



【智能联合建模平台】

产品操作手册

(V3.8)

目 录

1.1 产品简介	3
1.2 产品核心优势	3
1.3 产品使用场景	4
2 产品架构和配置要求	5
2.1 项目总体架构	5
2.2 系统网络结构	6
2.3 系统运行环境	7
3 智能联合建模平台功能介绍	8
3.1 Iam 账户权限管理	9
3.1.1 创建账号	9
3.1.2 应用接入及权限分配	9
3.2 系统界面	11
3.2.1 系统初始化	11
3.2.2 系统菜单	17
3.2.3 系统功能区	18
3.3 主页	18
3.4 联邦	19
3.4.1 成员列表	19
3.4.2 联邦资源	20
3.5 资源中心	21
3.5.1 我的资源	21
3.5.2 添加资源	24
3.6 合作中心	48
3.6.1 项目列表	48
3.6.2 创建机器学习项目	49
3.6.3 创建深度学习项目	96
3.6.4 创建 SQL 探索任务	101
3.7 全局设置	116
3.7.1 成员设置	116
3.7.2 专有网络设置	117
3.7.3 成员黑名单	118
3.7.4 系统设置	119
3.7.5 计算引擎设置	120

产品综述

1.1 产品简介

隐私计算是在大数据与人工智能产业下想实现数据共享及数据价值, 但又受数据监控及隐私保护法限制而深挖和孕育出的一大技术体系; 而它的技术体系是建立在联邦学习、多方安全计算、可信执行环境、差分隐私、同态加密、区块链技术等众多学科上实现的一项交叉融合技术。

不同技术的组合使用, 能保证原始数据安全及隐私性的同时, 实现数据计算和数据分析任务以实现对数据“可用、不可见”的目的; 在充分保护数据和隐私安全的前提下, 去实现数据价值的转化和释放。简单的说在政务及各行业上有数据价值交换的地方, 就有隐私计算需求的存在。

1.2 产品核心优势

隐私计算服务平台在隐私计算技术体系上已实现了联邦学习、多方安全计算、可信执行环境、差分隐私、同态加密、区块链等交叉融合技术; 它的存在能实现数据的可信挖掘和可信存证。

在系统架构上我们设计了一个计算引擎层, 包含的技术内容有“函数计算”“Spark”、“GPU 加速”, 而它们能大弧度提高我们的计算性能, 从而实现“数据+算法+算力+场景”的联合模式。

在系统数据存储上我们采用主流数据库 (MySQL、MongoDB、Oracle 等), 联盟链来存储数据样本及成员信息, 并使用 Union 区块链技术防止数据篡改。

1.3 产品使用场景

隐私计算服务平台能解决金融、政务、医疗、互联网等企业客户在联合风控、联合营销、联合分析等场景的数据安全及隐私保护需求，以实现数据价值极大释放。

隐私计算+金融

传统金融机构、互联网金融、金融科技公司及其他相关企业可以通过隐私计算分析各自的非结构化数据，来补充彼此之间多场景的用户数据信息（支付应用消费数据、出行应用的出行数据、运营商收集的客户行为数据、保险公司的投保、保险定价数据等多维度的数据特征），以此为基础进行信用画像评分来构建更为精准的联邦风控模型；形成精确客户画像而建立客户数据标签体系实现联邦营销；而私人商业银行通过隐私计算技术体系解决多方贷款检测问题。其次，电子银行可利用隐私计算技术体系保证敏感数据不共享，又能大大提升欺诈行为检测的准确率，提升银行反欺诈检测效果。

隐私计算+政务

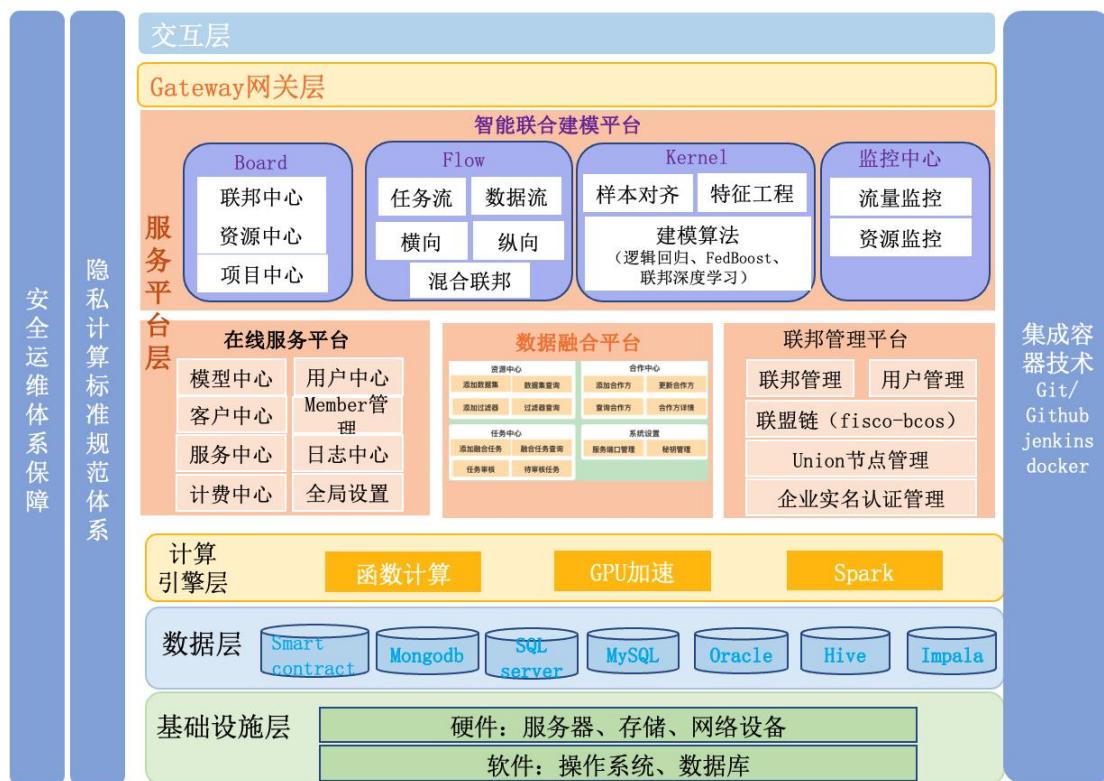
保证政务明文数据不出库，根据数据使用方的需求联合建模，将模型结果输出给数据使用方。

在明文数据不出库的前提下推进政府数据开放共享、提升数据资源的价值、加强数据资源整合。

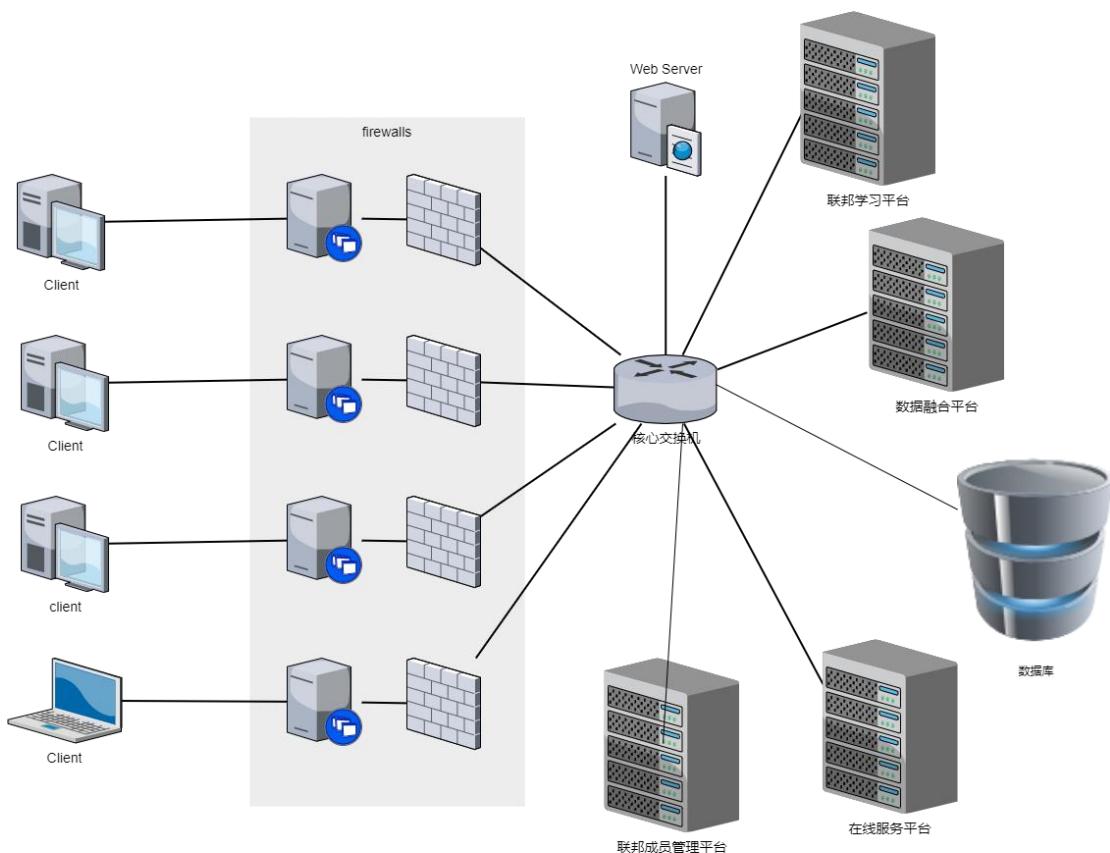
2 产品架构和配置要求

2.1 项目总体架构

隐私计算服务平台的总体架构分别由 4 个子服务平台（智能联合建模平台、在线服务平台、数据融合平台、联邦管理平台）、一个 IAM 统一认证平台、二个体系（联邦学习和安全多方计算）、一个集成容器技术、基础层、数据层、计算引擎层、网关层、交互层等组成。



2.2 系统网络结构



隐私计算平台网络拓扑图

Web Server: 实现 WEB 访问功能。

联邦学习平台服务器: 提供联邦建模任务功能。

数据融合平台服务器: 提供多个参与方计算交集功能。

在线服务平台服务器: 提供模型预测和多方安全计算功能。

联邦成员管理服务器: 提供联邦成员管理功能。

IAM 统一认证平台: 提供权限管理的基础服务。

数据库服务器: 储存业务数据。

防火墙: 保障数据安全。

核心交换机: 搭建内部局域网。

客户端: 浏览器客户端无须安装软件，通过浏览器完成客户端功能。

2.3 系统运行环境

硬件配置

推荐服务器配置: 32 核 128G 内存 500G 固态硬盘

最低服务器配置: 8 核 16G 内存 100G 固态硬盘

操作系统: Version: CentOS Linux release 7.5

软件配置

系统	服务	运行环境	说明
智能联合建模平台	Board-website	Nginx/OpenResty	可视化前端，用于查看和维护联邦学习相关作业。
	Board-service	JDK1.8	与前端对应的后台服务
	Flow	Python3	执行建模流程任务，可配置函数计算和spark 集群用于算法运算。
	Gateway	JDK1.8	多个联邦成员系统之间进行交互的出入口
数据融合平台	Fusion-service	JDK1.8	数据融合平台的后端服务
	Fusion-website	Nginx/OpenResty	可视化前端，用于进行数据融合的操作

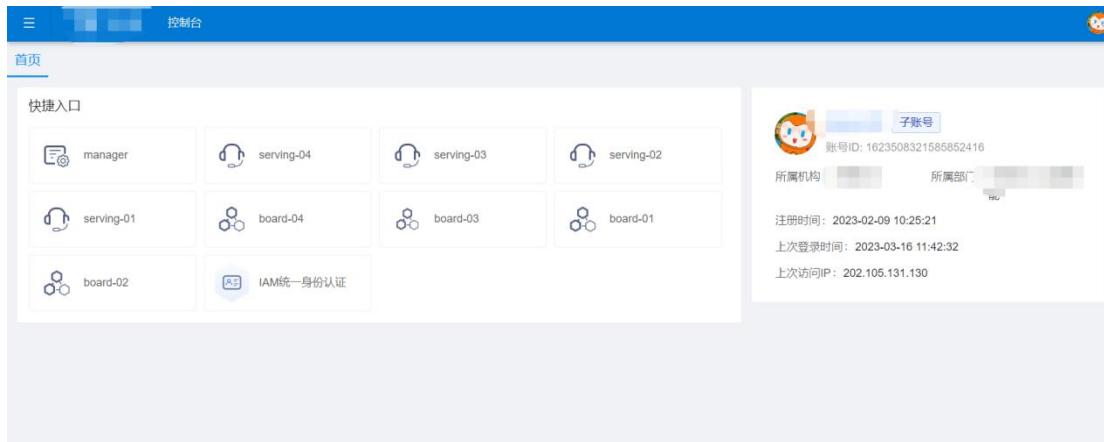
			作业
在线服务平台	Serving-service	JDK1.8	在线服务平台的后端服务
	Serving-website	Nginx/OpenResty	可视化前端, 用于进行在线服务平台的操作作业
联邦管理平台	Manager-service	JDK1.8	与前端对应的后端服务
	Manager-website	Nginx	可视化前端, 用于查看和维护联邦成员相关
	blockchain-data-sync	JDK1.8	实现了区块链数据到业务库(mongodb)的实时同步功能
	union-service	JDK1.8	提供对成员信息,数据集信息提供查询,登记服务;使用去中心化(区块链)存储
Iam 统一认证平台	iam-service	JDK1.8	iam 的后端服务
	iam-website	Nginx	iam 的前端运行环境
	Portal	Nginx	统一门户
	iam-gateway	JDK1.8	iam 的网关服务

3 智能联合建模平台功能介绍

本章提供联合建模平台各个功能模块详细的功能介绍说明, 系统平台包括系统界面、主页、联邦、数据中心、合作中心、用户管理和全局设置模块, 以下将介绍具体每个模块的功能点。

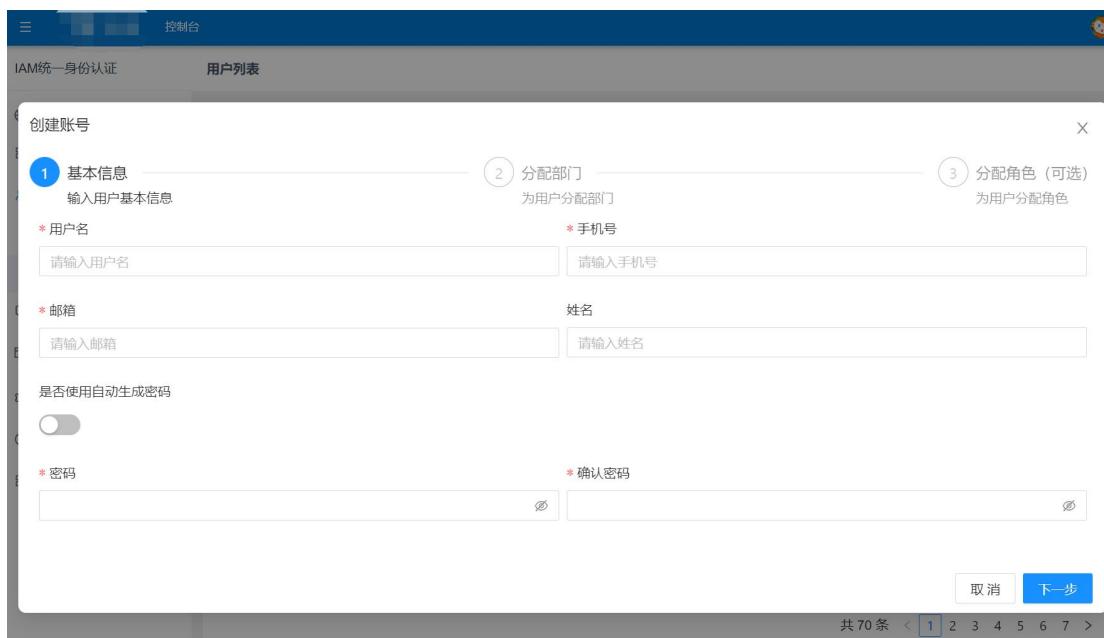
3.1 iam 账户权限管理

智能联合建模平台接入 iam 权限管理系统，统一管理系统账户及资源权限。



3.1.1 创建账号

由 IAM 管理员创建账号后可以登录 IAM 平台

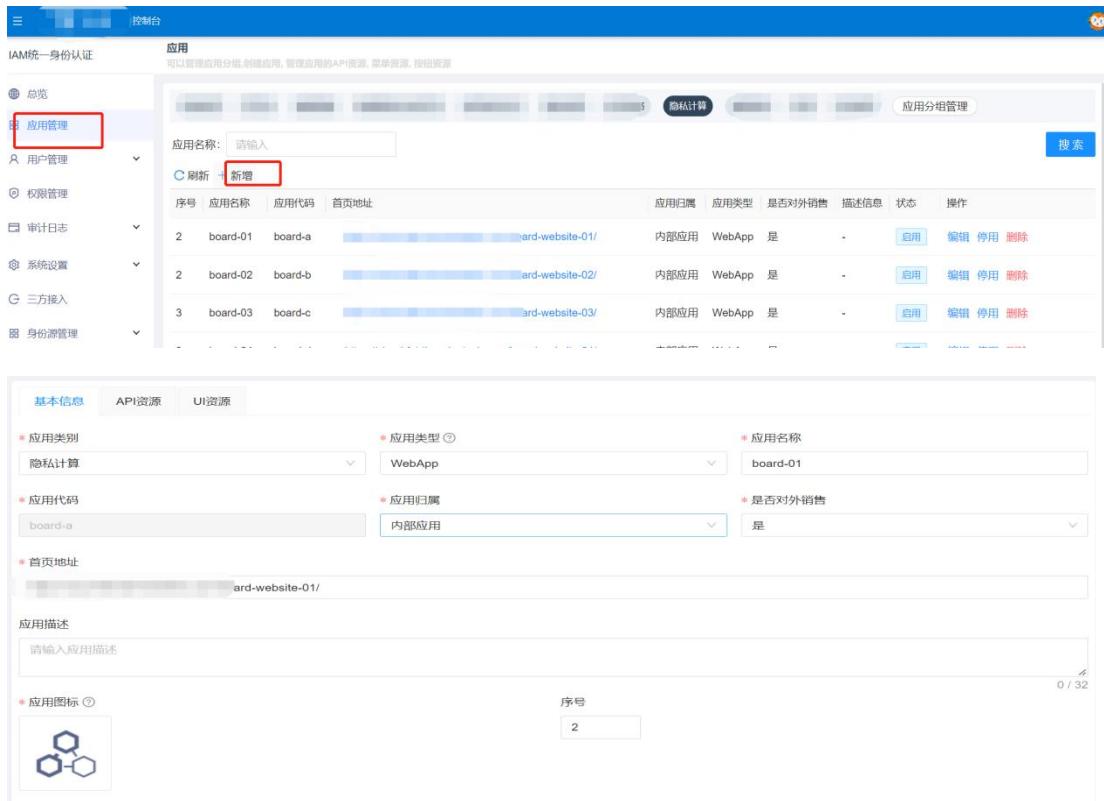


3.1.2 应用接入及权限分配

智能联合建模平台接入 iam 需要添加应用及配置系统菜单资源。

操作:

- 1) 添加应用



基本信息

应用类别: 隐私计算

应用类型: WebApp

应用名称: board-01

应用代码: board-a

应用归属: 内部应用

是否对外销售: 是

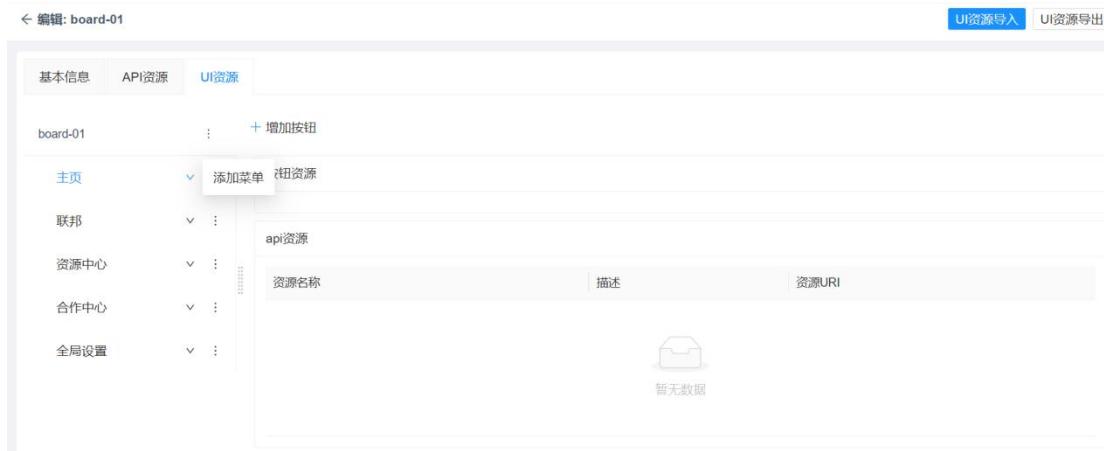
首页地址: ard-website-01/

应用描述: 请输入应用描述

应用图标: 

序号: 2

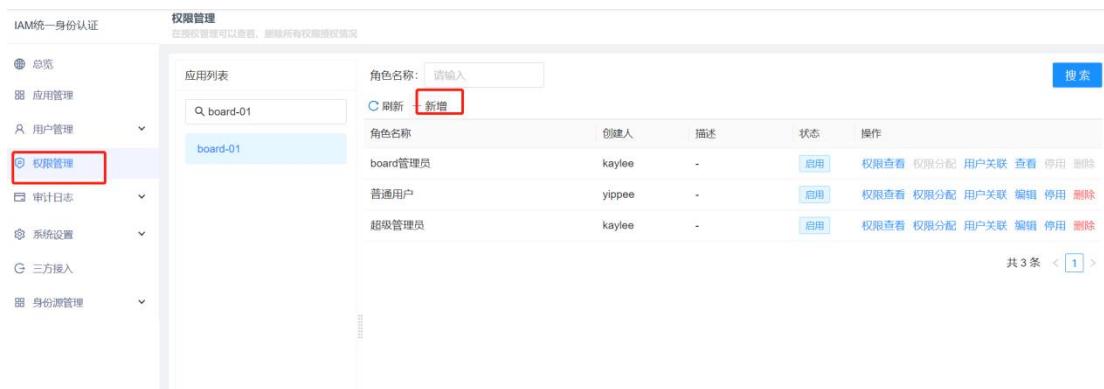
2) 配置系统菜单资源



资源名称	描述	资源URI
	暂无数据	

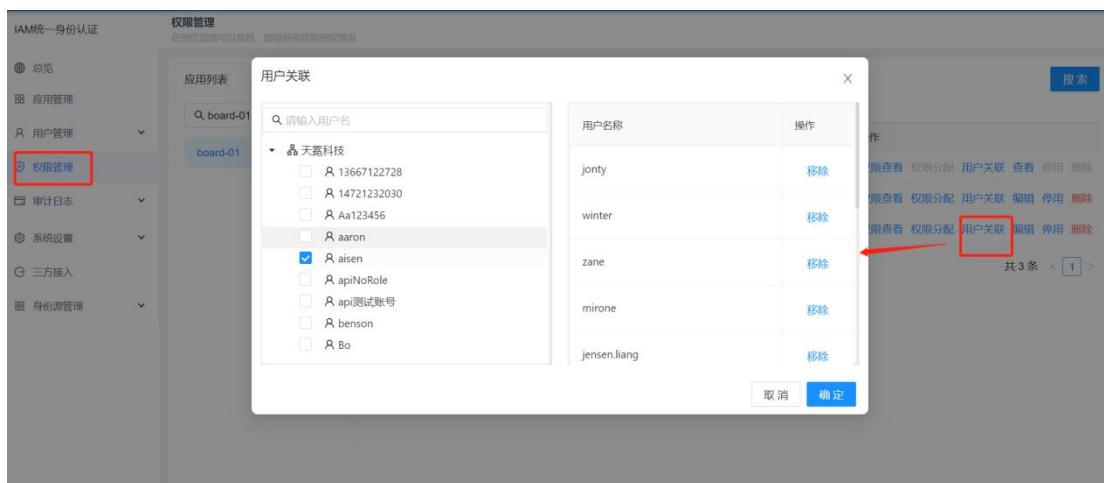
3) 创建应用管理员

创建应用管理员，并关联应用可访问的菜单权限。



4) 用户关联角色

给用户分配角色，则该用户拥有该角色下的资源访问权限。



3.2 系统界面

3.2.1 系统初始化

联邦学习系统网址:

[http\(s\)://xxxxxx/board-website-01/](http(s)://xxxxxx/board-website-01/)

操作:

1) 成员初始化

系统部署完成后第一个访问系统需要成员初始化，依系统要求输入联邦成员名称、邮箱、联系电话、是否允许对外公开数据集基础信息等，点击“加入联邦”

进行成员初始化。密钥类型支持 rsa 和 sm2 两种。

初始化系统 · 成为联邦成员

* 联邦成员名称:

0 / 12

邮箱:

联系电话:

是否允许对外公开数据资源基础信息:

是 否

密钥类型:

▼

加入联邦 !

2) 系统配置

进入系统后需要根据提示去 【全局设置】 - 【系统设置】 更新系统配置:



智能联合建模平台

主页

项目详情 × 流程详情 × 主页

✓ UnionService 联邦服务 Check

BoardService 控制台服务 Check

FlowService: 获取检查点 config value 失败, RuntimeException:请在[全局设置]-[系统设置]中对 flow 的内网地址进行设置

GatewayService 网关服务 Check

获取 GatewayService 服务可用性状态失败: gateway 内网地址不正确,请在[全局设置]-[系统设置]中设置 gateway 服务的内网地址, 格式为 HOST:PORT

FlowService 工作流服务 Check

获取 FlowService 服务可用性状态失败: 尚未设置 flow 内网地址, 请在[全局设置]-[系统设置]中设置 flow 服务的内网地址。

收起 ▾

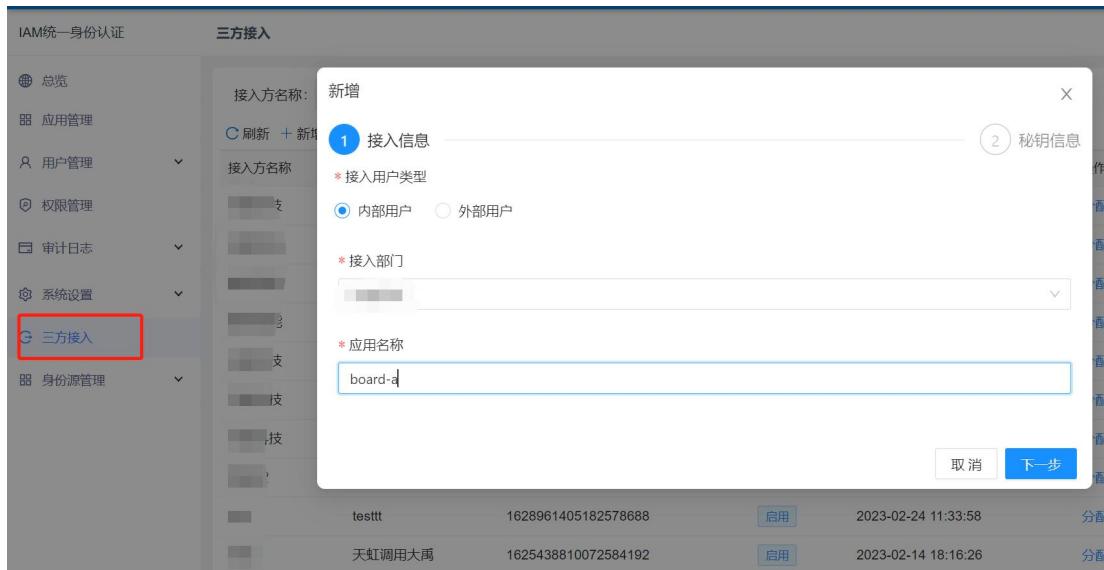
更新完成后查看首页各服务状态，运行正常。

3) IAM 配置

在【全局设置】 - 【系统设置】配置接入 IAM 权限:

(1) 配置 iam 三方接入应用

在 IAM 系统-【三方接入】，点击新增，选择接入部门，填写接入应用名称，点击下一步



选择 RSA 签名方式，点击【新增】 - 【编辑】，填写合作方公钥

publicKey	加密密钥	合作方公钥	创建时间	操作
MIGfMA0GCSqGSIb3...	rMt/nno/inD4HLmQU...	-	2023-05-18 13:18:19	编辑 校验 删除

在智能联合建模平台，生成 rsa 公钥私钥



并将公钥填写至 iam 的合作方公钥处，点击保存，提交成功如图：

接入方名称	应用名称	应用ID	状态	创建时间	操作
XXX科技有限公司	5	15905458660248007168	启用	2022-11-10 11:24:26	分配密钥 开通权限 权限查看 编辑 禁用 删除
XXX科技有限公司	4	1590545866047516672	启用	2022-11-10 11:24:20	分配密钥 开通权限 权限查看 编辑 禁用 删除
XXX科技有限公司	133	1590545813815848960	启用	2022-11-10 11:24:08	分配密钥 开通权限 权限查看 编辑 禁用 删除

(2) 配置 board 接入 iam 权限

IAM

check

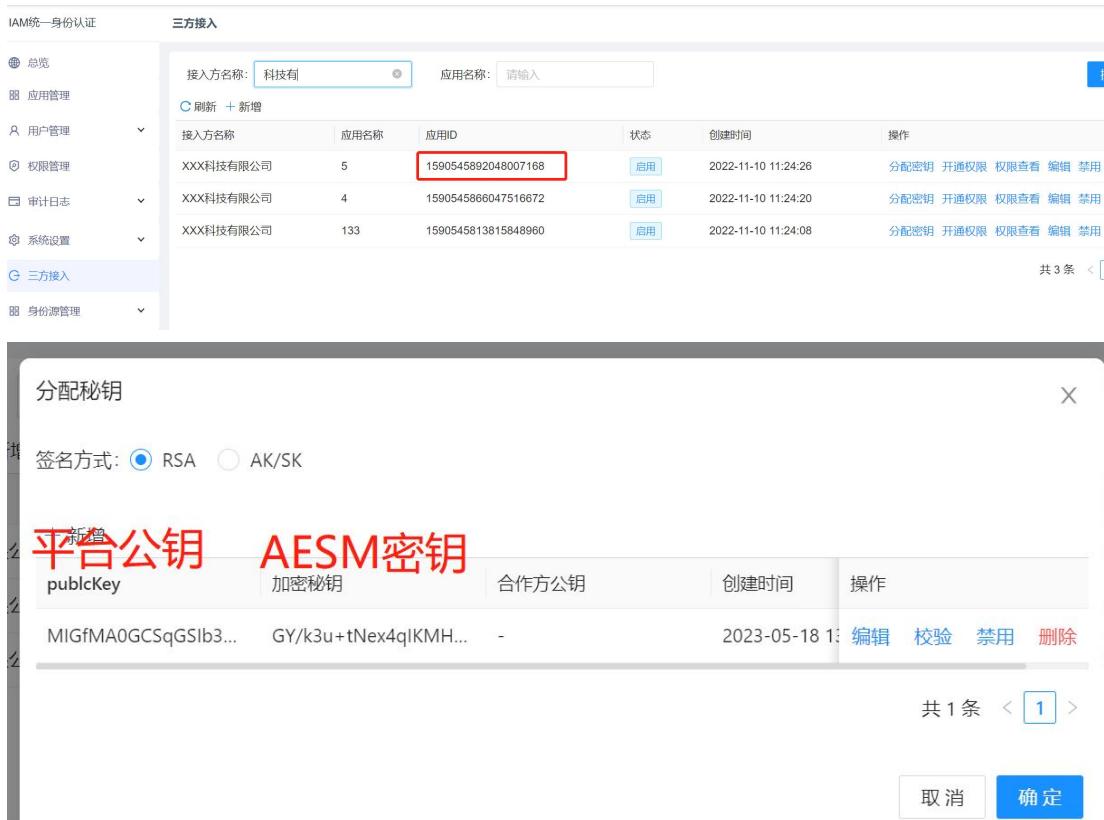
Board应用代码:	board-a
IAM后端地址(IAM分配):	.com
应用ID(IAM分配):	69767680
AES密钥(IAM分配):	*****
平台公钥(IAM分配):	----- -----GSIb3DQEBAQUAA4GNADCBiQK BgO----- -----VYrJ6eEU+G67vQoBpVCazFc -----EZ4312Yet3kyTxI6Xa2R72MV
RSA私钥:	*****
RSA公钥:	-----DEBAQUAA4GNADCBiQK -----QgkMFxNiKvO4I+6sx/E1 -----btFuOO+okxCF4zrYI0UigSVoD
在线生成RSA公私钥	

- 应用代码为 iam 配置应用的应用代码, 如图:



应用管理						
应用名称:		请输入				
刷新		新增				
序号	应用名称	应用代码	首页地址	应用归属	应用类型	是否对外销售
2	board-01	board-a	http://board-website-01/	内部应用	WebApp	是
2	board-02	board-b	http://board-website-02/	内部应用	WebApp	是
3	board-03	board-c	http://board-website-03/	内部应用	WebApp	是
9	board-04	board-d	http://board-website-04/	内部应用	WebApp	是

- Iam 后端地址填写为 iam 后端域名;
- 应用 id, AES 密钥, 平台公钥如图, 复制填写到 board 的 iam 配置中;



The screenshot shows the 'Three-party Access' section of the IAM product. On the left sidebar, under 'Three-party Access', there is a 'Secret Key Allocation' dialog box. The dialog box has a title '分配秘钥' (Allocate Secret Key) and a subtitle '签名方式: RSA' (Signature Method: RSA). It displays a table with one row of data:

公钥	密钥	合作方公钥	创建时间	操作
publicKey	AESM密钥	MIGfMA0GCSqGSIb3...	2023-05-18 1...	编辑 校验 禁用 删除

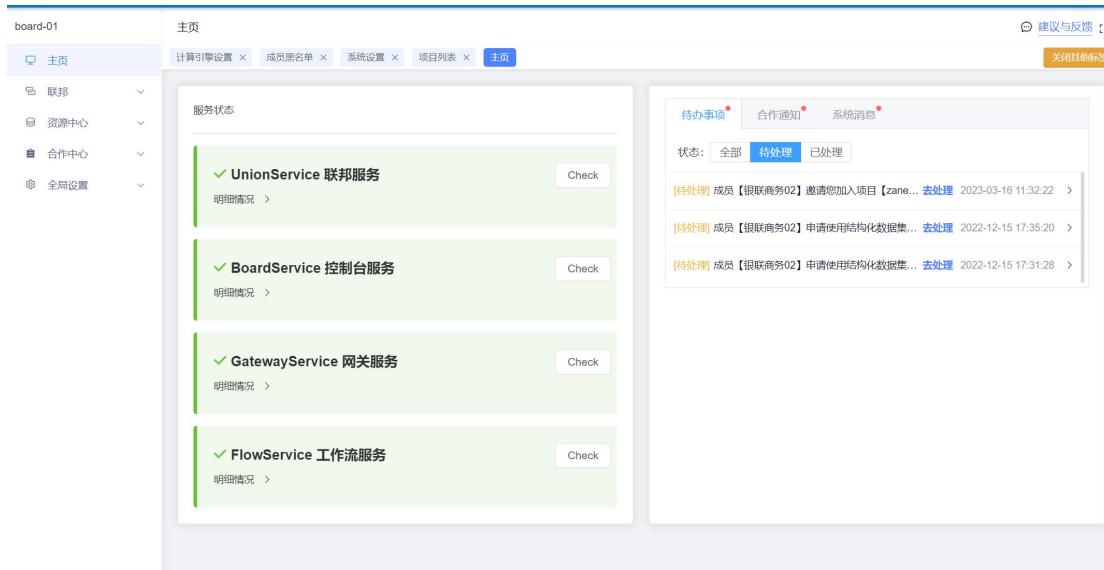
At the bottom right of the dialog box are two buttons: '取消' (Cancel) and '确定' (Confirm).

- RSA 公钥私钥为 board 在线生成的公钥私钥。

全部配置成功后，点击右上角【check】按钮校验配置是否可用，是，点击保存。

3.2.2 系统菜单

登录系统后进入主页，页面左侧展示为系统菜单栏，顶部展示页面导航栏。



The screenshot shows the main page of the system. At the top is a navigation bar with tabs: '计算引擎设置' (Compute Engine Settings), '成员黑名单' (Member Blacklist), '系统设置' (System Settings), '项目列表' (Project List), and '主页' (Home). To the right of the navigation bar are two buttons: '建议与反馈' (Suggestion and Feedback) and '关闭其他标签' (Close Other Tabs).

The left side features a vertical menu bar with sections: 'board-01', '主页' (Home), '联邦' (Federation), '资源中心' (Resource Center), '合作中心' (Collaboration Center), and '全局设置' (Global Settings).

The main content area contains several service status cards:

- UnionService 联邦服务**: Status: Normal, Last checked: 2023-03-16 11:32:22.
- BoardService 控制台服务**: Status: Normal, Last checked: 2023-03-16 11:32:22.
- GatewayService 网关服务**: Status: Normal, Last checked: 2023-03-16 11:32:22.
- FlowService 工作流服务**: Status: Normal, Last checked: 2023-03-16 11:32:22.

To the right of the service cards is a '待办事项' (Pending Tasks) panel. It includes tabs for '待办事项' (Pending), '合作通知' (Collaboration Notifications), and '系统消息' (System Messages). The '待办事项' tab is selected, showing three items:

- [待处理] 成员【银联商务02】邀请您加入项目【zane... 去处理 2023-03-16 11:32:22 >
- [待处理] 成员【银联商务02】申请使用结构化数据集... 去处理 2022-12-15 17:35:20 >
- [待处理] 成员【银联商务02】申请使用结构化数据集... 去处理 2022-12-15 17:31:28 >

3.2.3 系统功能区

在左侧功能区选择相应功能后，右侧区域展示该功能区的内容。



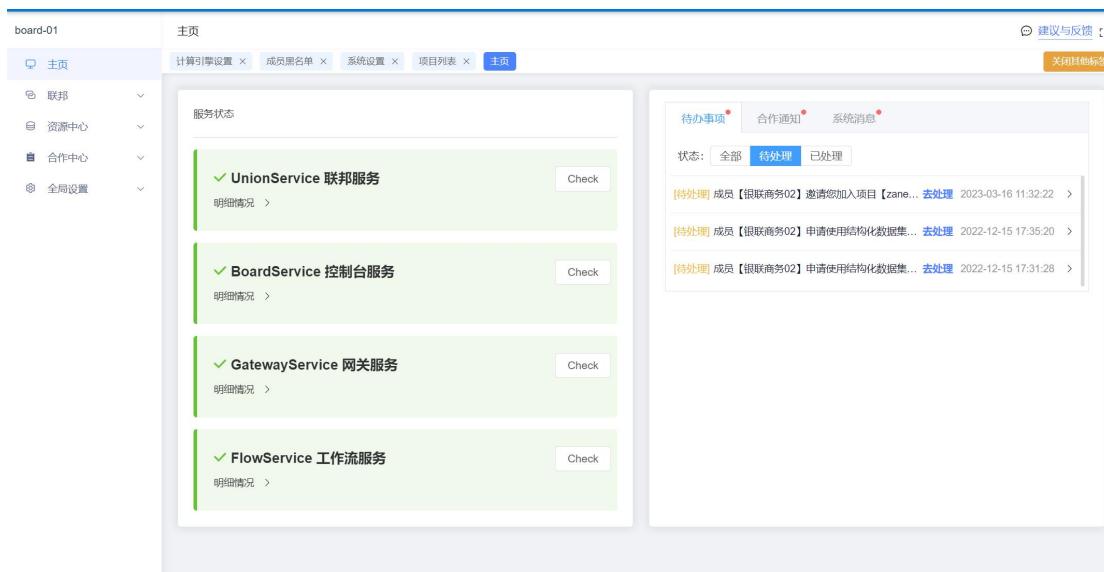
The screenshot shows the 'board-01' interface. On the left sidebar, under '我的资源', the '添加资源' button is highlighted with a red box. The main content area displays three resource categories: '结构化数据' (with a chart icon), '图像数据' (with a person interacting with a large screen icon), and '过滤器数据' (with a funnel icon). Each category has a brief description and a blue '进入' (Enter) button at the bottom.

3.3 主页

主页显示本地服务状态模块和 Message 信息模块。

1. 服务状态

显示服务状态包括：union、gateway、storage、flow 的状态信息。



The screenshot shows the 'board-01' homepage. The left sidebar includes '主页面' and other sections like '联邦', '资源中心', '合作中心', and '全局设置'. The main area features a '服务状态' section listing four services with green checkmarks: 'UnionService 联邦服务', 'BoardService 控制台服务', 'GatewayService 网关服务', and 'FlowService 工作流服务'. Each service has a '明细情况' link and a 'Check' button. To the right is a '待办事项' (Pending Tasks) module showing three items: '成员【银联商务02】邀请您加入项目【zane... 去处理」 (Status: 待处理), '成员【银联商务02】申请使用结构化数据集... 去处理」 (Status: 待处理), and '成员【银联商务02】申请使用结构化数据集... 去处理」 (Status: 已处理). There are also tabs for '合作通知' and '系统消息'.

2. Message 信息

代办事项包括需要审核的项目，需要审核的数据资源等；

合作通知包括合作方审核过的项目，数据资源等；

系统消息包括网关收发的异常消息。



3.4 联邦

在智能联合建模平台里各自企业注册成功后为成员方，所有成员方的信息及成员上传的数据样本（即所有可见权限的数据）存放在联邦模块通过 Union 服务进行管理。

3.4.1 成员列表

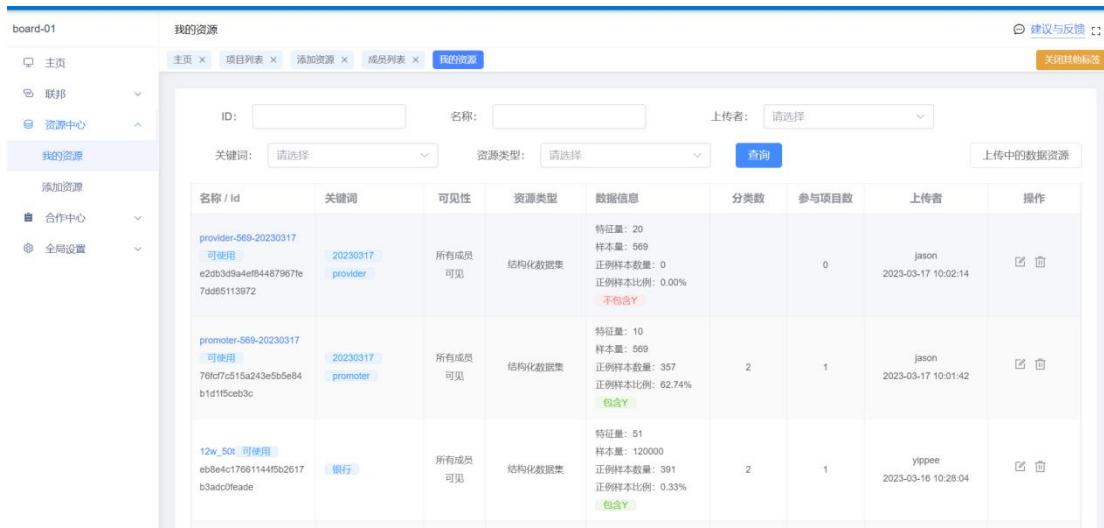
成员列表模块展示所有在智能联合建模平台里注册成功的成员方以及成员方的信息，包括：公司名称、公司 LOGO、邮箱、联系电话、最后活动时间等。

操作：

- 1) 在左侧菜单栏中点击“成员列表”，右侧展示联邦中所有成员信息；



- 2) 点击名片右下角“查看数据资源”可快捷进入“联邦资源”模块，方便联邦成员可快速查看对方企业的数据集信息。



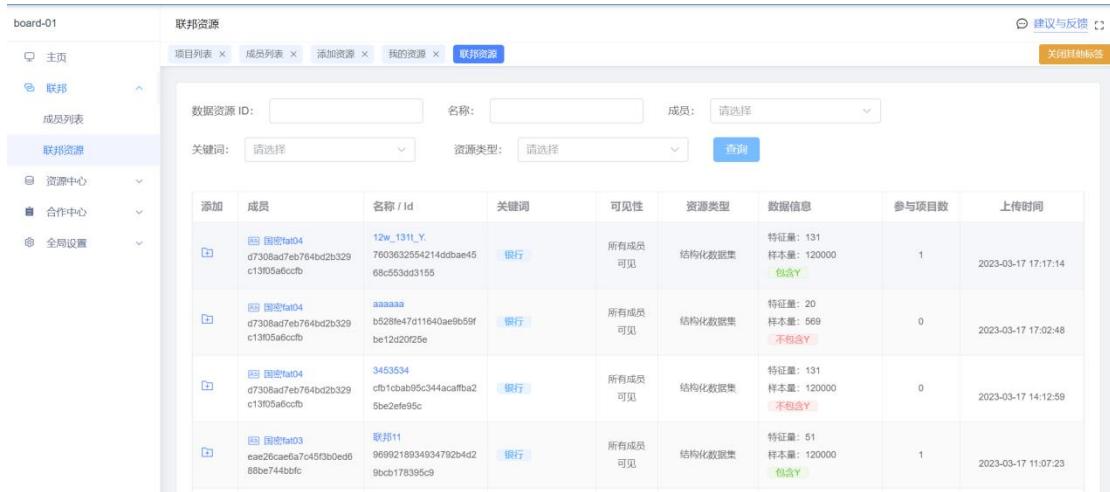
名称 / ID	关键词	可见性	资源类型	数据信息	分类数	参与项目数	上传者	操作
provider-569-20230317 可使用 e20b3d9a4ef84487967fe 7dd65113972	20230317 provider	所有成员可见	结构化数据集	特征量: 20 样本量: 569 正例样本数量: 0 正例样本比例: 0.00% 不含Y	0	1	jason	查看
promoter-569-20230317 可使用 78fc7c515a243e5b5e84 b1d1f5ce83c	20230317 promoter	所有成员可见	结构化数据集	特征量: 10 样本量: 569 正例样本数量: 357 正例样本比例: 62.74% 包含Y	2	1	jason	查看
12w_50t 可使用 eb8e4c17661144f5b2617 b3adcc0feade	银行	所有成员可见	结构化数据集	特征量: 51 样本量: 120000 正例样本数量: 391 正例样本比例: 0.33% 包含Y	2	1	yppee	查看

3.4.2 联邦资源

联邦资源模块是存放所有成员方上传的数据集(此数据权限设置为所有可见状态)，展示出成员名、数据集、关键词、数据量、参与项目、包含 Y、上传时间等信息。

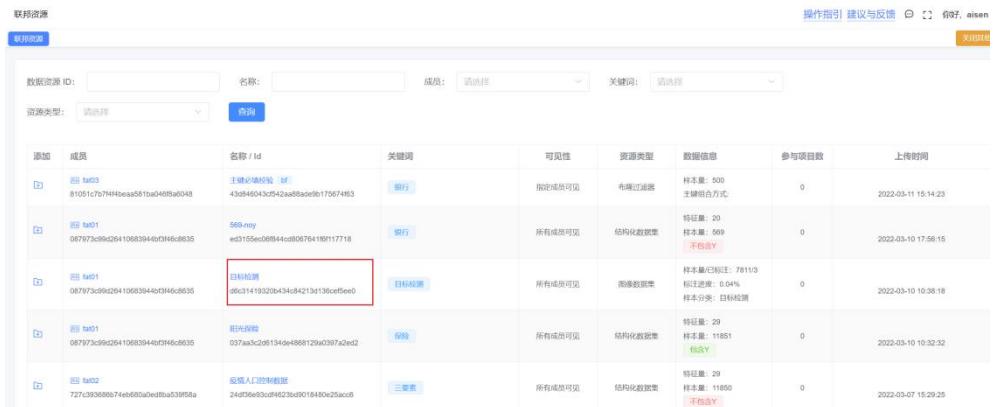
操作：

- 1) 点击左侧菜单栏“联邦资源”，右侧展示页面；



2) 进入右侧展示界面，可根据需求输入查询条件，包括数据 ID、数据集名称、成员名称和关键词，点击“查询”按钮，查询数据集。

1) 点击数据集名称，可进入数据集页面详细查看数据集信息。



进入联邦资源可预览数据集详情，包含样本量/特征量。

3.5 资源中心

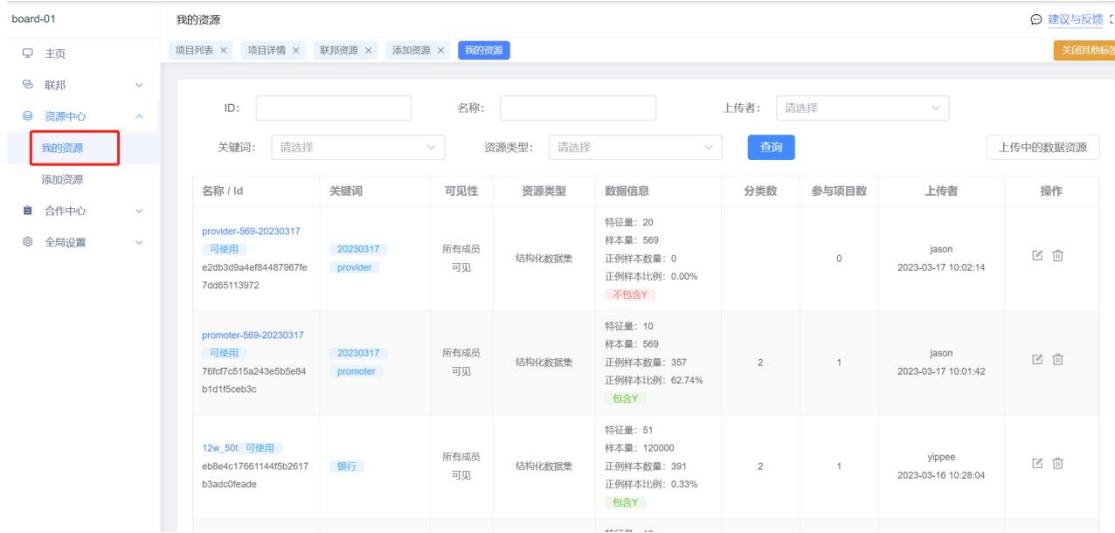
资源中心模块是成员管理自己本地数据的模块，可以对数据集进行上传及管理操作。

3.5.1 我的资源

1. 全部数据集

操作：

1) 点击左侧菜单栏“我的资源”，右侧页面展示数据。



The screenshot shows the 'My Resources' section of the software interface. On the left, there's a sidebar with various options like 'Main', 'Federal', 'Resource Center', and 'My Resources'. The 'My Resources' option is highlighted with a red box. The main area displays a table of resources. The columns include: Name / ID, Keyword, Visibility, Resource Type, Data Information, Category Count, Participation Project Count, Uploader, and Operation. There are three entries listed:

- provider-569-20230317**: 可使用, provider. Data: 特征量: 20, 样本量: 569, 正例样本数量: 0, 正例样本比例: 0.00%. 按钮: 编辑, 查看.
- promoter-569-20230317**: 可使用, promoter. Data: 特征量: 10, 样本量: 569, 正例样本数量: 357, 正例样本比例: 62.74%. 按钮: 编辑, 查看.
- 12w_50t 可使用**: 银行. Data: 特征量: 51, 样本量: 120000, 正例样本数量: 391, 正例样本比例: 0.33%. 按钮: 编辑, 查看.

2) 可根据需求条件进行查询。输入查询条件，点击“查询”。

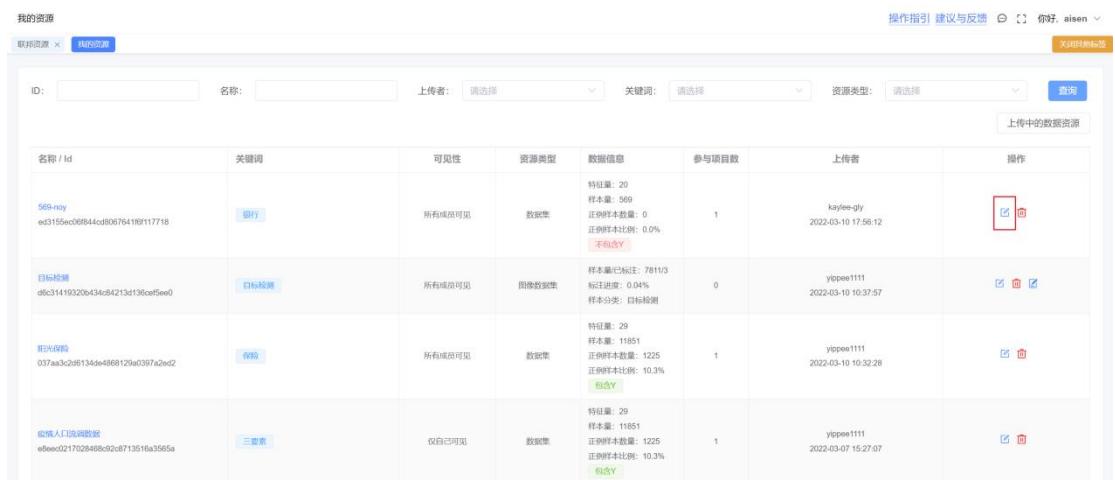


This screenshot shows the search interface for 'My Resources'. It features a search form with fields for 'ID', 'Name', 'Uploader', 'Keyword', 'Resource Type', and a 'Search' button. Below the search form is a table with columns: Name / Id, Keyword, Visibility, Resource Type, Data Information, Participation Project Count, Uploader, and Operation. A single entry is shown:

目标检测	d6c31419320b434c84213d136cef5ee0	所有成员可见	图像数据集	样本量/已标注: 7811/3 标注进度: 0.04% 样本分类: 目标检测	0	yippee1111 2022-03-10 10:37:57	编辑, 查看
------	----------------------------------	--------	-------	--	---	-----------------------------------	--------

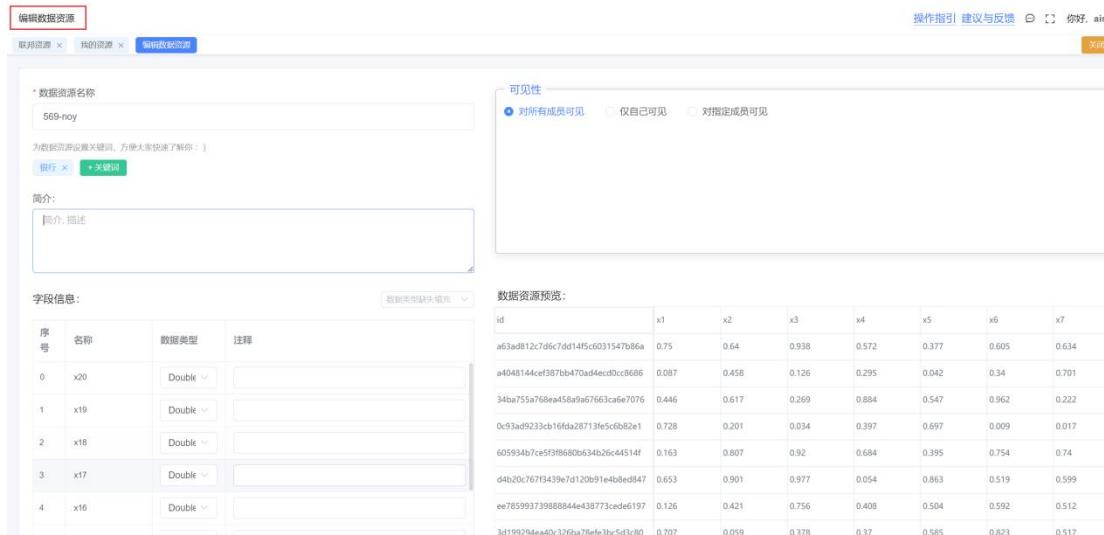
At the bottom, there are pagination controls: 共 1 条, 20条/页, <, 1, >, 前往, 1, 页.

3) 选中一条数据集，点击右侧的“编辑”按钮，进入“编辑数据源”界面，可对数据集名称、关键词、简介、字段信息、可见性等修改，点击“保存”提交。



This screenshot shows the 'Edit Data Source' interface. It has a search form with fields for 'ID', 'Name', 'Uploader', 'Keyword', 'Resource Type', and a 'Search' button. Below the search form is a table with columns: Name / Id, Keyword, Visibility, Resource Type, Data Information, Participation Project Count, Uploader, and Operation. Four entries are listed:

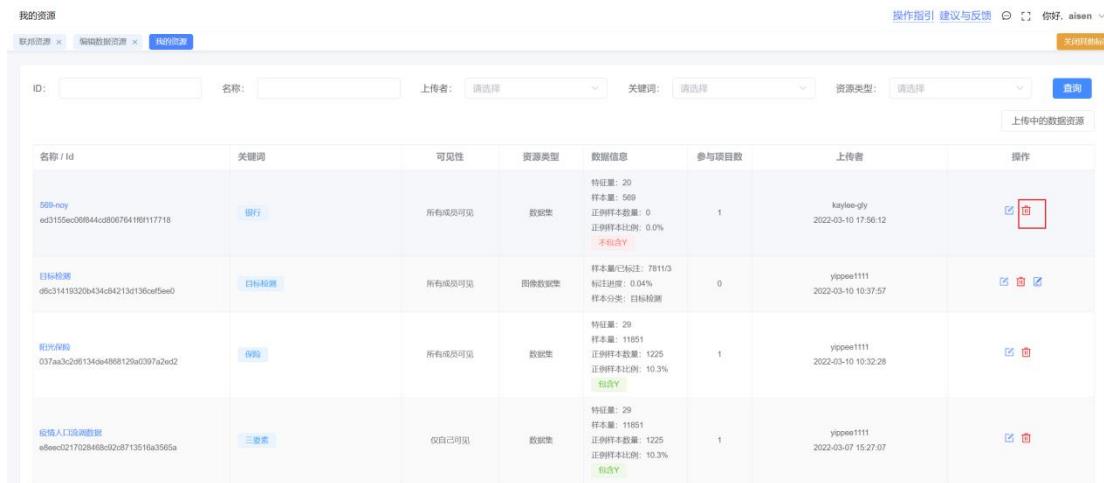
569-roy	ed3155ec06fb44cd80678410ff117718	所有成员可见	数据集	特征量: 20 样本量: 569 正例样本数量: 0 正例样本比例: 0.00% 不包含Y	1	kaylee-gly 2022-03-10 17:58:12	编辑, 查看
目标检测	d6c31419320b434c84213d136cef5ee0	所有成员可见	图像数据集	样本量/已标注: 7811/3 标注进度: 0.04% 样本分类: 目标检测	0	yippee1111 2022-03-10 10:37:57	编辑, 查看
阳光采购	037aa32e9f134ae4869129a0397a2e2	所有成员可见	数据集	特征量: 29 样本量: 11851 正例样本数量: 1225 正例样本比例: 10.3% 包含Y	1	yippee1111 2022-03-10 10:32:28	编辑, 查看
虚假人口洗钱数据	e6eedc217028468c92c8713516a395a	仅自己可见	数据集	特征量: 29 样本量: 11851 正例样本数量: 1225 正例样本比例: 10.3% 包含Y	1	yippee1111 2022-03-07 15:27:07	编辑, 查看



The screenshot shows the 'Edit Data Resource' interface. It includes fields for 'Data Resource Name' (569-noy), 'Description' (Bank), and a preview of the data table with columns x1 through x7.

ID	x1	x2	x3	x4	x5	x6	x7
a63ad812c7d6c7dd14f5c6031547b86a	0.75	0.64	0.938	0.572	0.377	0.605	0.634
a4048144ceef87b1b470a3d4ec3b0c8686	0.087	0.458	0.126	0.295	0.042	0.34	0.701
34ba755a768ea458a9a67663ca6e7076	0.446	0.617	0.269	0.884	0.547	0.962	0.222
0c93ad9233cb1fd428713fe5cf8b2e1	0.728	0.201	0.034	0.397	0.697	0.009	0.017
605934b7ceef3f68806543d026c44514f	0.163	0.807	0.92	0.684	0.395	0.754	0.74
d4b20c767f1439e7d120991e4b8ed947	0.653	0.901	0.977	0.054	0.863	0.519	0.599
ee785993739888844e43873cede197	0.126	0.421	0.756	0.408	0.504	0.592	0.512
3d199294ea40c3262a78e63bc5d3c80	0.707	0.059	0.378	0.37	0.585	0.823	0.517

4) 点击删除按钮，删除数据集，同时删除联邦数据集



The screenshot shows the 'My Resources' interface. It lists several data sets, including '569-noy' (Bank), '目标检测' (目标检测), '阳光保险' (阳光保险), and '疫情人口流动监测' (疫情人口流动监测). Each entry includes a 'Delete' button in the 'Operations' column.

Name / ID	关键词	可见性	资源类型	数据信息	参与项目数	上传者	操作
569-noy ed3155ec06844cd80676416ff117718	银行	所有成员可见	数据集	特征量: 20 样本量: 569 正例样本数量: 0 正例样本比例: 0.0% 不包含Y	1	kaylee-gjy 2022-03-10 17:56:12	<input checked="" type="checkbox"/> 删除
目标检测 d9c31419320b434d84213d136cef5ee0	目标检测	所有成员可见	图像数据集	样本量已标注: 7811/3 标记进度: 0.04% 样本分类: 目标检测	0	yippee1111 2022-03-10 10:37:57	<input checked="" type="checkbox"/> 删除
阳光保险 037aa3cd86134de4868129a0397a2ed2	保险	所有成员可见	数据集	特征量: 29 样本量: 11651 正例样本数量: 1225 正例样本比例: 10.3% 包含Y	1	yippee1111 2022-03-10 10:32:28	<input checked="" type="checkbox"/> 包含Y
疫情人口流动监测 e0beec0217024468cd2c8713516a03565a	三要素	仅自己可见	数据集	特征量: 29 样本量: 11651 正例样本数量: 1225 正例样本比例: 10.3% 包含Y	1	yippee1111 2022-03-07 15:27:07	<input checked="" type="checkbox"/> 包含Y

2. 上传中数据集

展示上传中的数据集详情。

操作:

点击“我的资源” — “上传中数据资源”，切换至上传中数据集 tab 页。

我的资源

操作指引 | 建议与反馈 | 你好, aisen | 关闭其他标签

ID:	名称:	上传者:	关键词:	资源类型:	操作
569-ray ed3155ec06fb44cd80676416ff1f7718	银行	所有成员可见	数据集	特征量: 20 样本量: 569 正例样本数量: 0 正例样本占比: 0.0% 不包含	1 kayjee-gly 2022-03-10 17:56:12
日标检测 d6c31419320b434c84213d136eaf5ed0	日标检测	所有成员可见	图像数据集	样本量: 7811/3 标注精度: 0.04% 样本分类: 日标检测	0 yippee1111 2022-03-10 10:37:57
人脸检测 037aa3c2d6134de4860129a0397a2e2	人脸	所有成员可见	数据集	特征量: 29 样本量: 11851 正例样本数量: 1225 正例样本占比: 10.3% 包含	1 yippee1111 2022-03-10 10:32:28
疫情人口监测数据 e6eec0217028486cd92d8713516a3585a	三要素	仅自己可见	数据集	特征量: 29 样本量: 11851 正例样本数量: 1225 正例样本占比: 10.3% 包含	1 yippee1111 2022-03-07 15:27:07

上传中的数据资源

名称 / Id	数据资源类型	上传进度	上传样本总量	已处理样本量	无效数据量	上传时间	耗时
b-n... 315... 后检... 2014... 化质... taa...	待处理	0%	特征量: 29	0	0	2022-03-10 10:37:57	00:00:00

哦豁，啥也没有~

3.5.2 添加资源

添加资源功能是平台主要的功能点，企业通过此功能把所有数据样本上传到平台上，点击左侧菜单栏“添加资源”，右侧展示页面，可以添加结构化数据，图形数据和联邦 sql 资源四种。

dev03

- 主页
- 资源中心
- 我的资源
- 添加资源**
- 联邦
- 合作中心
- 全局设置

结构化数据
用于机器学习解决二分类，多分类，回归等问题，挖掘数据中隐藏的模式，适用于信用评分、风控预测、广告推荐等场景。

联邦SQL资源
将数据源连接到联邦SQL查询引擎，使用SQL语句安全地对多方异构数据进行访问。

图像数据
定制图像分类，物体检测的AI模型，适用于图像内容检索定位、安防监控、工业质检等场景。

过滤器数据
将样本数据生成为二进制的布隆过滤器资源，生成后不可更改，如需变更需要重新上传。布隆过滤器资源只适用于样本对齐，无法参与训练模型。布隆过滤器资源仅用于RSA-PSI样本对齐算法。

3.5.2.1 添加结构化数据集

操作:

1) 点击“进入”上传数据操作，输入数据集名称、标签、简介。



注意：数据集名称输入至少为 4 个字符。

2) 选择文件

支持.csv .xls .xlsx 的文件；



支持三种上传格式：

(1) 选择数据集文件上传

添加数据集

主页 < 联邦数据集 < 联邦数据集详情 < 我的数据集 < 编辑数据集 < **添加数据集**

* 数据集名称:	data_01	简介:	data set 01																								
为数据集设置关键词, 方便大家快速了解你:)		<input checked="" type="checkbox"/> 信贷风控 <input type="checkbox"/> 政务 <input type="checkbox"/> 安防 <input type="checkbox"/> 保险 <input type="checkbox"/> 租车 <input type="checkbox"/> 广告推荐 <input type="checkbox"/> 教育 <input type="checkbox"/> 银行 <input type="checkbox"/> 医疗 按回车或 Tab 添加关键词 0/10																									
可见性:		<input checked="" type="radio"/> 对所有成员可见 <input type="radio"/> 仅自己可见 <input type="radio"/> 对指定成员可见																									
选择文件:		<input checked="" type="radio"/> 上传数据集文件 <input type="radio"/> 服务器本地文件 <input type="radio"/> 数据库 <div style="border: 1px solid #ccc; padding: 5px; margin-top: 5px;"> data_1.csv 57 KB 成功 X </div> <ul style="list-style-type: none"> • 主键字段必须是第一列, 并且会被自动 hash • 主键重复的数据会被自动去重, 仅保留第 1 条 • y 值列的列名必须为 y • csv 文件请使用 utf-8 编码格式 																									
数据集预览:																											
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">序号</th> <th style="text-align: left;">名称</th> <th style="text-align: left;">数据类型</th> <th colspan="4">注释</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>id</td> <td>Integer</td> <td colspan="4"></td> </tr> <tr> <td>1</td> <td>y</td> <td>Integer</td> <td colspan="4"></td> </tr> </tbody> </table>				序号	名称	数据类型	注释				0	id	Integer					1	y	Integer							
序号	名称	数据类型	注释																								
0	id	Integer																									
1	y	Integer																									
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">id</th> <th style="text-align: left;">y</th> <th style="text-align: left;">x0</th> <th style="text-align: left;">x1</th> <th style="text-align: left;">x2</th> <th style="text-align: left;">x3</th> </tr> </thead> <tbody> <tr> <td>133</td> <td>1</td> <td>0.254879</td> <td>-1.046633</td> <td>0.209656</td> <td>0.074214</td> </tr> <tr> <td>273</td> <td>1</td> <td>-1.142928</td> <td>-0.781198</td> <td>-1.166747</td> <td>-0.923578</td> </tr> <tr> <td>175</td> <td>1</td> <td>-1.451067</td> <td>-1.406518</td> <td>-1.456564</td> <td>-1.092337</td> </tr> </tbody> </table>				id	y	x0	x1	x2	x3	133	1	0.254879	-1.046633	0.209656	0.074214	273	1	-1.142928	-0.781198	-1.166747	-0.923578	175	1	-1.451067	-1.406518	-1.456564	-1.092337
id	y	x0	x1	x2	x3																						
133	1	0.254879	-1.046633	0.209656	0.074214																						
273	1	-1.142928	-0.781198	-1.166747	-0.923578																						
175	1	-1.451067	-1.406518	-1.456564	-1.092337																						

(2) 服务器本地文件上传

输入服务器地址后敲回车上传

选择文件

上传数据集文件 服务器本地文件 数据库

文件在服务器上的绝对路径

• 主键字段必须是第一列, 并且会被自动 hash
 • 主键重复的数据会被自动去重, 仅保留第 1 条
 • y 值列的列名必须为 y
 • csv 文件请使用 utf-8 编码格式

(3) 数据库上传

选择文件

上传数据集文件 服务器本地文件 数据库

数据源

请选择

查询语句

```
select * from table where hello = 'world'
```

**• 主键字段必须是第一列，并且会被自动 hash
• 主键重复的数据会被自动去重，仅保留第 1 条
• y 值列的列名必须为 y
• csv 文件请使用 utf-8 编码格式**

操作:

(1) 添加数据源，配置数据库地址

添加数据源

* 数据源名称

数据库类型

* IP

* 端口

* 库名

数据库用户名

数据库密码

选择数据源，输入 sql，点击"连接测试"按钮。

选择文件

上传数据资源文件 服务器本地文件 数据库

数据源
10.10.0.213

查询语句
select * from test

注意事项:

- 主键字段必须是第一列，并且会被自动 hash;
- 主键重复的数据会被自动去重，仅保留第 1 条;
- y 值列的列名必须为 y;
- csv 文件请使用 utf-8 编码格式。

注意事项:

- 主键字段必须是第一列，并且会被自动 hash;
- 主键重复的数据会被自动去重，仅保留第 1 条;
- y 值列的列名必须为 y;
- csv 文件请使用 utf-8 编码格式。

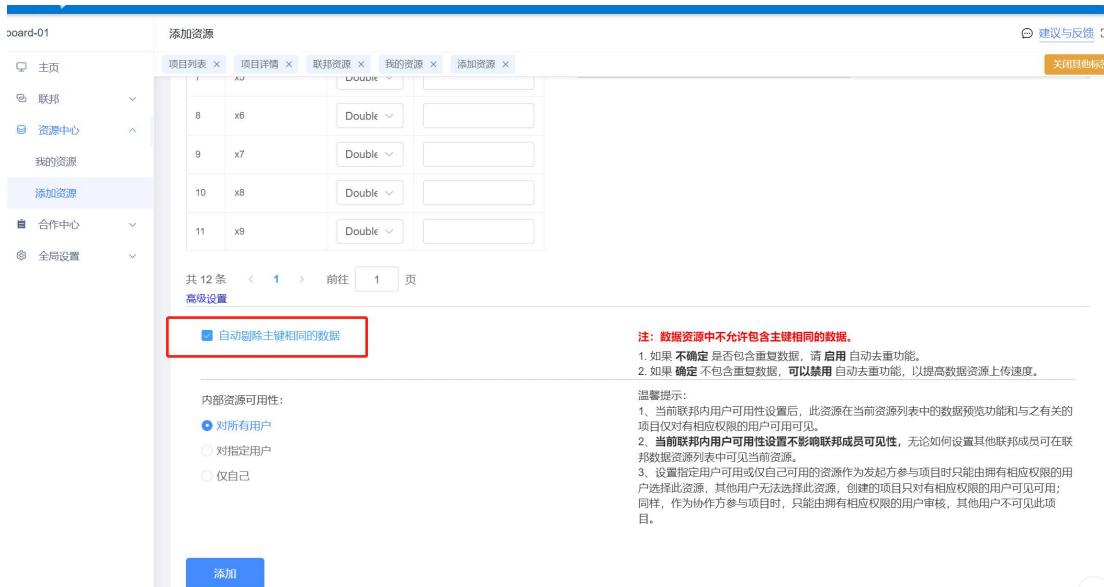
4) 修改字段信息

字段信息:

序号	名称	数据类型	注释
0	id	Intege	
1	y	Intege	
2	x0	Double	
3	x1	Double	
4	x2	Double	
5	x3	Double	

5) 自动剔除主键相同的数据

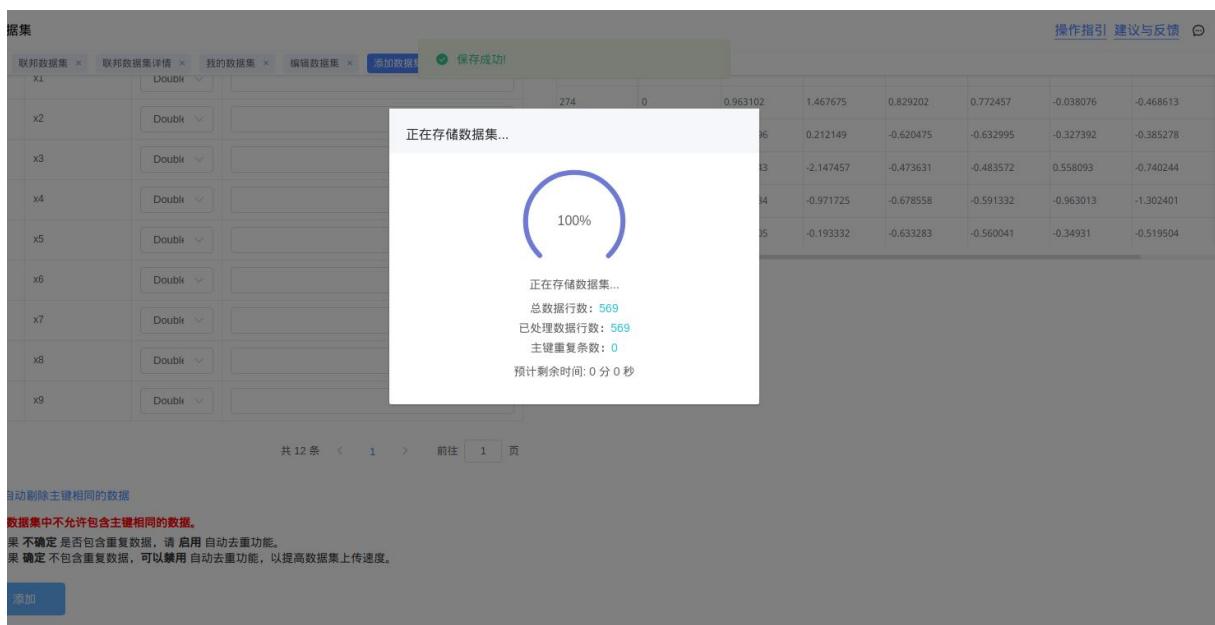
勾选此按钮会自动剔除主键重复的数据上传，否则不做去重操作上传。



The screenshot shows the 'Add Resource' interface. On the left sidebar, '添加资源' (Add Resource) is selected. In the main area, there is a table with 12 rows of data. Below the table, there is a checkbox labeled '自动删除主键相同的数据' (Automatically remove data with the same primary key). This checkbox is highlighted with a red rectangle. To the right of the checkbox, there is a note: '注：数据集中不允许包含主键相同的数据。' (Note: Data sets are not allowed to contain data with the same primary key.) and two sub-points: '1. 如果 不确定 是否包含重复数据，请 启用 自动去重功能。' (1. If uncertain about whether it contains duplicate data, please enable automatic deduplication function.) and '2. 如果 确定 不包含重复数据，可以禁用 自动去重功能，以提高数据资源上传速度。' (2. If it is determined that it does not contain duplicate data, you can disable the automatic deduplication function to improve upload speed.)

6) 添加

点击添加至上传进度页面，界面并在上传过程对数据进行校验分析，并会展示出上传进度及数据量等信息，上传成功跳转至数据详情页面。



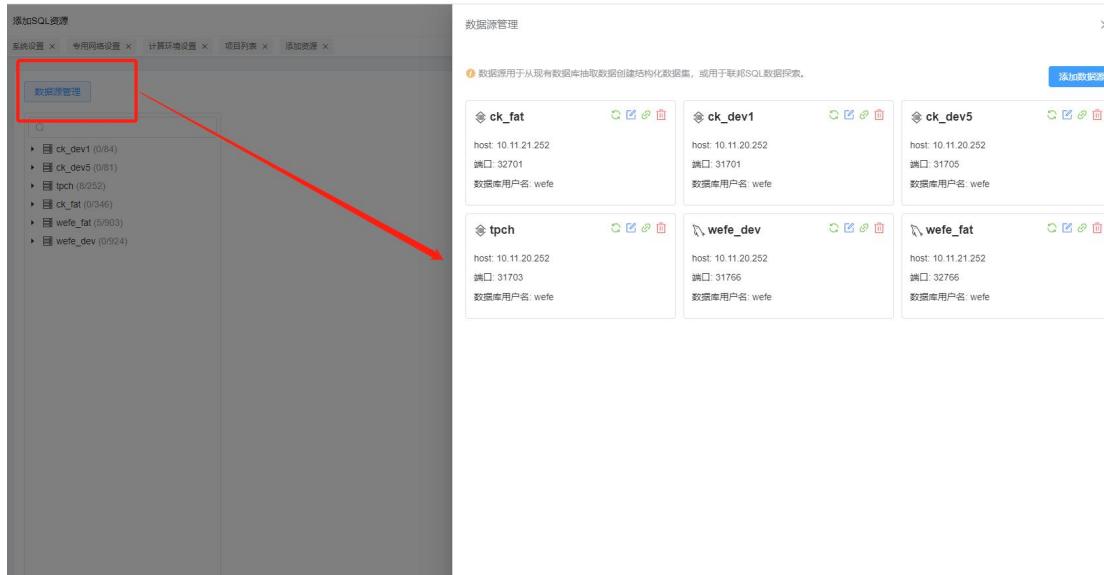
The screenshot shows the 'Upload Progress' page. At the top, a green success message box says '保存成功!' (Save successful!). Below it, there is a progress dialog box with a circular progress bar at 100%. The dialog also displays the following information: '正在存储数据集...' (Uploading dataset...), '总数据行数: 569' (Total data rows: 569), '已处理数据行数: 569' (Processed data rows: 569), '主键重复条数: 0' (Primary key duplicate count: 0), and '预计剩余时间: 0 分 0 秒' (Estimated remaining time: 0 minutes 0 seconds).

3.5.2.2 添加联邦 sql 资源

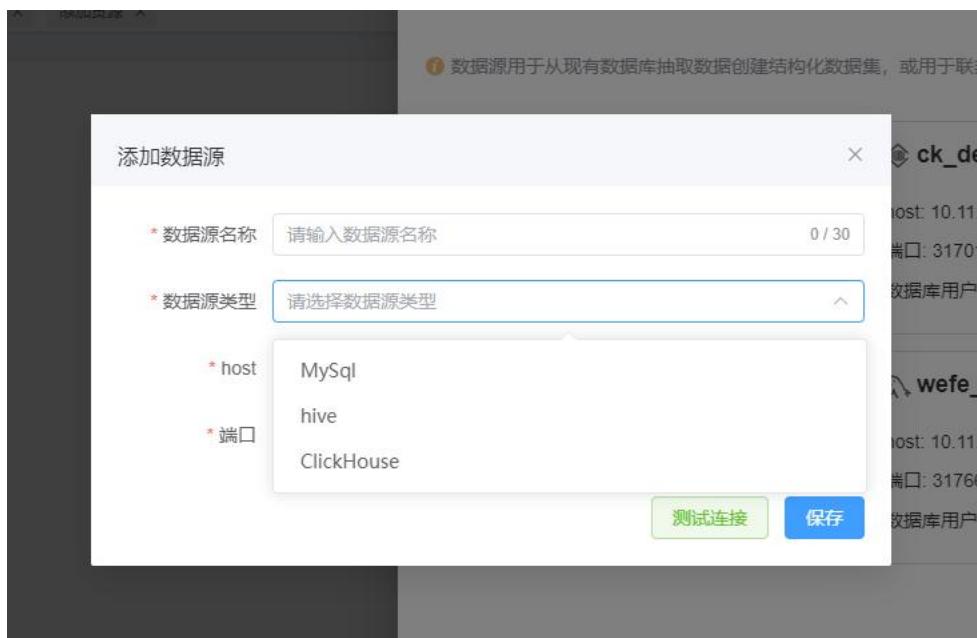
操作：

1) 数据源管理-添加数据源

点击“进入”进入添加sql资源页面，选择数据源



点击“添加数据源”，输入数据源名称，类型，host，端口，点击“保存”。



保存成功会自动拉取数据源的信息，展示库表，点击表名可查看表字段数据及预览信息。

数据源管理

- ▶ ck_dev1 (0/84)
- ▶ ck_dev5 (0/81)
- ▶ tpch (8/252)
- ▶ ck_fat (0/346)
- ▶ wefe_fat (5/903)
 - ▶ wefe_board_1 (1/86)
 - account
 - account_aaron
 - big_test_data
 - blacklist
 - bloom_filter
 - bloom_filter_column
 - bloom_filter_task
 - cert_info
 - cert_key_info
 - cert_request_info
 - chat_last_account
 - chat_unread_message
 - china_ums_account
 - cur_best_model
 - data_resource
 - data_resource_upload_t...
 - data_set
 - data_set_column

数据源管理

wefe_fat.wefe_board_1.big_test_data 的信息

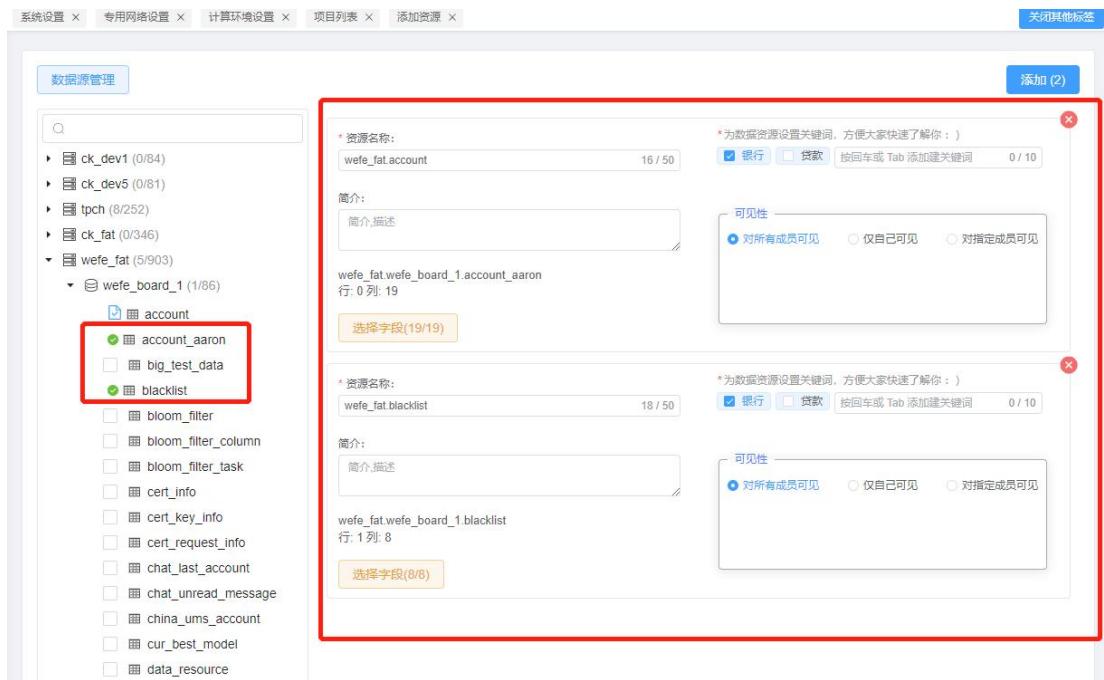
X

路径: wefe_fat.wefe_board_1.big_test_data
 数据量: 327680000 rows * 3 columns
 描述:

字段列表		数据预览		
字段名称	数据类型	id	user_id	create_time
id	integer	1	93e86fb5-a140-11ed-913c-fa163e465038	1675153424000
user_id	varchar(36)	2	93e8e644-a140-11ed-913c-fa163e465038	1675153424000
create_time	timestamp	3	93e94d62-a140-11ed-913c-fa163e465038	1675153424000
		4	93e9d63f-a140-11ed-913c-fa163e465038	1675153424000
		5	93ea327a-a140-11ed-913c-fa163e465038	1675153424000
		6	93eaf64f-a140-11ed-913c-fa163e465038	1675153424000
		7	93ebc807-a140-11ed-913c-fa163e465038	1675153425000
		8	93ec9303-a140-11ed-913c-fa163e465038	1675153425000
		9	93ecfa9c-a140-11ed-913c-fa163e465038	1675153425000
		10	93ed82d9-a140-11ed-913c-fa163e465038	1675153425000
		11	93edeccce-a140-11ed-913c-fa163e465038	1675153425000
		12	93ee5580-a140-11ed-913c-fa163e465038	1675153425000
		13	93eef38a-a140-11ed-913c-fa163e465038	1675153425000
		14	93ef116a-a140-11ed-913c-fa163e465038	1675153425000
		15	93ef712f-a140-11ed-913c-fa163e465038	1675153425000

2) 添加联邦 sql 资源

勾选要添加的表，编辑数据名称，设置表标签，编辑选择字段，点击添加数据集。



The screenshot shows the 'Add Resource' interface. On the left, there's a sidebar with project names like 'ck_dev1', 'ck_dev5', 'tpch', 'ck_fat', and 'wefe_fat'. Under 'wefe_fat', there's a tree view with 'account' expanded, showing 'account_aaron', 'big_test_data', and 'blacklist'. The 'blacklist' node is highlighted with a red box. On the right, two resource configuration panels are shown side-by-side. Both panels have 'wefe_fat.account' as the name and '银行' (Bank) as the label. The top panel has 'wefe_fat.wefe_board_1.account_aaron' selected under '选择字段(19/19)'. The bottom panel has 'wefe_fat.wefe_board_1.blacklist' selected under '选择字段(8/8)'. Both panels have sections for '简介' (Description), '可见性' (Visibility), and a note about keywords.

添加成功跳转到我的资源列表。



The screenshot shows the 'My Resources' list page. On the left, there's a sidebar with '主页', '资源中心', '我的资源' (selected), '添加资源', '联邦', '合作中心', and '全局设置'. The main area displays a table of resources. Two rows are highlighted with a red box: 'wefe_fat.blacklist' (ID: 59f4438) and 'wefe_fat.account' (ID: 7770a89). Both rows show '银行' as the label, '所有成员可见' (Visible to all members) as visibility, and 'SQL搜索' (SQL search) as the type. The table includes columns for '名称 / ID', '关键词', '可见性', '资源类型', '数据信息', '分类数', '参与项目数', '上传者', and '操作'. A note at the top right says '※ 上传中的数据资源'.

点击数据集名称跳转数据资源简介页面，展示数据量信息，字段信息，数据预览信息，以及参与的项目。

查看数据资源

我的资源 × 项目列表 × 项目详情 × SQL搜索 × 查看数据资源

version: 3.7.0 你好, admin ☰ 建议与反馈 ☰ 关闭其他标签

同步资源元数据

数据资源简介

wefe_fat.blacklist [编辑](#)

admin 上传于 2023-06-20 16:41:24, 在 0 个合作项目中, 参与了 0 次任务。

关键字:	银行
可见性:	对所有成员可见
数据源:	wefe_fat
数据库:	wefe_board_1
表名:	blacklist
数据量:	1 rows * 8 columns

数据资源信息

字段列表	数据预览	参与项目
字段名称	字段类型	注释
updated_time	timestamp	更新时间
member_id	varchar(32)	成员id
id	varchar(32)	全局唯一标识
created_time	timestamp	创建时间
blacklist_member_id	varchar(32)	被加入黑名单成员id
updated_by	varchar(32)	更新人

若数据表有字段及数据更新，可以同步数据源数据，同步操作如下：

➤ 同步资源元数据

在数据简介页面点击右上角“同步资源元数据”按钮。

dev03

主页

资源中心

我的资源

添加资源

联邦

合作中心

全局设置

数据资源简介

wefe_fat.blacklist [编辑](#)

admin 上传于 2023-06-20 16:41:24, 在 0 个合作项目中, 参与了 0 次任务。

关键字:	银行
可见性:	对所有成员可见
数据源:	wefe_fat
数据库:	wefe_board_1
表名:	blacklist
数据量:	1 rows * 8 columns

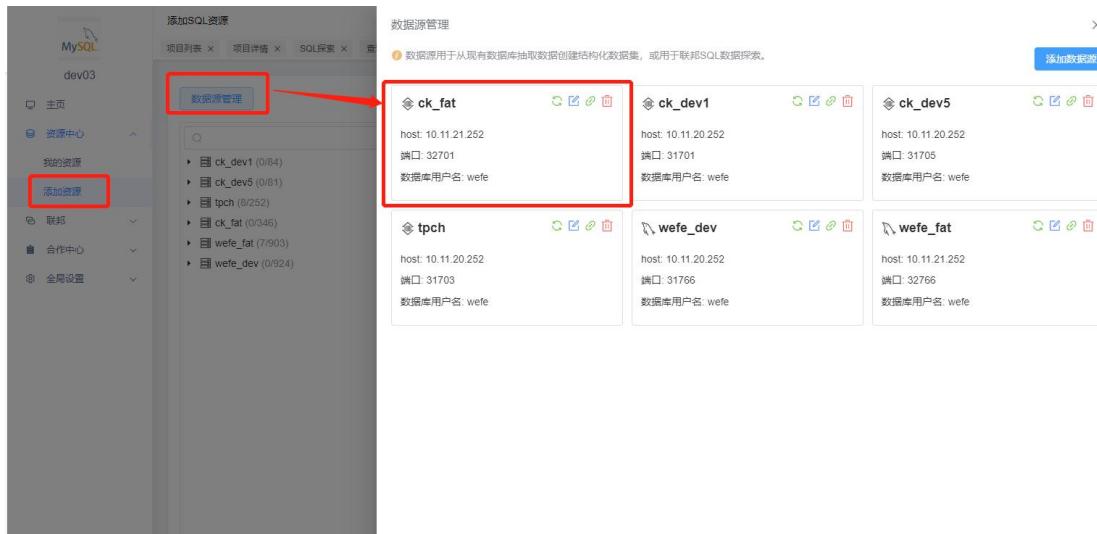
数据资源信息

字段列表	数据预览	参与项目
字段名称	字段类型	注释
updated_time	timestamp	更新时间
member_id	varchar(32)	成员id

同步资源元数据

➤ 刷新数据源

添加资源页面，点击“数据源管理”，弹出数据源管理面板，点击资源板块右上角的刷新按钮，可以刷新源数据信息。



3.5.2.3 添加布隆过滤器

操作:

- 1) 点击“进入”上传数据操作，输入数据集名称、标签、简介。



注意：数据集名称输入至少为 4 个字符。

- 2) 选择文件

布隆过滤器文件格式支持.csv .xls .xlsx 的表文件;



支持三种上传格式：

(1) 选择数据集文件上传

id	y	x0	x1	x2	x3	x4	x5
0	0	17.99	17.77	132.9	1326	0.08474	
1	1	20.57	17.77	132.9	1326	0.08474	18018790522
1	0	20.57	17.77	132.9	1326	0.08474	0.07864
2	0	19.69	21.25	130	1203		0.1599
3	0	11.42	20.38	77.58	386.1		0.2839
4	0	20.29	14.34	135.1	1297	0.1003	0.1328
5	0	12.45	15.7	82.57	477.1	0.1278	0.17
6	0	18.25	19.98	119.6	1040	0.09463	0.109

(2) 服务器本地文件上传

输入服务器地址后敲回车上传

文件在服务器上的绝对路径
• csv 文件请使用 utf-8 编码格式

(3) 数据库上传

选择文件

上传数据集文件 服务器本地文件 数据库

数据源

请选择

查询语句

```
select * from table where hello = 'world'
```

**• 主键字段必须是第一列，并且会被自动 hash
• 主键重复的数据会被自动去重，仅保留第 1 条
• y 值列的列名必须为 y
• csv 文件请使用 utf-8 编码格式**

操作:

(1) 添加数据源，配置数据库地址

添加数据源

* 数据源名称

数据库类型

* IP

* 端口

* 库名

数据库用户名

数据库密码

选择数据源，输入 sql，点击"连接测试"按钮。



注意事项:

- csv 文件请使用 utf-8 编码格式。

4) 设置主键 hash 方式

配置设置

内部资源可用性:

对所有用户
 对指定用户
 仅自己

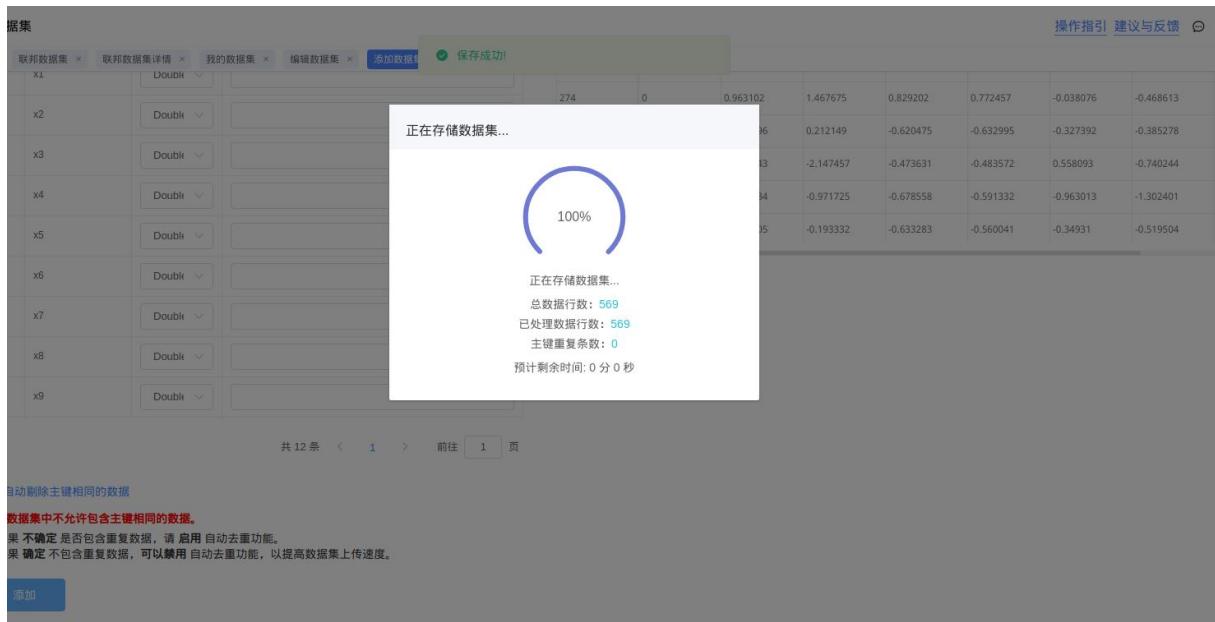
MD5
SHA1
SHA256
无

设置主键 hash 方式 (上传后不可更改):

x0:
x1:

6) 添加

点击添加至上传进度页面，界面并在上传过程对数据进行校验分析，并会展示出上传进度及数据量等信息，上传成功跳转至数据详情页面。



3.5.2.4 添加图像数据

根据图像资源类型分为图像分类和目标检测

图像分类的上传成功后需手动标注图片；

目标检测的根据上传的资源识别已标注图片。

* 数据资源名称：
0 / 30
* 为数据资源设置关键词，方便大家快速了解你：)
按回车或 Tab 键添加关键词 0 / 10

简介：
简介, 描述

* 数据资源类型：
 图像分类 目标检测

联邦资源可见性

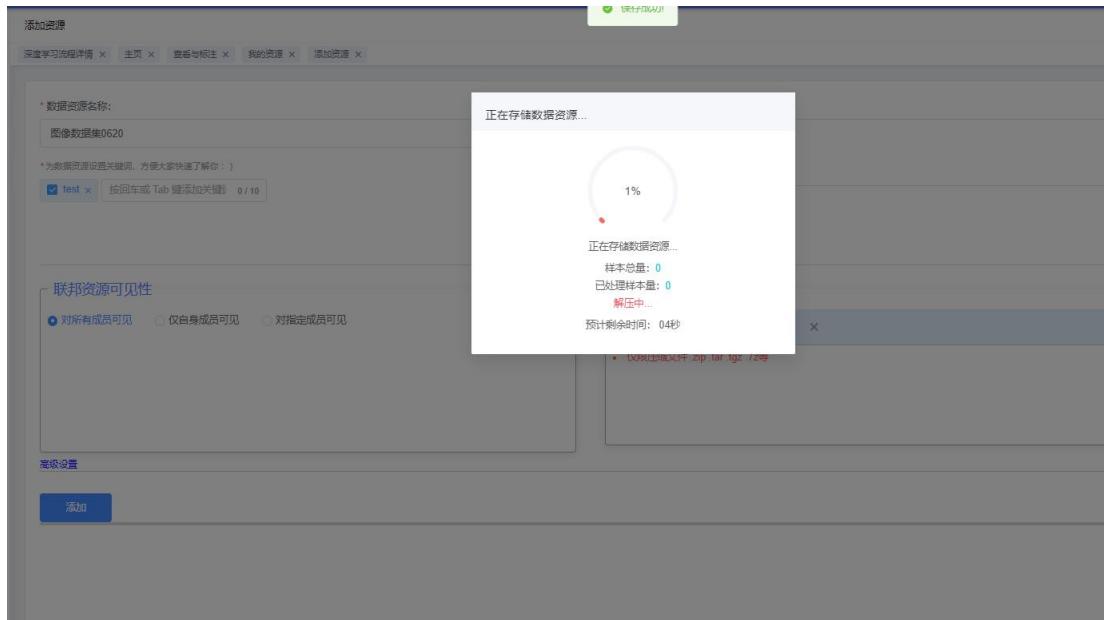
对所有成员可见 仅自身成员可见 对指定成员可见

选择文件

将文件 (.zip .tar .tgz .7z) 拖到此处
或 点击上传
• 仅限压缩文件 .zip .tar .tgz .7z 等

高级设置

点击添加，展示图片上传进度：

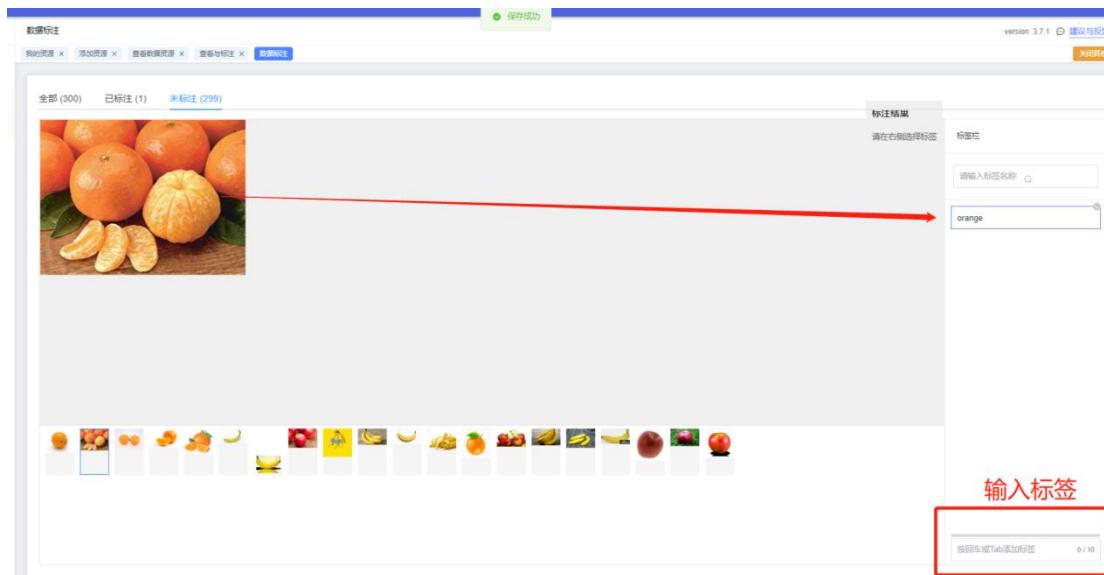


资源添加成功后可在我的资源列表查看，点击按钮插卡按与修改标注信息：

This screenshot shows the 'My Resources' list page. The sidebar on the left includes sections for '联邦', '资源中心', '我的资源' (selected), '添加资源', '合作中心', and '全局设置'. The main area displays a table of resources. Two rows are highlighted with red boxes around their edit icons (pencil and delete symbols) in the '操作' (Operation) column. The first row is for '数据标注测试 可使用' (Data Annotation Test Available) with ID '19d7cc592f224aa2867d0' and the second for 'mmmm 可使用' (mmmm Available) with ID 'c45a4455decc450c9009'. Both rows show details like '仅自身成员可见' (Visible to self-members) or '所有成员可见' (Visible to all members), '图像数据集' (Image Dataset), and file sizes.

This screenshot shows the 'Data Annotation' page for dataset '0620'. The top navigation bar includes '我的资源' (My Resources), '标注管理' (Annotation Management), '标注任务' (Annotation Tasks), and '标注历史' (Annotation History). The main content area shows a summary for 'kaylee' who uploaded the dataset on 2023-06-20 at 14:29:31. It lists '关键词' (Keywords: test), '标注状态' (Status: 进行中 - In Progress), '数据大小' (Data Size: 24.75M), and '标注进度' (Annotation Progress: 0.00%). A prominent red box highlights the '去标注' (Go Annotation) button. Below this, a section titled '数据资源信息' (Data Resource Information) shows a grid of 12 fruit images, each labeled '未标注' (Unlabeled). The entire page has a light gray background with dark blue header and button styling.

输入标签，选择图片标注标签。



异常情况:

- 如出现不符合条件的数据时，系统会给出提示信息并中断上传动作。

3.5.2.5 设置成员可见性

添加数据集设置可见性，此设置决定于成员方上传的数据集是否被其他成员可见。

1) 对所有成员可见：



添加数据集

主要 > 联邦数据集 > 联邦数据集详情 > 我的数据集 > 编辑数据集 > **添加数据集**

* 数据集名称：
data_1 6/30

为数据集设置关键词，方便大家快速了解你：)
 信贷风控 政务 安防 保险 租车 广告推荐 教育 银行
 医疗 按回车或 Tab 添加键入的标签 0/10

可见性
 对所有成员可见 仅自己可见 对指定成员可见

选择文件
 上传数据集文件 服务器本地文件 数据库
 将文件 (.csv .xls .xlsx) 拖到此处
 或 [点击上传](#)

将文件 (.csv .xls .xlsx) 拖到此处
 或 [点击上传](#)

• 主键字段必须是第一列，并且会被自动 hash
 • 主键重复的数据会被自动去重，仅保留第 1 条
 • y 值列的列名必须为 y
 • csv 文件请使用 utf-8 编码格式

自动删除主键相同的数据
注：数据集中不允许包含主键相同的数据。
 1. 如果 不确定 是否包含重复数据，请 启用 自动去重功能。
 2. 如果 确定 不包含重复数据，可以禁用 自动去重功能，以提高数据集上传速度。

[添加](#)

则添加的数据集可对所有联邦成员可见。

2) 设置仅自己可见:

则添加的数据集只有自己可见，其他联邦成员方不可见。

添加数据集

主页 < 联邦数据集 < 联邦数据集详情 < 我的数据集 < 编辑数据集 < 添加数据集

* 数据集名称: data_1 6/30	简介: date set 01
为数据集设置关键词，方便大家快速了解你：) <input checked="" type="checkbox"/> 信贷风控 <input type="checkbox"/> 政务 <input type="checkbox"/> 安防 <input type="checkbox"/> 保险 <input type="checkbox"/> 租车 <input type="checkbox"/> 广告推荐 <input type="checkbox"/> 教育 <input type="checkbox"/> 银行 <input type="checkbox"/> 医疗 按回车或 Tab 添加键关键词 0/10	
可见性 <input type="radio"/> 对所有成员可见 <input checked="" type="radio"/> 仅自己可见 <input type="radio"/> 对指定成员可见	选择文件 <input checked="" type="radio"/> 上传数据集文件 <input type="radio"/> 服务器本地文件 <input type="radio"/> 数据库 将文件 (.csv .xls .xlsx) 拖到此处 或 <input type="button" value="点击上传"/> <ul style="list-style-type: none"> • 主键字段必须是第一列，并且会被自动 hash • 主键重复的数据会被自动去重，仅保留第 1 条 • y 值列的列名必须为 y • csv 文件请使用 utf-8 编码格式
<small>注：数据集中不允许包含主键相同的数据。 1. 如果 不确定 是否包含重复数据，请 启用 自动去重功能。 2. 如果 确定 不包含重复数据，可以禁用 自动去重功能，以提高数据集上传速度。</small>	
<input type="button" value="添加"/>	

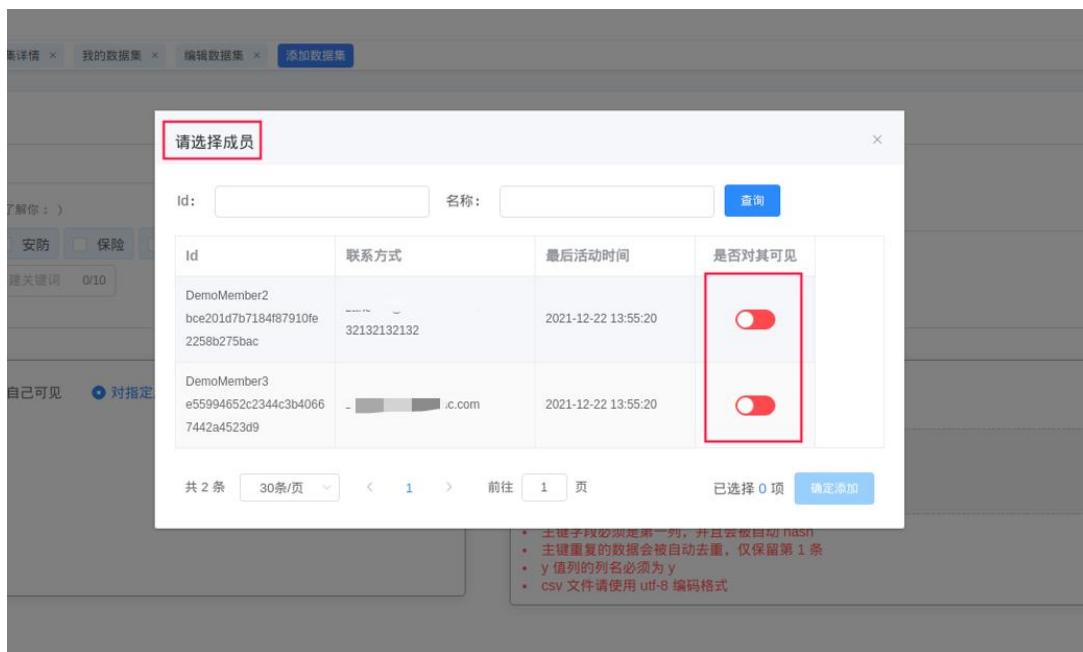
3) 对指定成员可见:

则对指定添加的成员可见其数据集，其他未添加的联邦成员不可见。

添加数据集

主页 < 联邦数据集 < 联邦数据集详情 < 我的数据集 < 编辑数据集 < 添加数据集

* 数据集名称: data_1 6/30	简介: date set 01
为数据集设置关键词，方便大家快速了解你：) <input checked="" type="checkbox"/> 信贷风控 <input type="checkbox"/> 政务 <input type="checkbox"/> 安防 <input type="checkbox"/> 保险 <input type="checkbox"/> 租车 <input type="checkbox"/> 广告推荐 <input type="checkbox"/> 教育 <input type="checkbox"/> 银行 <input type="checkbox"/> 医疗 按回车或 Tab 添加键关键词 0/10	
可见性 <input type="radio"/> 对所有成员可见 <input type="radio"/> 仅自己可见 <input checked="" type="radio"/> 对指定成员可见	选择文件 <input checked="" type="radio"/> 上传数据集文件 <input type="radio"/> 服务器本地文件 <input type="radio"/> 数据库 将文件 (.csv .xls .xlsx) 拖到此处 或 <input type="button" value="点击上传"/> <ul style="list-style-type: none"> • 主键字段必须是第一列，并且会被自动 hash • 主键重复的数据会被自动去重，仅保留第 1 条 • y 值列的列名必须为 y • csv 文件请使用 utf-8 编码格式
<small>注：数据集中不允许包含主键相同的数据。 1. 如果 不确定 是否包含重复数据，请 启用 自动去重功能。 2. 如果 确定 不包含重复数据，可以禁用 自动去重功能，以提高数据集上传速度。</small>	
<input type="button" value="添加"/>	



注意事项:

➤ 全局设置为公开

全局设置-成员信息页面，设置为对外公开数据集，则具体数据集可见性依据添加数据集时“可见性设置”操控。



➤ 全局设置为不公开

全局设置-成员信息页面，设置为不对外公开数据集，则即使添加数据集设置可见性为所有成员，成员添加的数据集在联邦数据集中也不可见。



* 成员名称:

在联邦中隐身:

是 否

是否允许对外公开数据集基础信息:

是 否

※ 其他成员目前不能查看到您的任何数据集

邮箱:

3.5.2.6 设置数据资源使用权限



主页

联邦

资源中心

我的资源

添加资源

合作中心

全局设置

联邦资源可见性

对所有成员可见 仅自身成员可见 对指定成员可见

选择文件

上传数据资源文件 服务器本地文件 数据库

将文件 (.csv .xls .xlsx) 拖到此处
或

- 主键字段必须是第一列
- 主键重复的数据会被自动去重，仅保留第 1 条
- y 值列的列名必须为 y
- csv 文件请使用 utf-8 编码格式

高级设置

自动剔除主键相同的数据

注: 数据资源不允许包含主键相同的数据。

1. 如果 不确定 是否包含重复数据，请 启用 自动去重功能。
2. 如果 确定 不包含重复数据，可以禁用 自动去重功能，以提高数据资源上传速度。

内部资源可用性:

对所有用户 对指定用户 仅自己

温馨提示:

- 当前联邦内用户可用性设置后，此资源在当前资源列表中的数据预览功能和与之有关的项目仅对有相应权限的用户可用可见。
- 当前联邦内用户可用性设置不影响联邦成员可见性，无论如何设置其他联邦成员可在联邦数据资源列表中可见当前资源。
- 设置指定用户可用或仅自己可用的资源作为发起方参与项目时只能由拥有相应权限的用户选择此资源，其他用户无法选择此资源，创建的项目只对有相应权限的用户可见可用；同样，作为协作方参与项目时，只能由拥有相应权限的用户审核，其他用户不可见此项目。

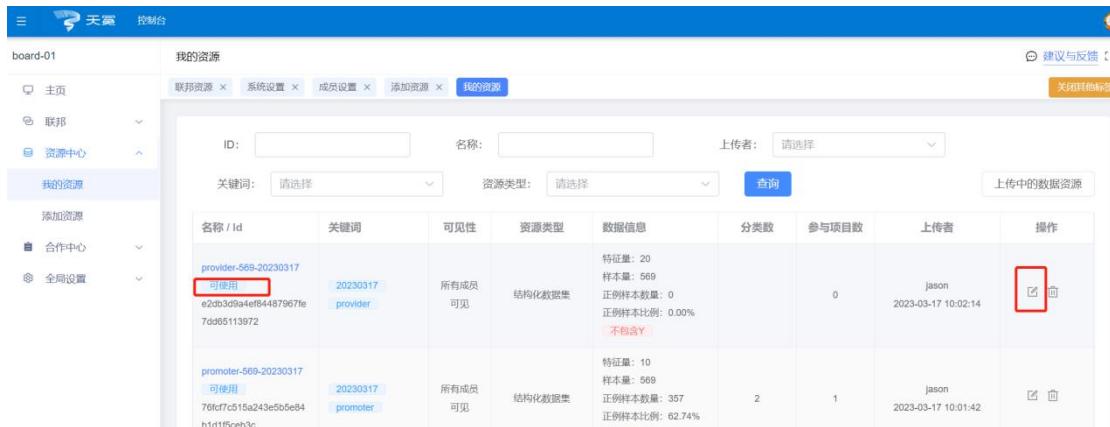
添加

设置该数据资源是否可被当前成员下其他用户所使用的权限。

1) 对所有用户可见

默认为此选项，勾选了此选项的数据资源所有用户都有使用权限。

(1) 该成员下所有用户：有使用权，标签为【可使用】包括在项目中/流程中使用此数据集资源，没有管理权限，不能编辑此数据集资源详情；



名称 / ID	关键词	可见性	资源类型	数据信息	分分类	参与项目数	上传者	操作	
provider-569-20230317 [可用]	20230317 provider	所有成员可见	结构化数据集	特征量: 20 样本量: 569 正例样本数量: 0 正例样本比例: 0.00% 包含Y	0	jason	2023-03-17 10:02:14		
promoter-569-20230317 [可使用]	20230317 promoter	所有成员可见	结构化数据集	特征量: 10 样本量: 569 正例样本数量: 357 正例样本比例: 62.74%	2	1	jason	2023-03-17 10:01:42	

(2) 创建者有使用权及管理权限：标签展示为【我拥有】，可编辑此数据资源，及在项目中/流程中使用此数据集；



名称 / ID	关键词	可见性	资源类型	数据信息	分分类	参与项目数	上传者	操作	
44444 我拥有 3d74b48ae1e45aeb2c0 51ac119ec5b2	银行	所有成员可见	结构化数据集	特征量: 10 样本量: 569 正例样本数量: 357 正例样本比例: 62.74% 包含Y	2	0	kaylee-fat	2023-03-14 10:44:03	
kaylee-fat 上传-指定用户 我拥有 af585a9f9a9142c5a441db 227506ac2b	银行	所有成员可见	结构化数据集	特征量: 10 样本量: 569 正例样本数量: 357 正例样本比例: 62.74% 包含Y	2	1	kaylee-fat	2023-03-13 15:25:23	
500-y-0310 可使用 13557e0259924c1a876c 19c34ce4aba	银行	所有成员可见	结构化数据集	特征量: 20 样本量: 500 正例样本数量: 258 正例样本比例: 51.60% 包含Y	2	0	kaylee	2023-03-10 17:26:03	

(3) 管理员有使用权及管理权限：标签展示为【可使用】，可编辑此数据资源，及在项目中/流程中使用此数据集；

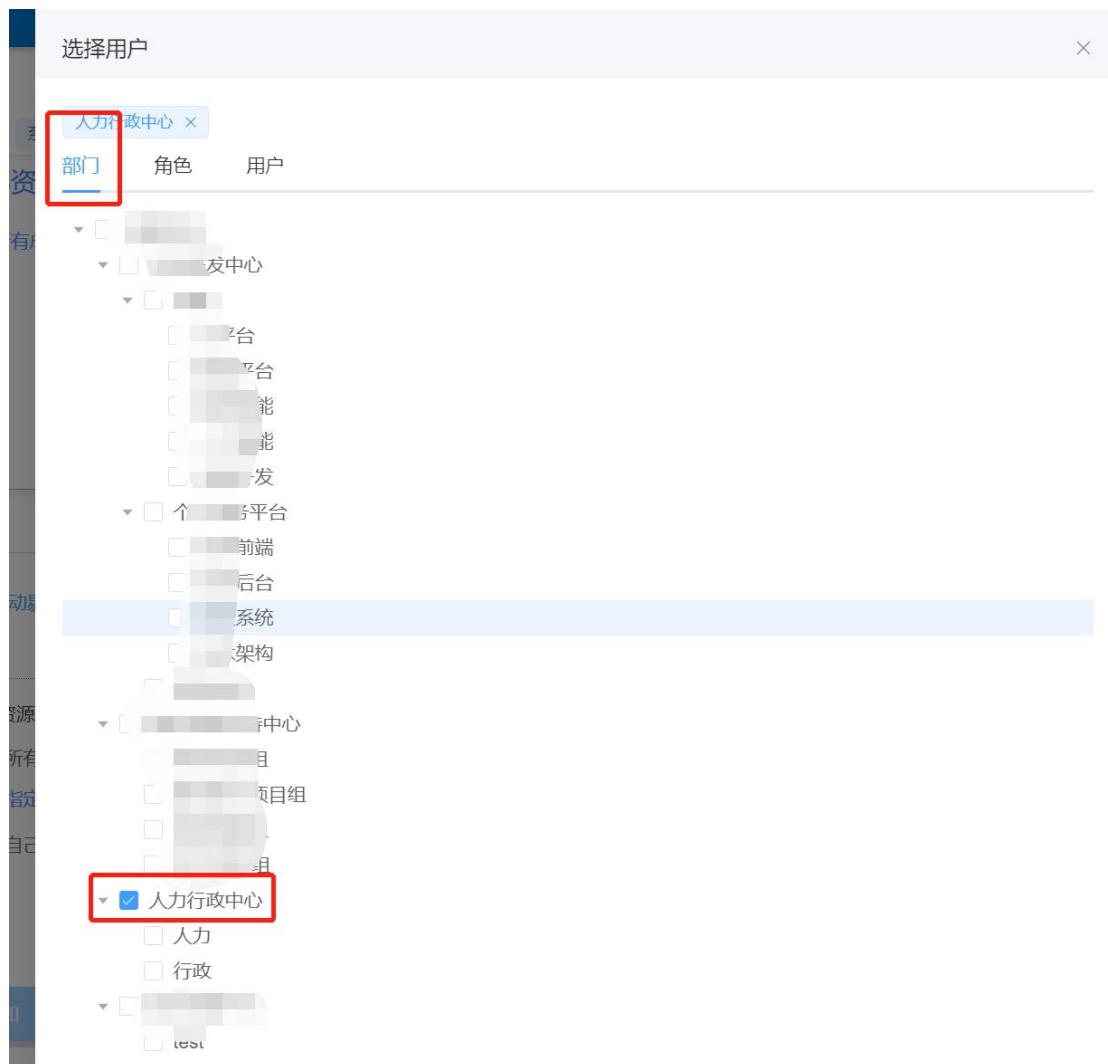


名称 / ID	关键词	可见性	资源类型	数据信息	分分类	参与项目数	上传者	操作
provider-569-20230317 [可用]	20230317 provider	所有成员可见	结构化数据集	特征量: 20 样本量: 569 正例样本数量: 0 正例样本比例: 0.00% 包含Y	0	jason	2023-03-17 10:02:14	

2) 对指定用户可见

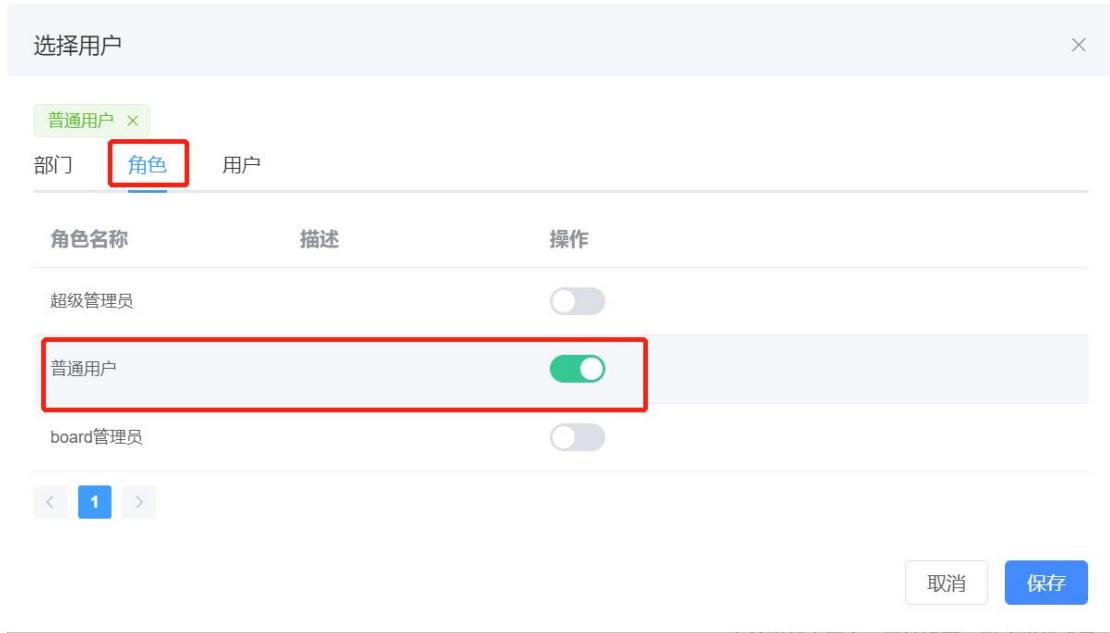
➤ 指定部门

数据资源权限指定部门可见，则这个部门下的所有用户都有这个资源的使用权限，标签为【可使用】，包括在项目中/流程中使用此数据集资源。



➤ 指定角色

数据资源权限指定部门可见，则这个角色下关联的所有用户都有这个资源的使用权限，标签为【可使用】，包括在项目中/流程中使用此数据集资源。



➤ 指定用户

3) 仅自己可见

指定仅自己可见，则该数据集只有创建者用户可使用。

[高级设置](#)

自动剔除主键相同的数据

注:

1. 如
2. 如

内部资源可用性:

温馨提示:
1、
项目
2、
邦数
3、
户选
同样
□

- 对所有用户
 对指定用户
 仅自己

其他用户（非超级管理员）：标签为【无权限】，没有使用数据集权限。

请选择数据资源

上传者:	请选择	名称:	kaylee	ID:		
资源类型:	结构化数据集	+ 1	查询			
名称 / Id	资源类型	关键词	数据信息	参与任务次数	上传者	选择数据资源
kaylee-fat上传 无权限 496425c3ec3a4c539cf09aa61946ca dc	结构化数据集	银行	特征量: 10 样本量: 569 正例样本数量: 357 正例样本比例: 62.7% 包含Y	0	kaylee-fat 2023-03-09 16:40:40	 
kaylee01添加资源 可使用 270753d8ec394b8faf9fc64a3971d48 39	结构化数据集	银行	特征量: 20 样本量: 500 正例样本数量: 258 正例样本比例: 51.6% 包含Y	0	kaylee01 2022-12-30 10:20:00	 
kaylee_poc8405758 可使用 b127f45e11c64348aeada49b371b6 0f2	结构化数据集	test	特征量: 10 样本量: 500 不包含Y	0	kaylee-gly 2022-09-07 15:20:00	 

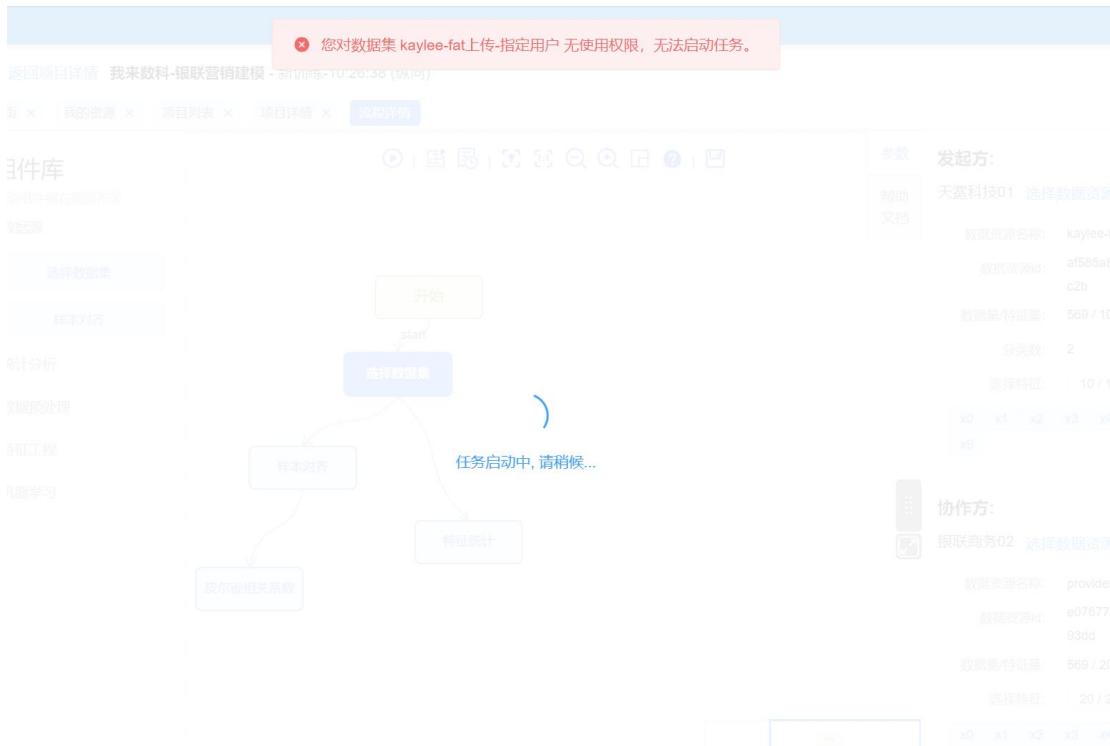
无权限者，不能在项目中添加这个资源，不能在流程中使用这个资源，如下图：

智能联合建模平台

项目详情

请选择数据资源

上传者:	请选择	名称:	ID:			
资源类型:	结构化数据集	+ 1	查询			
名称 / Id	资源类型	关键词	数据信息	参与任务次数	上传者	选择数据资源
kaylee-fat上传 无权限 496425c3ec3a4c539cf09aa61946ca dc	结构化数据集	银行	特征量: 10 样本量: 569 正例样本数量: 357 正例样本比例: 62.7% 包含Y	0	kaylee-fat 2023-03-09 16:58:40	 
特征统计分析测试 可使用 9c7c96ecd87034e8092ec6d1c3ecc5 c8	结构化数据集	特征统计分析	特征量: 10 样本量: 569 正例样本数量: 357 正例样本比例: 62.7% 包含Y	0	jason 2023-03-08 10:12:44	 
5000-III 无权限 5f0108bc745841f6a08becc6152e33 d8	结构化数据集	银行	特征量: 10 样本量: 569 正例样本数量: 357 正例样本比例: 62.7%	2	kaylee-fat 2023-03-07 15:52:53	 



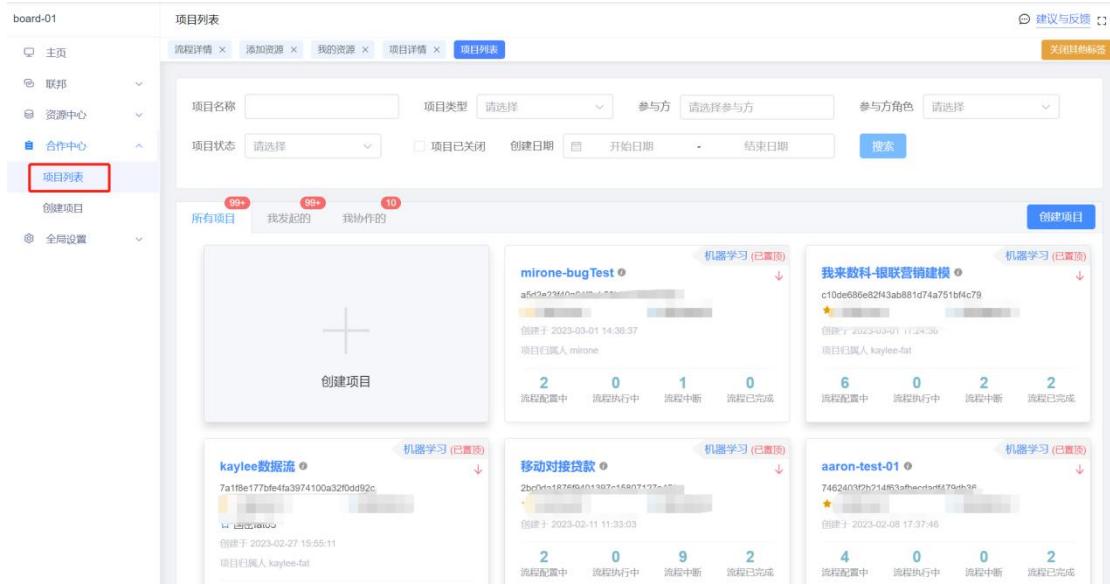
3.6 合作中心

合作中心模块是提供创建项目合作与协作方进行联合建模的功能模块, 支持机器学习建模和深度学习两类联合建模模式。

3.6.1 项目列表

项目列表功能模块是存放发起方与协作方创建的合作项目, 包括: 所有项目、我发起的、我协作的; 每个项目将以名片样式展示项目信息:

- **项目基本信息:** 项目名称、项目 ID、所有成员方角色;
- **项目流程状态信息:** 流程配置数量、流程执行中数量、流程已完成数量, 如需查看项目详细信息、使用操作可以直接点击项目名称进入。



The screenshot shows the 'Project List' page under the 'board-01' board. The left sidebar includes '联邦' (Federation), '合作中心' (Collaboration Center), and '全局设置' (Global Settings). The '项目列表' (Project List) tab is selected. The main area displays a grid of projects:

- mironne-bugTest**: 机器学习 (已置顶) - Created by mironne on 2023-03-01.
- 我来数据-报联营销建模**: 机器学习 (已置顶) - Created by kylee-fat on 2023-03-01.
- kaylee数据流**: 机器学习 (已置顶) - Created by kylee-fat on 2023-02-27.
- 移动对接贷款**: 机器学习 (已置顶) - Created by mironne on 2023-02-11.
- aaron-test-01**: 机器学习 (已置顶) - Created by aaron on 2023-02-08.

Each project card shows metrics: 流程配置中 (Workflow Configuration), 流程执行中 (Workflow Execution), 流程中断 (Workflow Interruption), and 流程已完成 (Workflow Completed).

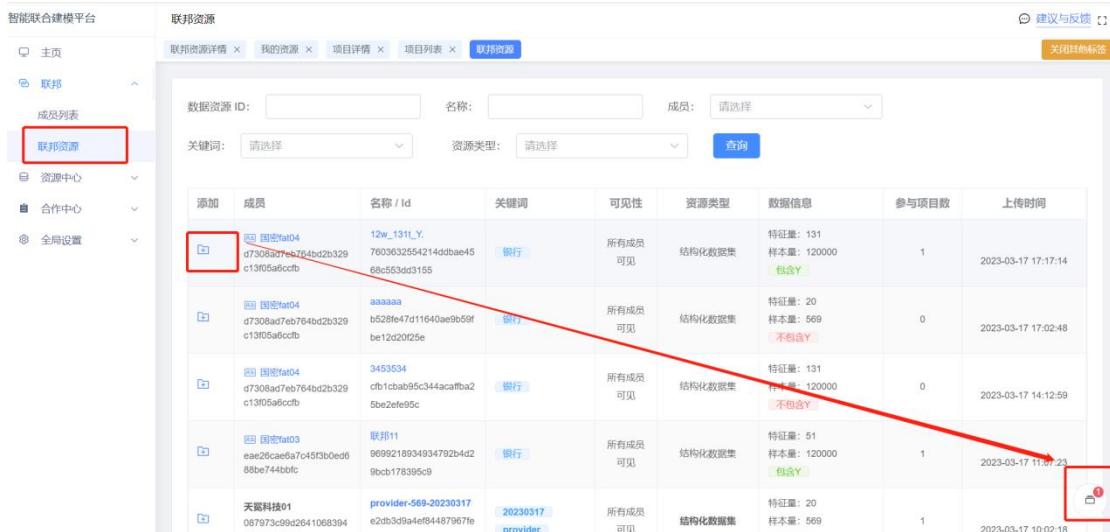
3.6.2 创建机器学习项目

3.6.2.1 初始化创建项目

支持两种创建方式：快捷创建项目，普通创建项两种方法。

1) 快捷创建项目

在联邦的联邦数据集模块页面点击左侧的  快速创建项目：



The screenshot shows the 'Federation Data Set' module under the 'Federation' section. The '联邦资源' (Federation Resources) tab is selected. A red box highlights the '快速创建' (Quick Create) button at the bottom right of the table. The table lists data sets with columns: 添加 (Add), 成员 (Member), 名称 / Id (Name / ID), 关键词 (Keywords), 可见性 (Visibility), 资源类型 (Resource Type), 数据信息 (Data Information), 参与项目数 (Number of Projects), and 上传时间 (Upload Time). One row is selected, showing details for '国密fat04'.

添加	成员	名称 / Id	关键词	可见性	资源类型	数据信息	参与项目数	上传时间
	国密fat04 d7308ad7eb764bd2b329 c13f05a6ccfb	12w_131L_Y. 7603632554214ddbae45 68c553d3d3155		所有成员可见	结构化数据集	特征量: 131 样本量: 120000 包含Y	1	2023-03-17 17:17:14
	国密fat04 d7308ad7eb764bd2b329 c13f05a6ccfb	aaaaaa b528fb47d11640ae9b59f be12d2025e		所有成员可见	结构化数据集	特征量: 20 样本量: 569 不包含Y	0	2023-03-17 17:02:48
	国密fat04 d7308ad7eb764bd2b329 c13f05a6ccfb	3453534 cf1bab95c344acaffba2 5be2fe95c		所有成员可见	结构化数据集	特征量: 131 样本量: 120000 不包含Y	0	2023-03-17 14:12:59
	国密fat03 ea26cae6a7c4f3b0ed6 88be744b0fc	联邦11 9699218934934797b4d2 9bcb178395c9		所有成员可见	结构化数据集	特征量: 51 样本量: 120000 包含Y	1	2023-03-17 11:07:23
	天器科技01 087973c99a2641068394	provider-569-20230317 e20b3d9a4ef84487967fe	20230317 provider	所有成员可见	结构化数据集	特征量: 20 样本量: 569	1	2023-03-17 10:02:18

创建者默认为发起方，选择数据集的其他成员为协作方。

联机资源

项目列表 < 快捷新建

数据资源 ID:	名称:	成员:	关键词:	资源类型:	数据信息	参与项目数
添加 成员	名称 / ID	关键词	可见性	资源类型	数据信息	参与项目数
 fa03 81051c7b7f4beaad581ba048fbaf048	43d846043cf542aa8badefb17567483	银行	指派或可见	布隆过滤器	样本量: 500 主键组合方式:	0
 fa01 087973cd99d26410683944bf346c8635	569-noy	银行	所有成员可见	结构化数据集	特征量: 20 样本量: 569 未包含Y	0
 fa01 087973cd99d26410683944bf346c8635	目标检测	目标检测	所有成员可见	图像数据集	特征量: 7811/3 标注量: 0.04% 样本分类: 目标检测	0
 fa01 087973cd99d26410683944bf346c8635	行人检测	行人	所有成员可见	结构化数据集	特征量: 29 样本量: 11851 包含Y	0
 fa01 727c030d68674eb680a0e08ba5395fa	新冠疫情数据	三要素	所有成员可见	结构化数据集	特征量: 20 样本量: 11850 未包含Y	0
 fa02 727c030d68674eb680a0e08ba5395fa	数据集简介_深圳福田区疫情防控人口流动数据	保险	所有成员可见	结构化数据集	特征量: 29 样本量: 11850 未包含Y	0
 fa01 087973cd99d26410683944bf346c8635	深圳市福田区疫情防控人口流动数据	三要素	所有成员可见	结构化数据集	特征量: 29 样本量: 11851 包含Y	0
 fa04 a94c7a7118c445a5a293a4ed266887	图像数据	图像分类	所有成员可见	图像数据集	样本量已标注: 3009 标注进度: 0.00% 样本分类: 图像分类	0
 fa03 81051c7b7f4beaad581ba048fbaf048	图像数据	图像分类	所有成员可见	图像数据集	样本量已标注: 3005 标注进度: 1.67% 样本分类: 图像分类	0

返回

点击创建生成快捷创建项目，跳转项目详情页面。

2) 普通创建项目

在合作中心的项目列表中创建项目：

智能联合建模平台

- 主页
- 联邦
- 资源中心
- 合作中心
- 项目列表
- 创建项目
- 全局设置

项目列表

联邦资源详情 < 我的资源 < 项目详情 < 联邦资源 < 项目列表

项目名称	项目类型	请选择
项目状态	请选择	<input type="checkbox"/> 项目已关闭
99+ 所有项目		
99+ 我发起的		
10 我协作的		
+ 创建项目		

进入创建项目菜单添加项目，填项目名称、项目类型和视觉处理，建模平台支持机器学习和视觉处理，选择项目类型：

创建项目

联邦资源 × 项目列表 × **创建项目**

1 发起项目

* 项目名称
0 / 40

* 项目类型
机器学习

机器学习
视觉处理

2 发起方 ① + 添加更多发起方

fat01 服务正常

+ 添加资源到此项目

3 协作方 + 添加更多协作方

保存项目

只有己方成员时可进行本地建模



添加参与方和协作方，可以在创建项目时添加多个参与方和多个协作方：

2 发起方 ① + 添加更多发起方

DemoMember1 服务正常

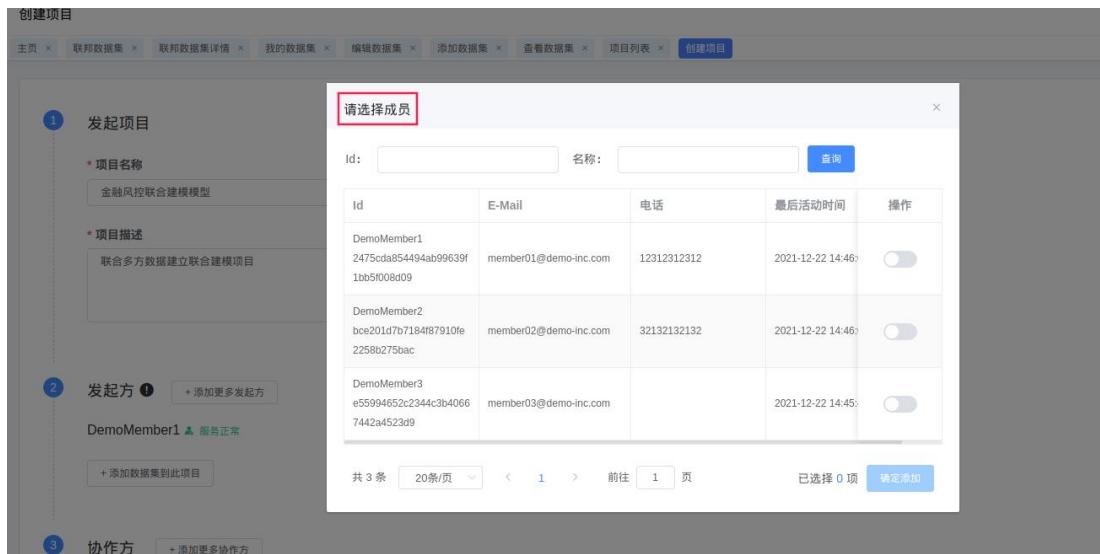
+ 添加数据集到此项目

3 协作方 + 添加更多协作方

保存项目

只有己方成员时可进行本地建模



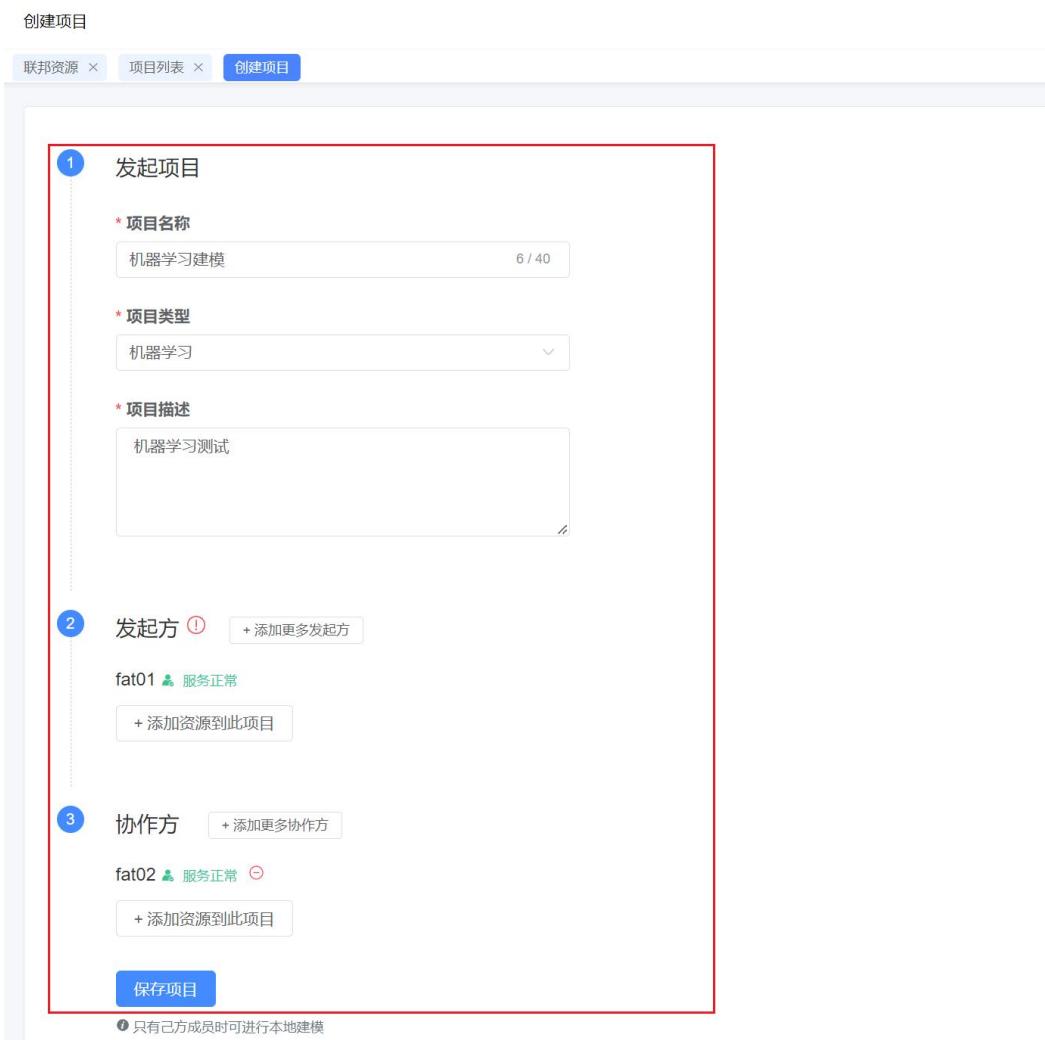


The screenshot shows the 'Create Project' interface. In the main area, there are three sections: 1. Initiator (发起项目) with fields for project name (金融风控联合建模项目) and description (联合多方数据建立联合建模项目). 2. Initiator Party (发起方) listing 'DemoMember1' as active. 3. Collaborator Party (协作方) with a '+ Add more' button. A modal window titled '请选择成员' (Select Member) is open, displaying a table of users:

ID	E-Mail	电话	最后活动时间	操作
DemoMember1 2475cda854494ab99639f 1bb5f008d09	member01@demo-inc.com	12312312312	2021-12-22 14:46	<input checked="" type="checkbox"/>
DemoMember2 bcc201d7b7184f87910fe 2258b275bac	member02@demo-inc.com	32132132132	2021-12-22 14:46	<input checked="" type="checkbox"/>
DemoMember3 e55994652c2344c3b4066 7442a4523d9	member03@demo-inc.com		2021-12-22 14:45	<input checked="" type="checkbox"/>

Modal controls include 'Search' (查询), pagination (共 3 条, 20条/页), and a 'Confirm Selection' (确定添加) button.

协作方添加完成，保存项目，创建项目成功后系统直接跳转至项目详情页面：



The screenshot shows the 'Create Project' interface after saving. A red box highlights the 'Initiator' section (步骤 1) of the form. The initiator information is filled out: project name '机器学习建模', project type '机器学习', and description '机器学习测试'. Below this, the 'Initiator Party' (步骤 2) lists 'fat01' as active, and the 'Collaborator Party' (步骤 3) lists 'fat02' as active. At the bottom is a large blue 'Save Project' (保存项目) button. A note at the bottom left states: '只有己方成员时可进行本地建模' (Only local modeling can be performed when only your own members are present).

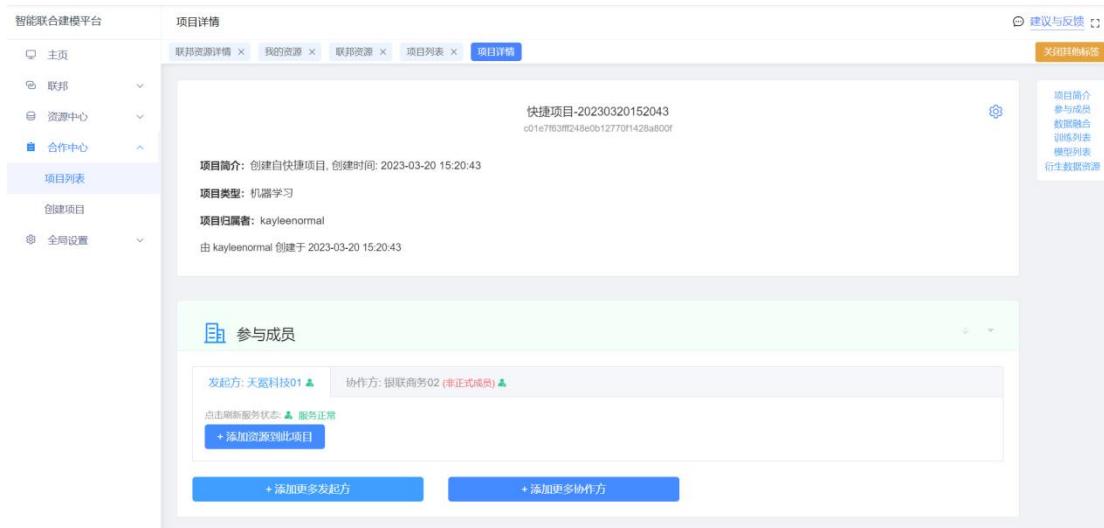
3.6.2.2 项目详情页

项目详情分为以下几个模块：

项目简介：创建项目时输入的项目介绍，若采用快捷方式创建项目其简介会自动生成。

项目类型：创建项目时选择的项目类型。

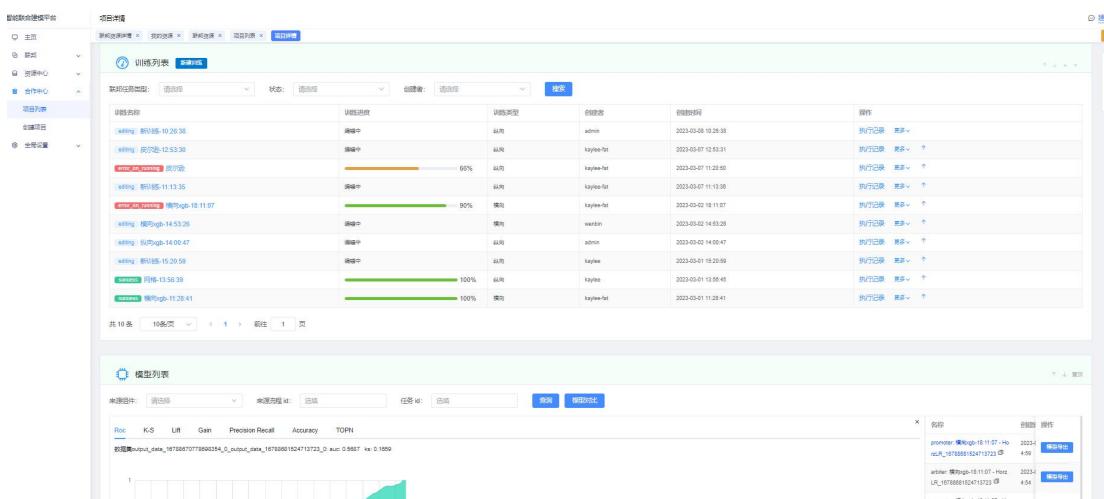
参与成员：创建项目时选择的所有参与成员以及参与成员的状态信息。初始创建的项目成员，只需要自己审核，否则非正式成员需要其他正式成员审核后方可加入项目。



The screenshot shows the 'Project Details' page. At the top, there's a navigation bar with tabs: '联邦资源详情', '我的资源', '联邦资源', '项目列表', and '项目详情'. The right side has a sidebar with links like '项目简介', '参与成员', '数据融合', '训练列表', '模型列表', and '衍生数据资源'. The main content area displays project details: '快捷项目-20230320152043' and its creation time '2023-03-20 15:20:43'. It also shows the project type as '机器学习' and the owner as 'kayleenormal'. Below this, the 'Participants' section lists '发起方: 天禹科技01' and '协作方: 银联商务02 (非正式成员)'. There are buttons to '+添加资源到此项目', '+添加更多发起方', and '+添加更多协作方'.

流程列表：历史建模训练记录。新增的训练流程。只有发起方可以编辑，创建者只能删除自己创建的流程。

模型列表：各算法组件生成的模型列表，可以查看模型的评估结果。



The screenshot shows the 'Model List' page. The top navigation bar includes '项目管理', '训练列表', and '模型列表'. The 'Training Log' section displays a table of training processes:

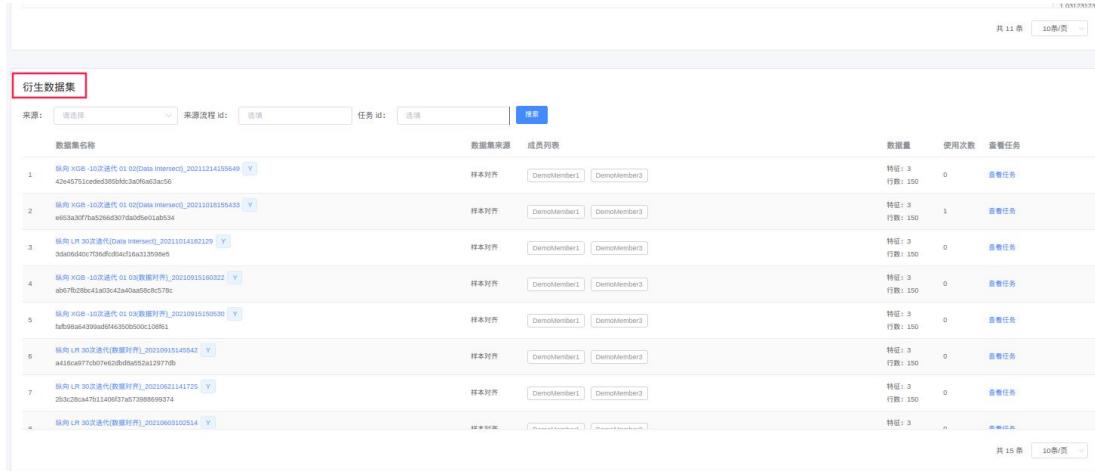
训练任务名	训练进度	训练类型	创建者	创建时间	操作
wleng_预测-10.26.35	完成	公用	kaylee-fst	2023-03-08 12:53:31	执行已终 正常 -
wleng_预测-12.53.30	完成	公用	kaylee-fst	2023-03-07 11:23:50	执行已终 正常 -
wleng_预测-10.07.07	65%	公用	kaylee-fst	2023-03-07 11:11:36	执行已终 正常 -
wleng_预测-11.11.07	90%	演示	kaylee-fst	2023-03-02 18:11:07	执行已终 正常 -
wleng_预测-14.53.26	演示	公用	weibin	2023-03-02 14:53:26	执行已终 正常 -
wleng_预测-14.09.47	演示	公用	admin	2023-03-02 14:09:47	执行已终 正常 -
wleng_预测-15.28.09	完成	公用	kaylee	2023-03-01 13:28:09	执行已终 正常 -
wleng_预测-13.58.39	100%	公用	kaylee	2023-03-01 13:58:45	执行已终 正常 -
wleng_预测-11.28.41	100%	演示	kaylee-fst	2023-03-01 11:28:41	执行已终 正常 -

The bottom section shows the 'Model List' with a table of models and their evaluation metrics:

模型组件	模型ID	任务ID	状态	评估指标	评估值		
Rec	K-S	Lift	Gain	Precision	Recall	Accuracy	TOPN
数据集: output_data_10738057775804254_0_output_data_1073805824713723_0 acc: 0.9887 vs 0.9895							

A detailed view of a specific model row is shown on the right, displaying its name, status, and evaluation results.

衍生数据集: 流程执行过程中生成的衍生数据集, 如样本对齐组件生成衍生数据集, 衍生数据集可直接用于其他流程做数据集。



The screenshot shows a list of derived data sets. Each entry includes a preview icon, a unique ID, a creation date, and a detailed view button. The columns are: '衍生数据集' (Derived Data Set), '数据集名称' (Dataset Name), '数据集来源' (Data Source), '成员列表' (Member List), '数据量' (Data Volume), '使用次数' (Usage Times), and '查看任务' (View Task). The data volume is consistently listed as 3 rows and 150 rows.

衍生数据集	数据集名称	数据集来源	成员列表	数据量	使用次数	查看任务
1 私用 XQB - 10次迭代_01_02[Data intersect]_20211214155640	42bf45751eeded3954cc5a0f6d3ac56	样本对齐	DemoMember1, DemoMember3	特征: 3 行数: 150	0	查看任务
2 私用 XQB - 10次迭代_01_02[Data intersect]_20211018155433	e6b3aa3cf7ba52e6d307d4d0fe1ab534	样本对齐	DemoMember1, DemoMember3	特征: 3 行数: 150	1	查看任务
3 私用 LR 30次迭代[Data intersect]_20210104148219	3dd46460c779d6bd4d4a13a3199865	样本对齐	DemoMember1, DemoMember3	特征: 3 行数: 150	0	查看任务
4 私用 XQB - 10次迭代_01_02[数据对齐]_20210915160322	a607fb28cc01d10c4340a50a58cf87bc	样本对齐	DemoMember1, DemoMember3	特征: 3 行数: 150	0	查看任务
5 私用 XQB - 10次迭代_01_02[数据对齐]_20210915150520	tafb98aa439ad9465050600c10861	样本对齐	DemoMember1, DemoMember3	特征: 3 行数: 150	0	查看任务
6 私用 LR 30次迭代[数据对齐]_20210915145042	a416ca0777007402bd8d552a2977db	样本对齐	DemoMember1, DemoMember3	特征: 3 行数: 150	0	查看任务
7 私用 LR 30次迭代[数据对齐]_20210903114175	2b5d28ca4101140ff74d739969969374	样本对齐	DemoMember1, DemoMember3	特征: 3 行数: 150	0	查看任务
... 私用 LR 30次迭代[数据对齐]_20210803102514		样本对齐	DemoMember1, DemoMember3	特征: 3 行数: 150	6	查看任务

3.6.2.3 成员审核

1、初始创建

发起方第一次创建项目邀请的协作方, 只需要协作方成员自己审核通过后加入项目。

操作:

- 1) 发起方创建项目菜单输入项目名称, 描述, 邀请协作方成员。

创建项目

联邦资源 < 项目列表 < 创建项目

1 发起项目

* 项目名称
机器学习建模 6 / 40

* 项目类型
机器学习

* 项目描述
机器学习测试

2 发起方 ① + 添加更多发起方

fat01 ▲ 服务正常
+ 添加资源到此项目

3 协作方 + 添加更多协作方

fat02 ▲ 服务正常
+ 添加资源到此项目

保存项目

● 只有己方成员时可进行本地建模

2) 协作方 02 成员自己审核通过加入项目，02 加入后此时项目含有 01, 02 两个正式成员。

系统设置 < 主页 < 流程详情 < 项目列表 < 项目详情

快捷项目-20230320152043
c01e7f63fff248e0b12770f1428a800f

项目简介: 创建自快捷项目, 创建时间: 2023-03-20 15:20:43

项目类型: 机器学习

创建于 2023-03-20 15:20:43

参与合作审核意见:

同意 **拒绝**

参与成员

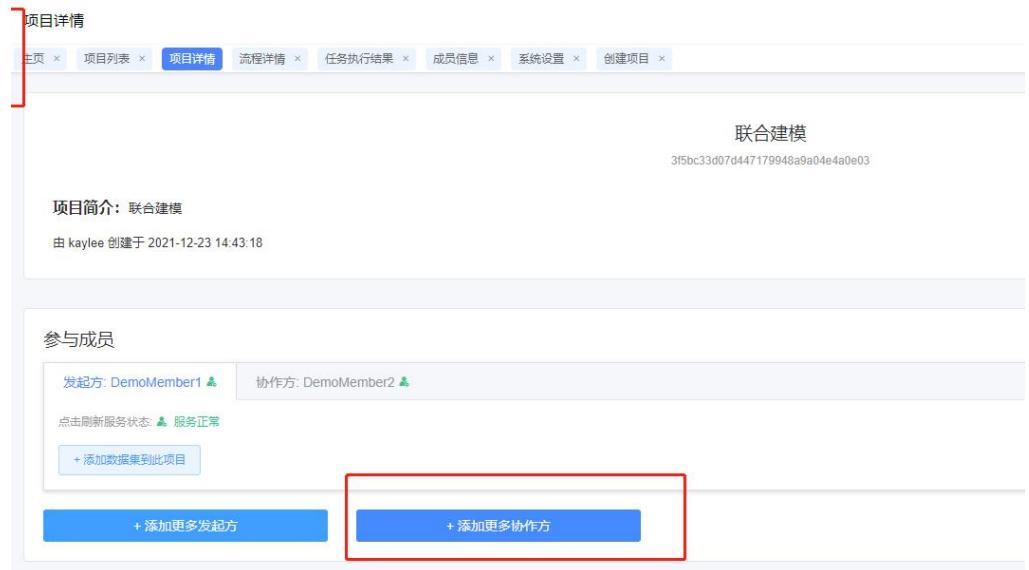
发起方: 天冕科技01 <small>▲ 服务正常</small>	协作方: 银联商务02 (非正式成员) <small>▲ 服务正常</small>
点击刷新服务状态: <small>▲ 服务正常</small>	
<small>非正式成员, 待审核</small>	

序号	数据资源	关键词	数据量	使用次数	是否包含 Y	状态	操作
1	(待审核) provider-569-20230317 可使用 65a6cfe0ede14486a176c00e249b4f3d	20230317 provider	特征量: 20 样本量: 569	3	否	等待审核	请先同意授权加入合作或 

3) 发起方邀请新成员加入项目

项目创建后，由发起方邀请成员进入项目；

项目详情页面，添加更多协作方-选择成员-点击确认。



项目详情

主页 × 项目列表 × 项目详情 × 流程详情 × 任务执行结果 × 成员信息 × 系统设置 × 创建项目 ×

联合建模
3f5bc33d07d447179948a9a04e4a0e03

项目简介：联合建模
由 kaylee 创建于 2021-12-23 14:43:18

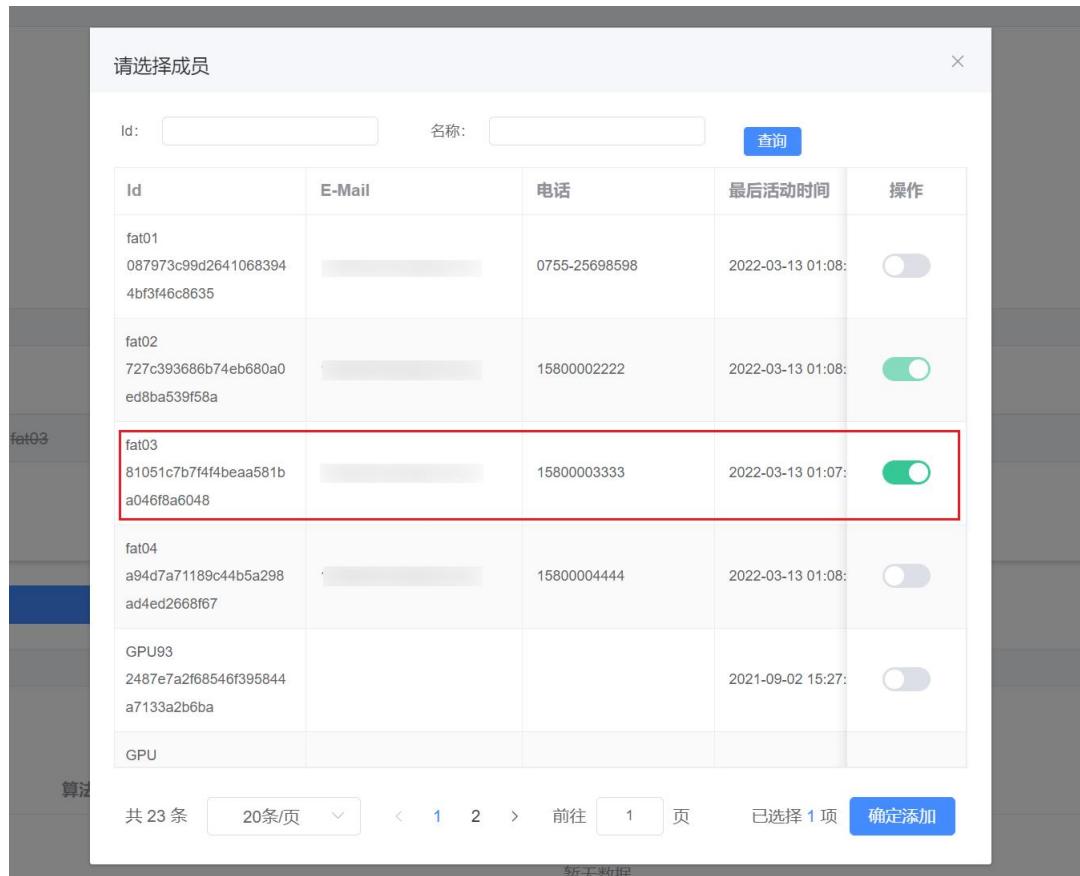
参与成员

发起方: DemoMember1 协作方: DemoMember2

点击刷新服务状态: 服务正常

+ 添加数据集到此项目

+ 添加更多发起方 + 添加更多协作方



请选择成员

Id: [] 名称: [] 查询

Id	E-Mail	电话	最后活动时间	操作
fat01 087973c99d2641068394 4bf3f46c8635	[]	0755-25698598	2022-03-13 01:08:	<input type="checkbox"/>
fat02 727c393686b74eb680a0 ed8ba539f58a	[]	15800002222	2022-03-13 01:08:	<input checked="" type="checkbox"/>
fat03 81051c7b7f4f4beaa581b a046f8a6048	[]	15800003333	2022-03-13 01:07:	<input checked="" type="checkbox"/>
fat04 a94d7a71189c44b5a298 ad4ed2668f67	[]	15800004444	2022-03-13 01:08:	<input type="checkbox"/>
GPU93 2487e7a2f68546f395844 a7133a2b6ba	[]		2021-09-02 15:27:	<input type="checkbox"/>
GPU	[]			

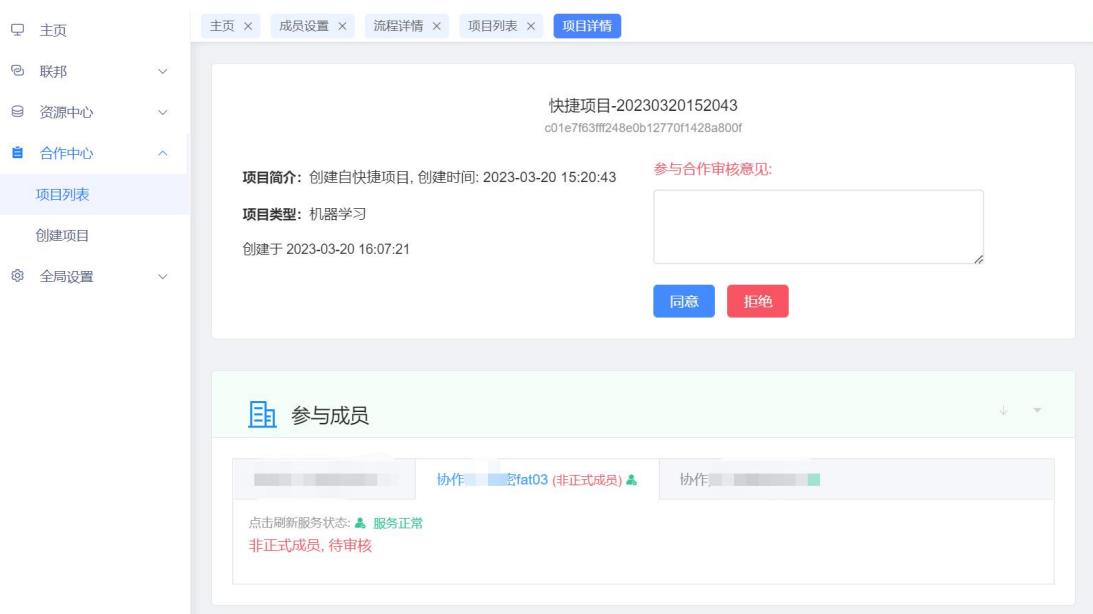
共 23 条 20条/页 < 1 2 > 前往 1 页 已选择 1 项 确定添加

暂无数据

此时 03 待审核状态，需要自己和其他正式成员审核。



4) 03 成员审核



项目简介: 创建自快捷项目, 创建时间: 2023-03-20 15:20:43
项目类型: 机器学习
创建于 2023-03-20 16:07:21

参与合作审核意见:

同意 拒绝

参与成员

点击刷新服务状态: 服务正常
非正式成员, 待审核

5) 其他人审核

自己审核通过后需要其他正式成员（这里指除了创建者之外的正式成员，创建者默认审核通过）审核。



项目详情

系统设置 × 主页 × 流程详情 × 项目列表 × 项目详情

项目归属者: Kaylee-tat
创建于 2023-03-20 15:20:43

参与成员

发起方: fat01 (绿) 协作方: fat02 (绿) 协作方: 国密fat03 (非正式成员) (红)

点击刷新服务状态: 服务正常
非正式成员, 待审核

天冕科技01的审核意见为: 邀请人, 自动通过。

参与合作审核意见:

同意 拒绝

其他正式成员审核通过后, 如下图所示:



参与成员

发起方: fat01 (绿) 协作方: fat02 (绿) 协作方: fat03 (绿)

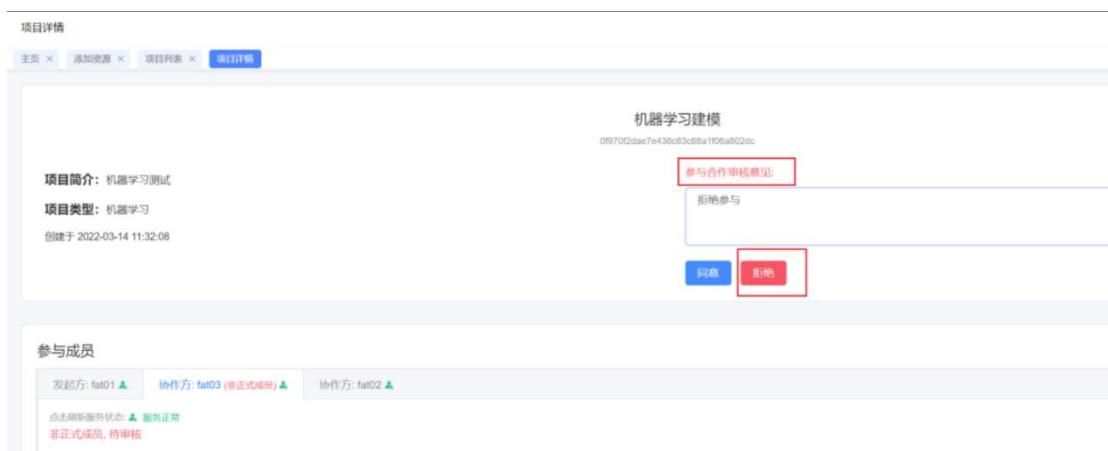
点击刷新服务状态: 服务正常

+ 添加资源到此项目

6) 审核被拒绝

操作:

(1) 成员自己拒绝加入项目, 那么审核流程终止



项目详情

主页 × 添加资源 × 项目列表 × 项目详情

机器学习建模
0f970f2dae7e436c83cd861f06a802dc

项目简介: 机器学习测试
项目类型: 机器学习
创建于 2022-03-14 11:32:08

参与成员

发起方: fat01 (绿) 协作方: fat03 (非正式成员) (红) 协作方: fat02 (绿)

点击刷新服务状态: 服务正常
非正式成员, 待审核

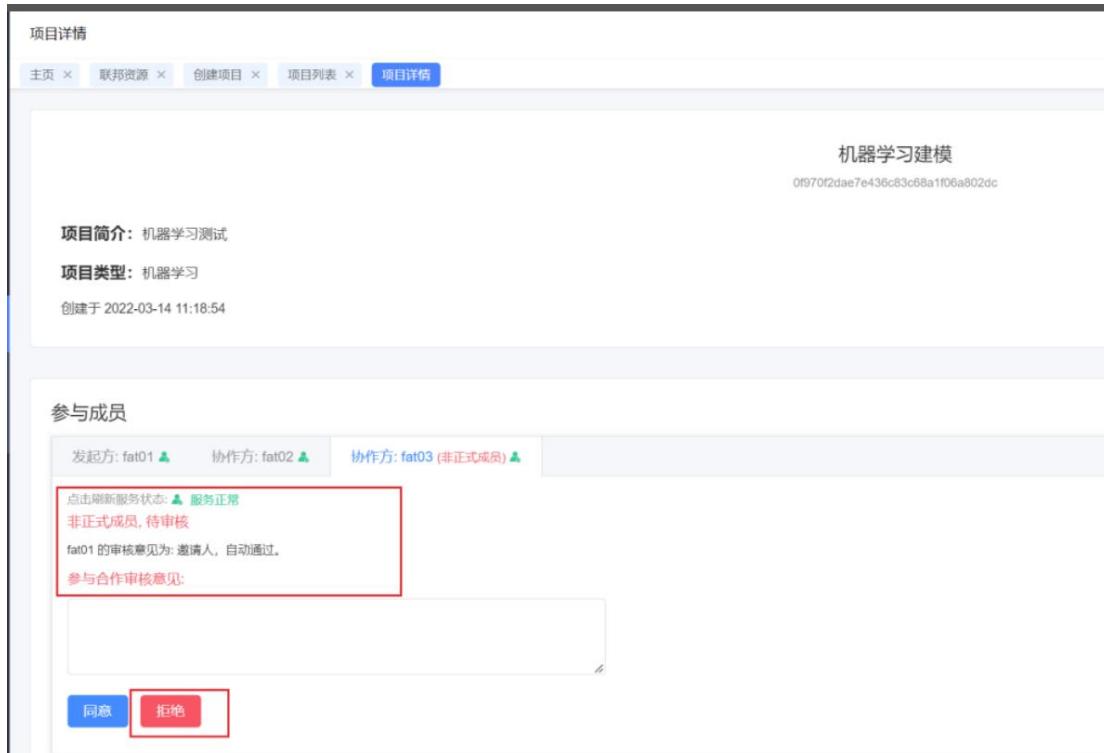
参与合作审核意见:
拒绝参与

同意 拒绝

流程终止。

(2) 成员被其他正式成员拒绝加入项目

其他正式成员审核不通过，那么审核流程终止。



项目详情

主页 × 联邦资源 × 创建项目 × 项目列表 × 项目详情

机器学习建模
0f970f2dae7e436c83c68a1f06a802dc

项目简介：机器学习测试

项目类型：机器学习

创建于 2022-03-14 11:18:54

参与成员

发起方: fat01 协作方: fat02 协作方: fat03 (非正式成员)

点击刷新服务状态: 服务正常
非正式成员, 待审核
fat01 的审核意见为: 邀请人, 自动通过。
参与合作审核意见:

同意 拒绝



参与成员

发起方: fat01 协作方: fat02 协作方: fat03

点击刷新服务状态: 服务正常
fat03 已退出该项目 (fat02意见为: 不同意03成员加入)

成员被拒绝加入后可先移除后重新邀请，则成员重新走审核流程。

7) 移除成员

只有发起方可以操作

操作:

点击移除按钮，点击确定。

项目详情

发起方: fat01 | 协作方: fat02 | **协作方: fat03**

机器学习建模
09/19/2022 14:36:03 (EST) 16608020ms

项目简介: 机器学习测试
项目类型: 机器学习
由 aisen 创建于 2022-03-14 11:18:54

参与成员

发起方: fat01 | 协作方: fat02 | **协作方: fat03**

点击刷新服务状态: 服务正常
+ 添加资源到协作方

+ 添加更多发起方 | + 添加更多协作方



机器学习建模

fat03

移除成员

注意: 成员被移除后其数据资源将变为不可用

移除

则成员退出该项目，发起方展示为成员被移除。

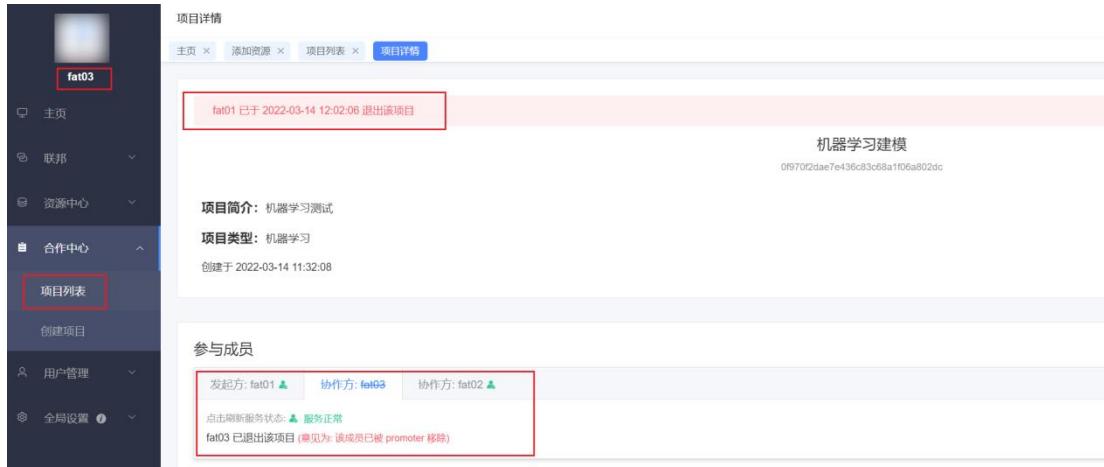
参与成员

发起方: fat01 | 协作方: fat02 | **协作方: fat03**

点击刷新服务状态: 服务正常
fat03 已退出该项目 (意见为: 该成员已被 promoter 移除)

+ 添加更多发起方 | + 添加更多协作方

协作方展示为成员已退出，退出项目后，成员无法再看到项目的变更。



3.6.2.4 数据集审核

1. 添加数据集

操作:

- 1) 项目详情页, 点击“添加资源到此项目”。



- 2) 选择数据集

请选择数据资源

上传者: 请选择 名称: ID:

资源类型: 结构化数据集 + 1

名称 / Id	资源类型	关键词	数据信息	参与任务次数	上传者	选择数据资源
provider-569-20230317 可使用 e2db3d9a4ef84487967fe7dd651139 72	结构化数据集	20230317 provider	特征量: 20 样本量: 569 不包含Y	0	jason 2023-03-17 10:02:14	 
promoter-569-20230317 可使用 76fcf7c515a243e5b5e84b1d1f5ceb3 c	结构化数据集	20230317 promoter	特征量: 10 样本量: 569 正例样本数量: 357 正例样本比例: 62.7% 包含Y	4	jason 2023-03-17 10:01:42	 
12w_50t 可使用 eb8e4c17661144fb2617b3adc0fea de	结构化数据集	银行	特征量: 51 样本量: 120000 正例样本数量: 391 正例样本比例: 0.3% 包含Y	0	yippee 2023-03-16 10:28:04	 
111111 我拥有			特征量: 10 样本量: 569			

支持数据集预览，最多只展示 100 条数据。

数据预览 >

id	y	x0	x1	x2	x3	x4	x5	x6	x7	x8
95cb335498e1c0e1509b97b692487983	1	-0.252473	-0.212873	-0.236834	-0.361925	0.58001	0.266129	-0.708059	-0.076228	-0.51
066c5b3b5ad3154af9d7cf7dcc51191	1	-0.561026	0.019993	-0.563882	-0.564436	0.474804	-0.489605	-0.536788	-0.790964	0.239
ac2cb830173b7ba37b48ce93679f3410	1	-0.608655	-0.033745	-0.543926	-0.620865	-0.160816	-0.186167	-0.242699	-0.062524	-1.09
2573c1c4409ffcc9ebdb25873d67df75	1	-1.033174	-0.825166	-1.064284	-0.861699	0.343297	-0.116191	-0.395261	-0.261383	-0.47
8c74d445bf9cbc8e47de5332af2e1e90	1	-0.505114	0.785359	-0.470652	-0.537716	-0.086295	-0.050669	-0.138113	0.089742	0.115
783e8c7097a12ba66bf81634ac3cede6	1	-0.527893	-1.427688	-0.592179	-0.535431	-0.796437	-0.361105	-0.61019	-0.854307	-0.751
0983bb52a74a8c803393cc7259a45a3	0	0.644194	0.871666	0.656444	0.49998	0.400283	1.350111	2.094174	1.677876	3.111
f1bce7fdb72c0c6944c8f886f0de597a	0	0.869914	-0.092369	0.763673	0.740814	0.413434	0.607736	0.413122	0.561767	-0.701
6bc46615dc82750706ee199118b31a81	0	1.88669	-1.359293	2.303601	2.001237	1.307686	2.616665	2.109526	2.296076	2.750
60acc15af1cad07a66531e8edb970d8	1	-0.606584	1.166414	-0.675579	-0.585004	-0.879725	-1.053734	-0.756514	-0.613574	-0.33
afb249520b5df858b4ba7425ae23ddcb	1	-1.062166	-0.009318	-1.083645	-0.87084	-0.393146	-0.636553	-0.693187	-0.746198	0.442
a09eae5ba5123702cf10617ecad19fdd	0	1.542933	1.664716	1.564912	1.482653	2.009061	0.825932	1.454664	1.105356	0.577
0c6c6dbc411adf0762f18f2198135b90	1	0.254879	-1.046633	0.209656	0.074214	-0.441366	-0.377645	-0.485934	0.347072	-0.28
b8c609d11ba4e23cb4feb258010729bd	1	-0.674921	-0.698148	-0.680345	-0.631237	-0.003007	-0.955896	-0.576127	-0.07166	-0.53

2. 数据集审核

发起方成员添加自己的数据集不需要审核，发起方成员添加其他成员的数据集需要该协作方成员自己审核(只有发起方可添加其他成员的数据集)。

操作:

- 1) 发起方添加协作方数据集

项目详情

查看数据资源 < 添加资源 < 我的资源 < 项目列表 < 项目详情

机器学习建模
0f970f2daef7e436cd3cd8a1106a802dc

项目简介: 机器学习测试
项目类型: 机器学习
由 aisen 创建于 2022-03-14 11:18:54

参与成员

序号	数据资源	数据类型	关键词	数据量	使用次数	是否包含 Y	状态	操作
1	资源已被删除或不可见 504f913e607e4d978ff203bdd2f2292	结构化数据集		特征量: 样本量:		否	已授权	

+ 添加更多发起方 + 添加更多协作方

2) 协作方审核数据集

若协作方拒绝授权数据集，发起方可重新添加数据集发起数据集审核。

项目详情

成员设置 < 我的资源 < 添加资源 < 项目列表 < 项目详情

机器学习建模
0f970f2daef7e436cd3cd8a1106a802dc

项目简介: 机器学习测试
项目类型: 机器学习
创建于 2022-03-14 11:18:54

参与成员

序号	数据资源	数据类型	关键词	数据量	使用次数	是否包含 Y	状态	操作
1	(待审核) docs 98b813029fc14019a2b6633b77946c75	结构化数据集		特征量: 9 样本量: 569	1	否	等待审核	

3. 数据集移除

移除项目内的数据集，移除后再次添加需要重新授权。

参与成员

序号	数据资源	数据类型	关键词	数据量	使用次数	是否包含 Y	状态	操作
1	机器学习数据集 13cf35140d247139209ce64b10823b6	结构化数据集		特征量: 14 样本量: 227	0	是	已授权	

4. 删除数据集

操作：

若在“资源中心”的“我的资源”中删除该数据集“删除”按钮，项目中已添加这个数据集后，成员方数据集被删除，展示不可见。

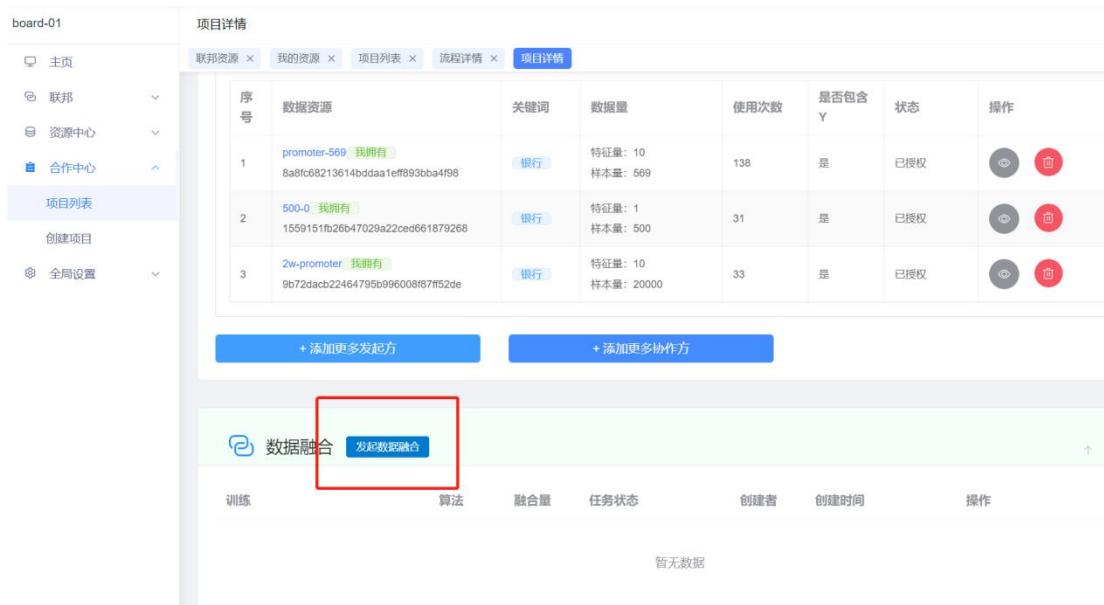


The screenshot shows a table of datasets. One row is highlighted with a red border, indicating it has been deleted. The columns include: 序号 (Index), 数据资源 (Data Resource), 关键词 (Keywords), 数据量 (Data Volume), 使用次数 (Usage Count), 是否包含 Y (Contains Y), 状态 (Status), and 操作 (Operations). The status for the deleted row is '已授权' (Granted).

序号	数据资源	关键词	数据量	使用次数	是否包含 Y	状态	操作	
1	资源已被删除 13cf05140d247139209be64b1082366	结构化数据集	特征量: 样本量:	138	否	已授权		

3.6.2.5 新增数据融合任务

数据融合是平台的子系统功能，在联邦平台中也支持融合项目的创建和执行。操作主要分为三个部分：添加数据融合任务基本信息，选择融合算法和选择发起方、协作方。



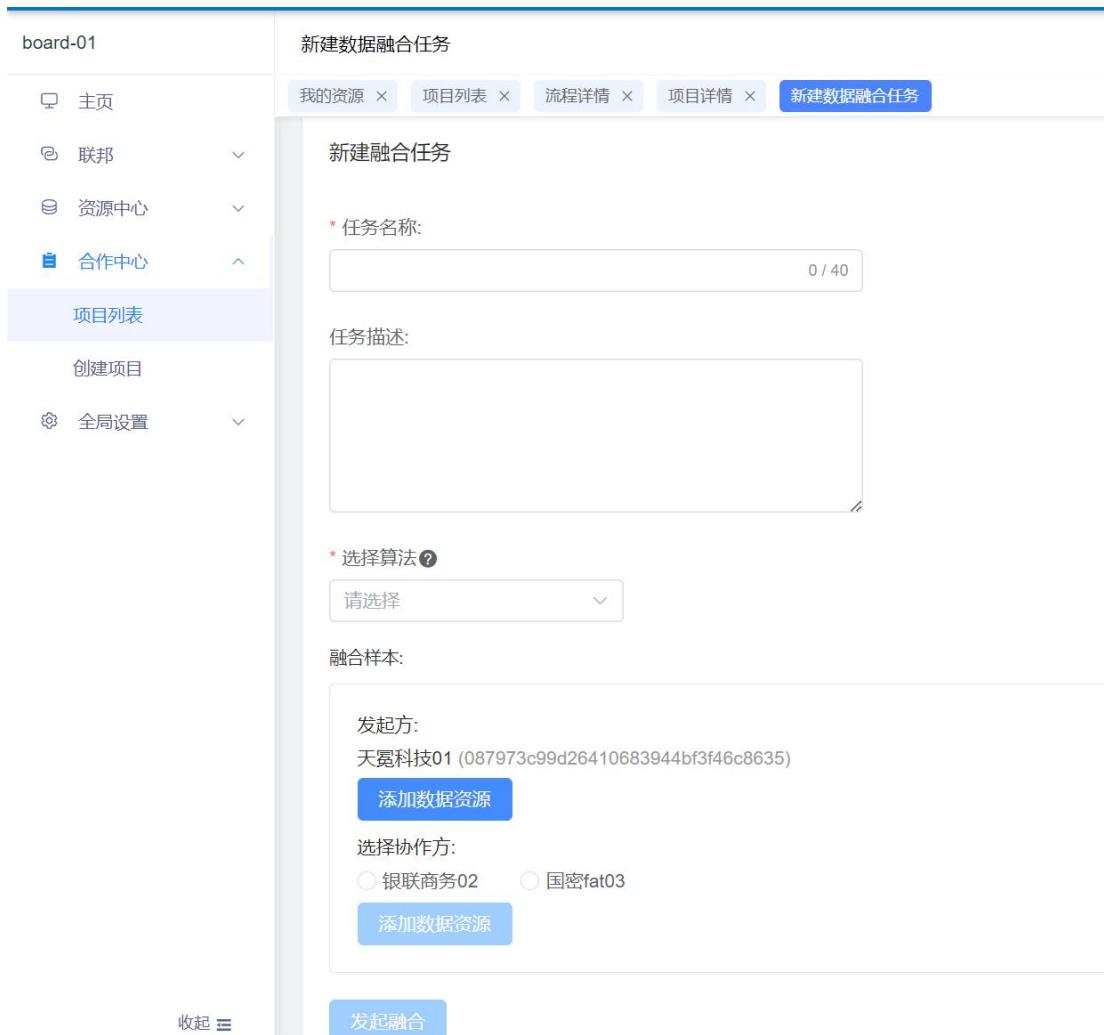
The screenshot shows the 'Data Fusion' section of a project details page. It includes a table of existing fusion projects and two buttons at the bottom: '+ 添加更多发起方' (Add more initiators) and '+ 添加更多协作方' (Add more cooperators). A red box highlights the '发起数据融合' (Initiate data fusion) button.

序号	数据资源	关键词	数据量	使用次数	是否包含 Y	状态	操作	
1	promoter-569 我拥有 8a8fcf68213614bddaa1eff893bba4f98	银行	特征量: 10 样本量: 569	138	是	已授权		
2	500-0 我拥有 1559151fb26b47029a22ced661879268	银行	特征量: 1 样本量: 500	31	是	已授权		
3	2w-promoter 我拥有 9672daac22464795b996008ff7ff52de	银行	特征量: 10 样本量: 20000	33	是	已授权		

+ 添加更多发起方 + 添加更多协作方

数据融合 发起数据融合

训练	算法	融合量	任务状态	创建者	创建时间	操作
暂无数据						



新建数据融合任务

我的资源 × 项目列表 × 流程详情 × 项目详情 × 新建数据融合任务

新建融合任务

* 任务名称:

0 / 40

任务描述:

* 选择算法?

请选择

融合样本:

发起方:

天冕科技01 (087973c99d26410683944bf3f46c8635)

添加数据资源

选择协作方:

银联商务02 国密fat03

添加数据资源

收起

发起融合

选择融合数据的协作方，分别添加在本项目中的数据融合数据集，在添加数据集时需要注意，数据资源类型选择“布隆过滤器”资源，另一方则必须为“数据资源”。

融合样本: RSA-PSI 算法要求至少一方需要选择布隆过滤器资源, 另一方则必须为数据资源资源

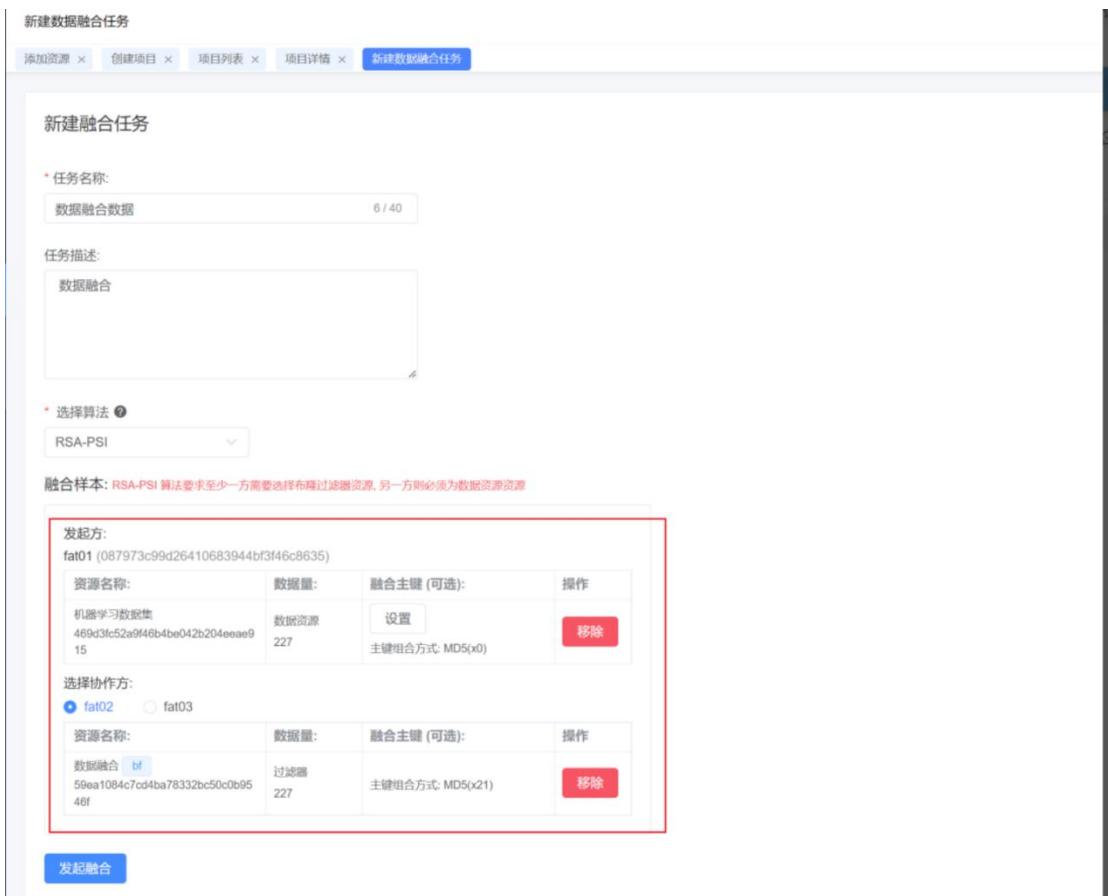
发起方:	fat01 (087973c99d26410683944bf3f46c8635)	
资源名称:	机器学习数据集 469d3fc52a9f46b4be042b204eeae9 15	
选择协作方:	<input checked="" type="radio"/> fat02 <input type="radio"/> fat03	
资源名称:	数据融合 bf 59ea1084c7cd4ba78332bc50c0b95 46f	
数据量:	数据资源 227	
融合主键 (可选):	操作	
<button>设置</button>	<button>移除</button>	
主键组合方式: 无		
数据量:	融合主键 (可选):	操作
过滤器 227	主键组合方式: MD5(x21)	<button>移除</button>

同时，数据资源方的主键设置，需要和布隆过滤器资源方的主键做同样的处

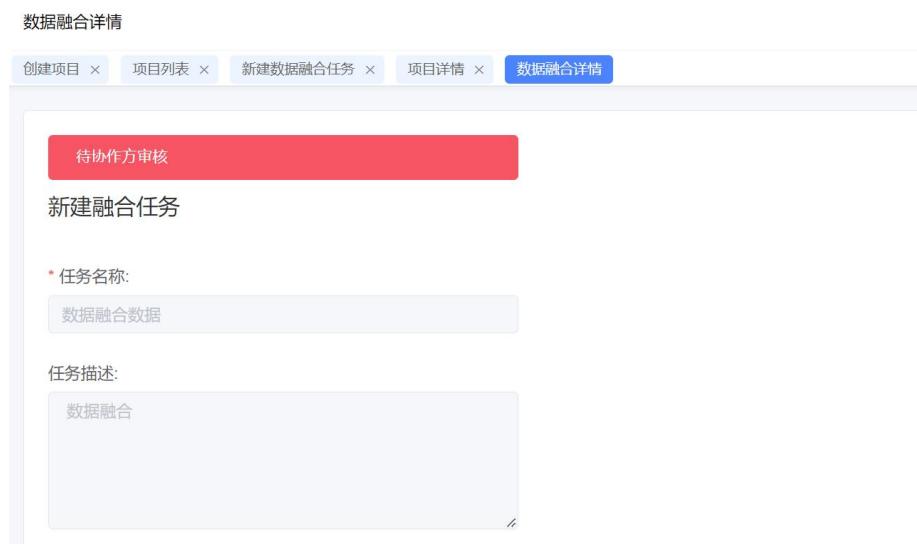
理:



点击确认“发起融合”任务:



任务发起后需要协作方审核同意后，任务开始执行：



协作方同意后任务开始执行，任务结束后可以导出融合的结果到数据库中：

新建融合任务

* 任务名称:

数据融合数据 6 / 40

任务描述:

数据融合

* 选择算法 

RSA-PSI

融合样本: RSA-PSI 算法要求至少一方需要选择布隆过滤器资源, 另一方则必须为数据资源资源

发起方:

fat01 (087973c99d26410683944bf3f46c8635)

资源名称:	数据量:	融合主键 (可选):	操作
机器学习数据集 469d3fc52a9f46b4be042b204eeae9 15	数据资源 227	设置 主键组合方式: MD5(x0)	

协作方:

fat02 (727c393686b74eb680a0ed8ba539f58a)

资源名称:	数据量:	融合主键 (可选):	操作
数据融合 bf 59ea1084c7cd4ba78332bc50c0b95 46f	过滤器 227	主键组合方式: MD5(x21)	

进度

融合量

耗时

操作

100%

0

00:00:02



3.6.2.6 新建训练流程

1. 新增联合建模流程

操作:

- 1) 新增模板创建流程。点击“新增训练流程”按钮，弹出选择模板框，可选择新增流程。



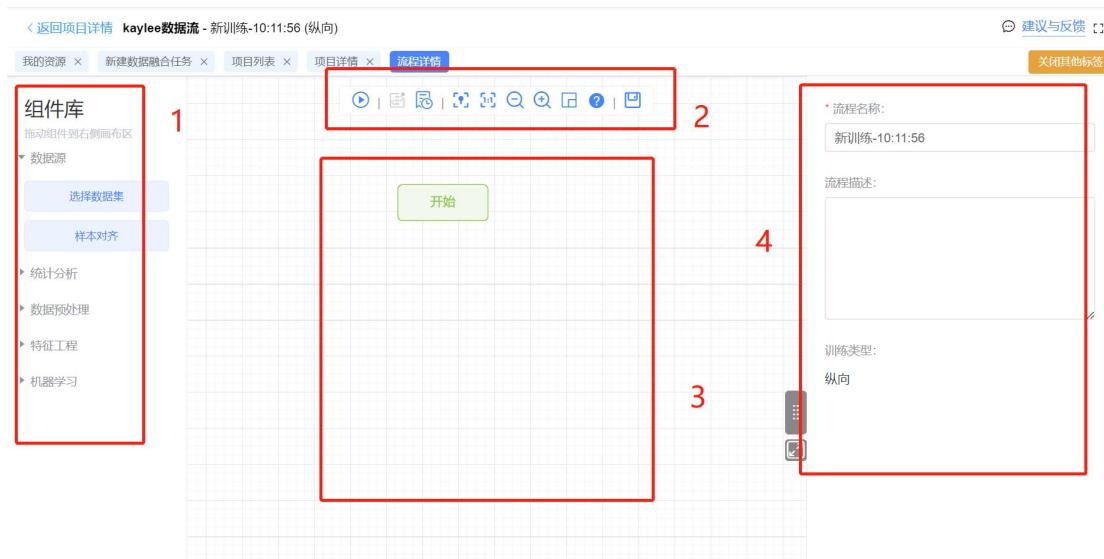
新增流程模板包含了空白流程和模板流程，用户可根据需要选择，新增纵向空白

流程为例：

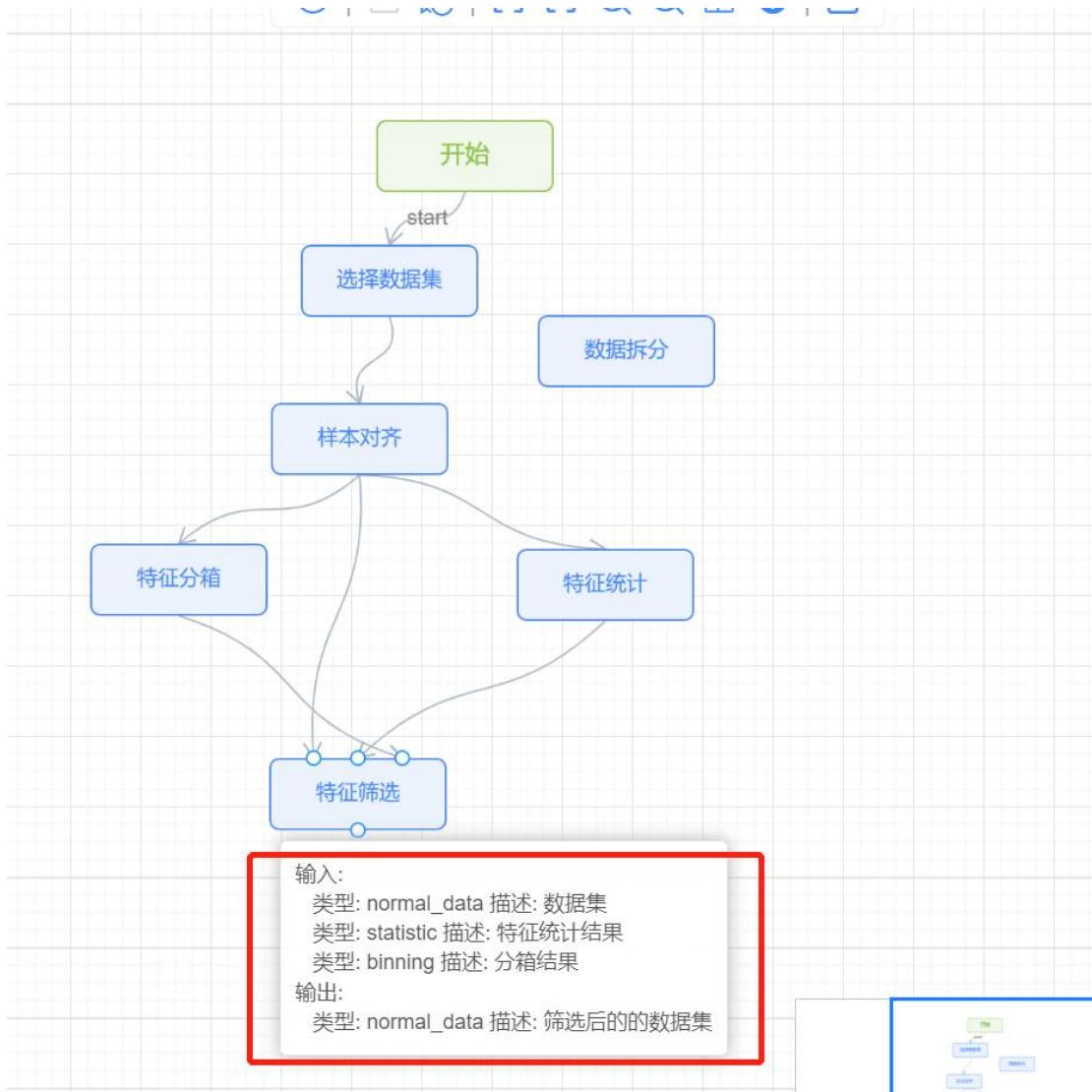


点击空白模板/流程模板创建流程后跳转流程详情页面：

1. 系统组件库；
2. 画布操作按钮，点击  可查看画布详细操作指引；
3. 画布区；
4. 组件参数编辑区。



每个组件都指定输入输出类型，每个输入有相同对应类型的锚点，只有相同类型的锚点才能连接：



2. 编辑流程

添加或修改模型组件，需要修改组件参数。

➤ 纵向流程:

1) 选择数据集

发起方只允许选择带 y 标签的数据集。

选择数据集

原始数据集 衍生数据集

名称 / id	授权情况	包含Y	关键词	数据信息	参与任务次数	选择数据集
breast_data 8ff523e95c084b8a847cba7a27a267b 6	已授权	是	医疗	特征量: 10 样本量: 569 正例样本数量: 357 正例样本比例: 62.7%	0	<input checked="" type="checkbox"/>

协作方选择数据集:

横向——协作方必须选择带 y 标签数据集; 纵向/混合——协作方需选择不带 y 标签数据集。

选择完成后点击保存:

返回项目详情 kaylee数据流 - 新训练-10:32:21 (纵向)

建议与反馈

关闭其他标签

组件库

拖动组件到右侧画布区

数据源

- 选择数据集
- 样本对齐

统计分析

数据预处理

特征工程

机器学习

发起方:

fat01 选择数据资源

数据资源名称: data-569-0321 y
数据资源id: b535f2875bc04c5ba40079b1df203799
数据量/特征量: 569 / 13
分类数: 2
选择特征: 13 / 13
x0 x1 x2 x3 x4 x5 x6 x7 x8 x9 x1_CP6001_test x2_CP6001_test x3_CP6001_test

协作方:

银联商务02 选择数据资源

数据资源名称: provider-569
数据资源id: e07677a7486c42ec8686aa63413793dd
数据量/特征量: 569 / 20
选择特征: 20 / 20
x0 x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13 x14 x15 x16 x17 x18 x19

国密fat03 选择数据资源

开始

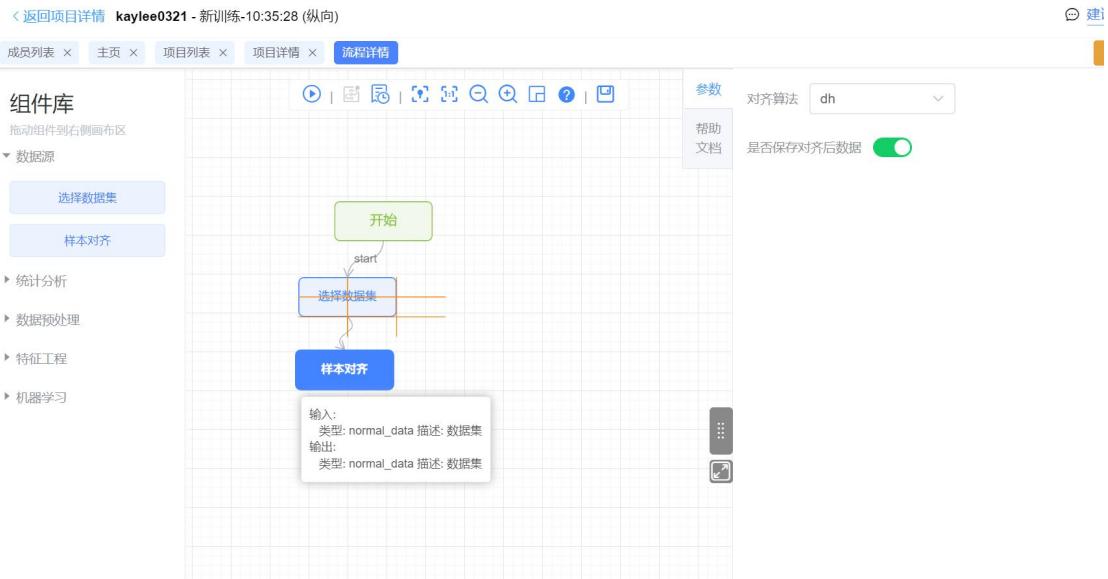
start

选择数据集

保存 取消

2) 样本对齐组件

默认为 dh 算法, 点击保存按钮, 此步骤是对成员方的数据集进行主键 id 对齐数据。



3) 特征筛选组件



特征筛选组件有三个锚点入参，类型分别为：

类型 normal_data 描述: 数据集

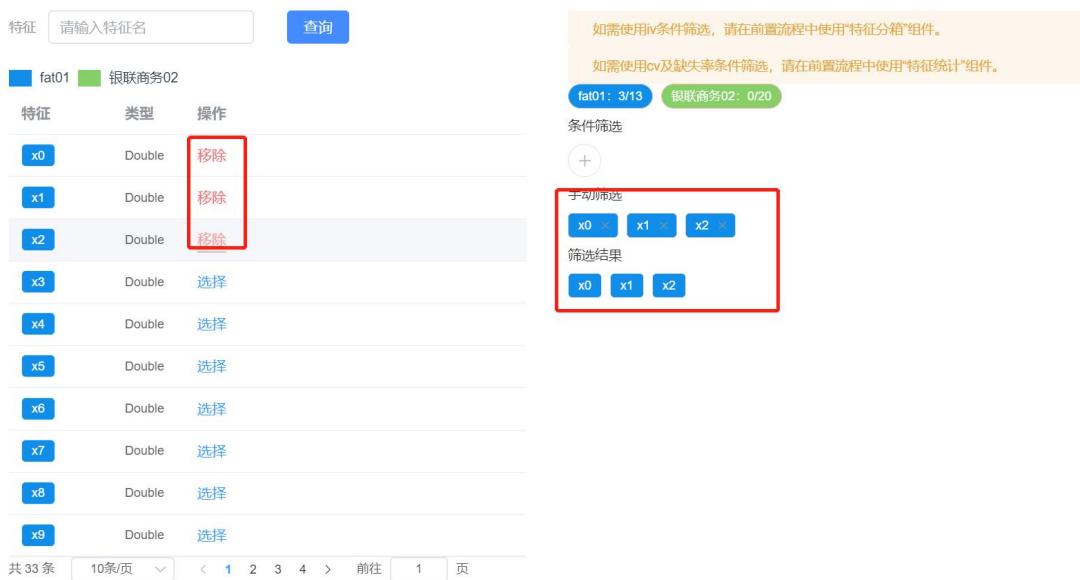
类型 statistic 描述: 特征统计结果

类型 binning 描述: 分箱结果

特征筛选组件支持 iv, cv, 缺失率的条件筛选，及手动筛选特征；Iv 值来源于特征分

箱组件, cv 及缺失率来源于特征统计组件, 需要前置特征分箱/特征统计组件入参; 若都没有入参可使用手动筛选。

➤ 手动筛选

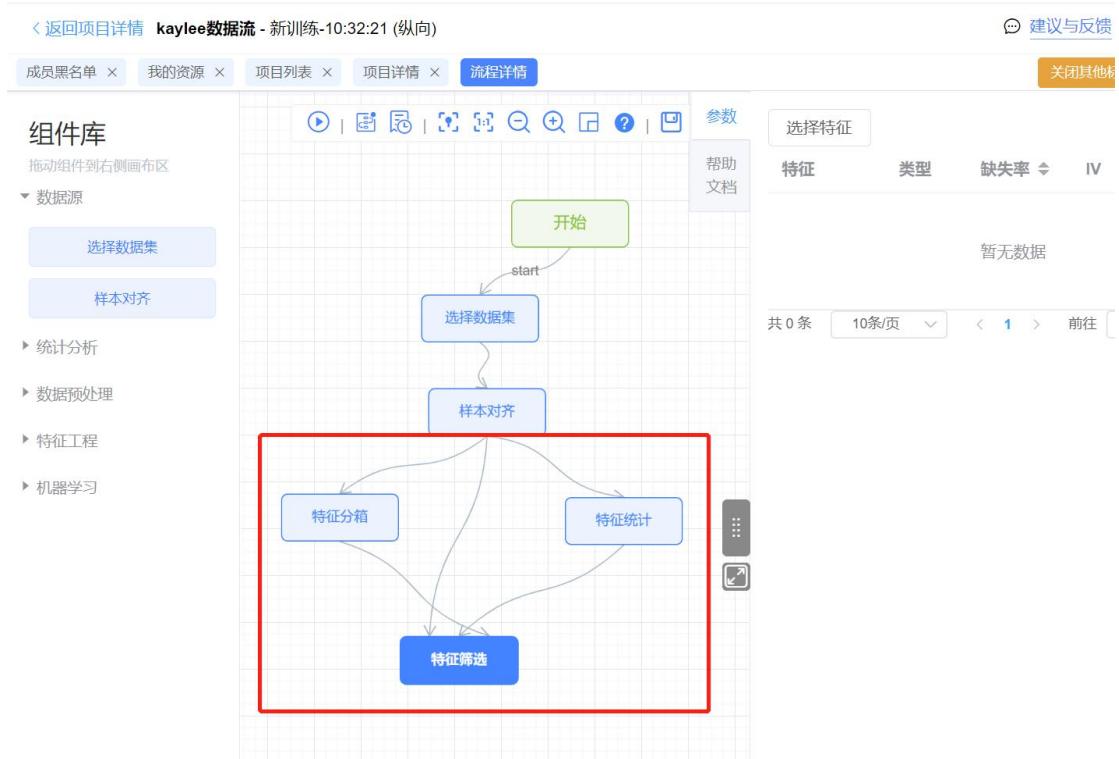



➤ 条件筛选

特征分箱及统计组件入参, 特征筛选组件列表展示各特征的缺失率, iv, cv

值，没有计算则为空。

条件筛选如图：



条件筛选：支持条件组合筛选，筛选结果展示条件筛选出的特征



特征		请输入特征名		查询		
<input checked="" type="checkbox"/>	fat01	<input checked="" type="checkbox"/>	银联商务02			
特征	类型	缺失率	IV	CV	操作	
x0	Double	0	6.064 3.151	3793333 3.151	已选择	
x1	Double	0	1.217 8.189	9999999 8.189	已选择	
x2	Double	0	6.393 0.900	10000000 0.900	已选择	
x3	Double	0	6.128 2.881	9483333 2.881	已选择	
x4	Double	0	0.963 9.667	9999999 9.667	已选择	
x5	Double	0	2.321 1.701	5172727 1.701	已选择	
x6	Double	0	3.625 2.482	10000000 2.482	已选择	
x7	Double	0	5.849 0.654	10000000 0.654	已选择	
x8	Double	0	1.084 3347058	3347058	已选择	

fat01: 10/13 银联商务02: 15/20

条件筛选

CV值大于等于 100 × and IV值大于等于 0.5 × +

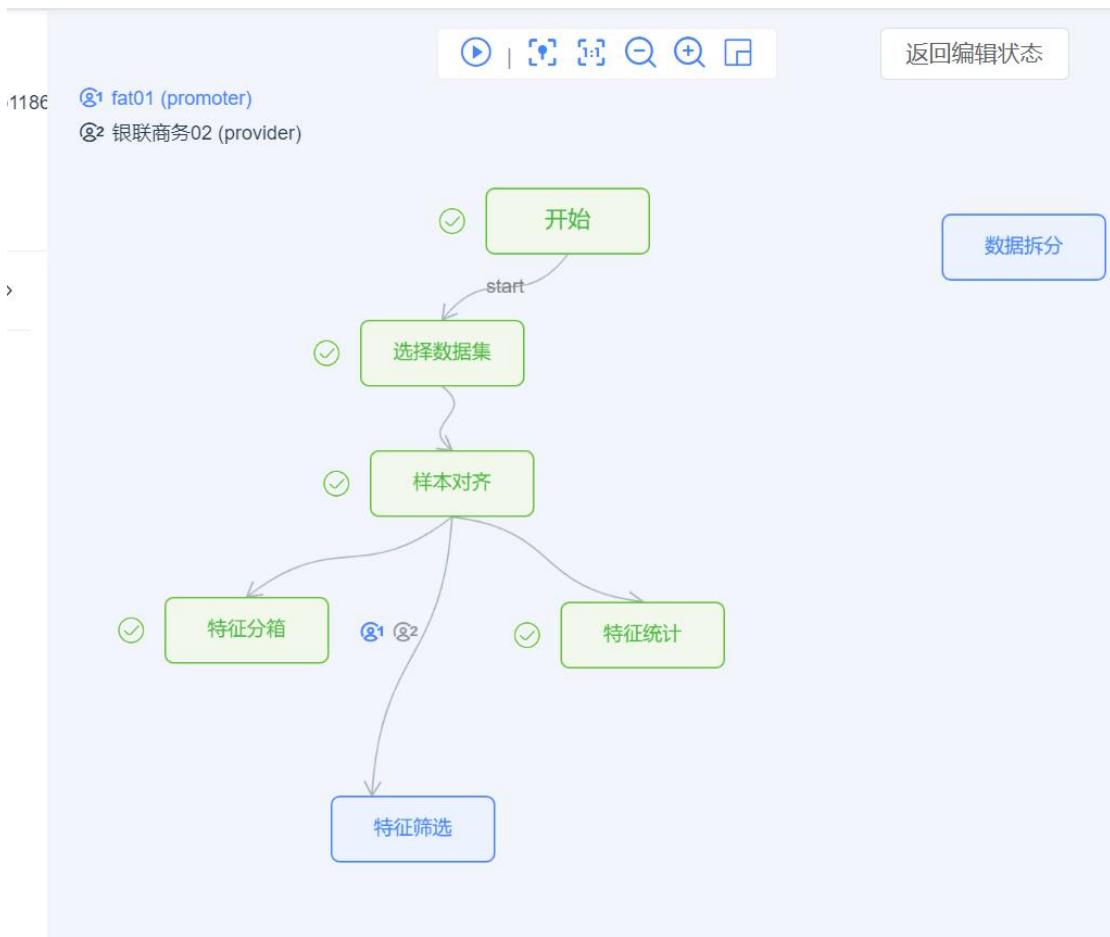
手动筛选

在左侧列表中选择需要的特征

筛选结果

x0 x1 x2 x3 x4 x5 x6 x7 x8 x9 x0 x1
x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13 x14 x15
x16 x17

注意：当特征筛选组件没有选择特征时，流程执行到特征筛选前停止。



4) 数据转换组件



数据转换组件三个锚点入参，标准化方法支持 z-score, min_max, woe 三种
数据标准化方法，z-score, min_max 依赖特征统计组件，woe 依赖特征分箱组件，
根据需要可选特征分箱/特征统计组件入参。

例:



5) 数据采样组件

数据采样组件对数据进行采样，提取部分数据，或者生成新数据。

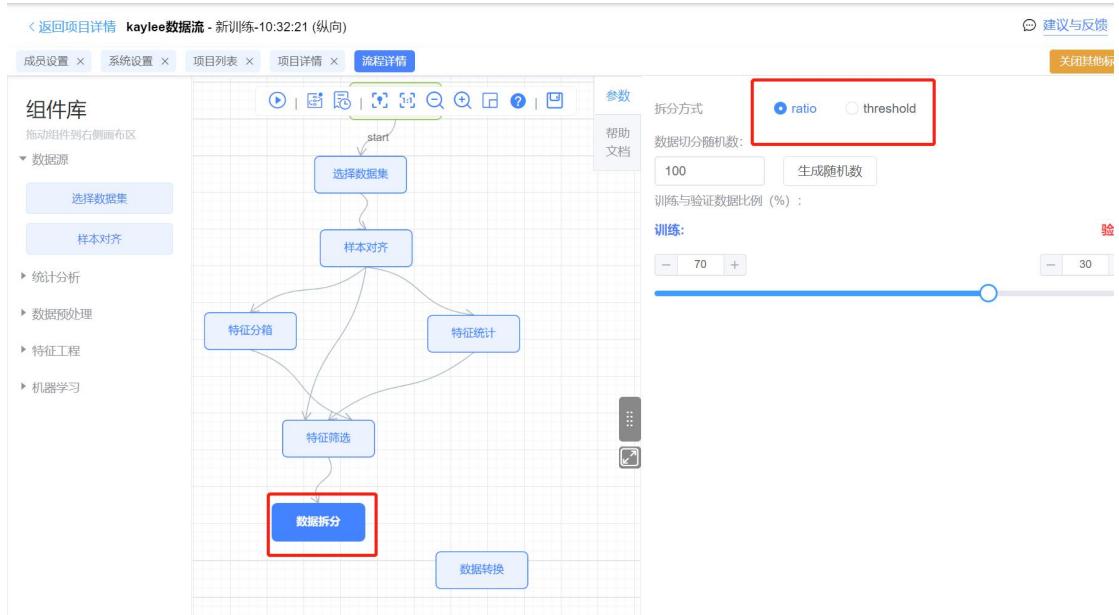
数据采样组件分为：分层抽样和随机抽样两种；采样方法分为：上采样和下采样。

输入标签数据（即 y 标签的数据），采样比例，选择下采样时，比率不能大于 1：

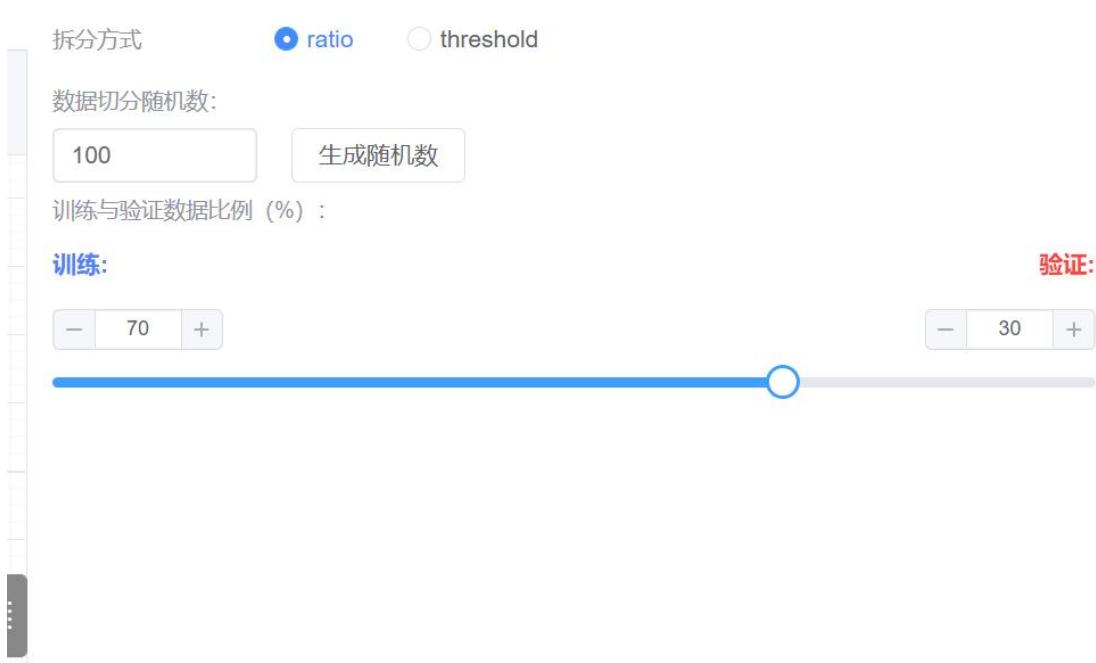


6) 数据拆分组件

数据拆分组件可分为按比率拆分或按阈拆分。



➤ 按比例拆分



This is a detailed view of the 'ratio' split configuration. It includes fields for 'Data切分随机数' (Random Seed) set to 100, and '训练与验证数据比例 (%)' (Training & Validation Ratio %) set to 70% for training and 30% for validation.

➤ 按阈拆分

支持日期格式的样本筛选，将样本数据切分为验证集和测试集，点击保存按钮。

拆分方式 ratio threshold

发起方

成员名称: fat01

特征列表: x0 x1 x2 x3 x4 x5 x6 x7 x8 x9 x1_CP6001_test x2_CP6001_test

x3_CP6001_test

* 过滤规则: 特征 x15 (Double) 操作符 > 值 0.5 (+)

x16 (Double)

x17 (Double)

x18 (Double)

x19 (Double)

协作方

成员名称:

特征列表: x20 (Double) x6 x7 X_8 x9 x10 x11 x12 x13 x14 x15 x16

x21 (Double)

查看更多

* 过滤规则: 日期 (DateTime) 操作符 日期 > 值 2019-07-10 (+)

日期>2019-07-10

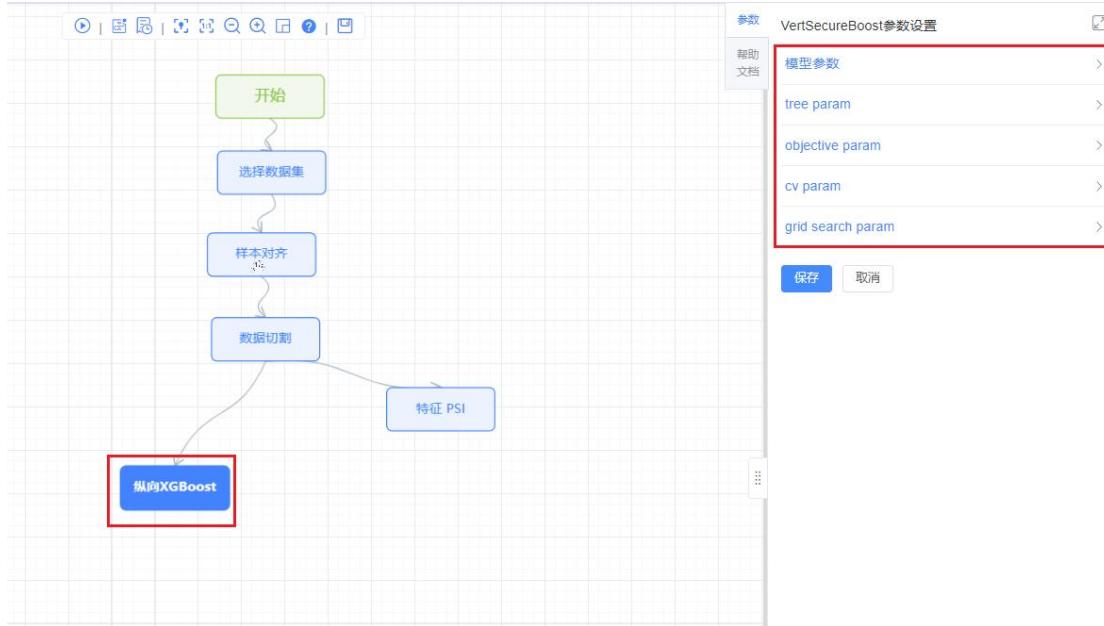
保存 取消

7) Xgb 组件



Xgb 组件有两个锚点入参，分别为训练集，测试集，入参可以只有一个 normal_data 训练集。

纵向 XGB 组件包括：模型参数，tree param，objective param，cv param 和 grid search param 参数：



可按需修改这几个参数调整算法：

① 模型参数

VertSecureBoost参数设置

模型参数

任务类型:	classification
学习率:	0.1
最大树数量:	10
树的最大深度:	3
特征随机采样比率:	1
n次迭代没变化是否停止:	<input checked="" type="radio"/> 是 <input type="radio"/> 否
收敛阈值:	0.0001
最大桶数量:	50
验证频次:	10
提前结束的迭代次数:	5
工作模式:	dp 模式
隐私预算:	3

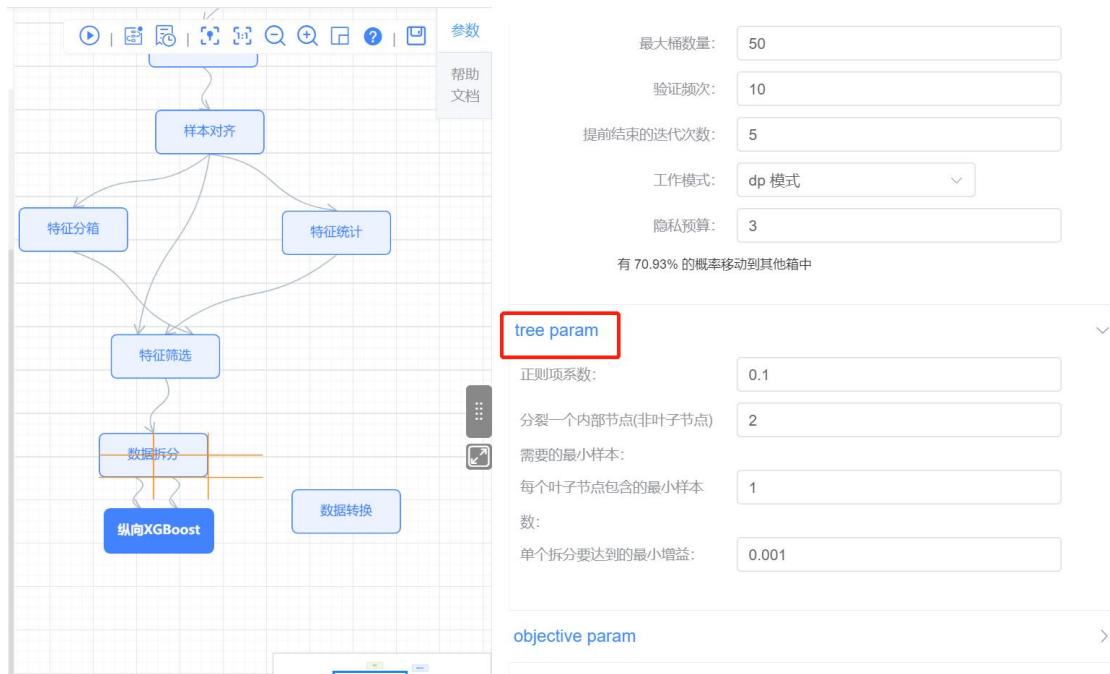
有 70.93% 的概率移动到其他箱中

参数说明可以参考“帮助文档”：



② tree param

设置了每棵树的基本参数:



参数

帮助
文档

最大桶数量: 50

验证频次: 10

提前结束的迭代次数: 5

工作模式: dp 模式

隐私预算: 3

有 70.93% 的概率移动到其他箱中

tree param

正则项系数: 0.1

分裂一个内部节点(非叶子节点) 2

需要的最小样本数: 1

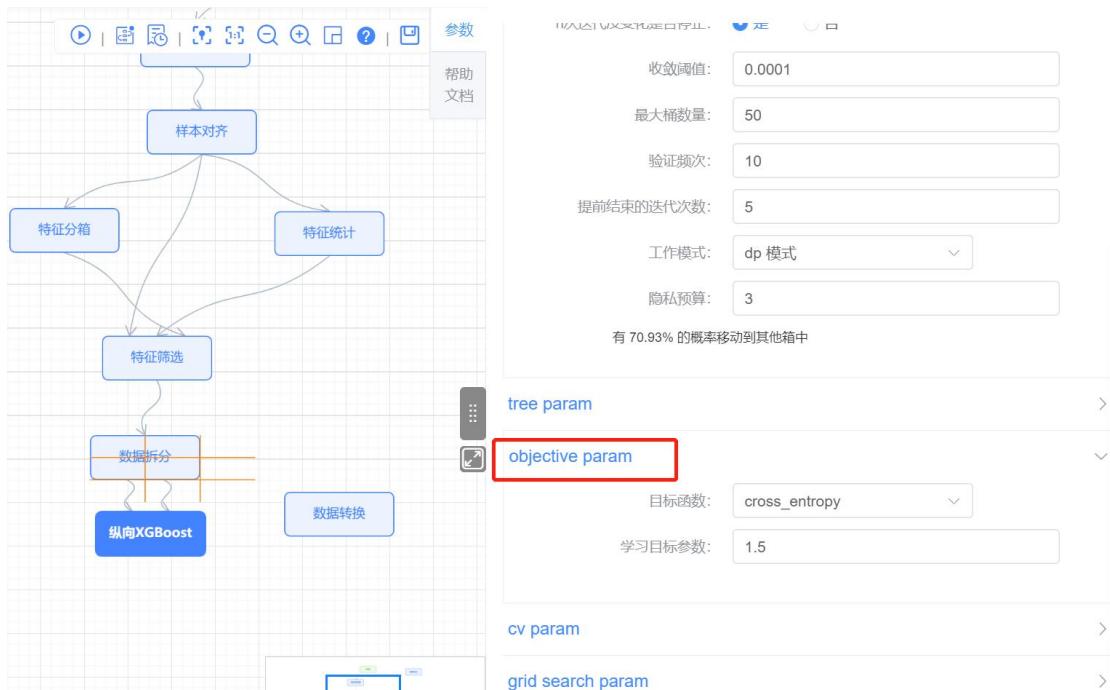
每个叶子节点包含的最小样本数: 1

单个拆分要达到的最小增益: 0.001

objective param

③ objective param

选择熵函数，支持 cross_entropy 函数:



参数

帮助
文档

收敛阈值: 0.0001

最大桶数量: 50

验证频次: 10

提前结束的迭代次数: 5

工作模式: dp 模式

隐私预算: 3

有 70.93% 的概率移动到其他箱中

tree param

objective param

目标函数: cross_entropy

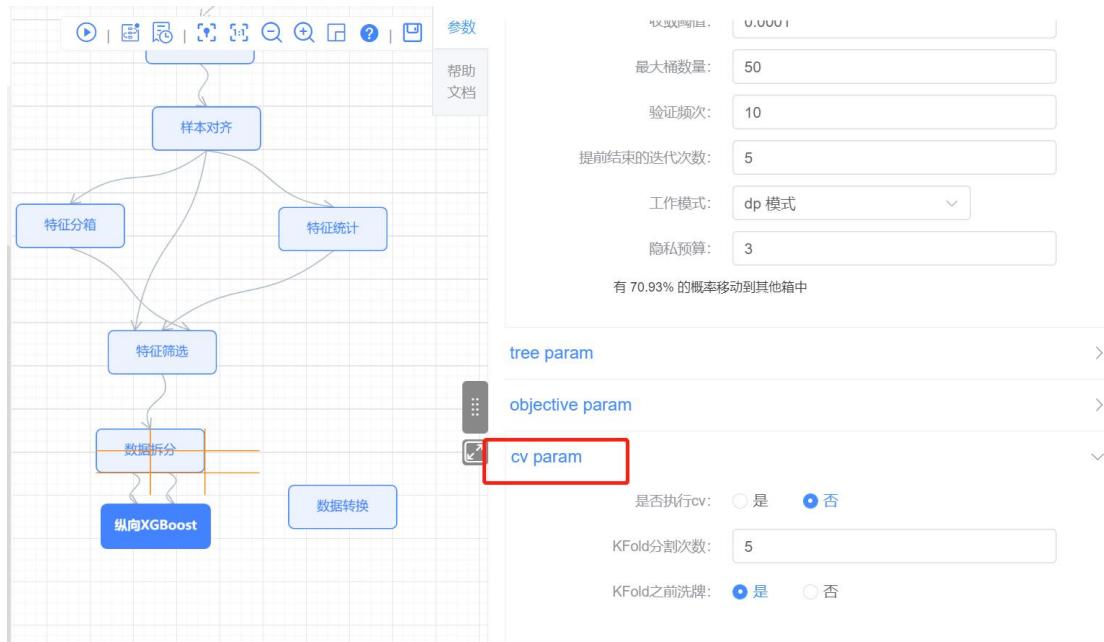
学习目标参数: 1.5

cv param

grid search param

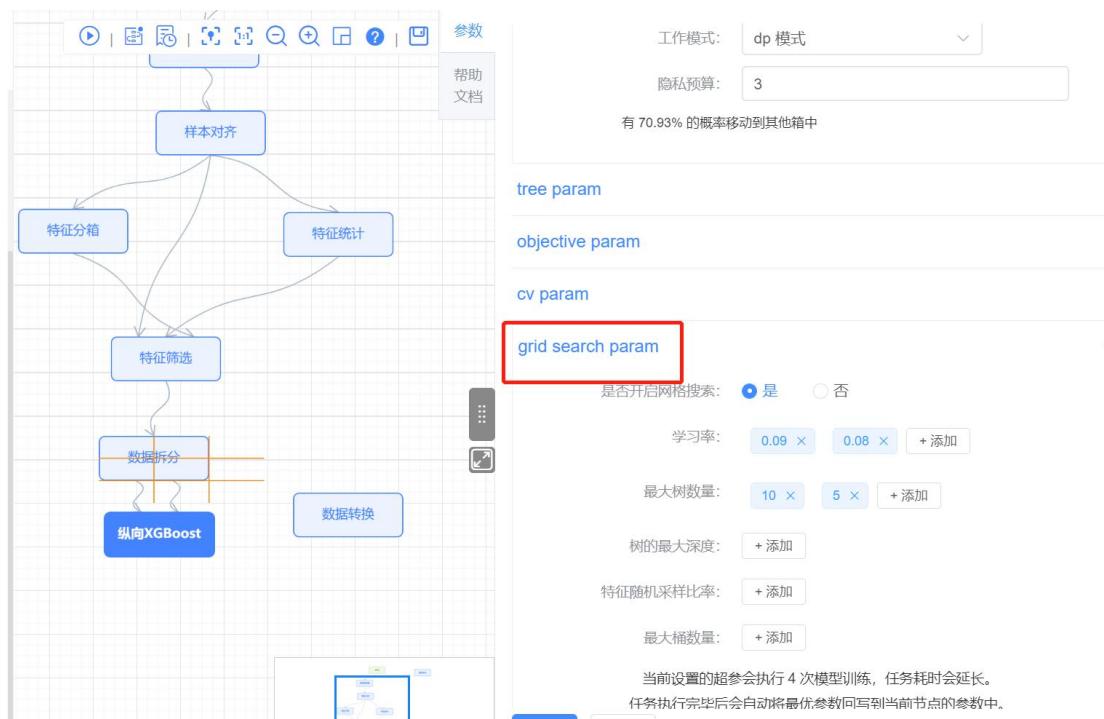
④ cv param

选择是否进行交叉验证:



⑤ grid search param

选择是否进行网格搜索，选择网格搜索时，组件会将所给的参数初始值组合入模，如下，学习率 2 个初始取值，最大树数量 2 个初始取值，树的最大深度 2 个初始取值，会进行 $2 \times 2 \times 2 = 8$ 次建模任务，建模任务会耗时会延长：



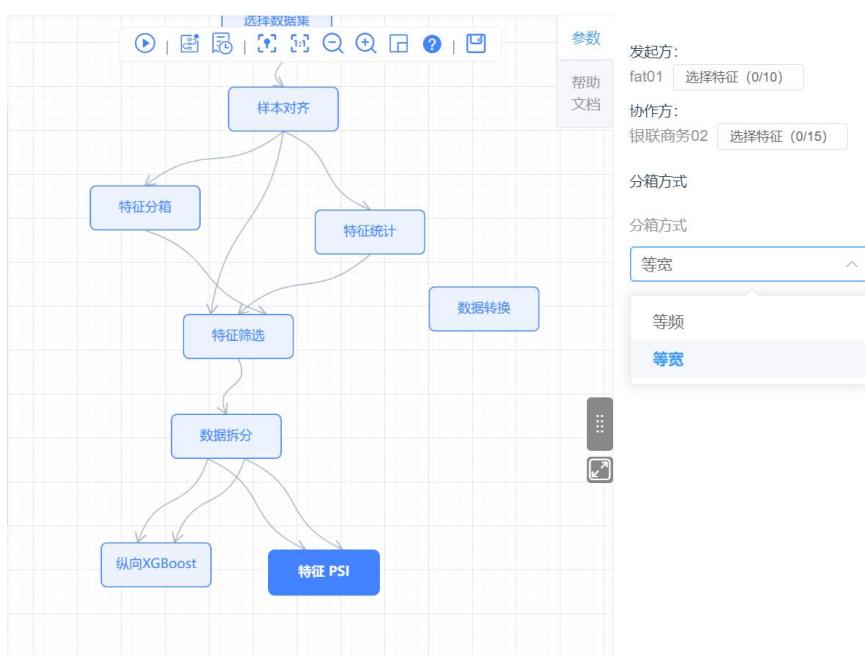
选择参数后点击保存按钮；组件执行完成后输出贡献度，特征重要性等结果

数据。

8) 特征 psi 组件



这个组件要求输入两个数据集，分别为训练集和验证集，可放在拆分组件后使用：

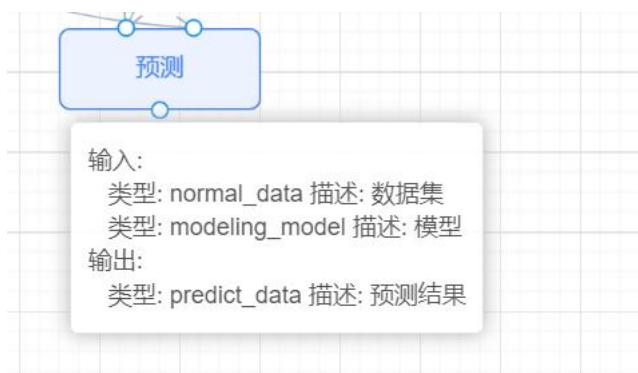


在右侧参数区域选择计算 PSI 的特征和计算 PSI 时需要的分箱方法和分箱箱数，分箱方法支持等宽和等频分箱方式，选择好参数后点击“保存”：

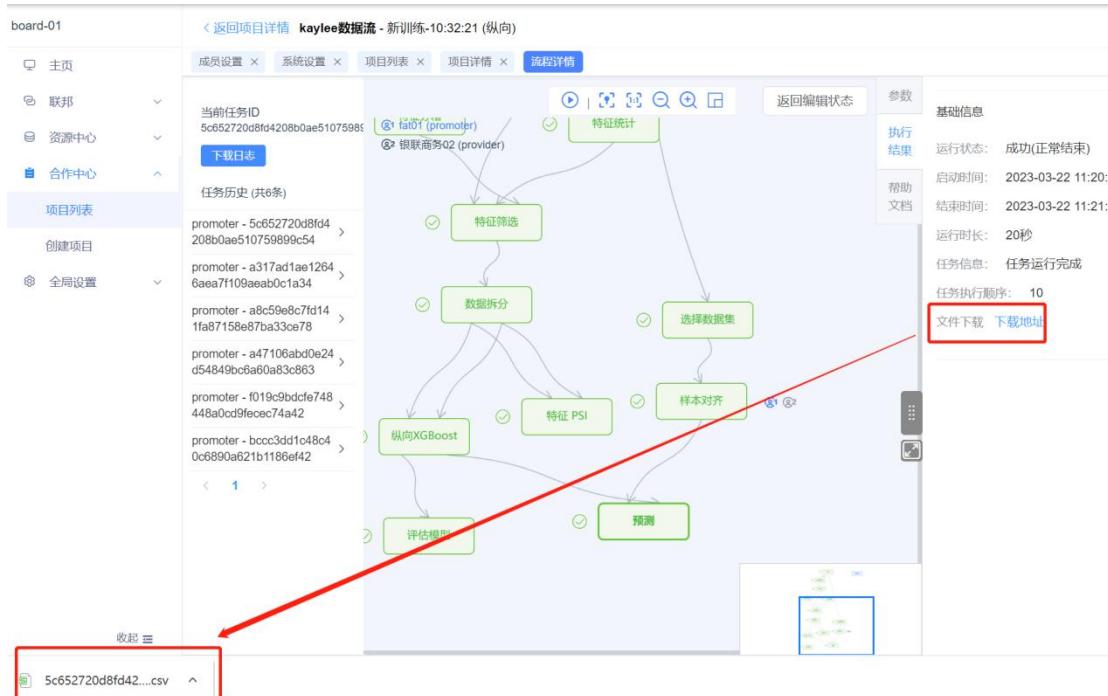


9) 预测组件

选择一份数据集，使用已经训练好的模型预测样本概率。



如图，训练完成可下载预测结果，展示各样本的预测概率：

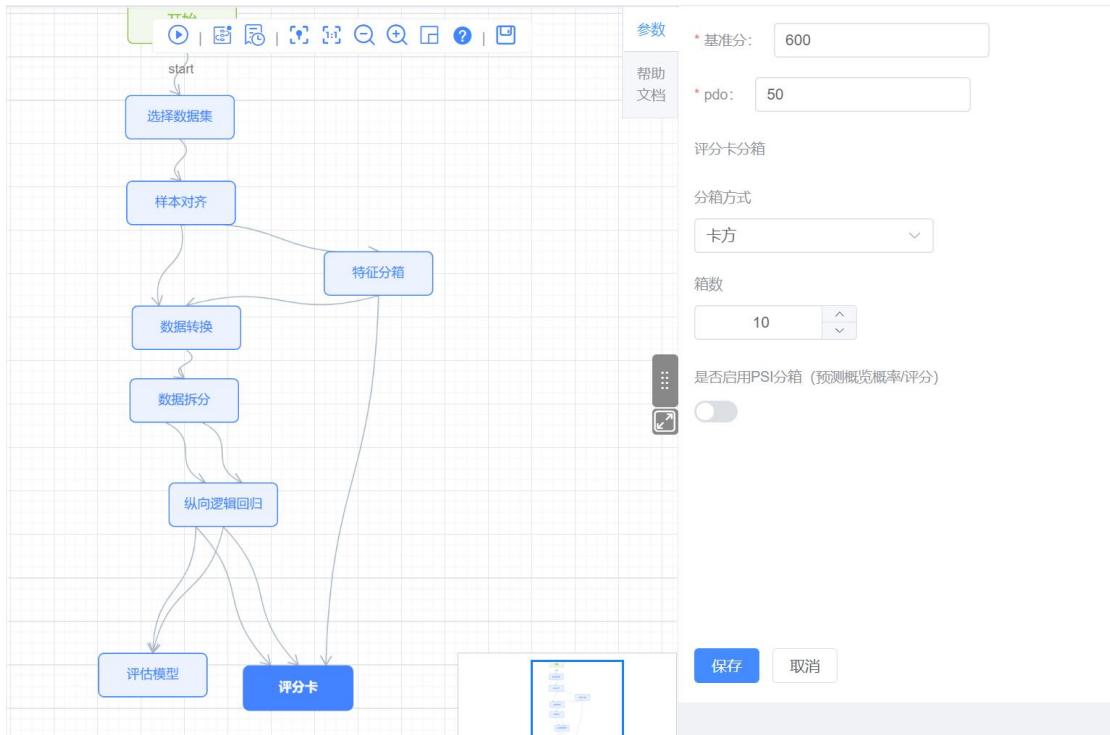


A	B
id	score
289	0.6663016
310	0.6663016
353	0.3363468
516	0.3363468
15	0.3363468
71	0.6663016
373	0.4022678
212	0.4022678
244	0.3363468
375	0.574052
451	0.3363468
483	0.6663016
66	0.6663016
160	0.6663016
194	0.4573185

10) 评分卡

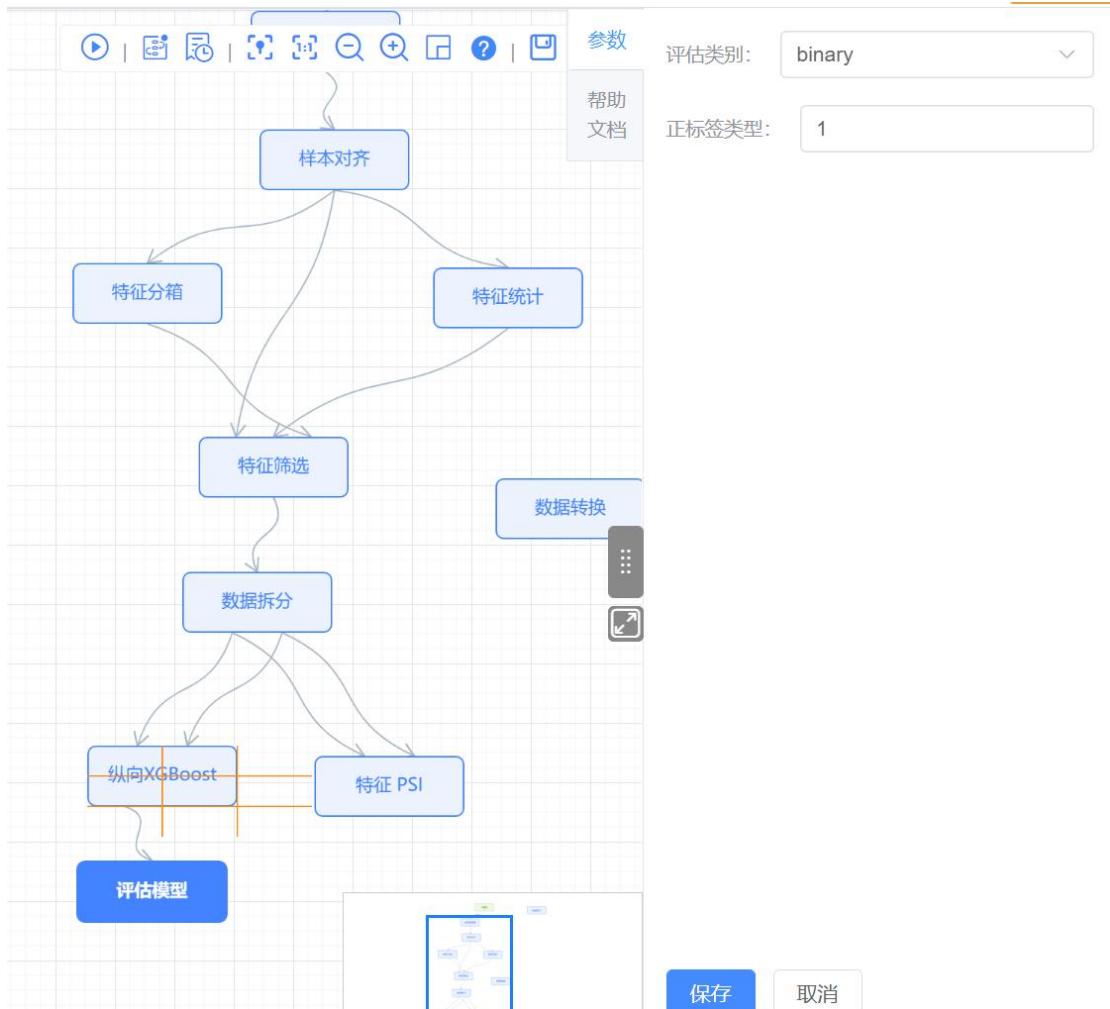
逻辑回归评分卡，将样本预测概率转换为分数。支持等频，等宽，卡方，自定义等分箱方式对转换成的分数进行分箱展示。





11) 模型评估组件

一般选择 binary (二分类)，点击保存



可按需添加其他组件，并根据实际情况配置参数。

3. 启动流程

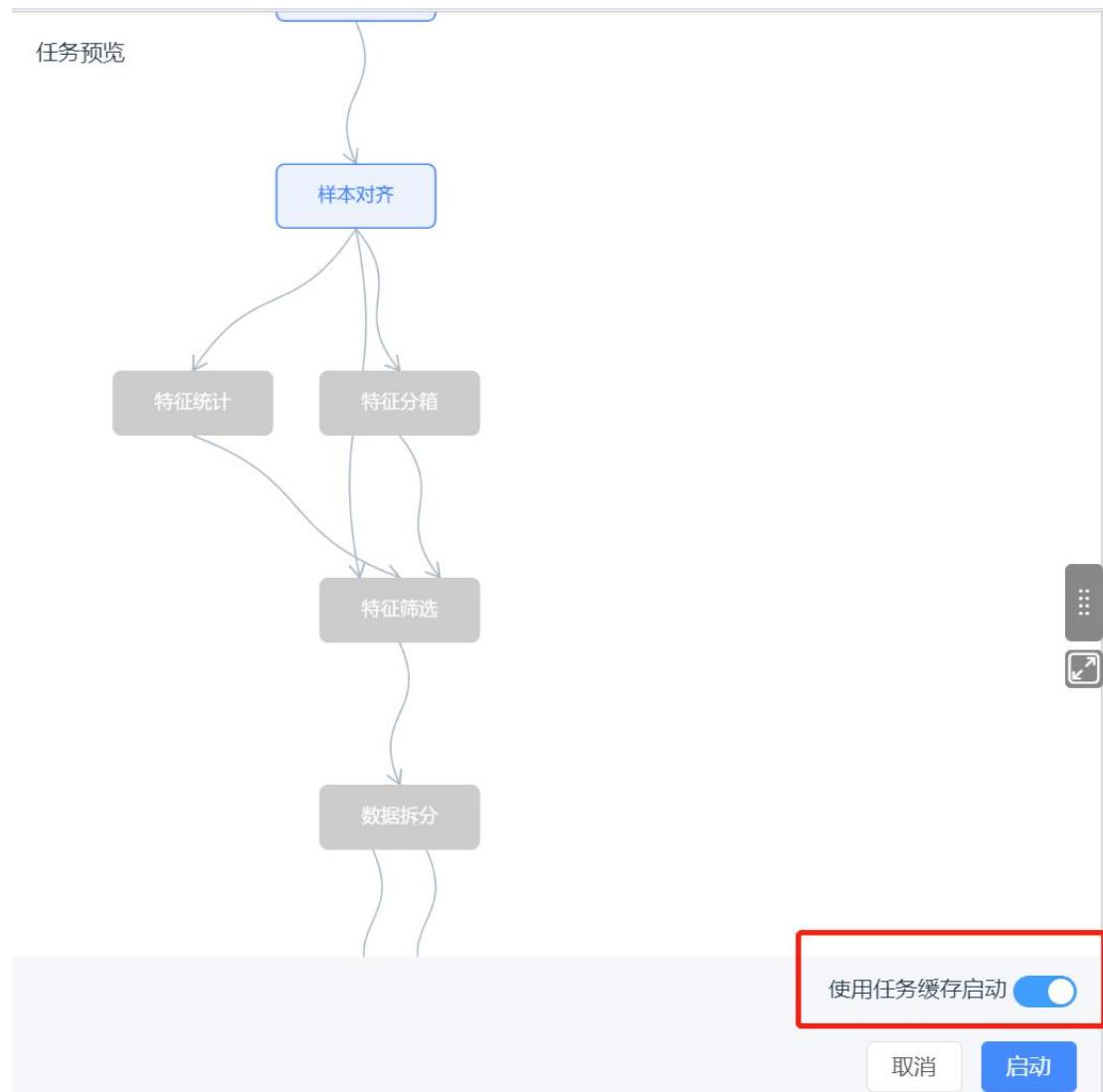
操作:

点击  按钮启动流程



1) 使用缓存启动流程

任务预览页面，勾选任务缓存启动，则启动任务默认跳过已缓存（上一次流程执行完成）的组件，从未缓存的组件往下执行流程



2) 不使用缓存启动流程

任务预览页面，不勾选任务缓存启动，则流程从头开始执行。

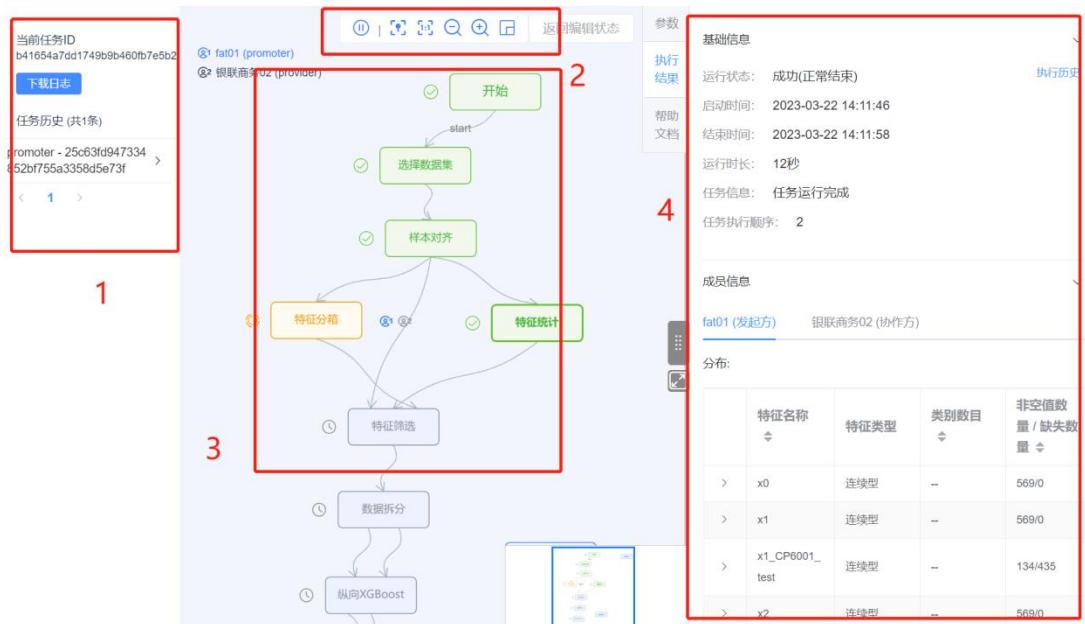


点击“启动”任务跳转任务详情面板

4. 流程执行中

操作:

流程详情页点击  跳转任务详情面板, 无任务详情, 则此按钮为置灰状态。



(1) 任务历史

包含当前执行的任务和历史，点击“查看更多”跳转流程执行结果页面；

promoter - 2a81f4f43b4d4d
588d50b529d8b2f638

创建者: kaylee
 开始时间: 2021-12-22
 间: 16:27:33
 任务进度: 60
 度:
 任务状态: running
 态:

[查看更多 ↗](#)

点击下载日志，可查看当前成员的详细执行日志；

当前任务ID
2a81f4f43b4d4d588d50b529
d8b2f638

[下载日志](#)

(2) 画布操作按钮

点击  可暂停流程；

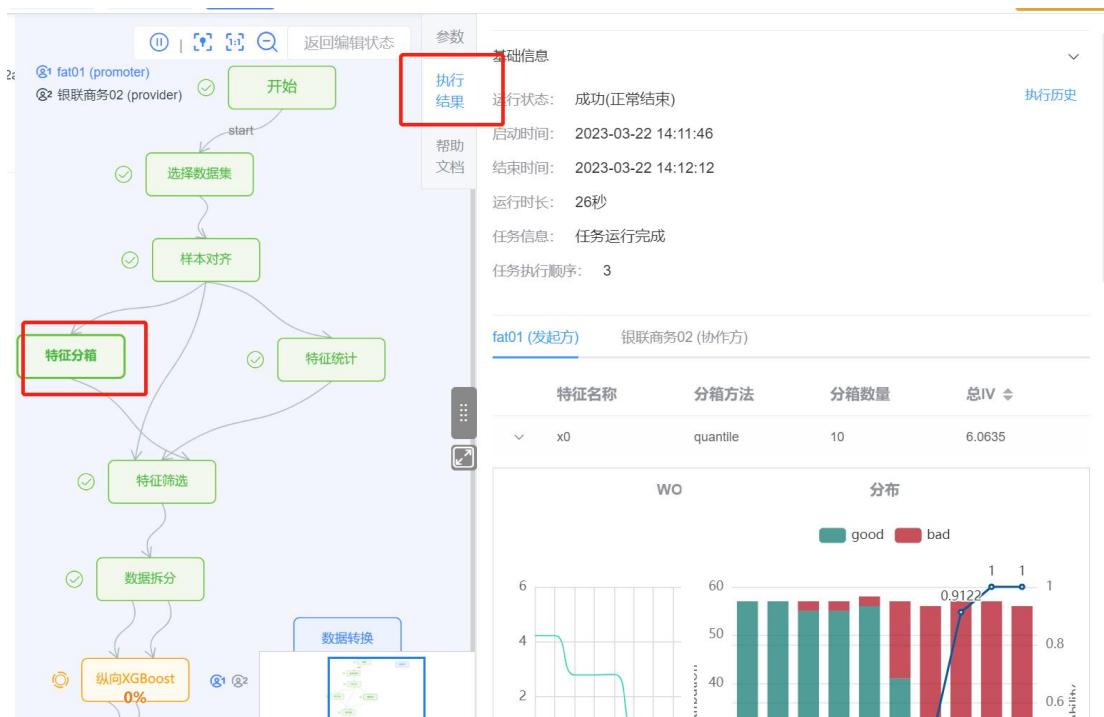
(3) 展示流程执行进度

(4) 展示组件的参数详情，执行结果等，不可编辑

5. 流程执行完成

操作:

流程执行完成，在流程详情页-组件-执行结果可查看具体组件的输出结果。

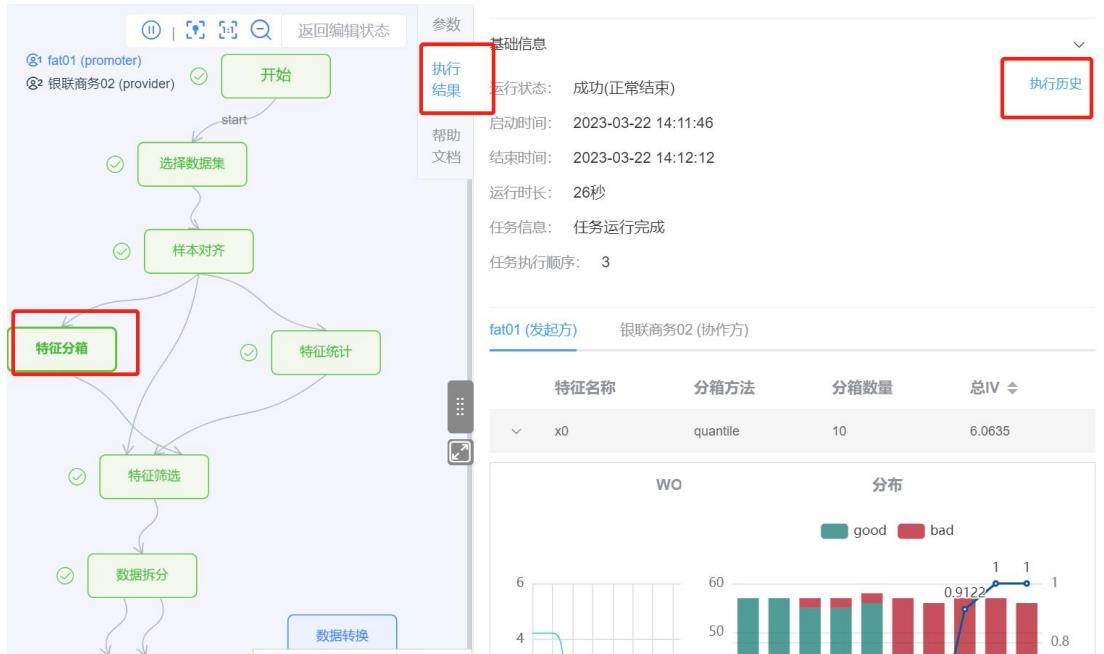


任务执行完成后，在项目详情页面-流程列表/模型列表/衍生数据集列表可查看生成的流程/模型/衍生数据集记录。

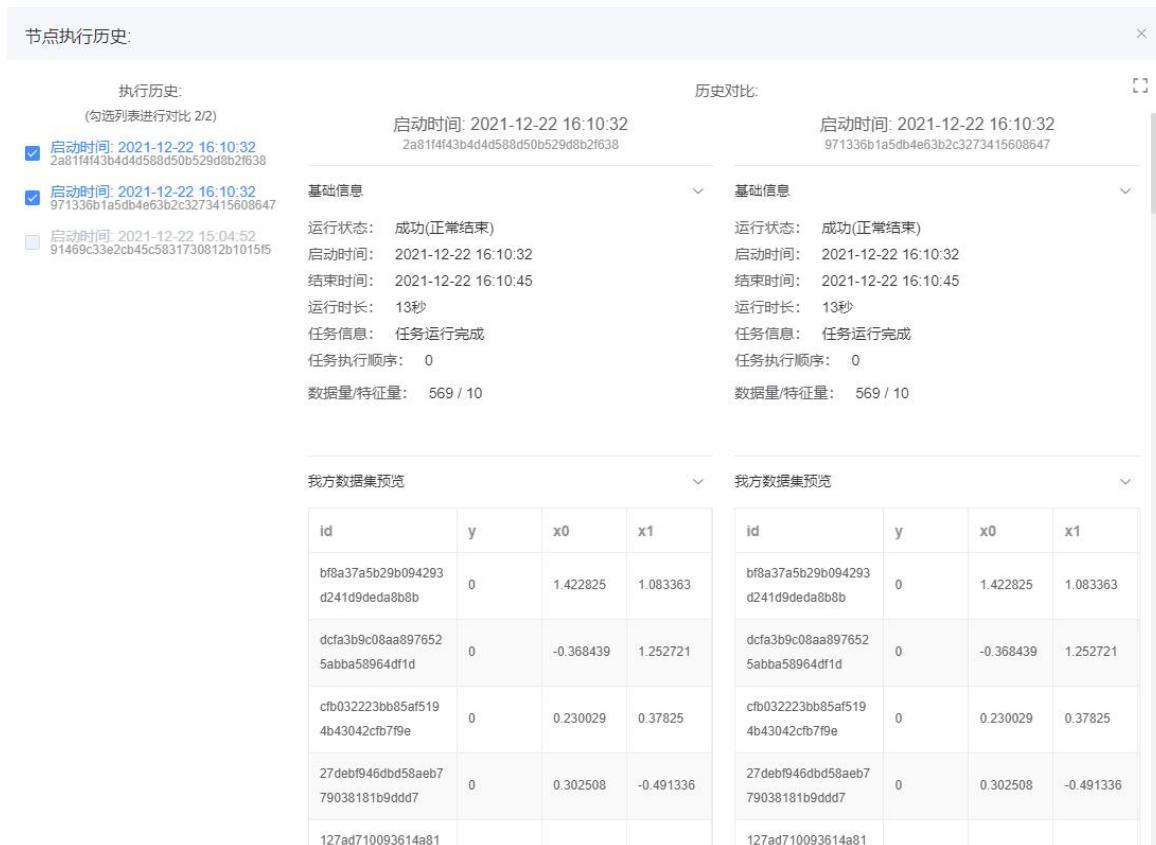
6. 节点执行历史

操作:

- 1) 任务详情面板-点击组件-执行结果-执行历史，打开节点执行历史窗口。



2) 可勾选节点执行历史做对比



This screenshot shows a comparison of execution histories for selected nodes. At the top, there is a header '节点执行历史:' (Node Execution History) with a close button '×'. Below it, a section titled '执行历史: (勾选列表进行对比 2/2)' lists two entries:

- 启动时间: 2021-12-22 16:10:32
2a81f4f43b4d4d588d50b529d8b2f638
- 启动时间: 2021-12-22 16:10:32
97136b1a5db4e63b2c3273415608647
- 启动时间: 2021-12-22 15:04:52
91469c33e2cb45c5831730812b1015f5

Below these, a '基础信息' (Basic Information) table compares the two selected runs:

启动时间	结束时间	运行时长	任务信息	任务执行顺序
2021-12-22 16:10:32	2021-12-22 16:10:45	13秒	任务运行完成	0
2021-12-22 16:10:32	2021-12-22 16:10:45	13秒	任务运行完成	0

A '历史对比' (History Comparison) section shows detailed comparison results for the two runs:

启动时间	结束时间	运行时长	任务信息	任务执行顺序
2021-12-22 16:10:32	2021-12-22 16:10:45	13秒	任务运行完成	0
2021-12-22 16:10:32	2021-12-22 16:10:45	13秒	任务运行完成	0

At the bottom, there are two tables for '我方数据集预览' (Local Data Set Preview):

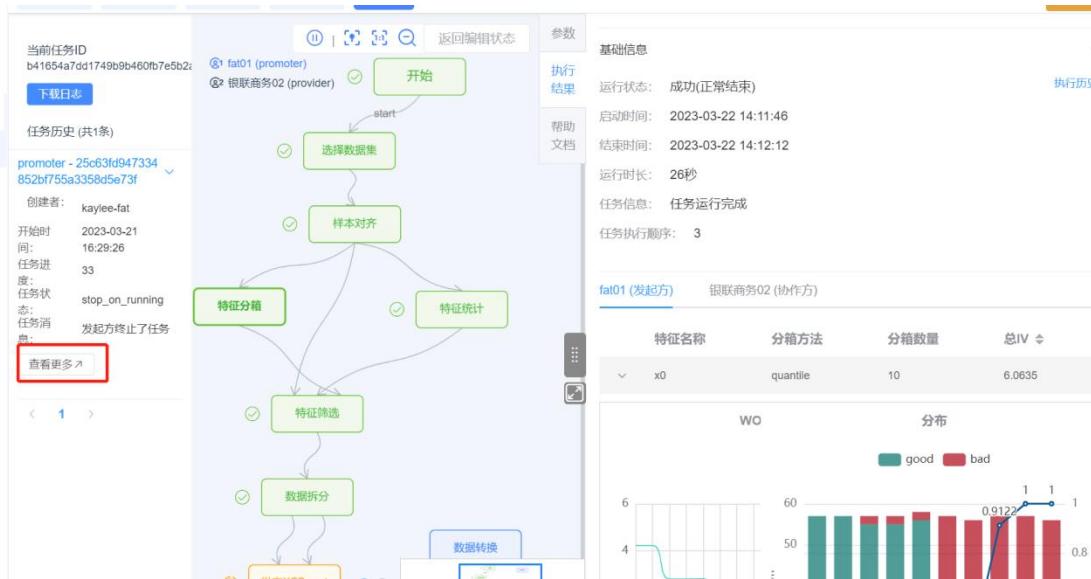
id	y	x0	x1
bf8a37a5b29b094293	0	1.422825	1.083363
d241d9deda8b8b	0	-0.368439	1.252721
dcfa3b9c08aa897652	0	0.230029	0.37825
5abba58964df1d	0	0.302508	-0.491336
127ad710093614a81	0	0.230029	0.37825

id	y	x0	x1
bf8a37a5b29b094293	0	1.422825	1.083363
d241d9deda8b8b	0	-0.368439	1.252721
dcfa3b9c08aa897652	0	0.230029	0.37825
5abba58964df1d	0	0.302508	-0.491336
127ad710093614a81	0	0.230029	0.37825

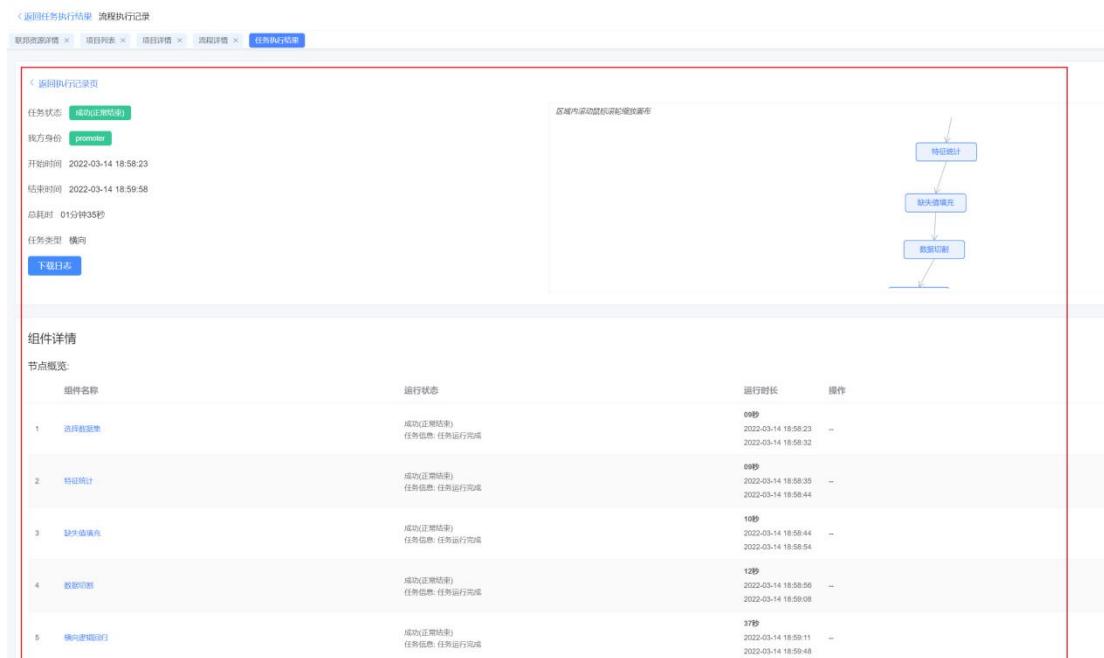
7. 任务执行结果

操作：

流程详情页-任务详情面板-任务历史-点击“查看更多”：

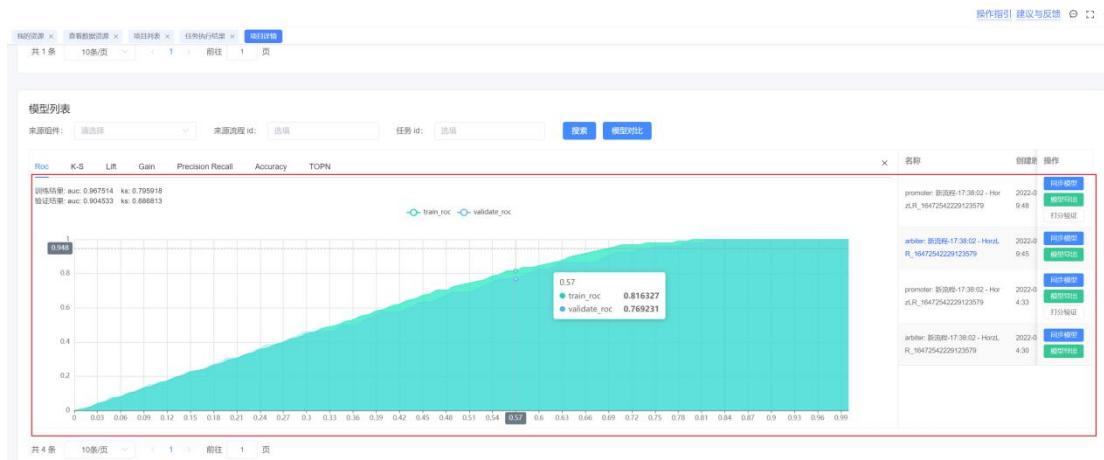


跳转任务执行结果页右侧是目录导航栏，左侧为结果展示，可查看所有历史结果展示：

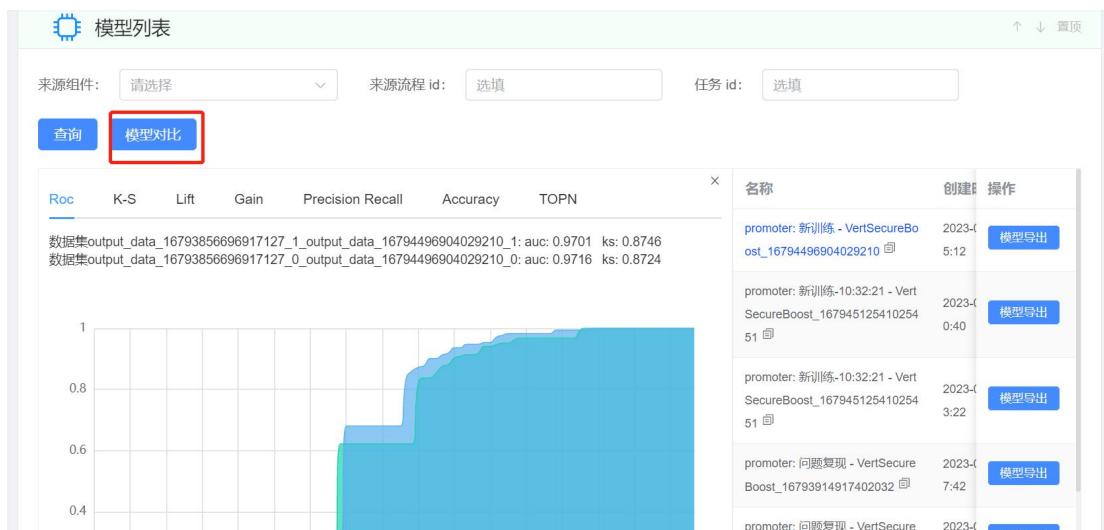


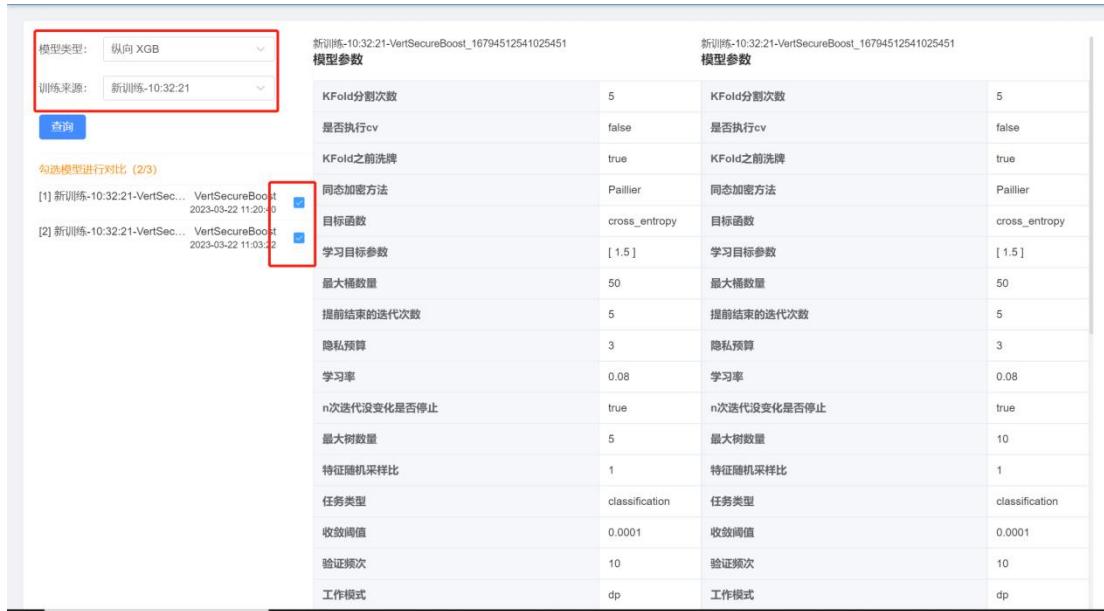
3.6.2.7 模型列表

在项目详情页面，根据流程生成模型列表，可视化的形式展示训练结果和验证结果：



点击模型对比按钮，跳转模型对比页，根据搜索条件，选择模型任务进行参数及模型结果的数据对比：





The screenshot shows a comparison of two XGBoost models. The left panel displays the 'Model Type' (縱向 XGB) and 'Training Source' (新训练-10:32:21). The right panel lists various model parameters such as KFold split count (5), execution method (Paillier), target function (cross_entropy), and learning rate (0.08).

参数	值	参数	值
KFold分割次数	5	KFold分割次数	5
是否执行cv	false	是否执行cv	false
KFold之前洗牌	true	KFold之前洗牌	true
同态加密方法	Paillier	同态加密方法	Paillier
目标函数	cross_entropy	目标函数	cross_entropy
学习目标参数	[1.5]	学习目标参数	[1.5]
最大桶数量	50	最大桶数量	50
提前结束的迭代次数	5	提前结束的迭代次数	5
隐私预算	3	隐私预算	3
学习率	0.08	学习率	0.08
n次迭代没变化是否停止	true	n次迭代没变化是否停止	true
最大树数量	5	最大树数量	10
特征随机采样比	1	特征随机采样比	1
任务类型	classification	任务类型	classification
收敛阈值	0.0001	收敛阈值	0.0001
验证频次	10	验证频次	10
工作模式	dp	工作模式	dp

3.6.2.8 生成衍生数据集

保存组件输出结果作衍生数据集，可用于其他流程使用。目前样本对齐，分箱，特征筛选和缺失值填充组件，执行完成均会生成衍生数据集：



The screenshot shows a chart comparing training and validation ROC curves and a list of generated derived datasets.

衍生数据集

衍生数据集名	来源	来源流程 id	任务 id	操作
train_roc	auc: 0.904533	ks: 0.686813		查看任务
validate_roc				查看任务

衍生数据资源

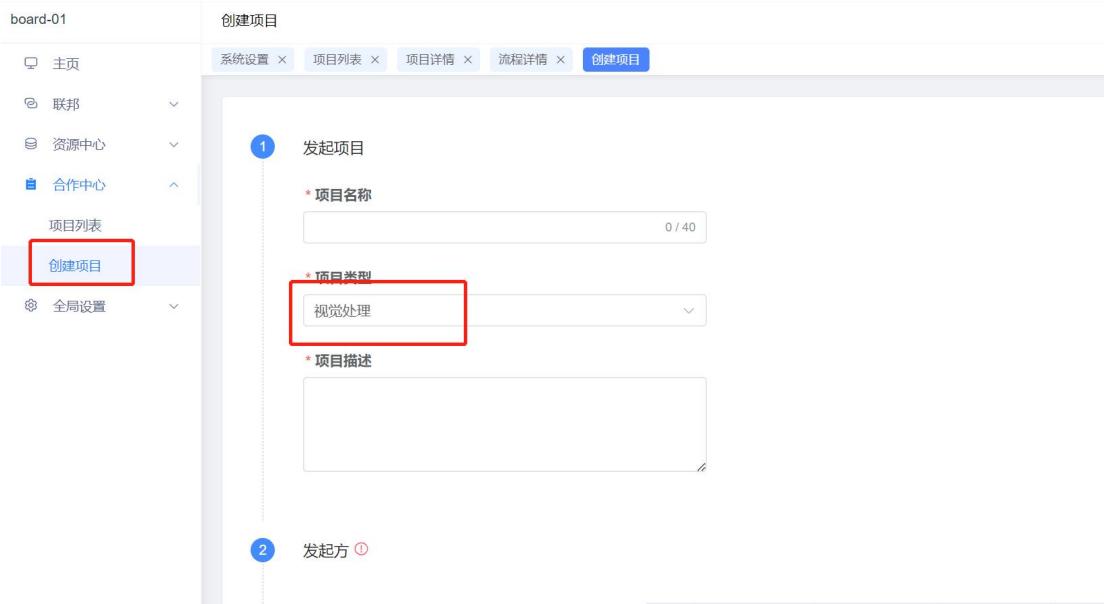
衍生数据资源名	数据资源来源	成员列表	数据量	使用次数	创建时间	查看任务	操作
衍生数据-17.38.02(FillMissingValue)_20220314180846	缺失值填充	fa01 fa02 fa03	特征量: 14 样本量: 227	0	2022-03-14 18:58:54	查看任务	删除
衍生数据-17.38.02(FillMissingValue)_20220314180530	缺失值填充	fa01 fa02 fa03	特征量: 14 样本量: 227	0	2022-03-14 18:53:36	查看任务	删除

3.6.3 创建深度学习项目

3.6.3.1 初始化创建项目

在创建项目时项目类型选择视觉处理，进入深度学习的项目创建，之后操作

与机器学习建模任务的创建流程一样，需要添加协作方后“保存项目”。



3.6.3.2 新增训练流程

项目创建成功后，向项目中添加资源深度学习的数据集，便可以进行建模操作，选择“新增数据流程”。

新建训练流程，选择流程类型，流程支持目标检测和图像分类。



选择模型类型进入，模型流程设置，流程设置分为三个步骤：

1. 配置基本设置；
2. 数据输入与处理；
3. 调整流程参数。

1. 填写基本设置

操作指引 建议与反馈 ⊖ 你好, aisen ▾ 关闭其他标签

我的资源 × 创建项目 × 项目列表 × 项目详情 × 深度学习流程详情

基本设置

* 流程名称:
新流程-19:01:09

流程描述:
请输入流程描述

训练类型:
图像分类

下一步

① 基本设置
② 数据输入与处理
③ 调整流程参数

2. 选择数据资源进入模型训练

数据输入与处理

参数

选择数据资源

发起方:

fat03 选择数据资源

数据资源名称: Nova02.26 ⓘ
数据资源id: 956fc8ab01ed4a68b2f530881a22a2ac
关键词: 目标检测
数据总量/已标注: 300 / 300
样本分类: 目标检测
标注标签: **apple** banana orange

协作方:

fat04 选择数据资源

数据资源名称: Nova的目标检测001 ⓘ
数据资源id: 79f037bf2daa48168caec27d0e3db232
关键词: nova
数据总量/已标注: 299 / 299
样本分类: 目标检测
标注标签: **apple** banana orange

数据切割

训练与验证数据比例 (%) :

训练: 验证:

- 70 + - 30 +

上一步 **下一步**

① 基本设置

② 数据输入与处理

③ 调整流程参数

3. 调整训练模型的参数，确认后点击“开始训练”。

调整流程参数

参数 **执行结果**

* 算法类型:
paddle_detection

* 模型名称:
yolov3_r34_voc

* 迭代次数:
2

* 聚合步长:
10

* 学习率:
0.00001

* 图片通道数:
 彩色 (3) 黑白 (1)

* 图片宽度(px):
608

* 图片高度(px):
608

* 批量大小:
10

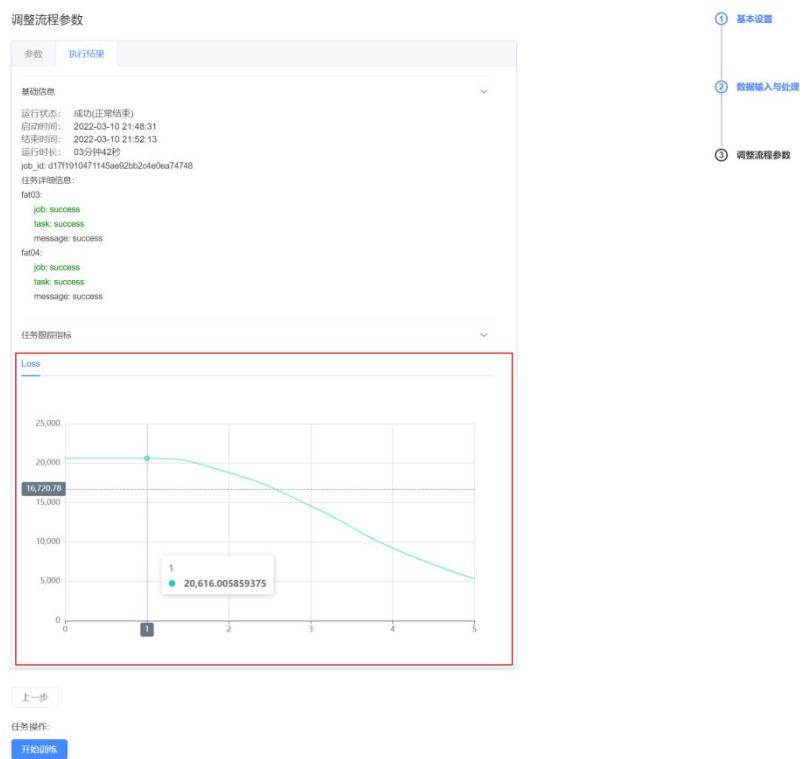
上一步

① 基本设置

② 数据输入与处理

③ 调整流程参数

运行成功，生成任务跟踪指标，完成训练。



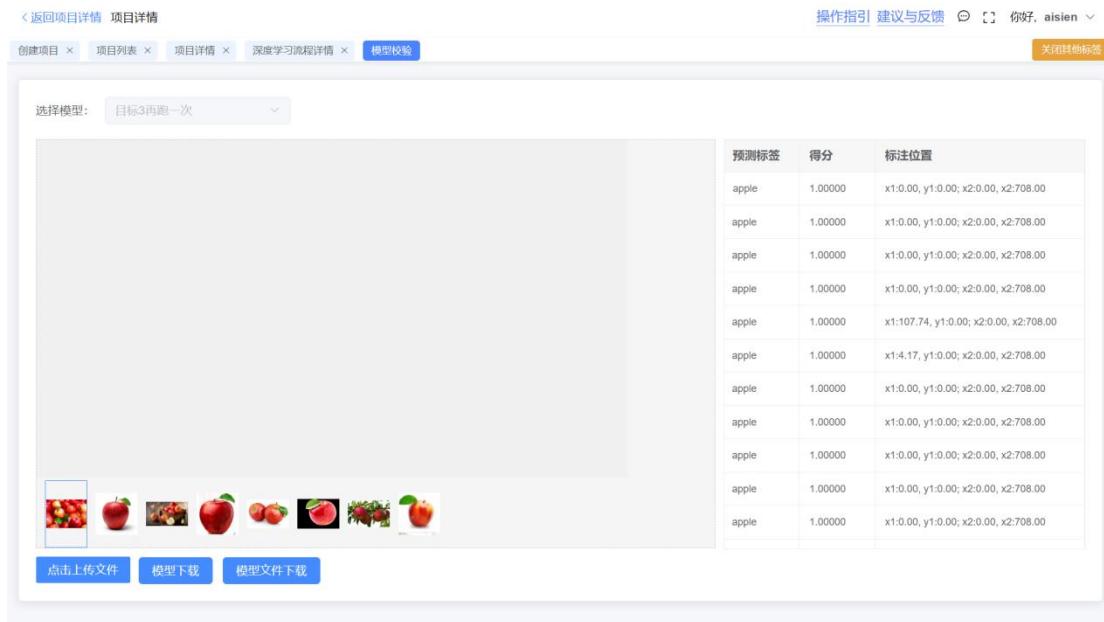
3.6.3.3 模型校验

模型成功后在模型训练列表中，可以针对此次的训练结果进行模型校验，选择“校验”，进入“模型校验”页面，上传需要预测的图片。

训练列表			
训练	创建者	创建时间	操作
running 目标1	03账号1	2022-03-15 16:28:25	查看 更多
editing 图像1	03账号1	2022-03-15 09:55:00	查看 更多
editing 新流程-17:36:52	03账号1	2022-03-11 17:36:52	查看 更多
editing 目标7	03账号1	2022-03-11 09:59:37	查看 更多
running 目标7	03账号1	2022-03-10 22:25:25	查看 更多
running 目标6	03账号1	2022-03-10 22:12:27	查看 更多
success 目标5	03账号1	2022-03-10 21:45:21	查看 校验 更多
success 目标5	03账号1	2022-03-10 15:11:01	查看 校验 更多
success 目标4	03账号1	2022-03-10 14:01:58	查看 校验 更多
success 目标1	03账号1	2022-03-09 18:23:05	查看 校验 更多



输出模型预测的结果:



3.6.4 创建 SQL 探索任务

联邦 SQL 是一种允许在多个数据库上执行查询的技术。它可以将多个数据库组合成一个虚拟数据库，使查询和分析变得更加简单和高效。通过联邦 SQL，我们可以在不同的数据库上执行复杂查询，而无需将数据复制到一个地方。这样可以减少数据冗余，提高数据的一致性和准确性。

3.6.4.1 初始化创建项目

在创建项目时项目类型选择 SQL 探索，添加协作方后“保存项目”。

创建项目

项目详情 × SQL探索 × 查看数据资源 × 添加资源 × **创建项目**

1 发起项目

* 项目名称: kaylee-sql探索 12 / 40

* 项目类型: SQL探索 (机器学习 视觉处理)
 使用 sql 语句安全地跨成员、跨数据源进行联合数据分析。

* 项目描述: kaylee-sql探索

2 发起方

	数据资源id	数据资源名称	数据库类
dev03	35efd2ef7c1541bab96a3b7275b6c6c7	服务正常	
	联系电话: 13168730657		
	邮箱: lyzane.it@gmail.com		
	gateway: 121.37.222.190:31053		

3 协作方 [+添加更多协作方](#)

添加成功后进入 SQL 探索的项目详情页:

项目详情

查看数据资源 × 添加资源 × 创建项目 × 项目列表 × [返回顶部](#)

version: 3.7.0 作者: admin ()

项目简介: kaylee-sql探索
项目类型: SQL探索
项目归属者: admin
由 admin 创建于 2023-06-16 11:47:46

SQL探索

关联查询SQL引擎, 用统一的语义SQL语句实现跨平台的数据查询与分析, 为了数据安全, 保护各成员权益:
 - 不允许查询其他成员的数据明细。
 - 已方如果允许查询明细, 但最多返回 100 行。
 - 多方统计查询结果最多返回 10000 行。

开始搜索

参与成员

发起方: dev03	协作方: dev004							
正在审核状态	▲ 刚刚正常							
+添加资源到此项目								
数据资源	关键词	数据量	数据库类型	数据源	库名	表名	状态	操作
lpch_supplier_local 746928862a34bd48a1b12e4ba61ca17	1	10000 rows * 7 columns	ClickHouse	lpch	lpch	supplier_local	已授权	预览 删除
wefe_board_1_account								

3.6.4.2 成员审核

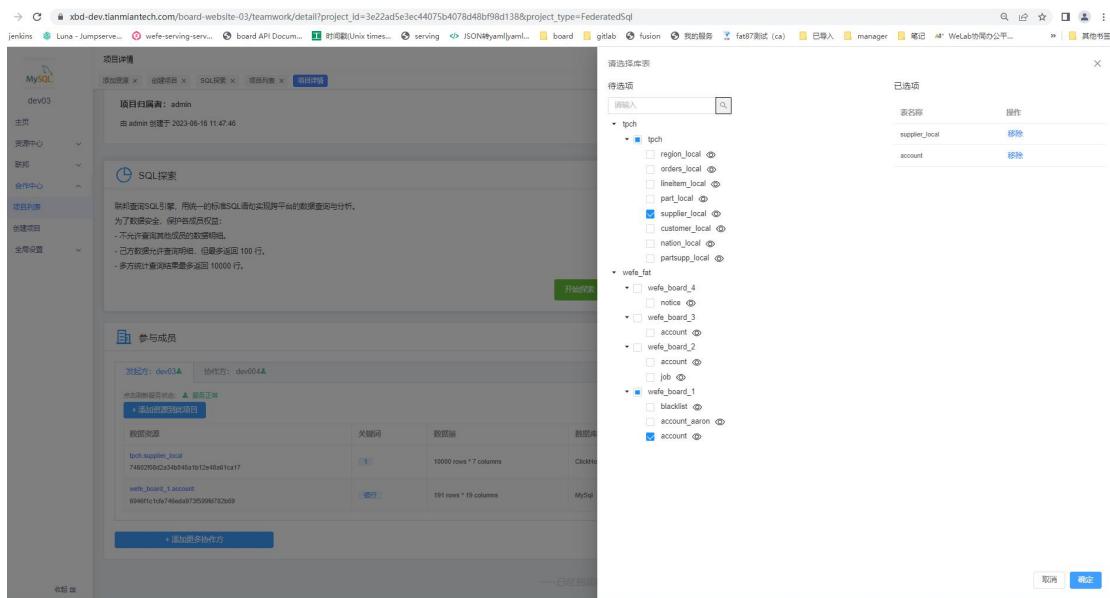
成员审核同 3.6.2.3 章节。

3.6.4.3 SQL 资源审核

1. 添加 sql 资源

操作:

1) 项目详情页面，点击“添加资源到此项目”，右侧弹出资源选择面板。



2) 选择面板，层级展示依次为数据源名称-库名-表名，勾选表，点击“确定”按钮，即可添加项目资源；

请选择库表

待选项

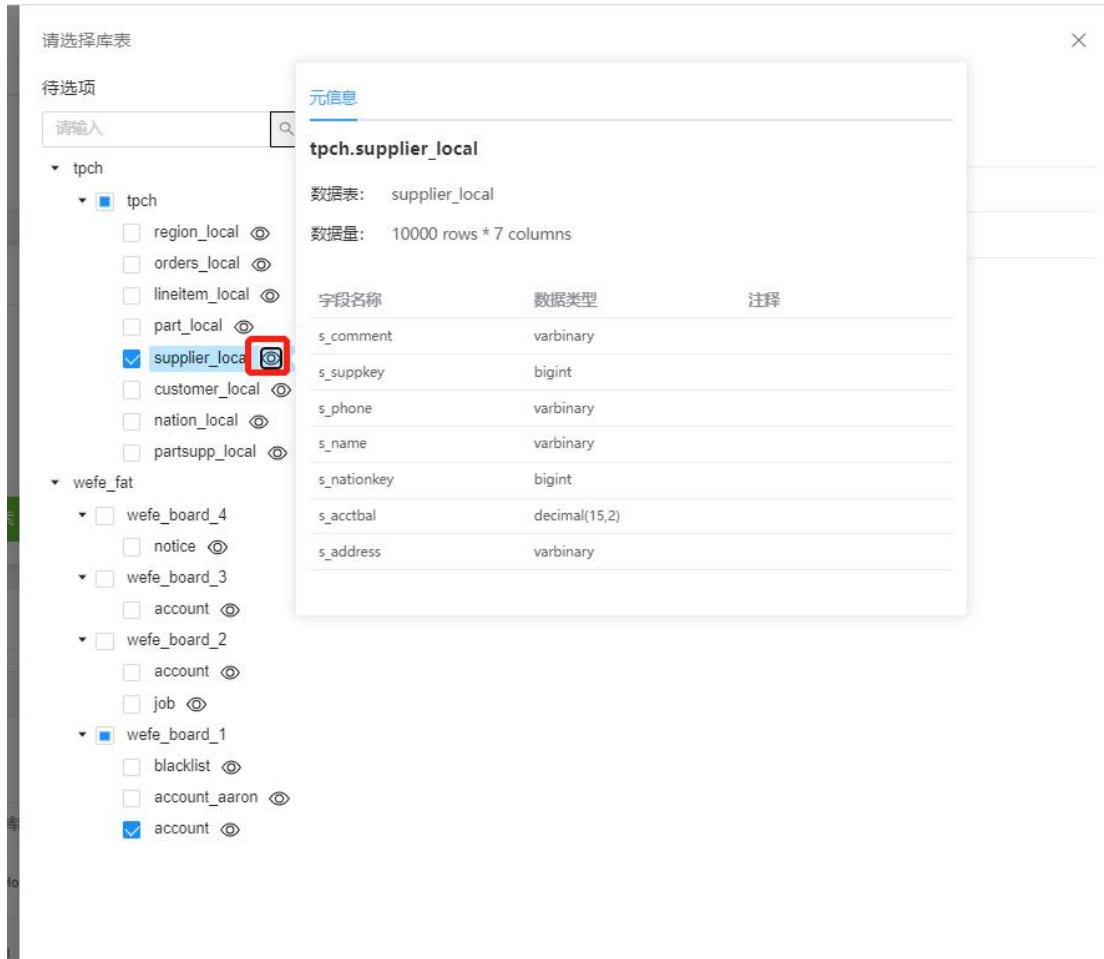
- ▼ tpch
 - region_local ◎
 - orders_local ◎
 - lineitem_local ◎
 - part_local ◎
 - supplier_local ◎
 - customer_local ◎
 - nation_local ◎
 - partsupp_local ◎
- ▼ wefe_fat
 - ▼ wefe_board_4
 - notice ◎
 - ▼ wefe_board_3
 - account ◎
 - ▼ wefe_board_2
 - account ◎
 - job ◎
 - ▼ wefe_board_1
 - blacklist ◎
 - account_aaron ◎
 - account ◎

已选项

表名称	操作
supplier_local	移除
account	移除

 取消  确认

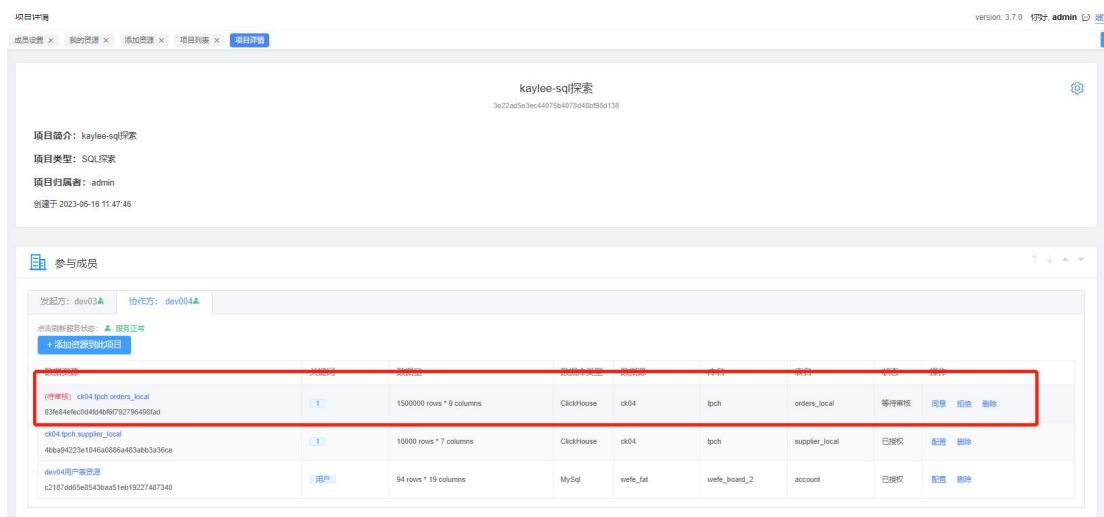
点击眼睛图标可以查看数据表的元数据。



2. SQL 资源审核

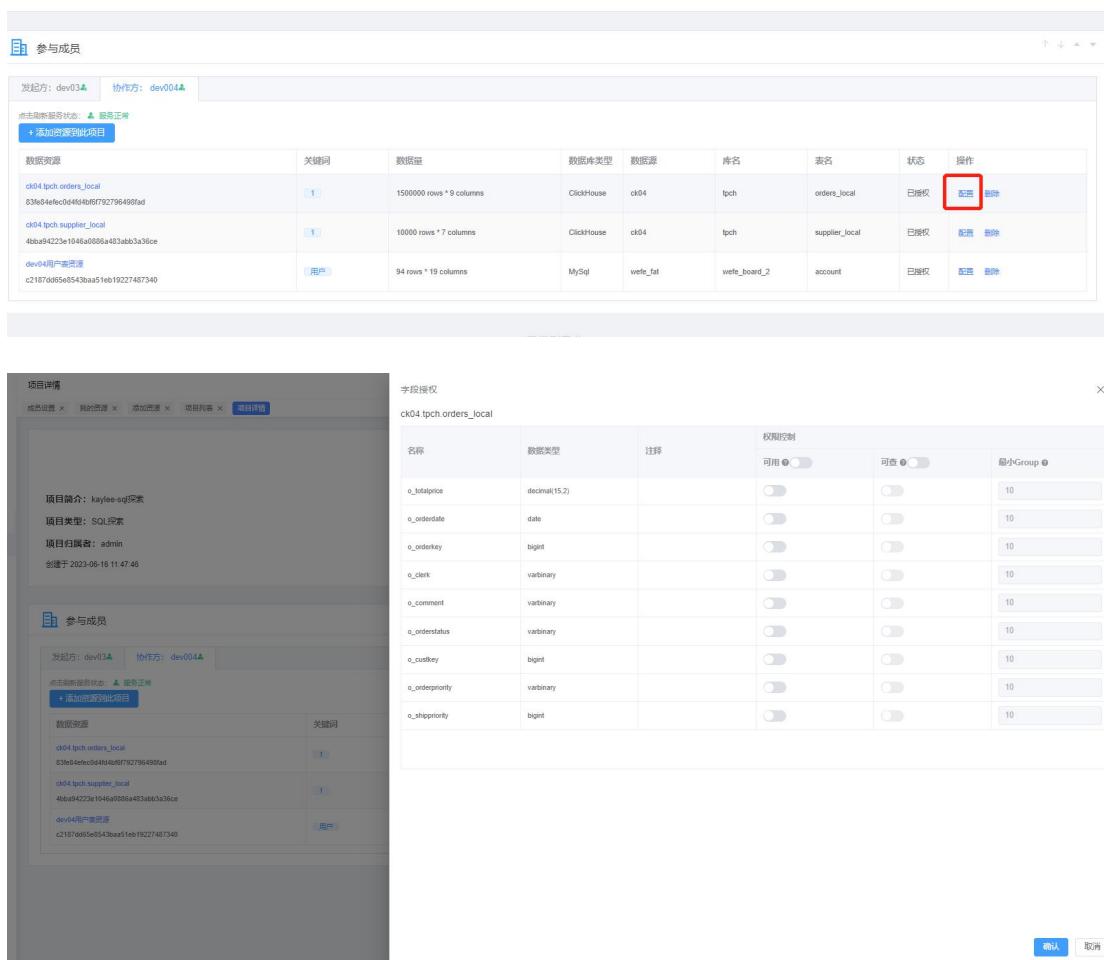
添加协作方的数据资源需要成员审核。其他数据集删除、移除等操作同

3.6.2.4 章节。



3. 配置数据资源可用字段

协作方授权 sql 数据资源可被项目使用后，需要配置项目内的 sql 可用字段，点击“配置”按钮，弹出字段授权选择面板。



数据资源	关键词	数据量	数据源类型	数据源	库名	表名	状态	操作
ck04_tpch.orders_local 83fe04fec0d4454b8f792796498fad	1	1500000 rows * 9 columns	ClickHouse	ck04	tpch	orders_local	已授权	配置 删除
ck04_tpch.supplier_local 4babb4223e1046a0886a453abb3a36ce	1	10000 rows * 7 columns	ClickHouse	ck04	tpch	supplier_local	已授权	配置 删除
dev04用户资源 c2187dd65e6543baa51eb19227487340	用户	94 rows * 19 columns	MySQL	wefe_fat	wefe_board_2	account	已授权	配置 删除

项目详情																																																										
成员设置	我的资源	添加资源	项目列表	项目详情																																																						
<p>项目简介: keylee-sql探索</p> <p>项目类型: SQL探索</p> <p>项目归属者: admin</p> <p>创建于: 2023-06-18 11:47:46</p> <p>参与成员</p> <p>发起方: dev03 协作方: dev04</p> <p>点击切换至正式态: ▲ 正常 + 混合模式</p> <p>数据资源</p> <p>ck04_tpch.orders_local 83fe04fec0d4454b8f792796498fad</p> <p>ck04_tpch.supplier_local 4babb4223e1046a0886a453abb3a36ce</p> <p>dev04用户资源 c2187dd65e6543baa51eb19227487340</p> <p>关键词</p>																																																										
<p>字段授权</p> <p>ck04_tpch.orders_local</p> <table border="1"> <thead> <tr> <th>名称</th> <th>数据类型</th> <th>注释</th> <th>权限控制</th> <th>最小Group</th> </tr> </thead> <tbody> <tr> <td>o_totalprice</td> <td>decimal(15,2)</td> <td></td> <td><input checked="" type="checkbox"/> 可用</td> <td>10</td> </tr> <tr> <td>o_orderdate</td> <td>date</td> <td></td> <td><input checked="" type="checkbox"/> 可用</td> <td>10</td> </tr> <tr> <td>o_orderkey</td> <td>bigint</td> <td></td> <td><input checked="" type="checkbox"/> 可用</td> <td>10</td> </tr> <tr> <td>o_clerk</td> <td>varbinary</td> <td></td> <td><input checked="" type="checkbox"/> 可用</td> <td>10</td> </tr> <tr> <td>o_comment</td> <td>varbinary</td> <td></td> <td><input checked="" type="checkbox"/> 可用</td> <td>10</td> </tr> <tr> <td>o_orderstatus</td> <td>varbinary</td> <td></td> <td><input checked="" type="checkbox"/> 可用</td> <td>10</td> </tr> <tr> <td>o_custkey</td> <td>bigint</td> <td></td> <td><input checked="" type="checkbox"/> 可用</td> <td>10</td> </tr> <tr> <td>o_orderpriority</td> <td>varbinary</td> <td></td> <td><input checked="" type="checkbox"/> 可用</td> <td>10</td> </tr> <tr> <td>o_shippriority</td> <td>bigint</td> <td></td> <td><input checked="" type="checkbox"/> 可用</td> <td>10</td> </tr> </tbody> </table>									名称	数据类型	注释	权限控制	最小Group	o_totalprice	decimal(15,2)		<input checked="" type="checkbox"/> 可用	10	o_orderdate	date		<input checked="" type="checkbox"/> 可用	10	o_orderkey	bigint		<input checked="" type="checkbox"/> 可用	10	o_clerk	varbinary		<input checked="" type="checkbox"/> 可用	10	o_comment	varbinary		<input checked="" type="checkbox"/> 可用	10	o_orderstatus	varbinary		<input checked="" type="checkbox"/> 可用	10	o_custkey	bigint		<input checked="" type="checkbox"/> 可用	10	o_orderpriority	varbinary		<input checked="" type="checkbox"/> 可用	10	o_shippriority	bigint		<input checked="" type="checkbox"/> 可用	10
名称	数据类型	注释	权限控制	最小Group																																																						
o_totalprice	decimal(15,2)		<input checked="" type="checkbox"/> 可用	10																																																						
o_orderdate	date		<input checked="" type="checkbox"/> 可用	10																																																						
o_orderkey	bigint		<input checked="" type="checkbox"/> 可用	10																																																						
o_clerk	varbinary		<input checked="" type="checkbox"/> 可用	10																																																						
o_comment	varbinary		<input checked="" type="checkbox"/> 可用	10																																																						
o_orderstatus	varbinary		<input checked="" type="checkbox"/> 可用	10																																																						
o_custkey	bigint		<input checked="" type="checkbox"/> 可用	10																																																						
o_orderpriority	varbinary		<input checked="" type="checkbox"/> 可用	10																																																						
o_shippriority	bigint		<input checked="" type="checkbox"/> 可用	10																																																						

➤ 字段授权字段参考如下:

```

select
    t2.admin_role,
    count(t2.admin_role) 可查
from
    promoter_0.wefe_fat.wefe_board_1.account as t1,
    provider_0.wefe_fat.wefe_board_2.account as t2
where
    t1.email = t2.email 可用
group by
    t2.admin_role;
  
```

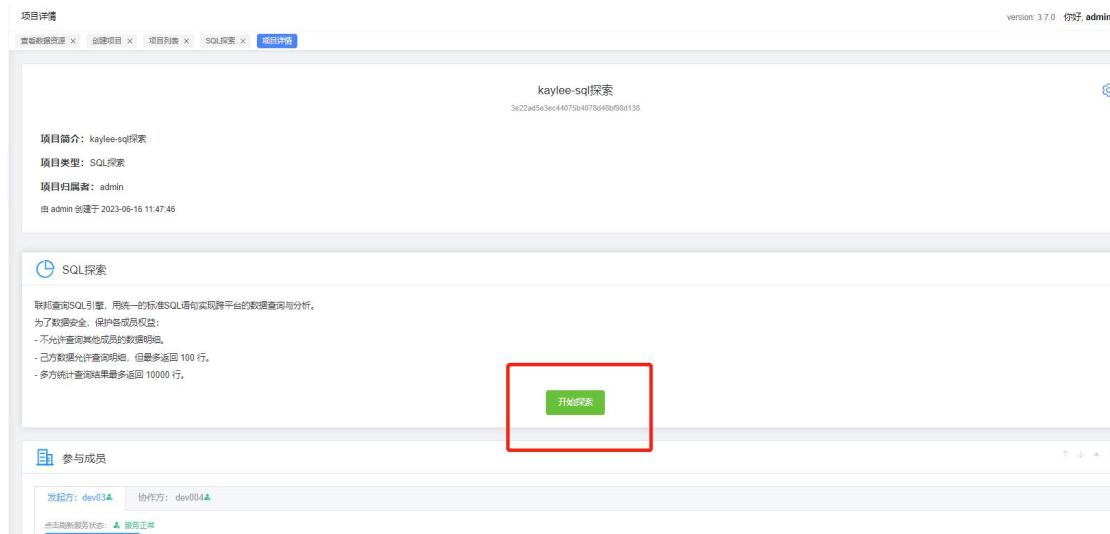
3.6.4.4 SQL 探索

sql 探索是用统一的标准 SQL 语句实现跨平台的数据查询与分析，为了数据安全，保护各成员权益：

- 不允许查询其他成员的数据明细。
- 己方数据允许查询明细，但最多返回 100 行。
- 多方统计查询结果最多返回 10000 行。

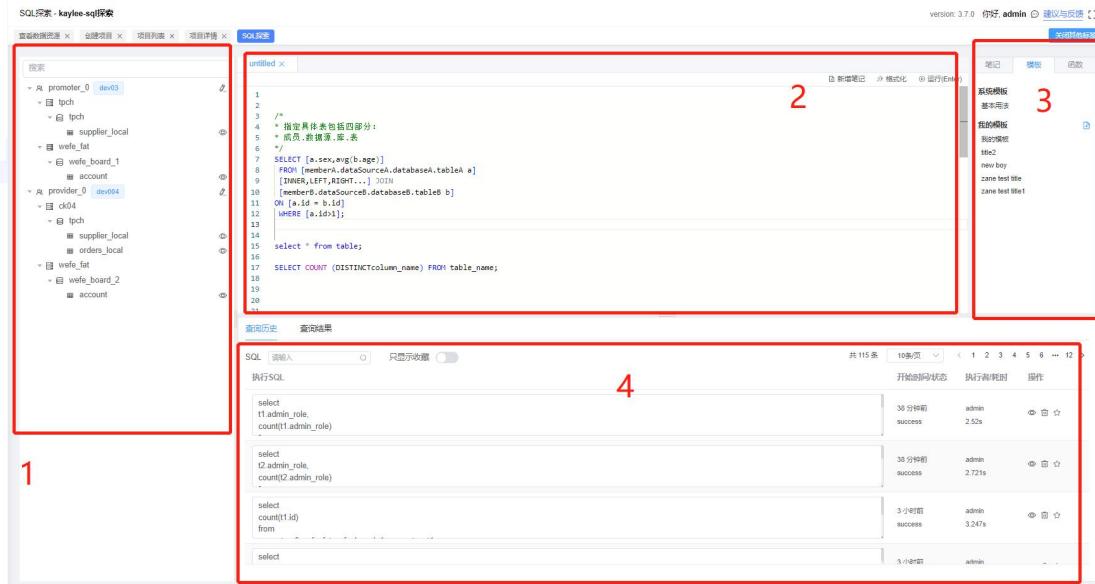
操作：

在项目详情页，点击“开始探索”按钮，进入 sql 探索页面。



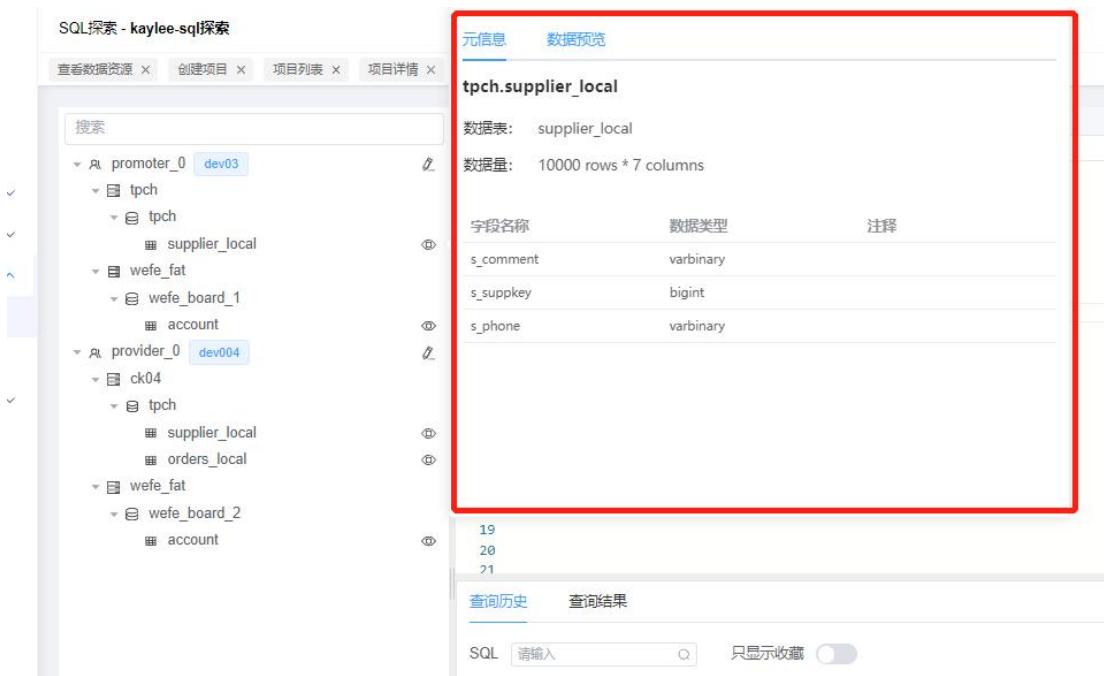
The screenshot shows the 'Project Details' page for a project named 'kaylee-sql探索'. The top navigation bar includes tabs for 'Project Details', 'Data Sources', 'Create Project', 'Project List', 'SQL Exploration', and 'Project Settings'. The right side of the header shows the version '3.7.0' and the user 'admin'. The main content area displays the project summary, including the project name, type ('SQL Exploration'), owner ('admin'), and creation date ('2023-08-16 11:47:46'). Below this, there's a section titled 'SQL Exploration' with a green button labeled 'Start Exploration' which is highlighted with a red box. At the bottom, there's a 'Participants' section showing 'Initiator: dev03' and 'Collaborator: dev004', along with a status message 'Last update status: Normal'.

Sql 探索页面分为以下 4 个区域，如下图：



1. 资源区

展示项目内可用数据库表，层级关系为成员-数据源名-数据库名-表名。点击已方的眼睛图标可以查看表元数据及数据预览信息；

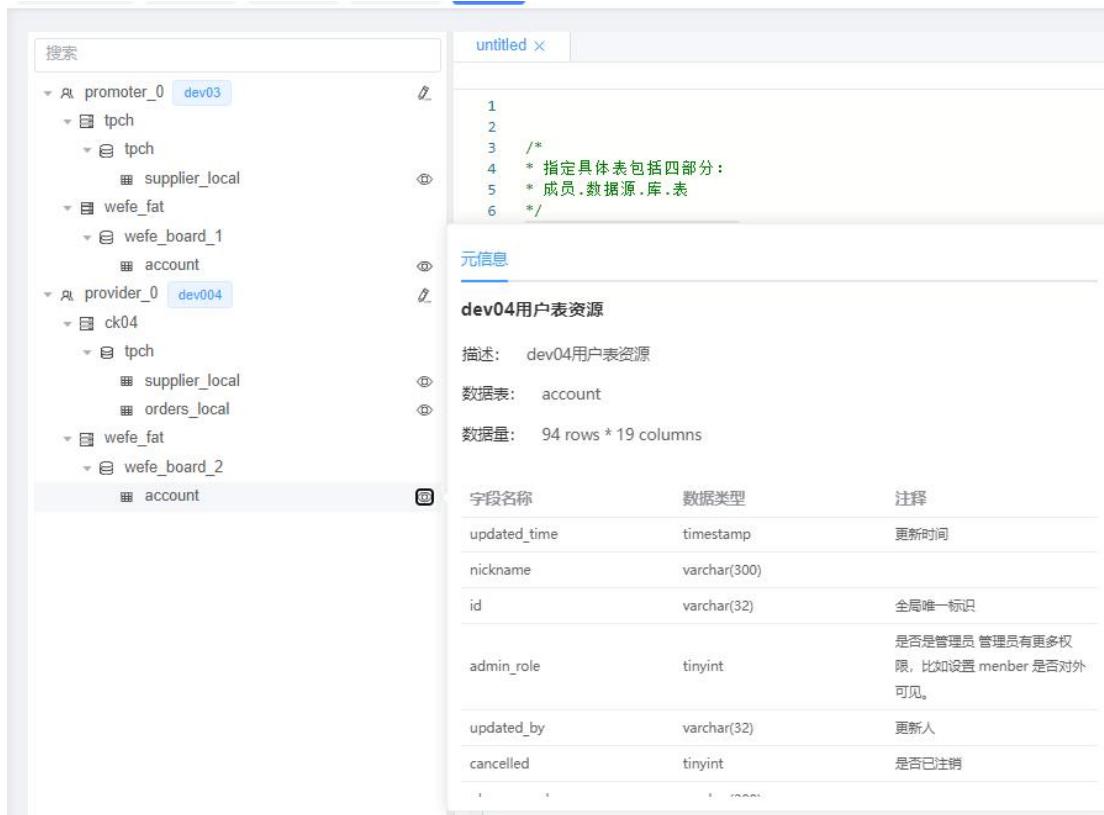


The screenshot shows the SQL Explorer interface with the '元信息' (Metadata) tab selected for the table 'tpch.supplier_local'. The interface includes:

- 左侧树状目录:** 显示了项目的树状结构，包括 'promoter_0'、'provider_0'、'ck04' 等。
- 右侧元信息视图:** 包含以下内容：
 - 数据表:** supplier_local
 - 数据量:** 10000 rows * 7 columns
 - 字段列表:**

字段名称	数据类型	注释
s_comment	varbinary	
s_suppkey	bigint	
s_phone	varbinary	
- 底部工具栏:** 包括 'SQL' 输入框、'只显示收藏' 按钮和 '查询历史' / '查询结果' 切换按钮。

点击协作方的眼睛图标可以查看对方表的元数据信息。



搜索

- + promoter_0 dev03
 - tpch
 - tpch
 - supplier_local
 - wefe_fat
 - wefe_board_1
 - account
 - provider_0 dev04
 - ck04
 - tpch
 - supplier_local
 - orders_local
 - wefe_fat
 - wefe_board_2
 - account

untitled x

```

1
2
3  /*
4  * 指定具体表包括四部分：
5  * 成员.数据源.库.表
6 */

```

元信息

dev04用户表资源

描述: dev04用户表资源

数据表: account

数据量: 94 rows * 19 columns

字段名称	数据类型	注释
updated_time	timestamp	更新时间
nickname	varchar(300)	
id	varchar(32)	全局唯一标识
admin_role	tinyint	是否是管理员 管理员有更多权限，比如设置 member 是否对外可见。
updated_by	varchar(32)	更新人
cancelled	tinyint	是否已注销

点击编辑按钮可修改成员名称。



搜索

- + promoter_0 dev03
 - tpch
 - tpch
 - supplier_local
 - wefe_fat
 - wefe_board_1
 - account
 - provider_k dev04
 - ck04
 - tpch
 - supplier_local
 - orders_local
 - wefe_fat
 - wefe_board_2
 - account

untitled x

```

1
2
3  /*
4  * 指定具体表包括四部分：
5  * 成员.数据源.库.表
6 */
7  SELECT [a.sex,avg(b.age)] FROM baseA AS tab1 WHERE provider_k = 'provider_k' GROUP BY a.sex;

```

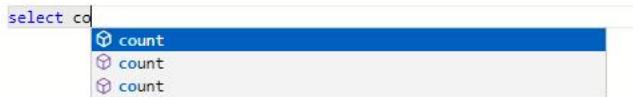
确认

2. sql 编辑区

Sql 编辑区是编写 sql 语句的区域，支持关键词提示，语句格式化，批量执行

sql 及操作快捷键等。

➤ 关键词提示:



➤ 语句格式化:

格式化前:

```
select t2.admin_role,count(t2.admin_role) from promoter_0.wefe_fat.wefe_board_1.account as t1,provider_0.wefe_fat.wefe_board_2.account as t2 where t1.email= t2.email group by t2.admin_role;
```

格式化后:

```
select
    t2.admin_role,
    count(t2.admin_role)
from
    promoter_0.wefe_fat.wefe_board_1.account as t1,
    provider_0.wefe_fat.wefe_board_2.account as t2
where
    t1.email = t2.email
group by
    t2.admin_role;
```

➤ 执行 sql:

支持单条执行及批量执行 sql 语句。

选中 sql, 点击运行, 执行选中的 sql 语句:

kaylee-test x | untitled x

```

1 select
2   t2.admin_role,
3   count(t2.admin_role)
4 from
5   promoter_0.wefe_fat.wefe_board_1.account as t1,
6   provider_0.wefe_fat.wefe_board_2.account as t2
7 where
8   t1.email = t2.email
9 group by
10  t2.admin_role;
11
12
13
14
15 select
16   count(t1.id)
17 from
18   promoter_0.wefe_fat.wefe_board_1.account as t1,
19   provider_0.wefe_fat.wefe_board_2.account as t2
20 where
21   t1.email = t2.email;

```

笔记 模板 函数

kaylee-sql探索

- kaylee-test
- untitled

terry测试

- untitled

sec_test

- 2
- 1
- untitled

sql_test

- terry测试
- tracy_test
- seclest
- terry
- jensen_0605
- zane

查询历史 搜索结果

④ 详情 ⑤ 下载

id_count
2429

不选中则全部执行笔记的 sql 语句:

kaylee-test x | untitled x

```

1 select
2   t2.admin_role,
3   count(t2.admin_role)
4 from
5   promoter_0.wefe_fat.wefe_board_1.account as t1,
6   provider_0.wefe_fat.wefe_board_2.account as t2
7 where
8   t1.email = t2.email
9 group by
10  t2.admin_role;
11
12
13
14
15 select
16   count(t1.id)
17 from
18   promoter_0.wefe_fat.wefe_board_1.account as t1,
19   provider_0.wefe_fat.wefe_board_2.account as t2
20 where
21   t1.email = t2.email;

```

笔记 模板 函数

kaylee-sql探索

- kaylee-test
- untitled

terry测试

- untitled

sec_test

- 2
- 1
- untitled

sql_test

- terry测试
- tracy_test
- seclest
- terry
- jensen_0605
- zane

查询历史 搜索结果

Query1 ✓ | Query2 ✓

④ 执行 ⑤ 取消

admin_role	admin_role_count
0	248
1	2181

➤ 操作快捷键

使用 Ctrl+Enter 执行 sql 语句。

3. 辅助区

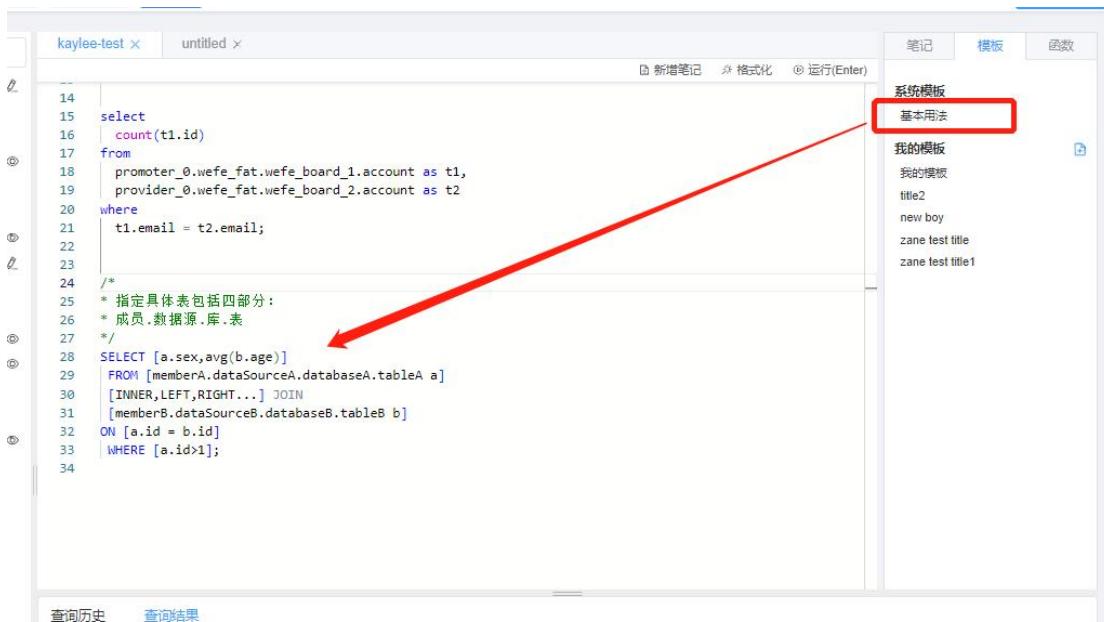
➤ 笔记

Sql 编辑区的文本会自动保存为笔记，在右侧辅助区可以查看自己的笔记，以及其他成员内用户创建的笔记。

➤ 模板

系统模板：

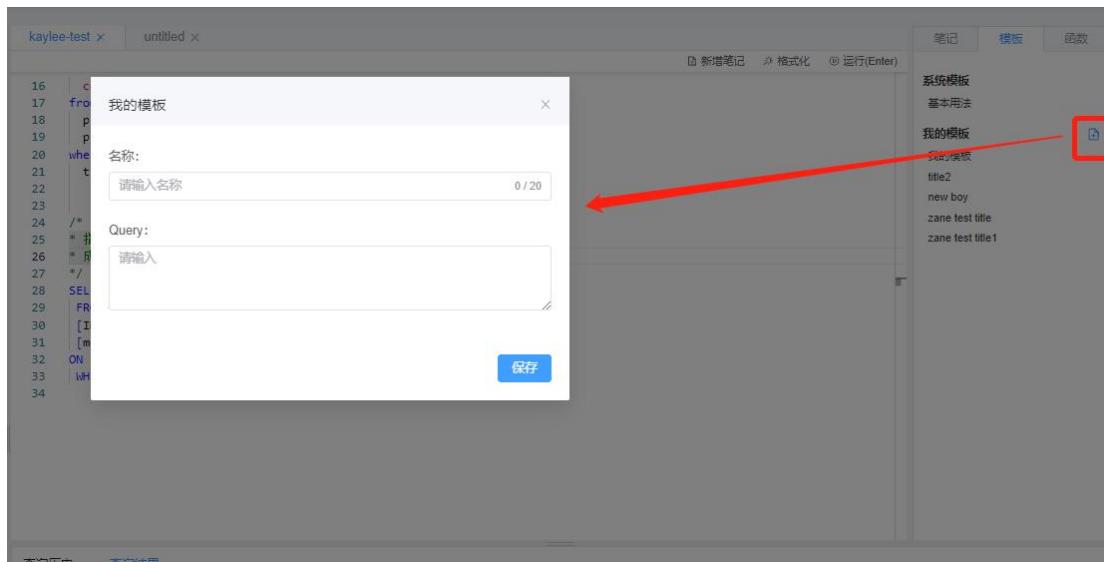
双击模板自动将 sql 同步至 sql 编辑区。



```
14
15 select
16   count(t1.id)
17   from
18     promoter_0.wefe_fat.wefe_board_1.account as t1,
19     provider_0.wefe_fat.wefe_board_2.account as t2
20   where
21     t1.email = t2.email;
22
23
24 /*
25 * 指定具体表包括四部分：
26 * 成员.数据源.库.表
27 */
28 SELECT [a.sex,avg(b.age)]
29   FROM [memberA.dataSourceA.databaseA.tableA a]
30   [INNER,LEFT,RIGHT...] JOIN
31   [memberB.dataSourceB.databaseB.tableB b]
32   ON [a.id = b.id]
33   WHERE [a.id>1];
34
```

自定义模板：

点击“+”号添加自定义模板，添加的模板该成员下其他用户都可见。



➤ 函数

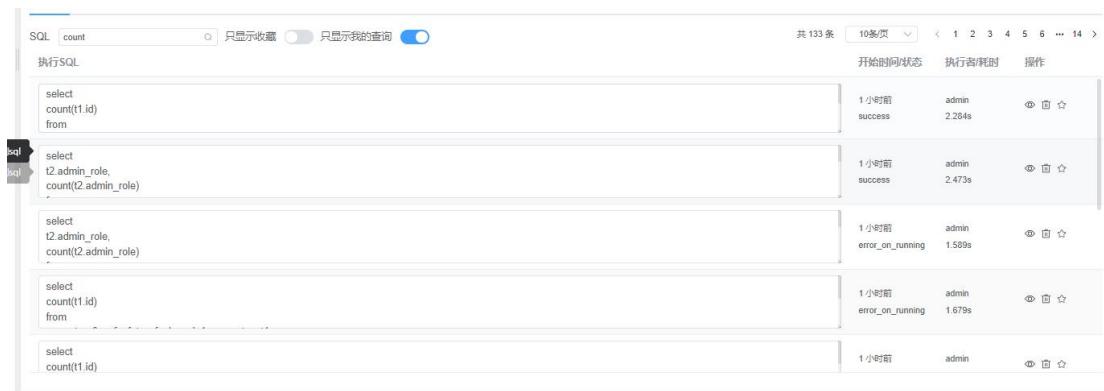
展示支持的函数，同时双击函数名，自动填充至编辑区；一期暂时支持下图几种函数，更多支持的函数测试中。



4. 结果区

➤ 查询历史

展示执行历史，执行时间，执行者，耗时，以及预览执行结果，删除执行记录，收藏执行记录的操作。



SQL	执行时间	状态	执行者	耗时	操作
select count(t1.id) from	1小时前	success	admin	2.284s	
select t2.admin_role, count(t2.admin_role)	1小时前	success	admin	2.473s	
select t2.admin_role, count(t2.admin_role)	1小时前	error_on_running	admin	1.589s	
select count(t1.id) from	1小时前	error_on_running	admin	1.679s	
select count(t1.id)	1小时前	success	admin	2.284s	

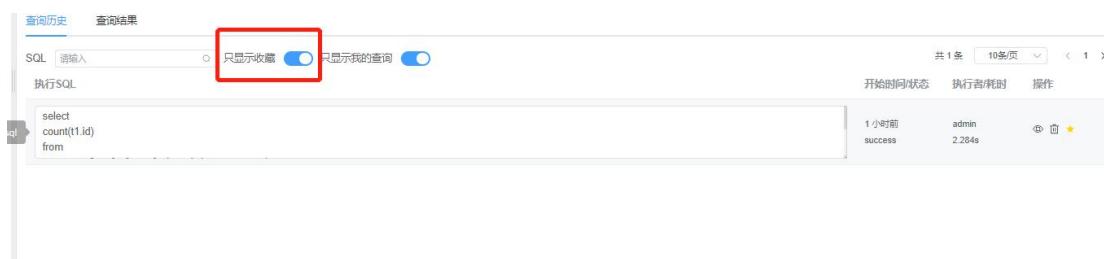
(1) 展示 sql 的执行结果，支持 sql 模糊查询。



(2) 执行的 sql 双击可复制，鼠标放至右下角可放大缩小展示框。

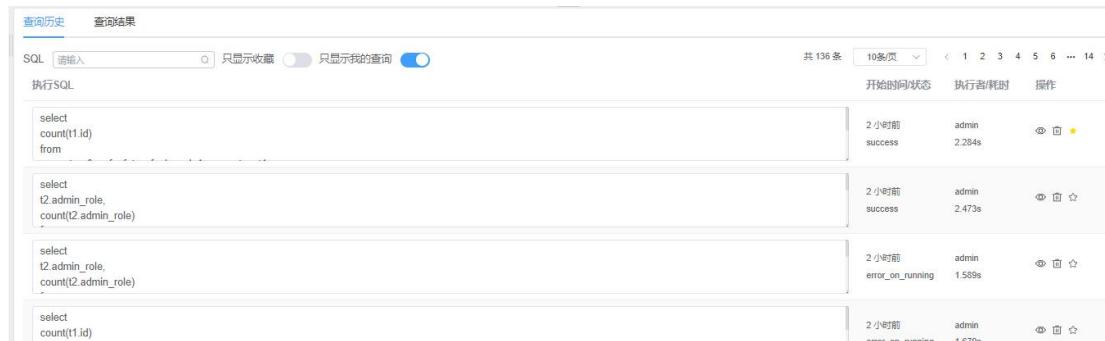


(3) 只展示收藏开启，只展示收藏的 sql 执行记录



(4) 只显示我的查询

按钮默认开启，只查询执行者是自己的执行历史。

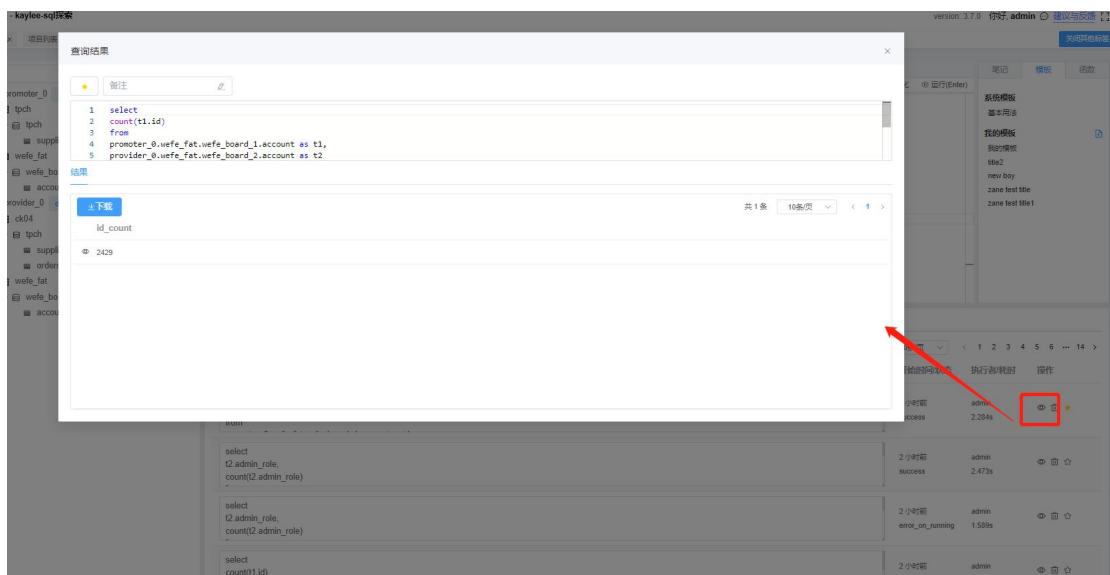


The screenshot shows a list of executed SQL queries in the 'Query History' section. Each query entry includes the SQL code, execution time, status, and the user who ran it. The status column indicates if the query was successful ('success') or had errors ('error_on_running'). The execution time is shown in seconds. The user column shows 'admin' for all entries.

SQL	开始时间	状态	执行者	耗时	操作
select count(t1.id) from	2小时前	success	admin	2.284s	
select t2.admin_role, count(t2.admin_role)	2小时前	success	admin	2.473s	
select t2.admin_role, count(t2.admin_role)	2小时前	error_on_running	admin	1.589s	
select count(t1.id)	2小时前	error_on_running	admin	1.870s	

(5) 操作

点击眼睛可查看查询结果。

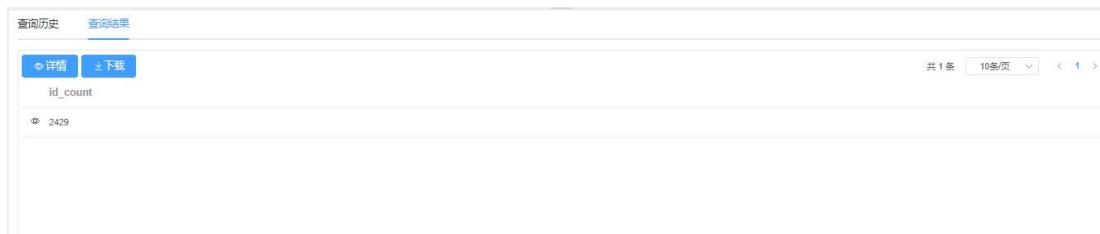


The screenshot shows a detailed view of a query result for a specific query. The result table has one row with the value 'id_count' and '2429'. A red arrow points from this result table to the '操作' (Operation) column in the main query history list, specifically to the delete button next to the first query entry.

点击“删除”按钮可删除sql执行记录。

➤ 查询结果

显示 sql 的查询结果，点击“详情”跳转查询结果页，点击下载按钮可将查询结果下载为 csv 文件。



The screenshot shows the detailed view of a query result for the query 'select id_count'. It displays the result table with one row 'id_count' and '2429'. There are two buttons at the top of this view: '详情' (Detail) and '下载' (Download). The '详情' button is highlighted in blue, indicating it is active. The '操作' (Operation) column in the main history list is also visible, showing the same query entry with its execution details.

3.6.4.5 Sql 笔记

Sql 使用笔记可参考以下附件。



sql笔记.sql

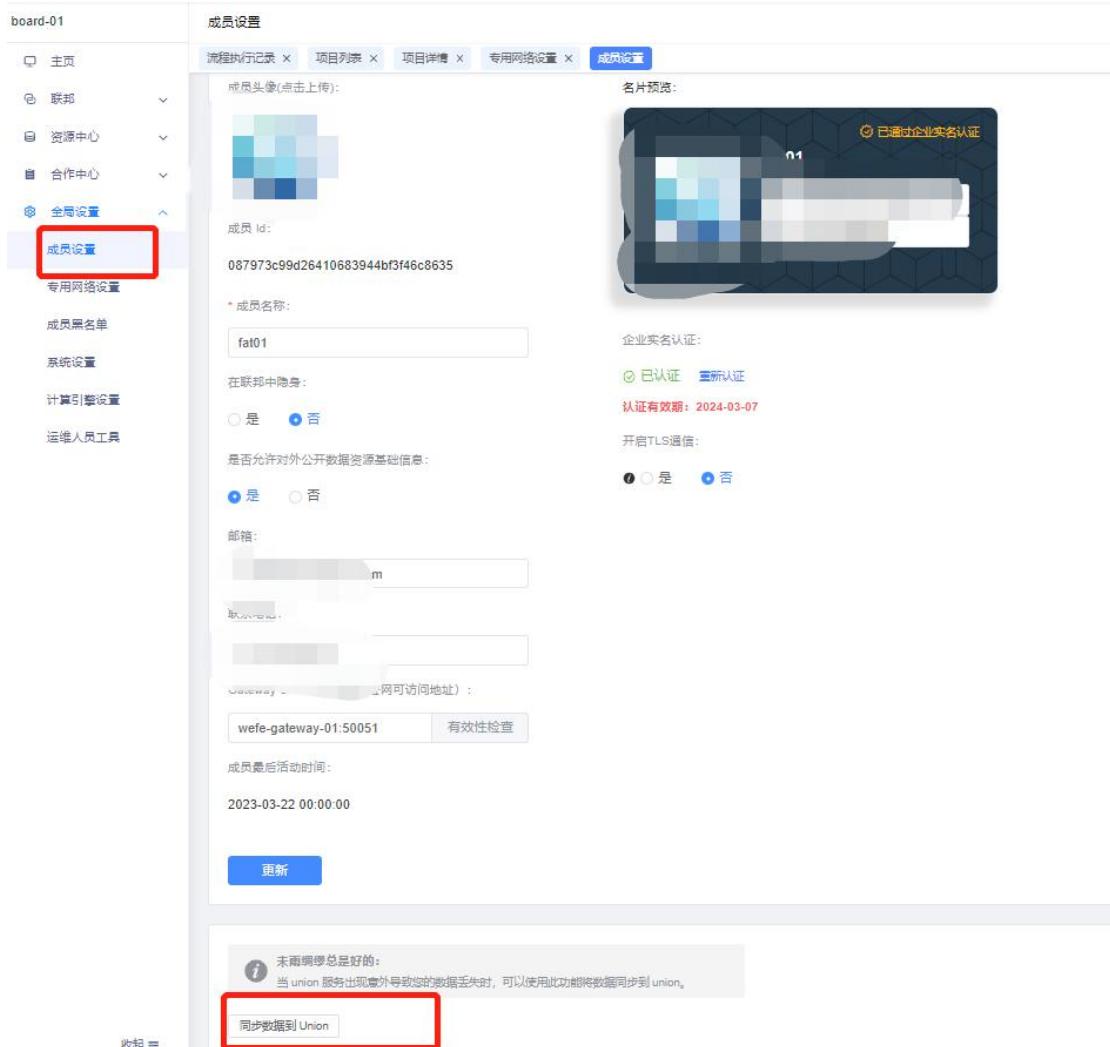
3.7 全局设置

3.7.1 成员设置

成员设置功能模块提供成员基本信息，包括：成员 ID、成员名称、E-mail、联系电话、Gateway Uri，只有管理员有修改信息权限，普通成员只有查看权限，不能修改。

操作：

- 1) 点击左侧菜单栏“成员信息”，右侧展示页面；
- 2) 输入要修改的信息，完成修改后，点击同步到 union。



3.7.2 专有网络设置

访问其他成员时，需要使用对方指定的专用网关地址，可以在[这里](#)设置对方成员及网关地址。

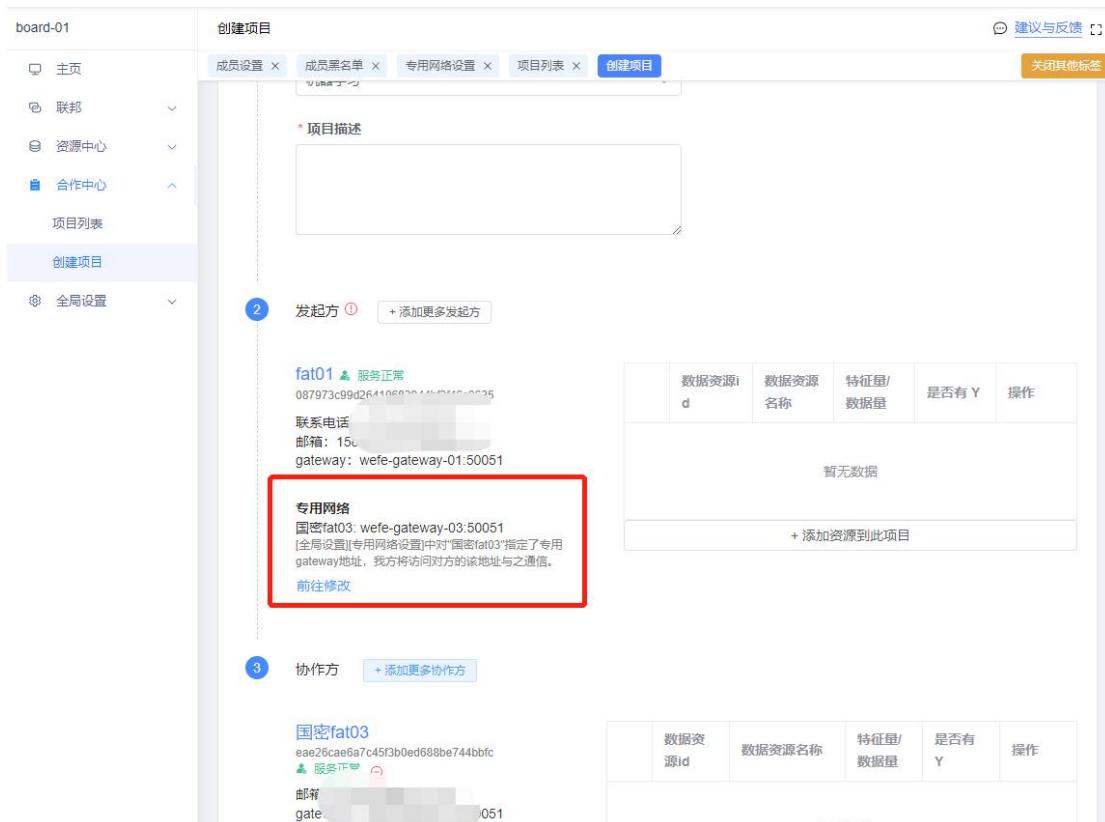


访问其他成员时，如果需要使用对方指定的专用gateway地址，请在此添加记录。

[添加配置](#)

成员	专用gateway地址	状态	操作
国密fat03	wefe-gateway-03:50051	正常	删除 测试

若访问对方成员已设置专用网络，添加项目时标识如下：



The screenshot shows the 'Create Project' interface. On the left sidebar, 'Create Project' is selected. In the main area, the 'Special Network Settings' tab is active. A note for member 'fat03' is highlighted with a red box:

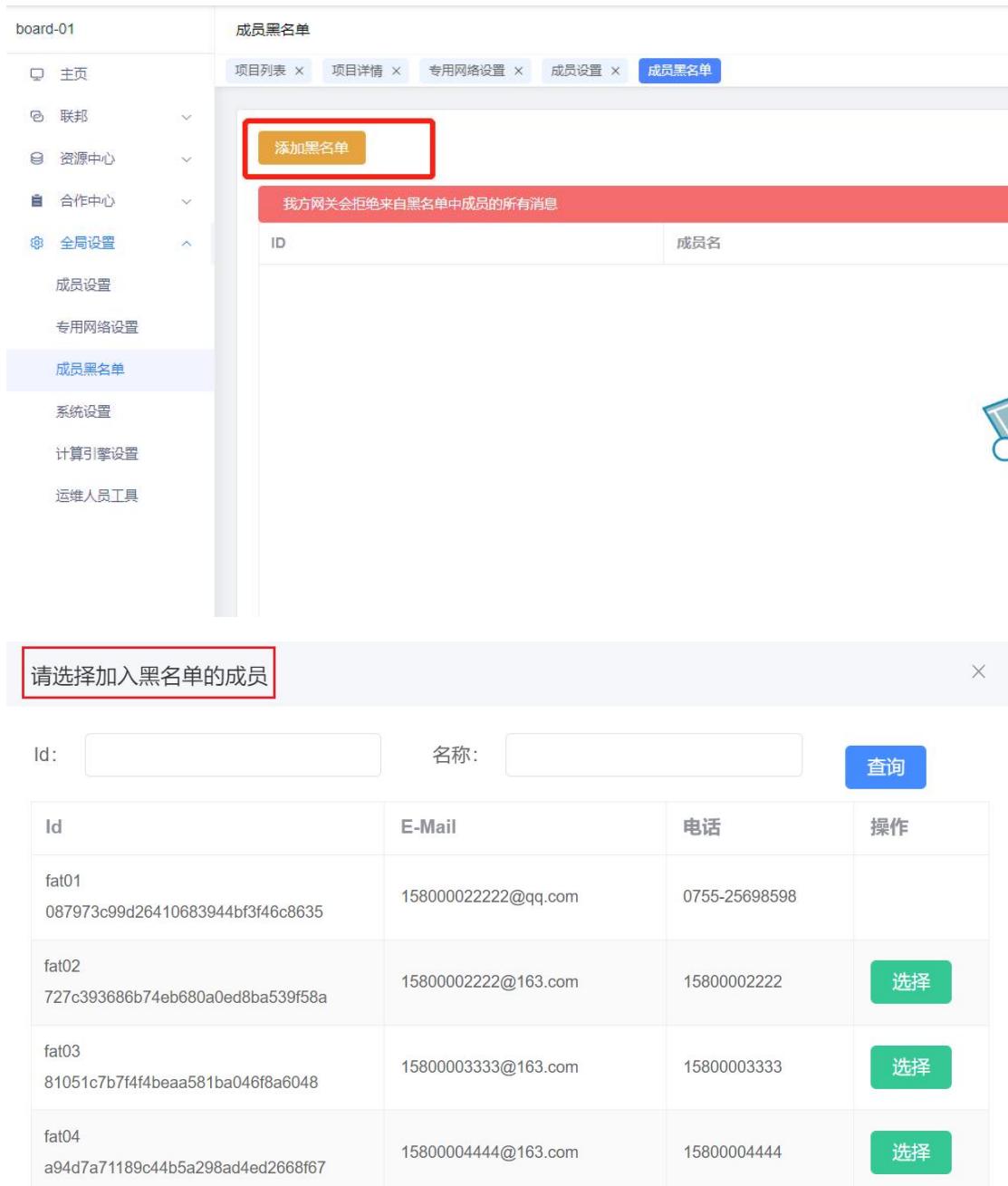
专用网络
国密fat03: wefe-gateway-03:50051
[全局设置][专用网络设置]中对“国密fat03”指定了专用 gateway地址，我方将访问对方的该地址与之通信。
[前往修改](#)

3.7.3 成员黑名单

成员黑名单模块提供参与者拉黑用户，对拉黑用户可以查看对方发送的数据，系统会自动拒绝拉黑用户发送的任务，双方将不能进行建模。

操作：

- 1) 点击左侧菜单栏“成员黑名单”，右侧展示页面；
- 2) 点击右侧“添加黑名单”按钮，进入添加黑名单操作。



成员黑名单

我方网关会拒绝来自黑名单中成员的所有消息

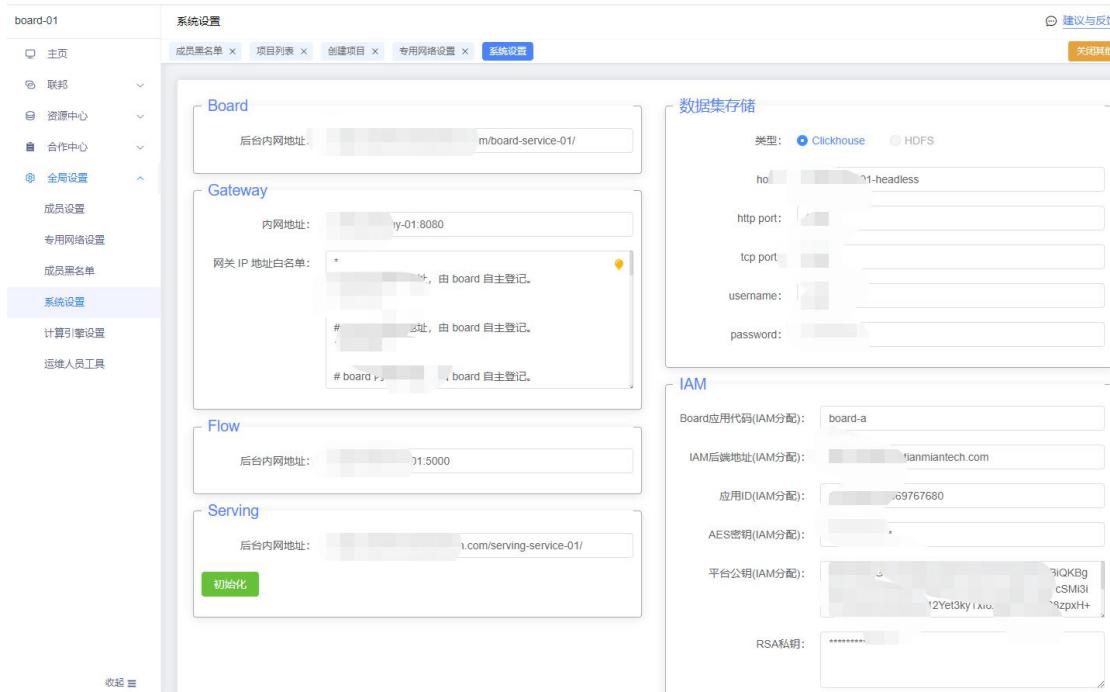
ID	成员名
fat01 087973c99d26410683944bf3f46c8635	
fat02 727c393686b74eb680a0ed8ba539f58a	
fat03 81051c7b7f4f4beaa581ba046f8a6048	
fat04 a94d7a71189c44b5a298ad4ed2668f67	

3.7.4 系统设置

系统配置模块可以针对 board 域名, board 端口, Gateway, Flow, Serving, ck 服务以及 iam 进行配置。

操作:

- 1) 点击左侧菜单栏“系统设置”，右侧展示页面；
- 2) 输入需要修改信息，完成修改后，点击“提交”已完成系统配置操作。



3.7.5 计算引擎设置

用户在此可以设置计算引擎，计算环境可以选择 spark 和函数计算 (FC) 两种方式，spark 包括单机版及集群版，函数计算支持阿里云及腾讯云两种服务；深度学习组件芯片支持 CPU 和 GPU 两种。

填写配置，通过切换计算引擎，再次启动流程，则使用对应的计算引擎执行流程。

计算引擎设置

主页 × 项目列表 × 流程详情 × 项目详情 × 计算引擎设置

机器学习相关配置

计算引擎: Spark 函数计算

Spark部署模式: LOCAL YARN

集群默认配置

driver.memory ⓘ 10g

driver.maxResultSize ⓘ 2g

executor.memory ⓘ 1g

核心数: 自动 手动

硬件加速: 无 GPU

视觉处理相关配置

芯片: CPU

VisualFL服务地址: http://10.90.0.86:10002

提交