累计访问人数（截止昨日）、昨日停留时长等基础指标，监测产品运营趋势和健康表现，数据通过指标卡、趋势图、表格的方式进行呈现。

12.1.3 业务概览

提供全局的核心数据看板，以数据统计、数据占比、趋势分析等运营统计手段，涵盖服务事项、用户、业务量等多维度的运营统计分析应用功能，从全局角度展现政务民生应用的整体运营情况。

12.1.4 关注看板

支持运营平台管理员关注数据报表，关注的报表将在应用首页的“我的关注”展示。

12.2 新闻政策

新闻政策整合数据中心相关的动态新闻，与领导工作动态及相关数据发布更新的信息。

12.2.1 政策依据内容详情内容

整合并结构化存储政策依据详情内容，并供政策依据详情页调用及政策文件下载调用。

12.2.2 下载文件管理信息

通过对下载文件的格式进行配置，实现满足不同工作场景需求的格式文档下载，文件格式配置的参数有如版面大小，字体类型，标题级别，行距等。

12.2.3 浏览下载纪录管理信息

用户在页面对新闻政策进行浏览下载，对这些用户访问痕迹进行存档管理，以供安全审计。

12.3 资源目录

通过省、市及县多级别的目录级联，实现多类型信息资源目录的整合呈现，并针对不同级别、不同类型的信息资源进行需求申请管理。

12.3.1 信息资源目录信息

提取不同级别、不同类型信息资源目录的公共管理信息字段，生成信息资源目录信息表，以供页面呈现的各项管理调用。

12.3.2 服务目录信息

提取不同级别、不同类型接口服务目录的公共管理信息字段，生成信息资源目录信息表，以供页面呈现的各项管理调用。

12.3.3 国家接口目录信息

抓取国家接口目录信息结构存储，并供页面呈现的各项管理调用。

12.3.4 省直单位目录信息

抓取省直单位目录信息结构存储，并供页面呈现的各项管理调用。

12.3.5 事项目录信息

抓取事项目录信息结构存储，并供页面呈现的各项管理调用。

12.3.6 人口库目录信息

抓取人口库目录信息结构存储，并供页面呈现的各项管理调用。

12.3.7 法人库目录信息

抓取法人库目录信息结构存储，并供页面呈现的各项管理调用。

12.3.8 信用库目录信息

抓取信用库目录信息结构存储，并供页面呈现的各项管理调用。

12.3.9 电子证照库目录信息

抓取电子证照目录信息结构存储，并供页面呈现的各项管理调用。

12.3.10 空间地理信息目录信息

抓取空间地理目录信息结构存储，并供页面呈现的各项管理调用。

12.3.11 数据共享需求申请管理

需求部门需求提出人员基于数据目录提出向数源部门提出共享需求，填写数据共享需求单。需求提出人员在统一门户的资源目录页提出需求，页面弹出需求申请单，需求提出员填写相关信息后点击提交，产生一张新的需求单，需求单状态更新为需求分析让需求分析人员处理。

12.3.12 数据分析需求申请管理

对数据分析有需求的部门人员，在统一门户发起数据分析工具的申请，若用户已通过统一认证，则直接进入数据分析平台进行工具申请。若用户未通过统一认证，则进入统一认证注册申请页面。认证成功后，再进入数据分析平台进行工具申请。

12.3.13 需求申请服务接口申请管理

实现对服务平台已发布的服务接口，可以对服务进行申请调用，填写申请理由、IP 地址，申请服务的开始使用时间、结束使用时间，以及需要申请服务的系统，信息填写完整后进行申请提交。

12.3.14 国家接口调用管理

在统一门户发国家接口调用，在国家相应部委批准国家接口申请需求后，需求分析人员进行实施信息的补充，再提交实施工单。

12.3.15 省直单位目录使用管理

对接省直单位的资源目录，用户在统一门户发起相关单位的信息资源使用申请，在本地生成需求申请单，接收该单位的数据授权反馈结果。

12.3.16 信息资源管理信息

用户在查询浏览信息资源目录时，对符合其应用需求的信息资源进行收藏，生成信息资源收藏信息。

用户在使用信息资源时，发目录中列举的信息资源存在错误的问题发起纠错流程，生成信息资源纠错信息。

12.3.17 目录应用接口

目录应用接口提供向各类外部应用开放的自动化使用目录的应用服务，包括但不限于目录查询接口、目录订阅接口、目录推送接口等，原则上所有可人为操作的功能都应开放应用接口，这样应用系统就可以通过调用目录应用接口直接发起目录查询、目录订阅、目录推送等业务操作。

各类目录应用接口具体说明如下：

1、目录查询接口：向委办局应用系统提供数据资源共享交换平台的政务资源目录信息应用接口和相关调用环境，完成自动化的目录查询业务操作。具体目录信息包括资源属性、涉密属性、共享属性和层级属性等。

2、目录订阅接口：向委办局应用系统提供数据资源共享交换平台的政务资源目录应用接口和相关调用环境在合法授权的前提下对整个平台政务资源目录信息订阅，完成自动化的目录查询业务操作。

3、目录推送接口：向委办局应用系统提供数据资源共享交换平台的政务资源目录应用接口和相关调用环境，完成自动化的目录推送业务操作。支持根据目录订阅情况对资源目录信息进行主动推送。

12.4 应用成效

应用成效页作为政务数据中心的 application 成效绩效统计汇总的页面呈现，通过对接生产系统的应用数据库，实时计算并生成如下定制的关键统计汇总结果：

通过配置统计相关参数，如涉及的表单、数据范围，选取的字段，各项统计项的公式等参数，最终生成多维度交叉的统计表单，该功能点具体生成及操作的信息如下：

12.4.1 数据资源指标配置管理

通过配置数据资源指标统计涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现数据资源多维汇总统计输出。

12.4.2 数据资源多维度交叉统计信息

根据数据资源指标统计配置参数，汇总出多维度交叉统计结果。

12.4.3 接口指标统计配置管理

通过配置接口指标统计涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.4 接口多维度交叉统计信息

根据接口指标统计配置参数，汇总出多维度交叉统计结果。

12.4.5 库表指标统计配置管理

通过配置库表指标统计涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.6 库表多维度交叉统计信息

根据库表指标统计配置参数，汇总出多维度交叉统计结果。

12.4.7 文件指标统计配置管理

通过配置文件指标统计涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.8 文件多维度交叉统计信息

根据文件指标统计配置参数，汇总出多维度交叉统计结果，供界面呈现调用。

12.4.9 人口库数据指标配置信息管理

通过配置人口库数据指标统计涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.10 人口库多维度交叉统计信息

根据人口库指标统计配置参数，汇总出多维度交叉统计结果。

12.4.11 法人库数据指标配置信息管理

通过配置法人库数据指标统计涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.12 法人库多维度交叉统计信息

根据法人库指标统计配置参数，汇总出多维度交叉统计结果。

12.4.13 电子证照库指标配置信息管理

通过配置电子证照库指标统计涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.14 电子证照库多维度交叉统计信息

根据电子证照库指标统计配置参数，汇总出多维度交叉统计结果。

12.4.15 空间地理库指标配置信息管理

通过配置空间地理库涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.16 空间地理库多维度交叉统计信息

通过空间地理库指标的涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.17 省级部门资源目录指标统计配置管理

通过省级部门资源目录指标涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.18 地市政府资源目录指标统计配置管理

通过配置地市政府资源目录涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.19 挂接指标统计配置管理

通过配置挂接指标统计涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.20 查询指标统计配置管理

通过配置查询指标统计涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.21 申请指标统计配置管理

通过配置申请指标统计涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.22 共享指标统计配置管理

通过共享指标统计涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.23 协同应用支撑指标统计配置管理

通过配置协同应用支撑指标涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.24 资源使用指标统计配置管理

通过配置资源使用指标统计涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.25 接口调用指标统计配置管理

通过配置接口调用指标统计涉及的表单、数据范围，选取的字段，各项统计项的公式等参数信息，以实现多维汇总统计输出。

12.4.26 人口库支撑部门及应用场景统计信息

通过获取数据共享需求中的人口库支撑的情况，实时按支撑的部门及应用场景的分组统计，并对比历史统计数据，进一步生成同比，环比，增长率等额外指标值。

12.4.27 法人库支撑部门及应用场景统计信息

通过获取数据共享需求中的法人库支撑的情况，实时按支撑的部门及应用场景的分组统计，并对比历史统计数据，进一步生成同比，环比，增长率等额外指标值。

12.4.28 信用库支撑部门及应用场景统计信息

通过获取数据共享需求中的信用库支撑的情况，实时按支撑的部门及应用场景的分组统计，并对比历史统计数据，进一步生成同比，环比，增长率等额外指标值。

12.4.29 电子证照库支撑部门及应用场景统计信息

通过获取数据共享需求中的电子证照库支撑的情况，实时按支撑的部门及应用场景的分组统计，并对比历史统计数据，进一步生成同比，环比，增长率等额外指标值。

12.4.30 空间地理库支撑部门及应用场景统计信息

通过获取数据共享需求中的空间地理库支撑的情况，实时按支撑的部门及应用场景的分组统计，并对比历史统计数据，进一步生成同比，环比，增长率等额外指标值。

12.5 个人中心

用户可以个人中心页查看自己的个人信息，查看已提出需求或相关需求的环节状态和处理人信息，可点击进行催办。

12.5.1 用户信息管理

对门户用户的基本进行管理，具体的内容为帐户、密码、所属机构等基本信息。

12.5.2 角色信息管理

对门户的用户角色进行管理，一般用户角色分成厅局编目员，厅局审核员，部门管理员及系统管理员，统一配置每一种角色的应用权限。

12.5.3 权限信息管理

对门户的各个页面的浏览及操作权限进行统一管理。

12.5.4 机构信息管理

对政府的多层级的机构信息进行统一维护管理。

12.5.5 任务信息管理

针对编目、挂接、需求进展等任务进展进行统一管理。

12.5.6 个人收藏夹信息管理

每一个用户在浏览目录时，可以收藏资源目录，统计对这些用户的个人收藏夹信息进行统一管理。

12.5.7 系统消息管理

对系统消息进行统一管理。

12.5.8 服务指引信息管理

系统根据用户在浏览门户内容或者是使用门户各项功能时可能所遇到问题的提供指引服务的信息的管理。

12.5.9 厅局联系方式信息

管理各厅局的联系方式信息。

三、关于德拓

这是一个颠覆的时代，伴随着数据爆炸式增长，我们的生活、工作方式都将被数据所改变。

上海德拓信息技术股份有限公司（简称 DATATOM）致力于“数据智能”的科学研究。

我们坚信数据才是信息社会的中心，正如同如同原子对于物理世界的重要性。一切层出不穷的应用都是产生数据的工具和手段，而唯有数据才具备用户核心价值。我们希望改变当前信息社会对于数据的组织和利用方式，重新以数据为中心建立便捷的应用框架，将数据的管理从独立的应用中剥离出来，放入到数据智能平台中，帮助用户智能化的收集、存储、分类、处理、分享、可视、连接和应用数据，降低用户信息化投入成本，提高数据使用效率，加速数据价值创新应用。

基于团队自 2003 年开始的数据管理领域实践基础，利用创新的超融合大数据技术，DATATOM 可以提供数据基础架构、超融合管理平台和数据开发者服务，帮助用户智能化的收集、存储、分类、处理、分享、可视、连接和应用数据，降低用户信息化投入成本，提高数据使用效率，加速数据价值创新应用。

截止到目前为止，采用 DATATOM 数据智能产品、方案和技术能力的用户遍及电信、金融、能源、媒体、涉密、教育、医疗、政府和企业等多个行业。

了解 DATATOM 更多信息请访问：<http://www.datatom.com/>