



数据安全专区

用户指南

天翼云科技有限公司

目录

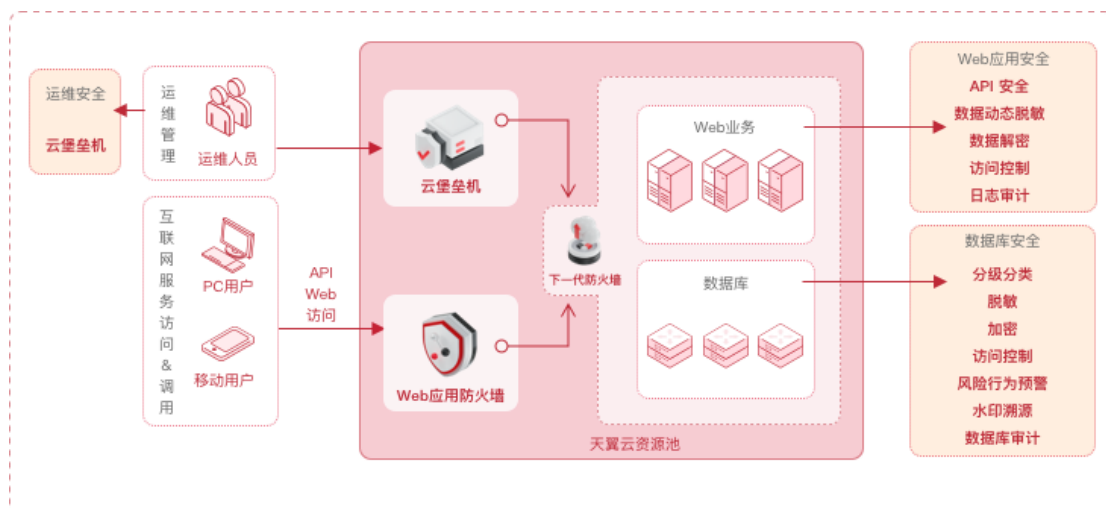
1	产品概述	4
1	1	4
1.1	产品介绍	4
1.2	功能特性	4
1.3	服务版本差异	5
1.4	产品优势	6
1.5	应用场景	7
1.6	使用限制	10
2	计费说明	11
2.1	计费项	12
2.2	计费模式	17
2.3	快速购买数据安全专区服务	19
2.4	续费与退订	20
3	快速入门	23
3.1	数据脱敏快速入门	23
3.2	数据分类分级快速入门	25
4	用户指南	26
4.1	数据脱敏操作指导	26
4.1.1	概述	26
4.1.2	数据资产管理	29
4.1.3	敏感数据发现	33
4.1.4	脱敏/水印任务	41
4.1.5	数据对比任务	54

4.1.6 水印数据溯源.....	57
4.1.7 系统统计报表.....	59
4.1.8 任务配置管理.....	59
4.1.9 系统配置管理.....	67
1.....	69
2.....	69
3.....	69
4.....	69
4.2 数据分类分级操作指导.....	69
4.2.1 概述.....	69
4.2.2 数据资产管理.....	71
4.2.3 分类分级任务.....	74
4.2.4 分类分级配置.....	77
4.2.5 分类分级统计.....	84
4.2.6 安全评估.....	87
4.2.7 可视化大屏.....	88
6 常见问题.....	89
6.1 产品类.....	89
6.2 产品使用类.....	91

1 产品概述

1.1 产品介绍

数据安全专区是一个一体化的云上数据安全防护产品方案,从数据采集、传输、存储、处理、共享、销毁的全流程建立安全防护体系,快速满足《数据安全法》等合规要求,帮助用户解决数据安全治理难、防护难的问题。



1.2 功能特性

数据分类分级

通过自动发现与多行业法规模板实现用户数据的精准识别和风险分级,支持自定义规则并按库表列维度生成可视化资产台账,助力精细化安全管理;

内置分类框架及多维条件组合规则,结合交互式图表报告直观呈现数据分布与敏感占比,确保合规治理与动态防护的有效平衡。

数据脱敏与水印

采用 240 余种算法防止敏感数据共享泄露，支持 40 余类数据库及文件的数据源脱敏，并嵌入水印实现泄漏溯源；

依托内容识别与自定义规则自动发现敏感信息，突破元数据依赖，覆盖主流关系库、国产库及非结构化文件的全场景防护体系。

数据安全服务

数据安全咨询规划服务：依据《数据安全法》《个人信息保护法》《网络数据安全条例》（征求意见稿）等法规要求，为客户制定针对性的数据安全咨询服务内容，提供数据安全咨询服务报告 。

数据分类分级服务：依据各行业数据安全分类分级标准规范，为客户建立统一的数据资产安全等级标准，在企业内部各部门的协同配合下形成涵盖业务、运营、财务、人事、合规等各类数据资产安全等级清单 。

1.3 服务版本差异

数据分类分级实例

数据安全专区的数据分类分级实例仅有标准版可选择。

版本	支持纳管的数据库数量	支持的数据库字段数	产品能力
标准版	4/8/16/32	3 万/7 万/15 万/30 万字段以内	提供流程化的数据分类分级能力

数据脱敏与水印实例

数据安全专区的数据脱敏与水印实例支持标准版、高级版、企业版三种规格供您选择，具体支持的数据库版本

版本	支持纳管的数据库数量	数据脱敏能力	水印溯源能力	基本数据库	国产数据库	大数据组件
标准版	4/8/16/32	√	√	√	×	×
高级版	4/8/16/32	√	√	√	√	×
企业版	4/8/16/32	√	√	√	√	√

1.4 产品优势

多维度的数据资产快速梳理

快速识别网络中的个人信息、重要数据、行业敏感数据，掌握数据的分布情况、存储方式和数据量级，生成数据保护对象清单。

采用自动发现、人工录入等方式，对数据库、数据表、数据字段进行管理，支持对敏感数据分类分级自动梳理。

采用自动发现、人工录入等方式，对应用系统资产、API 接口进行管理，并形成数据链路刻画。

采用自动发现、人工录入等方式，对网络主机资产信息进行统一梳理。

丰富的数据源支持

系统适配关系型数据库、国产数据库、MPP 数据库、数仓、大数据组件等合计超 40+ 数据源类型，同时支持 word、pdf、txt 等非结构化数据的分类分级。

基于业务系统的数据流动全链路监控

围绕业务场景开展的数据流动安全监管，将业务场景分解到业务系统上的数据流和数据节点进行细粒度的监控，实现基于业务规则的数据流动安全监管。包括全局性敏感数据的违规访问、调用、泄露等；基于业务特性的数据异常情况监控。

智能灵活的分析建模引擎

规则引擎内置 200+ 模型规则，涵盖常见安全攻击行为特征；支持规则模型、统计模型、关联模型以及智能 AI 模型，并可通过图形化方式进行配置编排，可通过不同规则组合以满足复杂的告警场景需求。

支持多款脱敏水印算法

支持伪列算法、伪行算法、内容修改水印算法。

1.5 应用场景

敏感数据梳理场景

数据自动发现

基于网络嗅探技术，自动发现网络环境中数据源，并结合通过内置的 70+ 的敏感数据识别

算法快速、准确的梳理出核心重要数据，同时在数据库漏洞、数据库配置、账号权限等方面进行全面风险梳理，并给出加固建议。

丰富的敏感数据识别算法

包括正则表达式、关键字、机器学习、NLP、文档指纹等先进 AI 技术创建识别规则、实体识别模型，支持对复合字段进行识别。

智能推荐模型

产品包含使用 Apriori 或 FP-growth 关联规则等算法，训练出的推荐模型，针对命中多个规则与分级分类包的情况，给出分类分级推荐。

智能聚类梳理

产品基于聚类算法对相似表、相似字段进行分析，汇总信息后进行展示，并进行批量确认，通过具智能梳理向导功能减少总工作量，大幅提升梳理效率。

敏感数据访问活动重点监测

对数据库中的对象可被哪些用户访问的情况进行归纳总结，特别是对包含了敏感列的表或者敏感度评分较高的对象，着重监测其访问情况。

图表分类显示

对敏感数据的分布情况进行图形化分类显示，直观、快速掌握家底

数据库访问数据安全

防外部攻击

提供数据库“虚拟补丁”，对数据库漏洞利用行为进行检测，结合内置安全规则，实现数据库的防护。

账号权限管

建立唯一数据库运维通道，提供账号准入、多重认证能力；

提供独立于数据库账号权限以外的第三方权控能力，回收、限制部分特权用户权限，防越权操作。

内部恶意数据篡改泄漏

提供细粒度访问控制能力，从访问源、访问目标、访问行为、访问结果集等多维度限制数据库风险操作、误操作等行为；

提供系统表、敏感表、敏感字段防护能力，防止敏感数据泄漏、篡改；

提供动态脱敏能力，防止运维过程造成敏感数据泄漏；

提供风险运维审批能力，当运维人员必须进行某些危险性操作或者访问敏感数据时，提交临时授权工单，由安全管理员进行逐级审批后方可进行操作。

数据脱敏与水印

数据共享交换灵活策略

为了确保数据安全，数据提供方需要按照需求方的实际用途对数据进行脱敏处理并添加水印信息。在数据开放时，提供方需要筛选出可开放字段，对不可开放字段进行脱敏处理。可通

过数据脱敏与水印基于用户业务情况制定不同的脱敏策略，灵活遮蔽、清洗各类敏感数据。

开发测试环境规避风险

在业务系统开发测试过程中，为确保数据质量存在直接引用生产数据的情况。可通过数据脱敏与水印一致性脱敏算法，将生产数据经过脱敏处理后导入开发测试环境，在保证数据的“真实性”的同时规避生产数据泄漏的风险。

共享数据泄漏溯源追责

当数据被共享给第三方后，第三方有责任对数据进行有效保护。可通过数据脱敏与水印将三方单位信息、共享时间等信息以数据水印的方式嵌入到共享数据中。一旦发生数据泄漏，用户只须提供少量的泄漏数据样本就可以通过嵌入的水印信息溯源泄漏源头，明确责任方并采取相应的追责措施。

数据分析计算隐私保护

数据脱敏与水印支持保留计算特征的脱敏方式，在剔除数据敏感性的同时最大程度保留 AI 建模相关数据的可用性。可通过数据脱敏与水印对个人识别信息进行脱敏处理，可完成数据去标识化且不影响统计分析结果，确保个人隐私安全。

1.6 使用限制

数据脱敏与水印系统支持的数据类型

数据类型	支持的数据
传统型数据库	MYSQL、GOLDENDB、TIDB、MYCAT、DBLE、DRDS、TELEDB、TDSQL_FOR_MYSQL、ORACLE、OCEANBASE、 POSTGRESQL、TELEPG、MsSQL(SQLServer)、MARIADB、INFORMIX、SYBASE、DB2、CHCHEDB、HANA

数据类型	支持的数据
非关系型数据库	ElasticSearch、MongoDB、ClickHouse
云数据库+大数据组件	HIVE、星环(INCEPTOR)、IMPALA、GREENPLUM、HBASE、ODPS(Maxcompute)、TERADATA、VERTICA、DORIS
国产数据库	南大通用(GBASE)、达梦(DM)、人大金仓(KINGBASE)、高斯 DB-T(Gauss100)、高斯 DB-A(Gauss200)、神通(Oscar)、瀚高(HighGo)、优炫(UXDB)
StructFile	FTP (支持 csv/xls/xlsx/txt/del/dbf 格式)、SFTP(同 FTP)、LOCAL(同 FTP)、ORACLEDMP、MYSQLDMP

数据分类分级系统支持的数据类型

数据类型	支持的数据
传统型数据库	Oracle、DB2、MySQL、MariaDB、SQLServer (MsSQL)、PostgreSQL、Informix、Sybase、CacheDB、HANA
非关系型数据库	ElasticSearch、MongoDB、ClickHouse
云数据库+大数据组件	ODPS(MaxCompute)、GaussDB、Teradata、Greenplum、Hive、Hbase、CDH、TDH、IMPALA、DRDS、RDS、Vertica、PolarDB
国产数据库	高斯 DB-T (GAUSS100)、高斯 DB-A (GAUSS200)、达梦 (DM)、南大通用 (GBASE)、人大金仓 (KINGBASE)、OceanBase、TiDB、优炫数据库 (UXDB)
文件传输协议	FTP、SFTP、Local (本地上传)

数据安全专区支持的资源池

数据安专区目前仅支持在以下资源池部署：华东 1。

IPv6 支持情况

仅数据分类分级实例中，MySQL、PG 数据源支持添加 IPv6 地址。

2 计费说明

2.1 计费项

数据安全专区仅支持包周期和一次性计费的内容。

产品套餐计费内容

产品名称	功能简述	版本	资产数	价格
数据分类 分级	提供流程化、开放化的数据分类分级能力	标准 版	4	4000 元/ 月
			8	7600 元/ 月
			16	14800 元 /月
			32	29200 元 /月
数据脱敏 与水印	提供数据脱敏与水印溯源能力，仅允许脱敏基本数据库类型	标准 版	4	3000 元/ 月
			8	5800 元/ 月
			16	11400 元 /月

产品名称	功能简述	版本	资产数	价格
	提供数据脱敏与水印溯源能力，允许脱敏基本数据库+国产数据库	高级版	32	22600 元/月
			4	4500 元/月
			8	8900 元/月
			16	17700 元/月
			32	35300 元/月
	提供数据脱敏与水印溯源能力，允许脱敏基本数据库、国产数据库、大数据组件	企业版	4	6000 元/月
			8	11600 元/月
			16	22800 元/月
			32	45200 元/月

说明：数据分类分级和数据脱敏与水印均享受如下优惠：包年购买，每年的年付价格为 10 个月的月付价格总和。

服务套餐计费内容

注意：数据安全专区服务套餐请您在咨询客户经理后，确认无误再下单。

产品名称	产品简述	版本	价格
数据安全 咨询规划 服务	依据《数据安全法》《个人信息保护法》等法规要求，为客户制定针对性的数据安全咨询服务内容，提供数据安全咨询服务报告	标准 版	80000 元/次
数据分类 分级服务	依据各行业数据安全分类分级标准规范，为客户建立统一的数据资产安全分类分级目录、清单	标准 版	80000 元/次

数据安全咨询规划服务内容说明

名称	服务子项	描述	交付物
数据安全 全管理 体系建	数据安 全管理 现状调 研	基于《银行保险机构数据安全管理办法》、《中国人民银行业务领域数据安全管理办法》等标准规范，参考合规监管要求，针对 XX 机构的数据安全管理现状进行调研，全	数据安全 管理现状 调研过程 件

名称	服务子项	描述	交付物
设咨询 服务方 案		面了解数据安全管理工作目标、数据安全管理工作组织架构、数据安全管理制度建设、数据安全管理工作流程情况等。	
	数据安全 全管理 组织建 设	梳理 XX 机构内部数据安全管理工作组织架构，包括决策层、管理层、执行层、监督层，针对实际情况设计具体职责。 针对具体的组织架构及职能，明确各角色部门之间的协调工作机制，保障数据安全体系的持续化建设工作顺利开展。	《数据安全组织架 构与角色 职责设 计》
	数据安全 全管理 制度编 制	结合数据安全管理工作现状调研结果，针对 XX 机构目前数据安全制度的缺失情况，设计与建设全面的数据安全管理制度，提供全面的、系统化的数据安全框架，通过规范、标准和流程的建设，保障其数据安全，包括数据全生命周期、数据访问控制、数据备份和恢复、数据加密、数据应急预案等方面。	数据安全 管理制度 初稿
	数据安全 全管理 制度评 审及修 订	针对已制定的制度内容进行评审，对制度内容存在不完善的方面进行修订和完善。	数据安全 管理制度 终稿

名称	服务子项	描述	交付物
	数据安全管理制度培训及试运行	制度试运行阶段包括针对制度体系的培训宣贯、制度执行过程中的内容答疑，以及结合实际落实情况针对制度进行进一步的审核修订。	/

数据分类分级服务内容说明

名称	服务阶段	工作任务内容	交付物
数据分类分级服务方案	数据分类分级产品部署	数据分类分级工具的安装调试，系统联调与测试，保障平台工具正常运行	数据资产梳理过程件：
	数据资产调研方案制定	制定数据资产调研工作方案，内容包括调研范围、参与对象、工作目的、时间进度、配合分工等	《数据分类分级逻辑框架》
	数据资产梳理	针对 XX 组织单位业务部门开展调研，梳理业务部门对应的关键业务流程以及业务活动，分析各业务流程涉及的关键敏感数据	《数据资产分类分级清单表》

名称	服务阶段	工作任务内容	交付物
	制定数据分类分级逻辑框架	根据数据安全标准和法律法规要求，调研梳理 XX 组织单位的业务条线，制定数据安全分类分级逻辑框架，为数据分类分级实施奠定基础	《数据分类分级指南》
	数据分类分级实施	针对 XX 组织单位试点系统范围内的数据，协调系统业务管理部门，开展数据发现梳理盘点、数据分类分级、归类归级工作，数据分类分级结果讨论形成一致意见，通过自动化工具整理数据分类分级结果	
	制定数据分类分级指南	基于数据分类分级逻辑框架基础以及数据分类分级工作持续实施的结果，优化数据分类分级工作流程及方法，输出数据分类分级指南	

2.2 计费模式

数据安全专区目前仅支持包年包月计费和一次性计费两种形式

数据脱敏与水印、数据分类分级支持包年包月计费。

数据安全咨询规划服务、数据分类分级服务支持一次性计费。

包年包月适用场景

包年包月是一种计费模式，适用于多种场景，尤其是需要稳定资源并长期使用的情況：

1.长期稳定使用

常驻服务： 适用于需要长期稳定运行的服务，如网站托管、数据库服务器、企业应用等。

2.成本可控和节约

成本控制： 对于需要长期使用的资源，包年包月可以降低长期成本，相比按需付费更为经济。

预算规划： 有助于企业预算规划，避免长期高额的变动费用，提供更稳定的费用支出。

3.稳定资源需求

资源保障： 对于有稳定资源需求的业务，如特定配置的服务器、存储或计算资源，包年包月提供稳定的资源保障。

包年包月约束与限制

包年包月的计费模式虽然有其优势，但也存在一些约束和限制，这些限制可能因服务提供商和具体服务套餐而异：

1.长期绑定

合约期限： 通常需要按照一定期限购买，可能是一年或更长时间的合约，无法随时更改或取消。

费用一次性支付： 一次性支付包年包月费用，无法根据实际使用情况灵活调整费用。

2.非弹性和限制

固定资源： 购买后资源固定，无法根据实际需求灵活调整，可能导致资源过剩或不足。

限制升级： 一些服务可能在包年包月期间限制资源的升级，只能在合同到期后进行。

3.限制服务范围

部分服务限制： 某些服务或特性可能无法包含在包年包月套餐内，需要额外付费或选择其他计费模式。

4.资费约束

费用固定性： 购买后的费用是固定的，无法根据实际使用情况进行变动或调整。

2.3 快速购买数据安全专区服务

若您需要购买云密评专区的相关业务，可参考本章节进行选购。

选购步骤

- 1.登录数据安全专区管理控制台。
- 2.单击右上角的“订购数据安全专区”，跳转至选购页面。
- 3.选择需要购买的业务及各项配置。

配置项	配置项说明
地域	选择您业务所需要部署的资源池。
可用区	选择业务需要部署的可用区。
虚拟私有云	选择业务所在的 VPC 组。 注意 VPC 选定后暂不支持更换，请谨慎选择。
子网	选择当前 VPC 内子网。 说明 子网选择必须在 VPC 的网段内。
配置清单	选择要购买的数据安全专区配置，具体可参考： 服务版本差异 和 计费项 章节。
购买时长	选择业务购买时长，支持选择 1-12 个月，2-5 年。
自动续订	选择是否自动续订业务。 说明 自动续订的时长会和您业务购买的时长保持一致。

4.选购完成后，勾选“我已阅读，理解并接受《天翼云数据安全专区服务协议》”。

5.单击“立即购买”跳转至支付页面。

2.4 续费与退订

控制台续费

为保证用户正常使用数据安全专区服务，在数据安全专区许可证到期前或使用许可到期后 15 天内，用户可通过“续费”操作增加授权使用期限。

在数据安全专区到期前，可以通过“续费”操作延长到期时间。

在数据安全专区到期后，若未及时续费，则进入“保留期”，不能登录数据安全专区系统。

“保留期”为 15 天，到期仍未续订或充值，存储在数据安全专区中的数据将被删除、资源将被释放。

前提条件

已获取控制台的登录账号与密码。

操作步骤

登录管理控制台。

选择“安全 > 数据安全专区”，进入数据安全专区实例管理页面。

选择待续费的实例，单击“更多 > 续费”，进入“续费”配置页面。

根据需要选择续费时长。

单击“去支付”，在支付页面完成付款。

返回数据安全专区实例列表页面，在“数据安全专区实例”列表查看授权后最新到期时间。

管理中心续订

登录天翼云官网，进入管理中心，选择续订管理，点击“手动续订”或者“开通自动续订”按钮，即可完成实例的续订。

退订前提条件

已获取管理控制台的登录账号与密码。

已使用的数据安全专区，需停止系统所有操作，解绑 EIP。

操作步骤

登录管理控制台。

选择“安全 > 数据安全专区（原生版）”，进入数据安全专区实例管理页面。

选择待退订的实例，单击所在行“操作”列的“更多 > 退订”，弹出的退订实例对话框。

确认实例信息无误后，单击“确定”，进入控制中心>退订管理>退订申请，选择退订原因，勾选确认退订金额，点击退订按钮完成退订。

到期与欠费

包周期资源开通成功后，如果没有按时续费，云平台会提供一定的保留期。

保留期：指宽限期到期后客户的包年/包月资源仍未续订或按需资源仍未缴清欠款，将进入保留期。保留期内客户不能访问及使用云服务，但对客户存储在云服务中的数据仍予以保留。

云服务进入保留期后，天翼云将会通过邮件、短信等方式向您发送提醒，提醒您续订或充值。

保留期到期仍未续订或充值，存储在云服务中的数据将被删除、云服务资源将被释放。

欠费后，可以查看欠费详情。为防止相关资源被停止或者释放，请及时进行充值，账号将进入欠费状态，需要在约定时间内支付欠款。

3 快速入门

3.1 数据脱敏快速入门

一、新增数据源

1.使用系统管理员账号登录数据脱敏实例。

2.在左侧目标栏选择“数据资产管理 > 数据源/目标”，进入“数据源/目标”管理页面。

3.单击页面左上角的“新增”按钮，在右侧的窗口中数据需要新增数据源的相关信息。

注意：

不同的数据源类型会存在不同的参数填写情况。

若您选择的数据源类型为 StructFile 下的 LOCAL/FTP/SFTP，且文件类型为 csv/txt/del，则可以在高级属性中填写表头行数：若表头行数为 0，将自动填充表头；若行数大于 1，则仅保留最后一行作为新表头，此配置可以使数据扫描更加准确。

4.填写完成后，单击窗口左下角的“测试连接”按钮，测试数据源的连通性。

5.连接成功后，若需要再新增目标请单击“保存并新增数据目标”重复上述操作，否则请单击“保存”按钮。

二、配置脱敏任务

1.使用系统管理员登录数据脱敏实例。

2.在左侧导航栏选择“脱敏/水印任务 > 结构化任务”，进入“结构化任务”页面。

3.单击页面左上角的“新增”按钮，进入“任务新增”界面。

4.填写完成后，单击“下一步”即可完成脱敏规则的基础配置。

三、配置脱敏规则

在完成基础配置后，会开始脱敏规则的创建。

配置数据脱敏任务时，您可以选择手动配置所需字段对应的脱敏规则，或者一键选择数据安全专区提供的脱敏模板。

若您希望创建自己的脱敏模板，可单击“管理”按钮，进入对应配置页面配置自定义脱敏模板，单击“保存”后单击“刷新”，即可在下拉框中选择新创建的自定义模板。如下图所示。

四、运行脱敏任务

配置数据脱敏任务时，您需要基于使用场景选择所需的数据脱敏类型。

选择完成后，单击“保存并运行”后，脱敏任务即可启动。

3.2 数据分类分级快速入门

步骤	步骤说明	参考链接
1.添加数据源	新增需要分类分级的数据源，并完成数据源同步	数据源管理
2.执行分类分级任务	为数据源新增分类分级任务，跟随流程向导完成分类分级任务	分类分级任务

步骤	步骤说明	参考链接
3.获取分类分级结果	查看和导出分类分级结果，提供分类分级报告、报表、可视化大屏功能	分类分级统计 可视化大屏

4 用户指南

4.1 数据脱敏操作指导

4.1.1 概述

数据安全专区-数据脱敏能力具有广泛的数据库支持、出色的性能和良好的可扩展性。它采用独特的脱敏与水印/溯源算法，能够高效处理敏感数据，实现去标识化和匿名化。多种处理方式如固定值替换、置空、乱序、保留统计特征等确保了在不改变现有业务逻辑的前提下，脱敏后的数据仍能保留原有业务逻辑特征。

数据安全专区支持软硬件一体机或虚拟化部署（保证与生产环境和开发 第三方测试环境前后路由可达即可）。生产库中的原始数据，经过数据脱敏清洗后离线分发至测试环境，所有敏感数据全部在内存中处理，可保证整个环节敏感数据不落地。

主要功能

敏感数据自动发现

数据脱敏系统内置多种发现规则，支持基于数据内容的识别，不仅依赖于用户的元数据管理系统、字段名或字段注释等信息。支持用户基于自身需求自定义规则，实现对敏感数据的自动发现和标识。

同时数据脱敏系统可和数据分级分类平台进行对接，可基于行业或法律法规对敏感数据进行分类分级与梳理，便于用户按照不同级别进行脱敏算法的配置，有效避免数据的过度保护。支持采用随机采样的方式，在采样的过程中过滤空值及脏数据。如在业务初期数据逻辑定义不规范往往会产生较多脏数据，而传统的敏感数据检索方式是简单过滤表中前若干条数据，无法有效避免此类脏数据的干扰。数据脱敏系统的随机采样是基于表总行数进行随机抽取，可有效避免脏数据的干扰，提高敏感数据识别的准确性。数据发现算法除常规的正则表达式、字典等，还加入了 NLP 算法，开箱即用 70 多种通用敏感字段识别算法，大幅提升了敏感数据内容的识别和处理能力。有效避免敏感数据的误报、漏报。

丰富的脱敏算法支持

数据脱敏系统支持 24 大类 240+ 种脱敏算法，适用各种业务场景主要算法支持情况如下：

置空：直接将待脱敏的信息以填充空字符或者删除的形式抹除。

乱序：在结构化数据（例如数据库）中颇为常用，对于待脱敏的列，不对列的内容进行修改，仅对数据的顺序进行随机打乱。

遮蔽：保留数据一些位置上的信息，对于敏感位置的信息使用指定的字符进行替换。

分割：保留数据一些位置上的信息，对于敏感位置的信息进行删除。

替换：使用固定值或字典映射表对敏感数据进行替换。

取整：对数值类型和日期时间类型的数据进行取整操作。

哈希：将哈希（或哈希加盐）编码后的数据作为脱敏结果输出。

仿真（保留业务含义）：考虑到业务含义，则生成的数据需符合核验规则，主要包括长度、取值范围以及校验规则和校验位的计算等。

密码学：根据所选参数指定的密码学加密算法对数据进行加密。支持的加密算法有：。

编码：根据所选参数指定的编码方式对数据进行编码。

保留类别频次特征：主要指的是反映事物类别的数据类型，具有有限个无序的值，或枚举类型，脱敏后各个类型出现的频次可保持不变。

保留数值统计特征：支持分布重建、平均、区间、标准化、归一化、添加噪声、一致性等算法（对数值型数据在脱敏后可保留数据中的高阶统计特征及数据分布特征，仍满足趋势分析要求）。

关联关系保留：支持计算关系保留、乱序关联保留、身份信息关联保留等算法（对数据在脱敏后可保留其原始的字段间关联关系）。

保留原数据：对指定的列的数据不做脱敏处理。

溯源算法：对数据本身添加仿真水印，做到精准溯源。

灵活的数据抽取

用户可基于自身需要，可灵活配置数据抽取的增量条件，数据脱敏系统可支持的增量条件包括但不限于：基于时间戳、基于主键或唯一键、基于分区或分片。

水印溯源

支持多种数据水印格式，包括并不限于伪行水印、伪列水印、脱敏水印、内容修改水印、零

宽水印：

伪行 伪列水印：通过添加配置比例的数据行或数据列来插入水印信息，当数据泄漏后，通过提取伪行 伪列水印信息来进行泄漏追溯。

脱敏水印：支持基于数据本身特征脱敏水印方式，不影响业务逻辑的同时能有效的避免了水印数据被绕开。

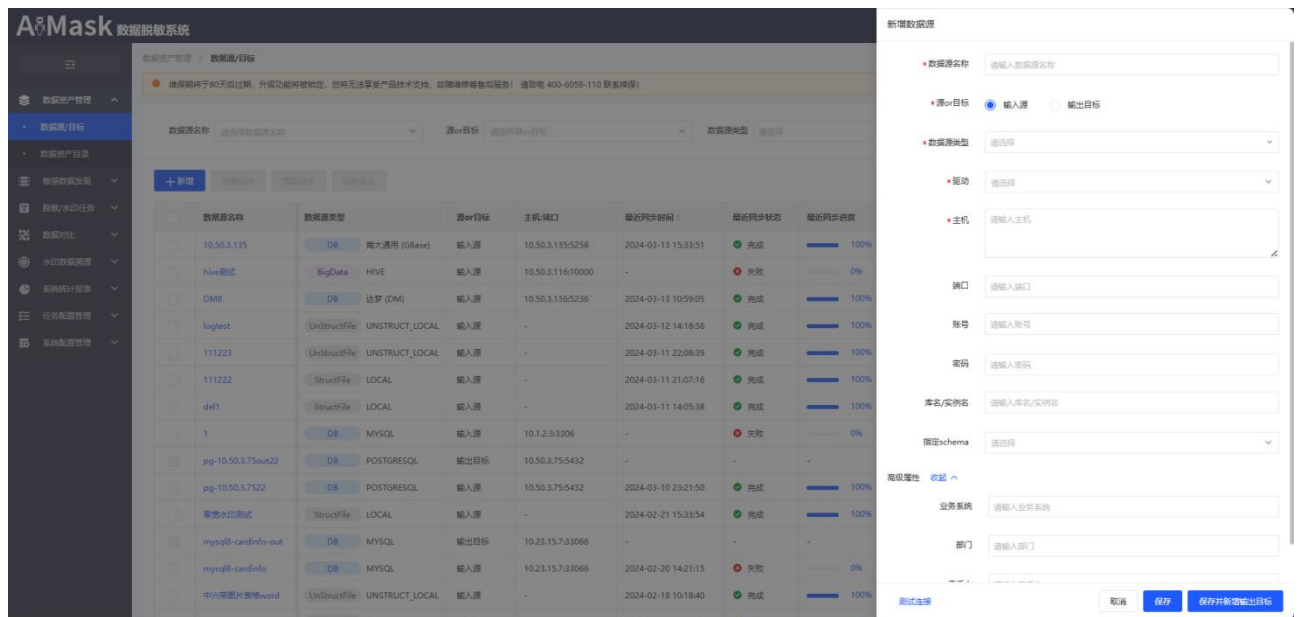
内容修改水印算法：适用于不能影响数据的业务含义并具有较强的业务使用或分析需求的场景，同时隐蔽性要求高。

零宽水印算法：在常见办公软件（office 等）或数据库管理软件（dbeaver 等）打开时是不可见的，在可视宽度上是不可感知的，不易单独选中。更适用于以文件形式发放数据，在文件中的数据内容中添加水印的场景。

4.1.2 数据资产管理

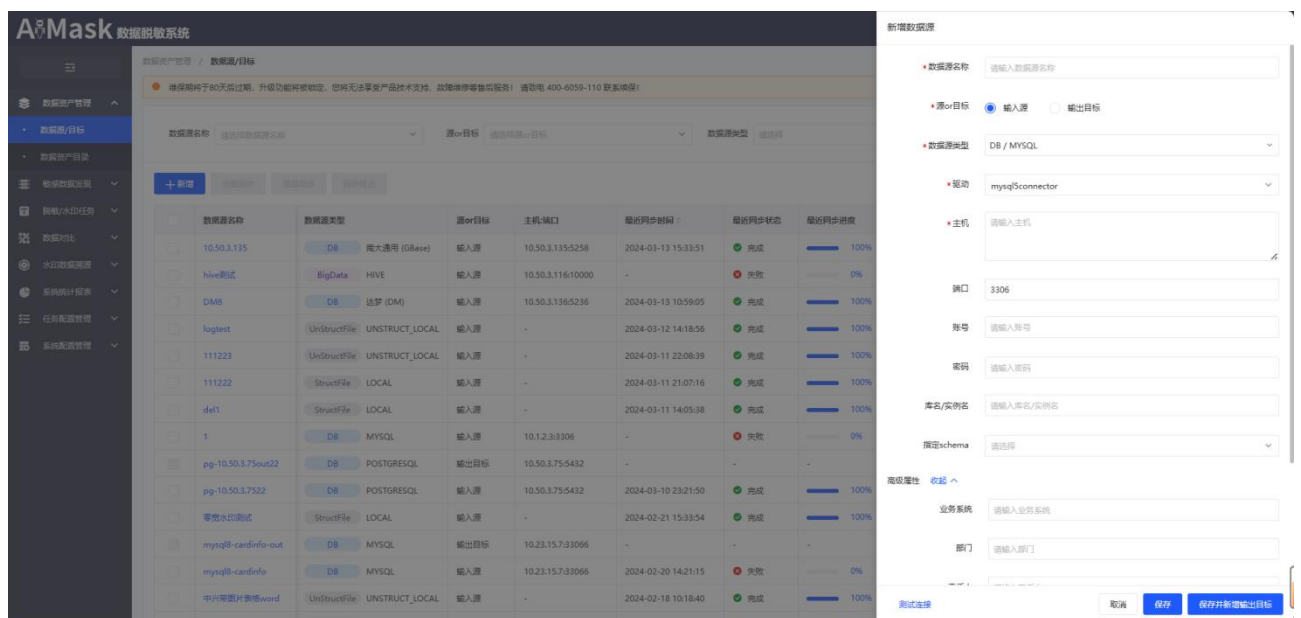
新增数据源

用户可以在该模块新增数据源/输出目标。点击<新增>输入数据源/输出目标的相关信息后点击<保存>或<保存并新增输出目标>即可新增，点击后者可以在保存本页面输入源后打开一个新的抽屉页供用户新增输出目标。为确保信息准确无误，建议在保存前点击<测试连接>进行连接测试。如下图所示。

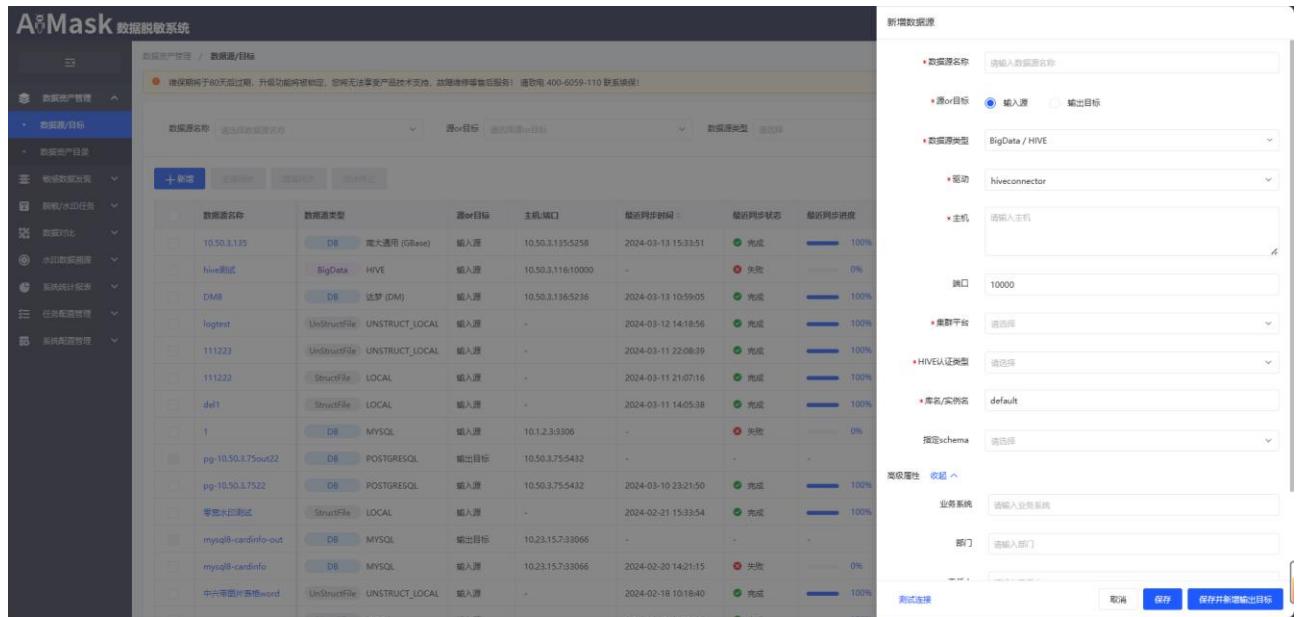


字段前带*的为必填字段, 字段前不带*的为选填字段。不同的数据库类型, 必填项有所差异, 请在选择“数据源类型”后根据页面要求仔细核对, 例如‘主机’填写内容为数据源的主机地址, 一般可填写为 IP 地址。如下图所示。

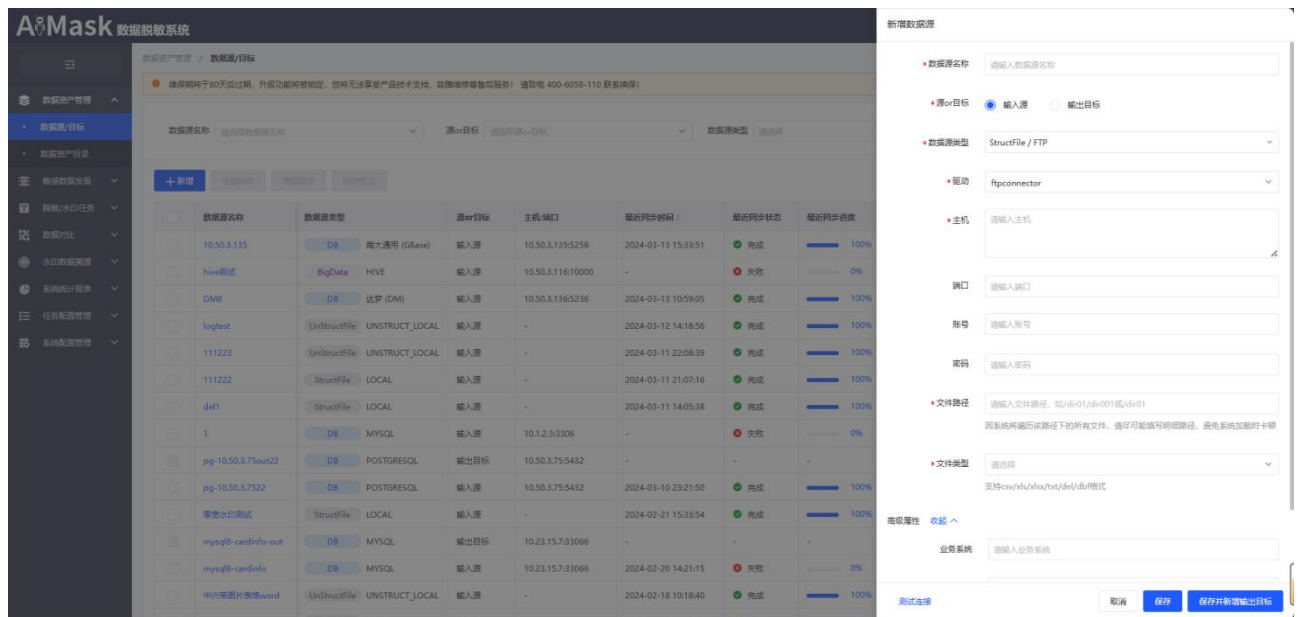
下图示例了 Mysql 类型数据库所需配置项清单。



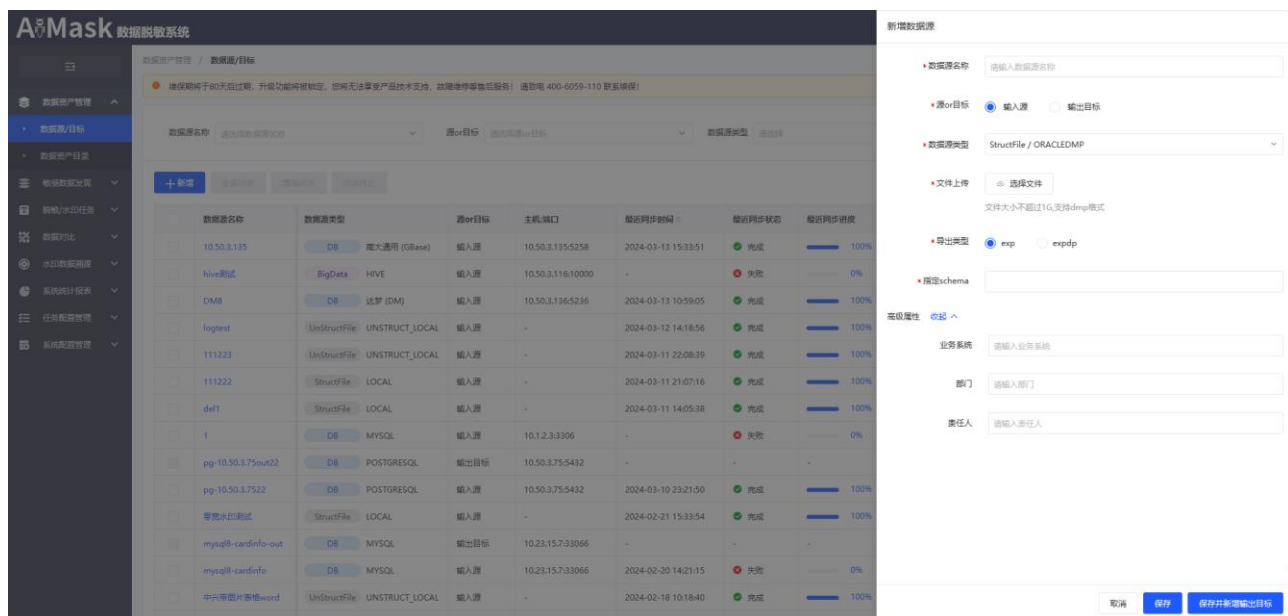
下图示例了 Hive 类型数据库所需配置项清单。



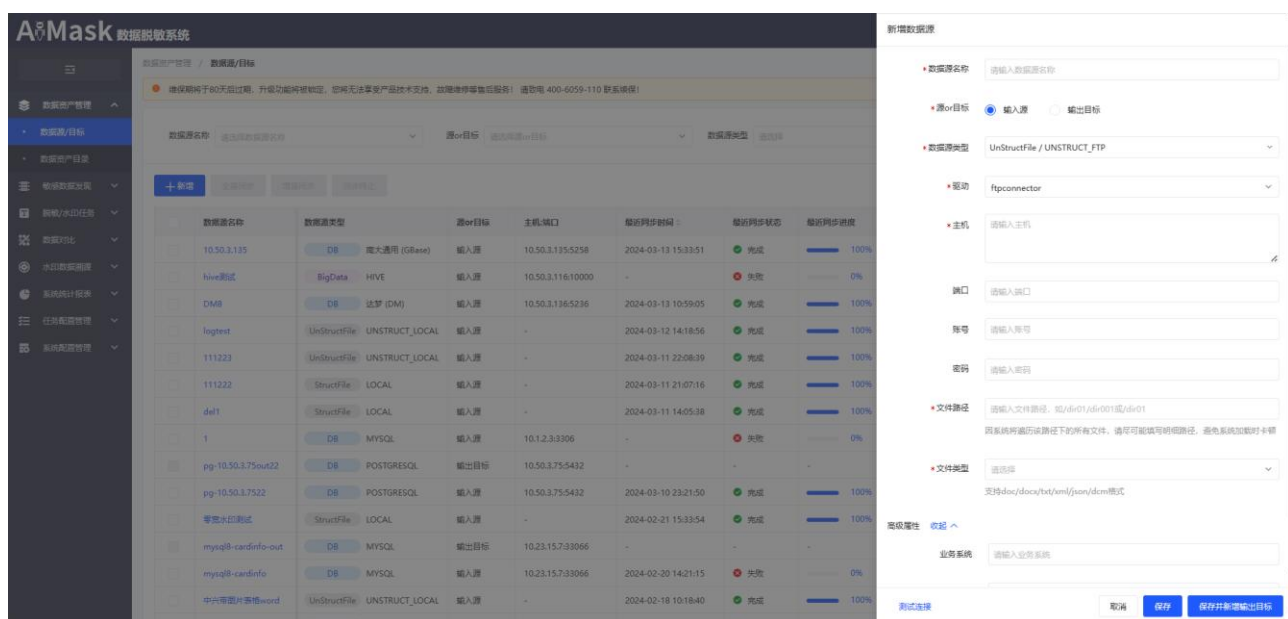
下图示例了结构化 FTP 类型数据所需配置项清单。



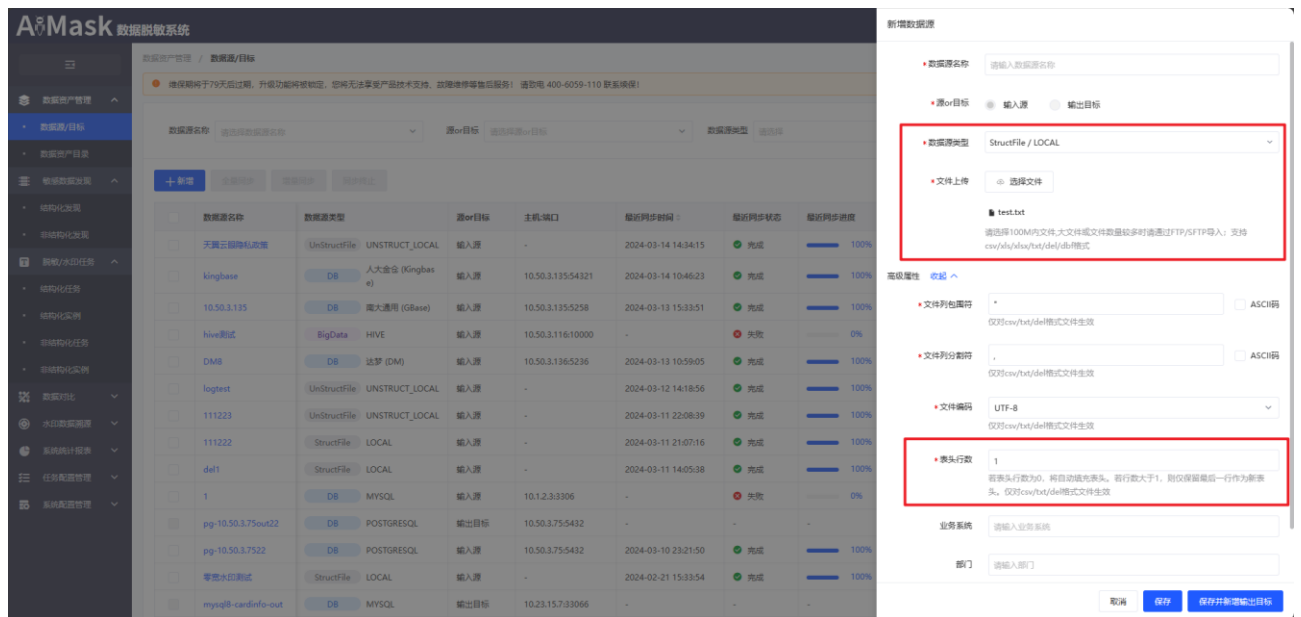
下图示例了 DMP 类型数据所需配置项清单。



下图示例了非结构化 FTP 类型数据所需配置项清单。



新增数据源时，用户若选择的数据源类型为 StructFile 下的 LOCAL/FTP/SFTP，且文件类型为 csv/txt/del，则可以在高级属性中填写表头行数：若表头行数为 0，将自动填充表头；若行数大于 1，则仅保留最后一行作为新表头。该步骤可以使数据扫描更加准确。



同步数据源

点击<元数据同步>，可以进行同步操作，会获取数据源的数据库、表、字段及账号信息。

数据库数据源同步之后，在页面内会更新该数据源的同步状态。

4.1.3 敏感数据发现

结构化发现

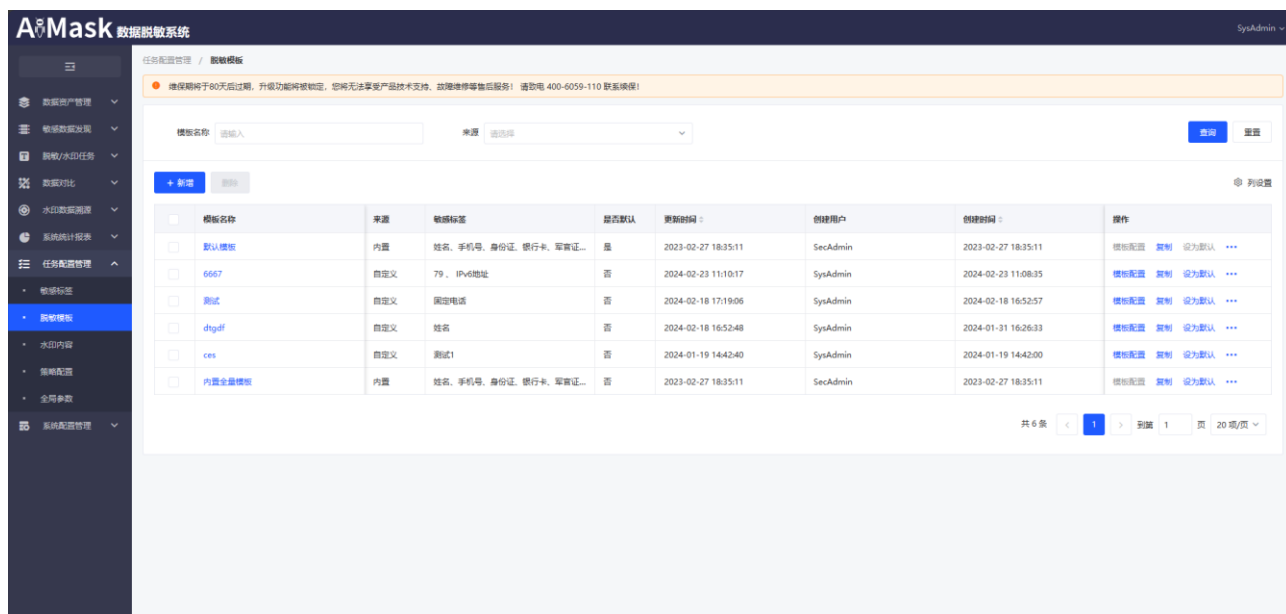
该模块可用于发现结构化源数据中的敏感数据，即点击<新增>来新增敏感发现任务。

A³Mask 数据脱敏系统										SysAdmin
敏感数据发现 / 结构化发现										
● 维保期将于80天后过期，升级功能将被锁定，您将无法享受产品技术支持、故障维修等售后服务！ 请致电 400-6059-110 联系维保！										
任务名称		数据源名称		运行状态		选择运行状态				
+ 新增		运行		禁止		删除				
⊞ 列设置										
☐	任务名称	数据源名称	敏感统计	最近运行时间	最近运行状态	最近运行进度	创建用户	创建时间	操作	
☐	[D] 111222_240311210727	111222	推荐敏感字段: 0 总字段: 1	2024-03-11 21:07:41	已确认	100%	SysAdmin	2024-03-11 21:07:41	运行	监控 详情 ...
☐	[D] 111222_240311210643	111222	推荐敏感字段: 0 总字段: 1	2024-03-11 21:06:53	已确认	100%	SysAdmin	2024-03-11 21:06:53	运行	监控 详情 ...
☐	[D] del1_240311140546	del1	推荐敏感字段: 12 总字段: 20	2024-03-11 14:05:45	已确认	100%	SysAdmin	2024-03-11 14:05:45	运行	监控 详情 ...
☐	[D] file_240311140230	del1	推荐敏感字段: 总字段:	2024-03-11 14:02:31	异常停止	0%	SysAdmin	2024-03-11 14:02:31	运行	监控 确认 ...
☐	[D] pg-10.50.3.7522_240310232156	pg-10.50.3.7522	推荐敏感字段: 67 总字段: 157	2024-03-10 23:22:33	待确认	100%	SysAdmin	2024-03-10 23:22:33	运行	监控 确认 ...
☐	[D] mysql-10.50.3.110_240308164537	mysql-10.50.3.110	推荐敏感字段: 2 总字段: 2	2024-03-08 16:47:04	已确认	100%	SysAdmin	2024-03-08 16:45:42	运行	监控 详情 ...
☐	[D] mysql-10.50.3.73_240308103948	mysql-10.50.3.73	推荐敏感字段: 1 总字段: 2	2024-03-08 10:39:48	已确认	100%	SysAdmin	2024-03-08 10:39:48	运行	监控 详情 ...
☐	[D] mask脱敏_240307091408	mask脱敏	推荐敏感字段: 2 总字段: 5	2024-03-07 09:14:15	已确认	100%	SysAdmin	2024-03-07 09:14:15	运行	监控 详情 ...
☐	[D] mask脱敏_240306182535	mask脱敏	推荐敏感字段: 2 总字段: 5	2024-03-06 18:26:23	已确认	100%	SysAdmin	2024-03-06 18:25:46	运行	监控 详情 ...
☐	[D] mysql-10.50.3.73_240227154041	mysql-10.50.3.73	推荐敏感字段: 117 总字段: 902	2024-03-05 13:17:05	已确认	100%	SysAdmin	2024-02-27 15:41:01	运行	监控 详情 ...
☐	[D] pg-10.50.3.75_240223112207	pg-10.50.3.75	推荐敏感字段: 24 总字段: 40	2024-02-23 11:22:26	已确认	100%	SysAdmin	2024-02-23 11:22:26	运行	监控 详情 ...
☐	[D] pg-10.50.3.75_240223111504	pg-10.50.3.75	推荐敏感字段: 15 总字段: 31	2024-02-23 11:15:13	已确认	100%	SysAdmin	2024-02-23 11:15:13	运行	监控 详情 ...
FTX 10.50.3.75_240223111										

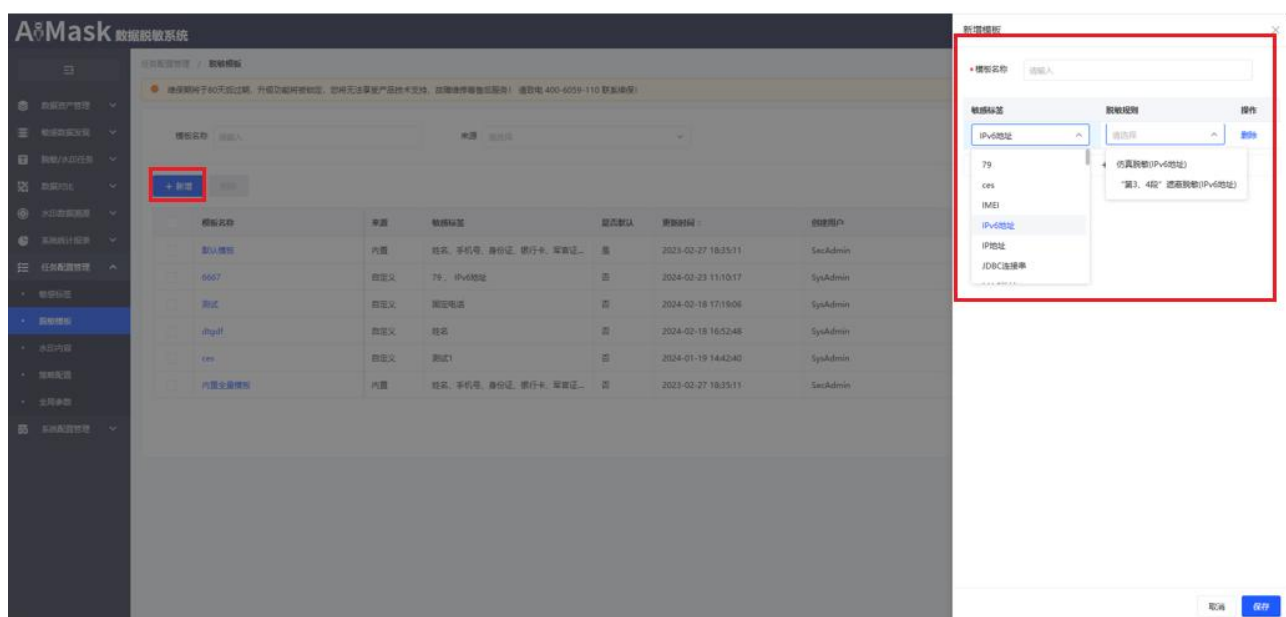
A³Mask 数据脱敏系统		SysAdmin
敏感数据发现 / 结构化发现 / 创建发现任务		
● 维保期将于80天后过期，升级功能将被锁定，您将无法享受产品技术支持、故障维修等售后服务！ 请致电 400-6059-110 联系维保！		
基本信息		
* 任务名称	[D] 10.50.3.135_240313163310	
* 输入源	10.50.3.135 数据库类型: GBASE 主机: 10.50.3.135	
* 脱敏模板	默认模板 刷新 管理	
* 发现方式	☒ 全量发现 ☐ 增量发现 全量发现会采样所有数据表对应的元数据，增量发现只采样未进行过敏感数据发现且未指定敏感标识的元数据。	
* 输入数据	☒ 手动选择 ☐ 全量所有 全量输入源内所有可选项，若您数据源内schema数量超过10个，请勿使用本功能。 ☐ 文件导入	
aisort		
待选表格 已选0/25项 已选表格 0项		
☐ aisort.car_info ☐ aisort.car_info_copy3 ☐ aisort.car_info_copy4 ☐ aisort.car_info_copy5 ☐ aisort.new_fagui		

用户可以在‘新增’页面填写敏感数据发现任务的相关参数，例如任务名称、发现方式等。刚打开本页面时，系统会自动将最近同步完成的结构化输入源名称填入“输入源”参数，若用户需要进行敏感发现操作的数据源并非此数据源，可以点击下拉框更换选择。

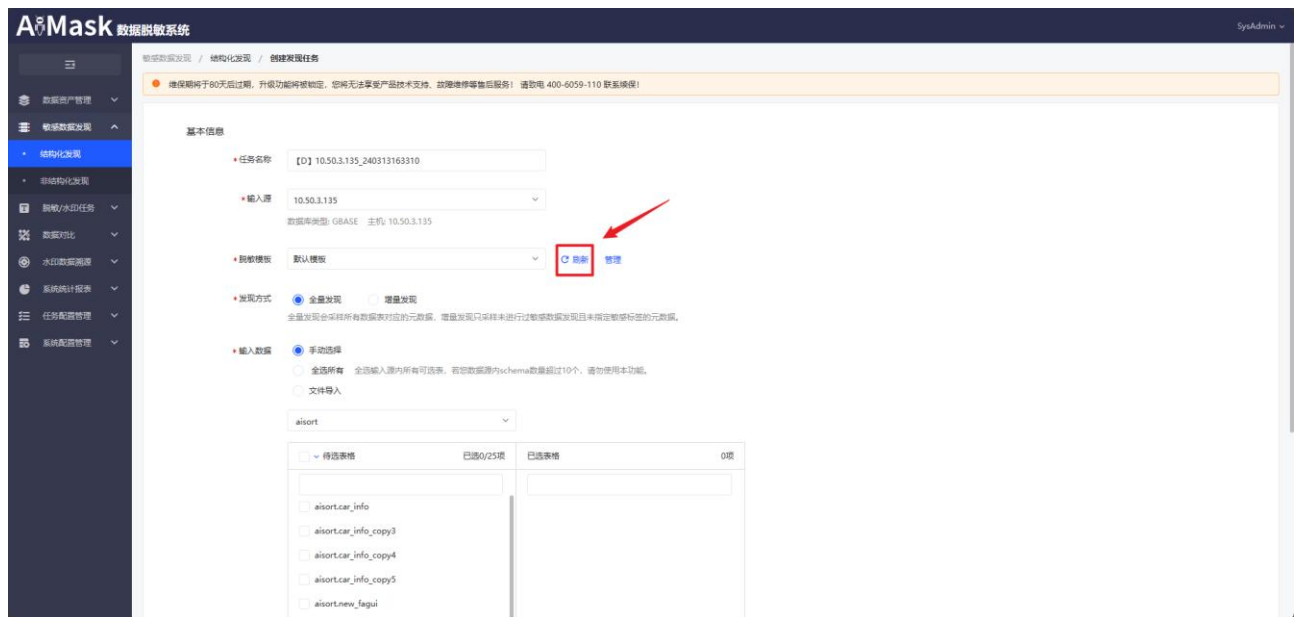
用户可以在“脱敏模板”参数选择进行敏感数据发现的模板，系统内置了默认模板供用户在常规场景下直接使用。如果用户需要自定义或者管理脱敏模板，可点击新增页面中的<管理>，然后系统会弹出一个新的名为‘脱敏模板’的页面供用户管理脱敏模板，如下图所示。



简单来说，脱敏模板就是我们去做敏感数据发现规则的集合。AiMask 默认模板是我们提前将客户常用的 49 种敏感标签，比如说姓名、手机号、身份证、银行卡等，它们的敏感发现规则归类整理合成的一个模板。他里面的内容就是，我用什么方式去识别姓名这个敏感类型，我用什么方式去识别身份证这个敏感类型等等。点击‘脱敏模板’的<新增>按钮后，会弹出抽屉页面提供用户新增自定义配置脱敏模板，如下图所示。



用户配置完毕后点击<保存>，回到先前配置敏感数据发现任务的页面并点击<刷新>即可在该任务中使用刚保存的自定义脱敏模板，如下图所示。



“发现方式”参数包括全量发现和增量发现 2 种。用数据举例，假设里面一共有 4 个表，如果我选择的是全量发现，那他无论之前这个敏感发现任务有没有运行过，会全部重新对这 4 个表做一次敏感数据发现。如果我选择的是增量发现，他只会去处理这 4 个表中跟上次这个敏感发现任务运行时相比新增的数据列，原来老的数据列 AiMask 会认为我已经发现过了，就不用再去发现一次了。增量和全量相比，因为处理的数据少很多，所以会快很多，那如果我要增量发现，我第一次配置任务选择全量还是增量？答案是增量，AiMask 会认为这个任务上次运行时这些列都不存在，因此会变相的全量发现一次。

“输入数据”参数包括手动选择、全选所有、文件导入 3 种。如果用户是手动选择的，他首先可以在这个栏切换 schema。



基本信息

任务名称 [D] Oracle_test_240627100805

输入源 Oracle_test
数据库类型: ORACLE 主机: 10.50.3.110

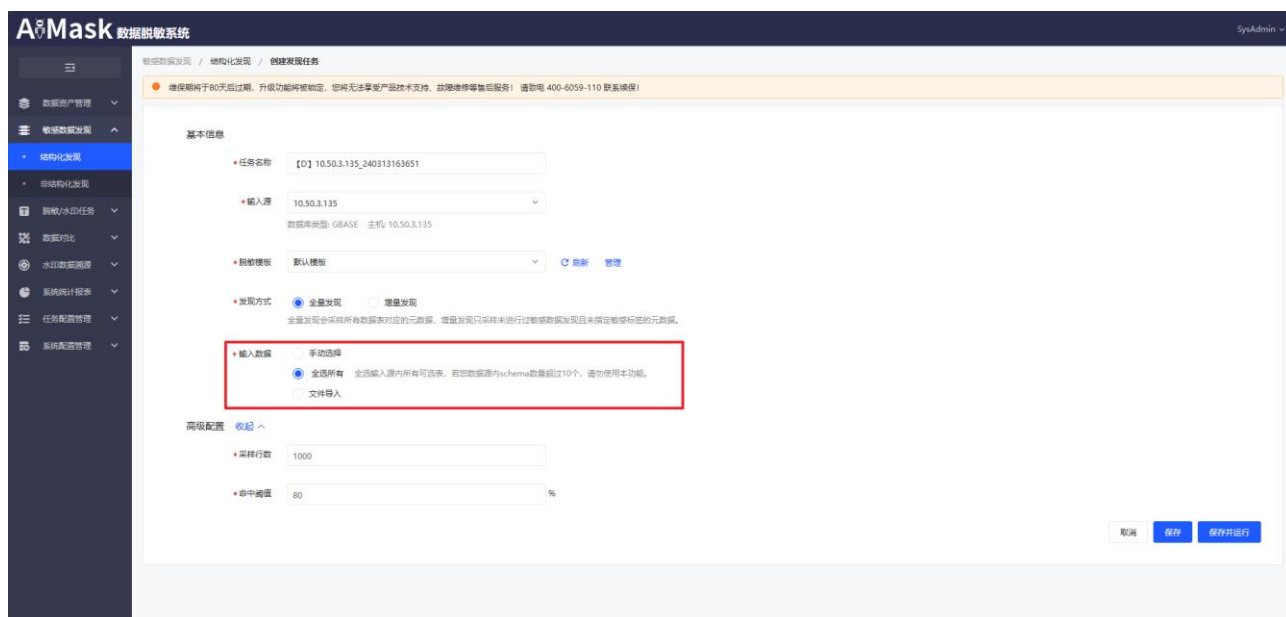
敏感模板 默认模板 [刷新](#) [管理](#)

发现方式 ☒ 全量发现 ☐ 增量发现
全量发现会采样所有数据表对应的元数据，增量发现只采样未进行过敏感数据发现且未指定敏感标签的元数据。

输入数据 ☒ 手动选择
☐ 全选所有 全选输入源内所有可选表，若您数据源内schema数量超过10个，请勿使用本功能。
☐ 文件导入

TESTUSER3
TEST
TEST_ZHANG
AUTOAISORTUSER
ZCH
AISORTCYDAUTO
TEST1

然后勾选他要敏感发现哪些表格的数据，我们提供了全选当页、全选所有、清空已选这些功能（均基于单个 schema）。如果用户是全选所有的，则默认全选输入源内所有 schema 下的全部可选表；若数据源内 schema 数量超过 10 个，请勿使用本功能。如果用户是文件导入的，直接按照我们提供的模板填写信息后上传即可，这个功能是为了方便用户批量操作。用户可以选择直接在新增界面手动选择输入源中的输入数据，或者按照系统要求选择通过文件批量导入。选择<全选所有>或<文件导入>时的界面如下图所示。



APMask 数据脱敏系统 SysAdmin

敏感数据发现 / 结构化发现 / 创建发现任务

● 维保期将于80天后过期，升级功能将被锁定，您将无法享受产品技术支持、故障维修等售后服务！请致电 400-6059-110 联系维保！

基本信息

任务名称 [D] 10.50.3.135_240313163651

输入源 10.50.3.135
数据库类型: GBASE 主机: 10.50.3.135

敏感模板 默认模板 [刷新](#) [管理](#)

发现方式 ☒ 全量发现 ☐ 增量发现
全量发现会采样所有数据表对应的元数据，增量发现只采样未进行过敏感数据发现且未指定敏感标签的元数据。

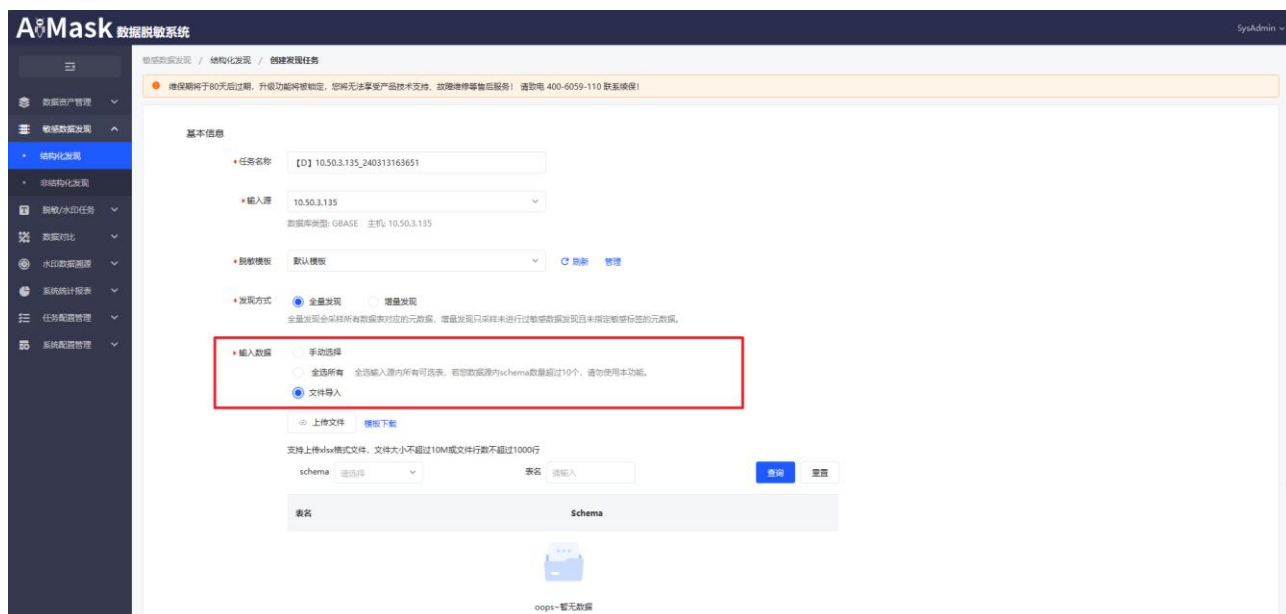
输入数据 ☒ 手动选择
☐ 全选所有 全选输入源内所有可选表，若您数据源内schema数量超过10个，请勿使用本功能。
☐ 文件导入

高级配置 [收起](#)

采样行数 1000

命中阈值 80 %

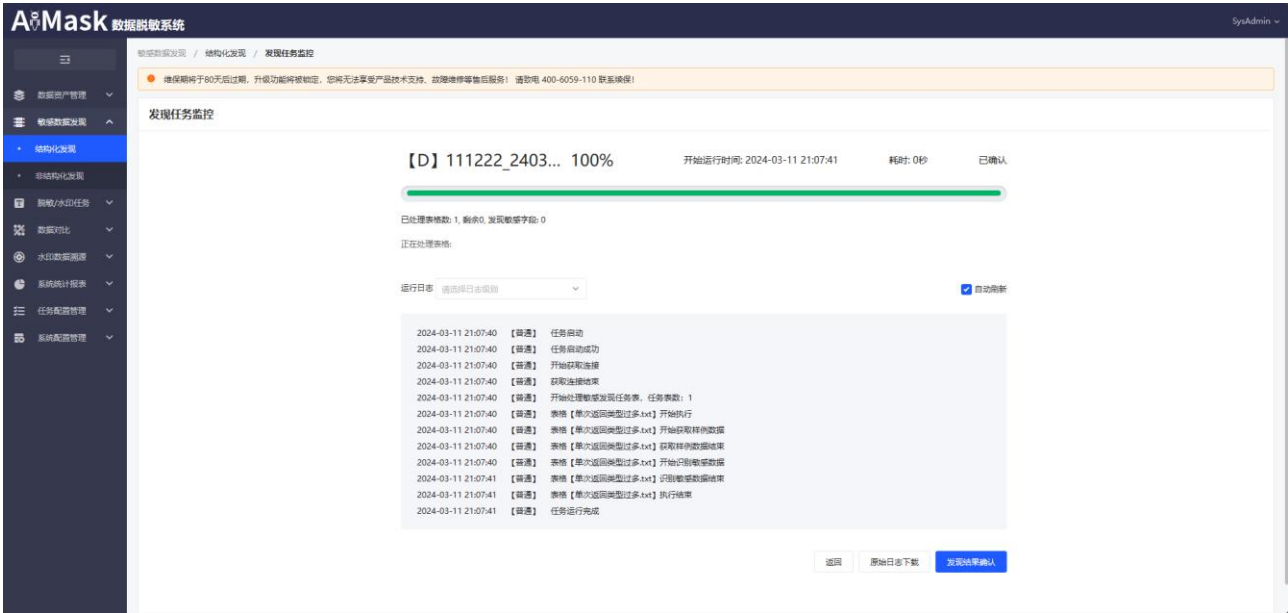
[取消](#) [保存](#) [保存并运行](#)



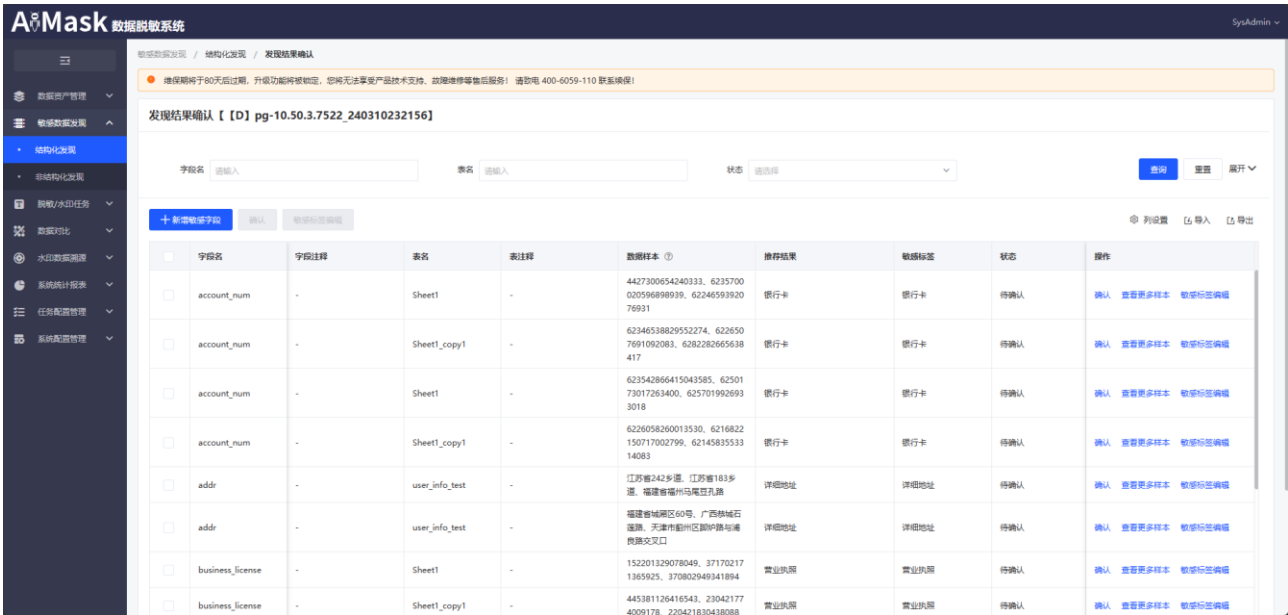
“高级配置”是我们系统默认的一些配置信息，一般是不需要修改的。我分别解释一下采样行数和命中阈值的概念。AiMask 做敏感数据发现时，一般是对数据内容或者数据包含的列名进行解析。如果我们要判断一个列是否是手机号码，一般情况下我们要拿到客户数据库里的一部分数据内容的。具体拿多少就是用采样行数来控制的，我们系统默认推荐的 1000 行是我们在大量客户实践过程中经验所得的一个相对合适的值。命中阈值跟客户的数据质量有关，如果客户数据质量很高，比如说他手机号码这一列的数据大多都是非常正规的手机号码，那我们设置的命中阈值可以高一些，比如 90%或者 95%。如果客户的数据质量比较差，我们建议设置的低一些。它内含的逻辑是我们从客户数据中取了 1000 行数据，如果其中有 800 行以上的数据符合设定的手机号码的敏感发现规则，我们就认为这一列的敏感标签是手机号码。

此时我们可以点击“保存”或“保存并运行”按钮。如果我们选择的是后者，他会自动跳到“发现任务监控”页面去实时展示这个敏感发现任务的运行细节，包括进度、运行状态、耗时、报错、日志等。我们也可以去切换日志级别来关注重点。如果是选择前者，他会自动返回上级页面，那我们点击“运行”按钮就可以运行任务了，但是此时是不会自动跳出监控页面的，需

要用户手动去点击“监控”按钮查看。我们也提供了日志下载按钮和终止任务按钮等功能。点击<监控>可实时观察敏感发现任务的日志信息、进度、报错，也可下载原始日志。



此外，本模块提供<确认>按钮供用户查看数据样本或手动对字段对应的敏感标签进行修改，如下图所示。



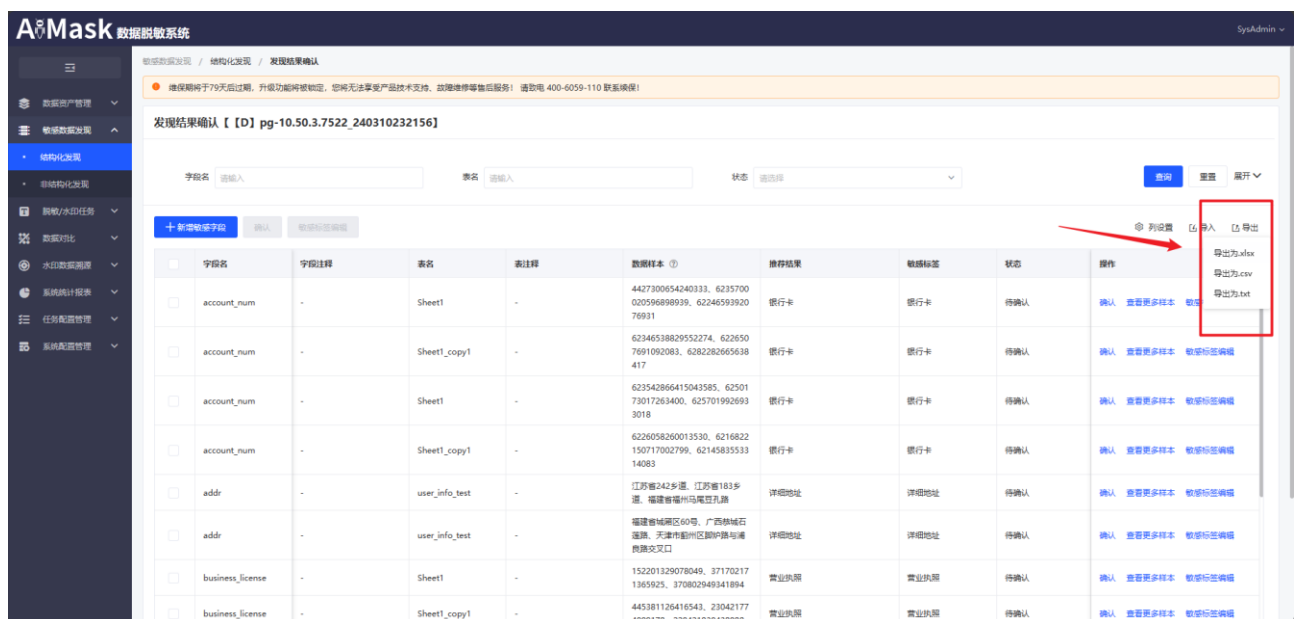
导入导出敏感字段

在结构化发现列表中，点击操作列<确认>或<详情>，进入发现结果确认页。页面支持对敏

感字段进行导入或导入操作，用户可在离线文件中快速修改和确认字段敏感标签。具体操作

如下：

- ◆ 点击<导出>，默认导出全部敏感字段，用户可自行选择以 xlsx、csv 或 txt 格式导出；
- ◆ 基于上述导出文件，用户可使用<导入>功能将修改后的敏感发现结果重新导入系统（其中，用户手动新增数据视为新增敏感字段，手动修改数据将覆盖系统原有结果，手动删除数据不影响系统原有结果）；
- ◆ 导入后，所有已导入字段的状态列自动变更为“已确认”。

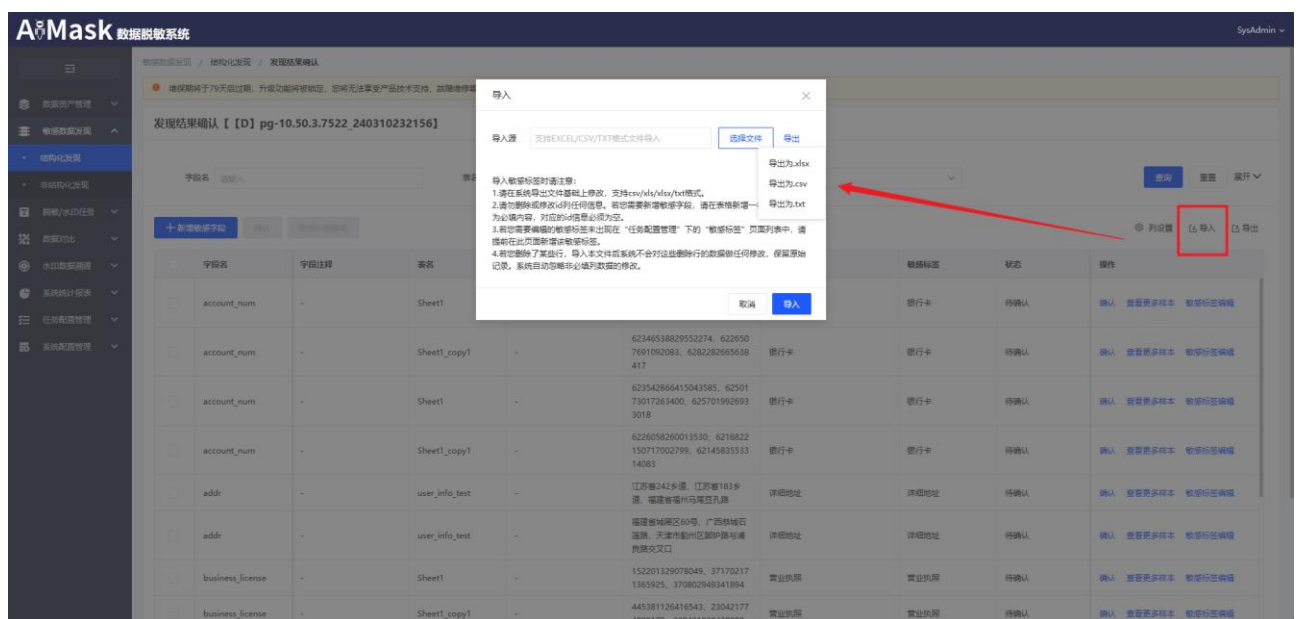


发现结果确认 [[D] pg-10.50.3.7522_240310232156]

字段名 请输入 表名 请输入 状态 请选择 查询 重置 展开

新增敏感字段 确认 敏感标签编辑

复选框	字段名	字段注释	表名	表注释	数据样本	推荐结果	敏感标签	状态	操作
☐	account_num	-	Sheet1	-	4427300654240333, 6235700020596899939, 6224659392078931	银行卡	银行卡	待确认	确认 查看更多样本 敏感标签编辑
☐	account_num	-	Sheet1_copy1	-	62346538829552274, 6226507691092083, 628228265638417	银行卡	银行卡	待确认	确认 查看更多样本 敏感标签编辑
☐	account_num	-	Sheet1	-	623542866415043585, 6250173017263400, 6257019926933018	银行卡	银行卡	待确认	确认 查看更多样本 敏感标签编辑
☐	account_num	-	Sheet1_copy1	-	6226058260013530, 6216822150717002799, 6214583553314083	银行卡	银行卡	待确认	确认 查看更多样本 敏感标签编辑
☐	addr	-	user_info_test	-	江苏省242乡道, 江苏省183乡道, 福建省福州市马尾区孔路	详细地址	详细地址	待确认	确认 查看更多样本 敏感标签编辑
☐	addr	-	user_info_test	-	福建省城南区60号, 广西桂林石道渠, 天津市静海区团泊新城南交口	详细地址	详细地址	待确认	确认 查看更多样本 敏感标签编辑
☐	business_license	-	Sheet1	-	152201329078049, 371702171365925, 370802949341894	营业执照	营业执照	待确认	确认 查看更多样本 敏感标签编辑
☐	business_license	-	Sheet1_copy1	-	445381126416543, 230421774009178, 220421830438088	营业执照	营业执照	待确认	确认 查看更多样本 敏感标签编辑



发现结果确认 [[D] pg-10.50.3.7522_240310232156]

字段名 请输入 表名 请输入 状态 请选择 查询 重置 展开

新增敏感字段 确认 敏感标签编辑

导入 支持EXCEL/CSV/TXT格式文件导入 选择文件 导出

导出为xlsx 导出为csv 导出为txt

导入敏感标签时请注意！

- 1.请在系统导出文件基础上修改，支持csv/xlsx/xls/txt格式。
- 2.请勿删除或修改已存在任何敏感数据，敏感数据新增敏感字段，请在表格新增一列为敏感字段，对应敏感数据必须存在。
- 3.若您需要编辑的敏感标签非出现在“任务配置管理”下的“敏感标签”页面列表中，请提前在此页面新增敏感标签。
- 4.若您删除了某一行，导入本文件或系统不会对已删除行的数据做任何修改，保留原始记录，系统会自动标记为敏感数据的修改。

取消 导入

导入敏感标签时请注意：



- ◆ 请在系统导出文件基础上修改，支持 csv/xls/xlsx/txt 格式。
- ◆ 请勿删除或修改 id 列任何信息。若您需要新增敏感字段，请在表格新增一行内容，带*列为必填内容，对应的 id 信息必须为空。
- ◆ 若您需要编辑的敏感标签未出现在“任务配置管理”下的“敏感标签”页面列表中，请提前在此页面新增该敏感标签。
- ◆ 若您删除了某些行，导入本文件后系统不会对这些删除行的数据做任何修改，保留原始记录。系统自动忽略非必填列数据的修改。

4.1.4 脱敏/水印任务

结构化任务

‘结构化任务’页面用于对结构化脱敏/水印任务进行管理。

A9Mask 数据脱敏系统

SysAdmin

数据源管理

敏感数据发现

脱敏/水印任务

结构化任务

结构化案例

非结构化任务

非结构化案例

数据资产化

水印数据溯源

系统统计报表

任务配置管理

系统配置管理

脱敏/水印任务 / 结构化任务

● 维保期将于79天后过期，升级功能将被锁定，您将无法享受产品技术支持、故障维修等售后服务！ 请致电 400-6059-110 联系维保！

任务名称

数据源名称

运行状态

新增

运行

删除

详情

重置

展开

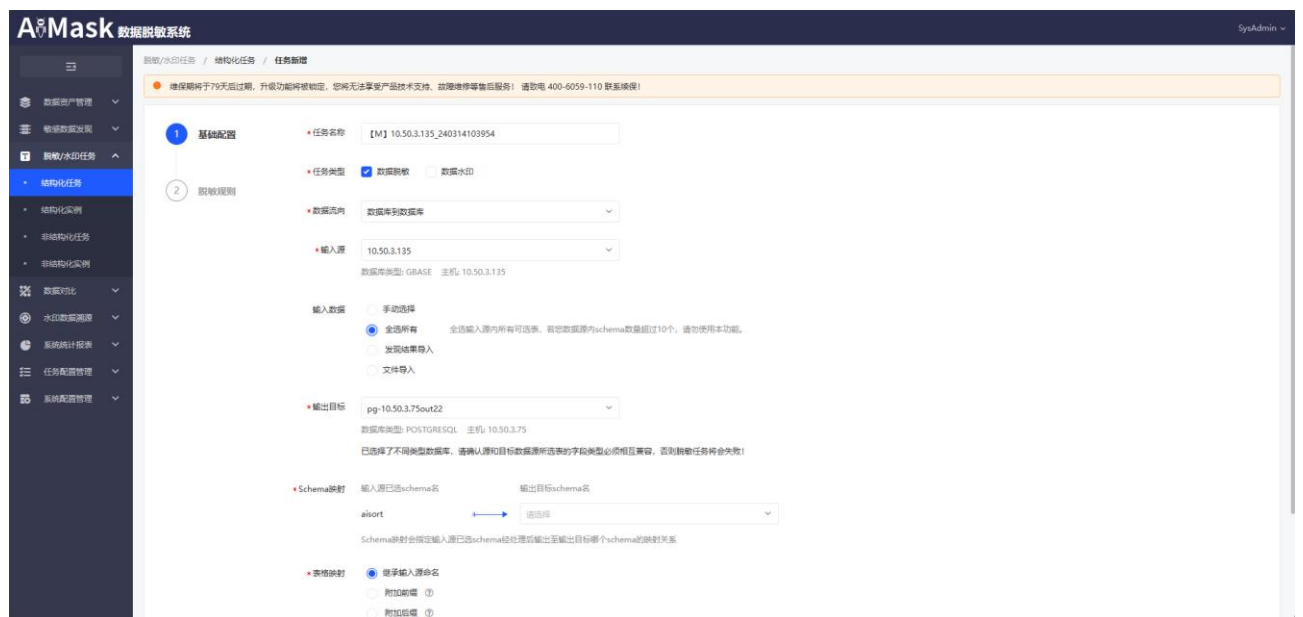
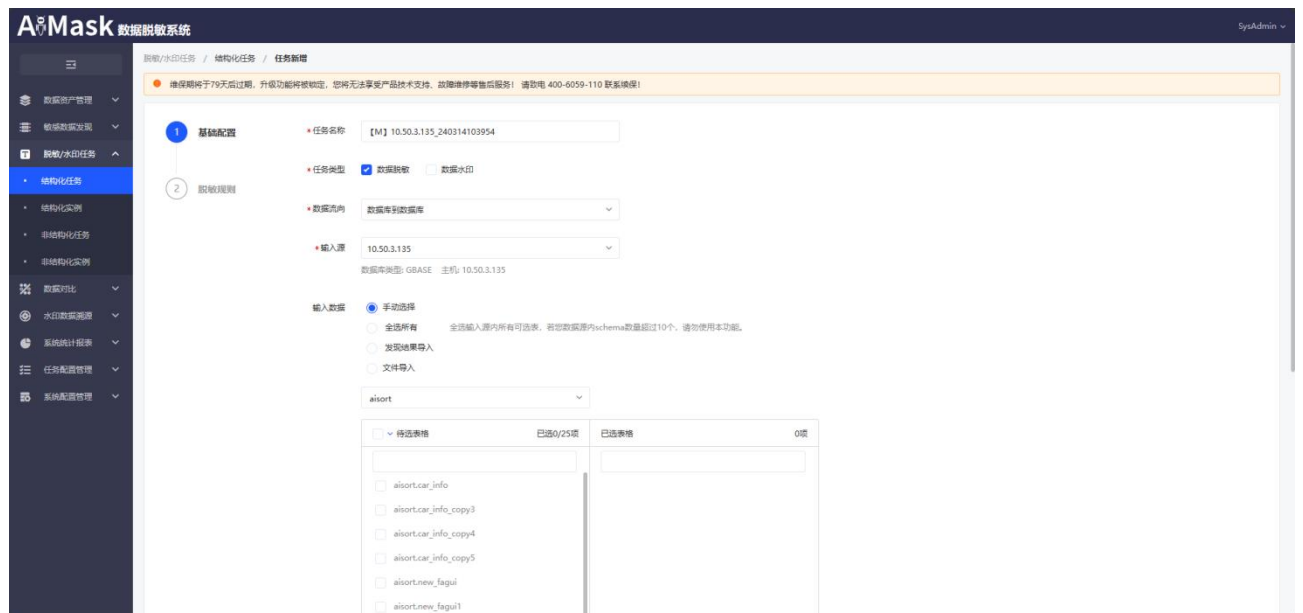
任务名称	任务类型	数据流向	数据源名称	敏感统计	执行周期	最近运行时间	最近运行状态	创建用户	创建时间	操作
☐ [M] DM8_2403111100591	数据脱敏	数据库 → 数据库	DM8	敏感字段：0 总字段：13	关闭	-	新建中	SysAdmin	2024-03-13 1	运行 环境检测 执行策略 ...
☐ 对mask	数据水印	数据库 → 数据库	mysql-10.50.3.73_240311143918	敏感字段：1 总字段：2	关闭	-	新建中	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
☐ [M] mysql-10.50.3.73_240311143918	数据脱敏	数据库 → 文件	mysql-10.50.3.73	敏感字段：1 总字段：2	关闭	-	新建中	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
☐ 同沙表对象	数据水印	数据库 → 数据库	mysql-10.50.3.73	敏感字段：1 总字段：2	关闭	2024-03-11 14:28:38	完成	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
☐ 5	数据脱敏	数据库 → 数据库	mysql-10.50.3.73	敏感字段：1 总字段：2	关闭	-	待运行	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
☐ 2	数据水印	文件 → 文件	del1	敏感字段：12 总字段：20	关闭	-	待运行	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
☐ 1	数据脱敏	文件 → 文件	del1	敏感字段：12 总字段：20	关闭	-	新建中	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
☐ [M] del1_240311141352	数据脱敏	文件 → 文件	del1	敏感字段：12 总字段：20	关闭	-	待运行	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
☐ 库到库水印	数据水印	数据库 → 数据库	mysql-10.50.3.73	敏感字段：1 总字段：2	关闭	-	待运行	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
☐ del1del1	数据脱敏和水印	文件 → 文件	del1	敏感字段：12 总字段：20	关闭	2024-03-11 14:08:07	待运行	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
☐ del1file	数据脱敏	数据库 → 文件	mysql-10.50.3.73	敏感字段：1 总字段：2	关闭	2024-03-11 13:50:51	完成	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
☐ 数据库到水印	数据脱敏和水印	数据库 → 文件	mysql-10.50.3.73	敏感字段：1 总字段：2	关闭	-	待运行	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
☐ [M] mysql-10.50.3.110_240308164537	数据脱敏	数据库 → 数据库	mysql-10.50.3.110	敏感字段：2 总字段：2	关闭	2024-03-08 16:46:17	完成	SysAdmin	2024-03-08 1	运行 环境检测 执行策略 ...
☐ [M] mysql-10.50.3.73_240308164537	数据脱敏	数据库 → 数据库	mysql-10.50.3.73	敏感字段：1 总字段：2	关闭	2024-03-08 10:40:46	新建中	SysAdmin	2024-03-08 1	运行 环境检测 执行策略 ...

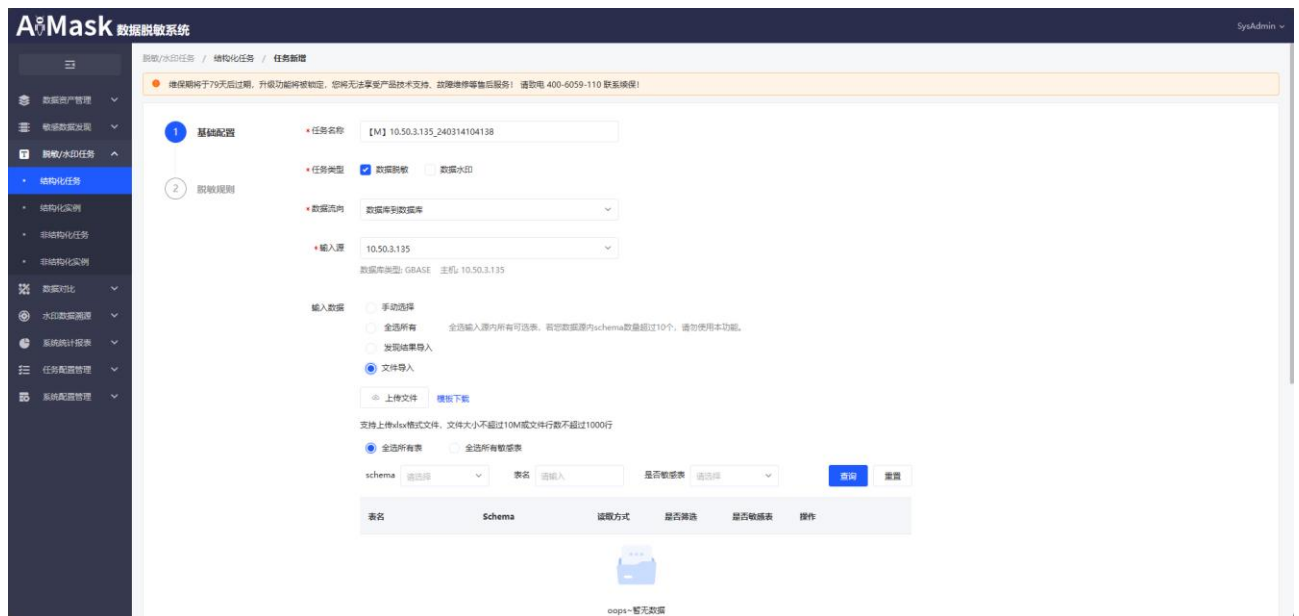
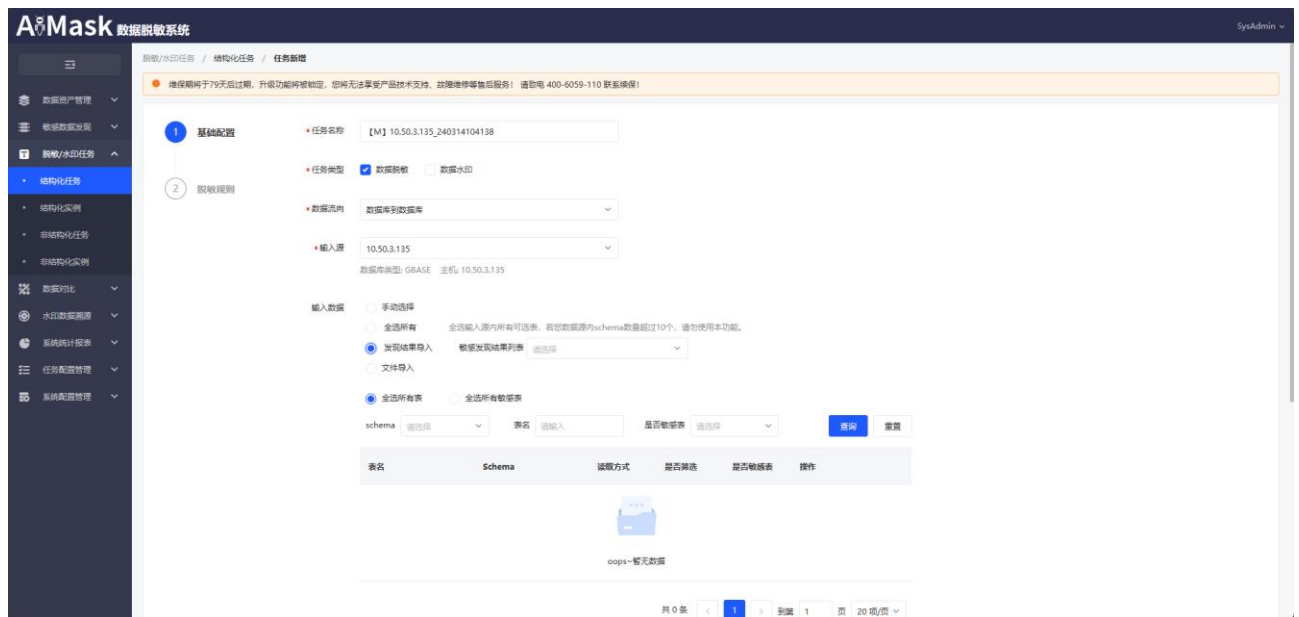
基础配置

用户可以通过<新增>增加脱敏和水印任务，具体可以在任务类型中对“数据脱敏”和“数据水印”类型多选。刚打开本页面时，系统会自动将最近敏感发现完成的结构化输入源名称填入“输入源”参数，若用户需要进行脱敏或水印操作的数据源并非此数据源，可以点击下拉框更

换选择。

数据流向提供了 4 种组合，此外 AiMask 提供 4 种输入数据的方式，分别为手动选择、全
选所有、发现结果导入和文件导入。





点击蓝色高亮的任务名称可查看任务的详细信息，如下图所示。

数据资产管理

数据数据发现

数据/水印任务

结构化任务

非结构化任务

数据资产北

水印数据源

系统统计报表

任务配置管理

系统配置管理

数据/水印任务 / 结构化任务

任务名称 任务类型 数据流向 数据源名称 敏感统计 执行周期 最近运行时间 最近运行状态 创建用户 创建时间 操作

任务名称

任务类型

数据流向

数据源名称

敏感统计

执行周期

最近运行时间

最近运行状态

创建用户

创建时间

操作

[M] DM8_2403131100591	数据脱敏	数据库 -> 数据库	DM8	敏感字段: 0 总字段: 13	关闭	-	新建中	SysAdmin	2024-03-13 1	运行 环境检测 执行策略 ...
刘建mask	数据水印	数据库 -> 数据库	mysql-10.50.3.73_240311143918	敏感字段: 1 总字段: 2	关闭	-	新建中	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
[M] mysql-10.50.3.73_240311143918	数据脱敏	数据库 -> 文件	mysql-10.50.3.73	敏感字段: 1 总字段: 2	关闭	-	新建中	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
同步表对象	数据水印	数据库 -> 数据库	mysql-10.50.3.73	敏感字段: 1 总字段: 2	关闭	2024-03-11 14:28:38	完成	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
s	数据脱敏	数据库 -> 数据库	mysql-10.50.3.73	敏感字段: 1 总字段: 2	关闭	-	待运行	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
2	数据水印	文件 -> 文件	del1	敏感字段: 12 总字段: 20	关闭	-	待运行	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
1	数据脱敏	文件 -> 文件	del1	敏感字段: 12 总字段: 20	关闭	-	新建中	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
[M] del1_240311141352	数据脱敏	文件 -> 文件	del1	敏感字段: 12 总字段: 20	关闭	-	待运行	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
库到库水印	数据水印	数据库 -> 数据库	mysql-10.50.3.73	敏感字段: 1 总字段: 2	关闭	-	待运行	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
del1del1	数据脱敏和水印	文件 -> 文件	del1	敏感字段: 12 总字段: 20	关闭	2024-03-11 14:08:07	待运行	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
到file	数据脱敏	数据库 -> 文件	mysql-10.50.3.73	敏感字段: 1 总字段: 2	关闭	2024-03-11 13:50:51	完成	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
数据库到水印	数据脱敏和水印	数据库 -> 文件	mysql-10.50.3.73	敏感字段: 1 总字段: 2	关闭	-	待运行	SysAdmin	2024-03-11 1	运行 环境检测 执行策略 ...
[M] mysql-10.50.3.110_240308164557	数据脱敏	数据库 -> 数据库	mysql-10.50.3.110	敏感字段: 2 总字段: 2	关闭	2024-03-08 16:46:17	完成	SysAdmin	2024-03-08 1	运行 环境检测 执行策略 ...
[M] mysql-10.50.3.73_240308164557	数据脱敏	数据库 -> 数据库	mysql-10.50.3.73	敏感字段: 1 总字段: 2	关闭	2024-03-08 10:40:46	新建中	SysAdmin	2024-03-08 1	运行 环境检测 执行策略 ...

数据/水印任务 / 结构化任务 / 任务详情

任务名称: [M] DM8_2403131100591

任务类型: 数据脱敏

数据流向: 数据库到数据库

输入源: DM8

输入数据:

表名	Schema	读取方式	是否筛选	是否敏感表	操作
API_STORE_ARCHIVE	AHDB	全量	否	未处理	增量配置 同步配置

输出目标: pg-10.50.3.75out22

Schema映射: 输入源已选schema名 AHDB 输出目标schema名 htest

策略映射: 继承输入源命名

覆盖方式: 清空数据 (truncate)

共 1 条 < 1 > 到第 1 页 20 页/页

取消 下一步

下方以某任务示例，展示基础配置的一种填写方式。

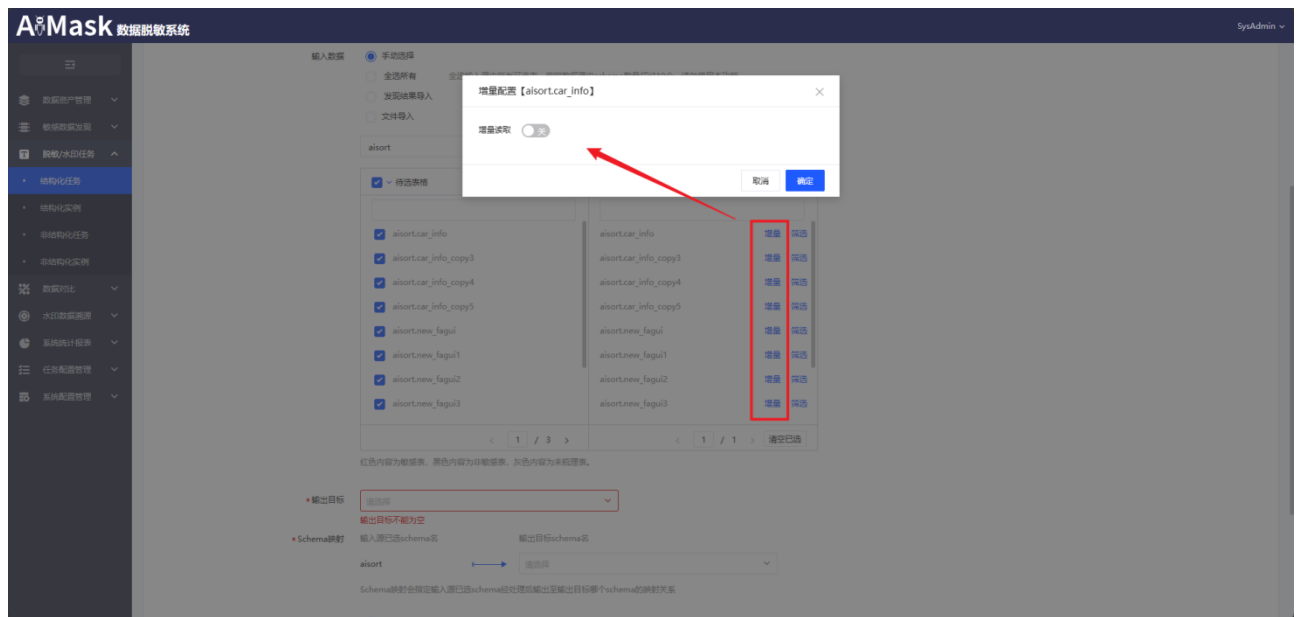


当数据流向为“数据库到数据库”时，用户可以自主设置是否启用“增量读取”，开启后可以以增量方式处理源表更新数据。启用该功能后，可进一步选择增量方式：

- ◆ “插入”指增量数据以插入方式写入目标，若目标数据存在，则插入失败或重复插入，用户若选择插入则需指定自增主键或时间作为增量标识；
- ◆ “更新”指增量数据以更新方式写入目标，若目标数据存在，则更新成最新的数据，用户若选择更新则需同时指定主键和时间作为增量标识。



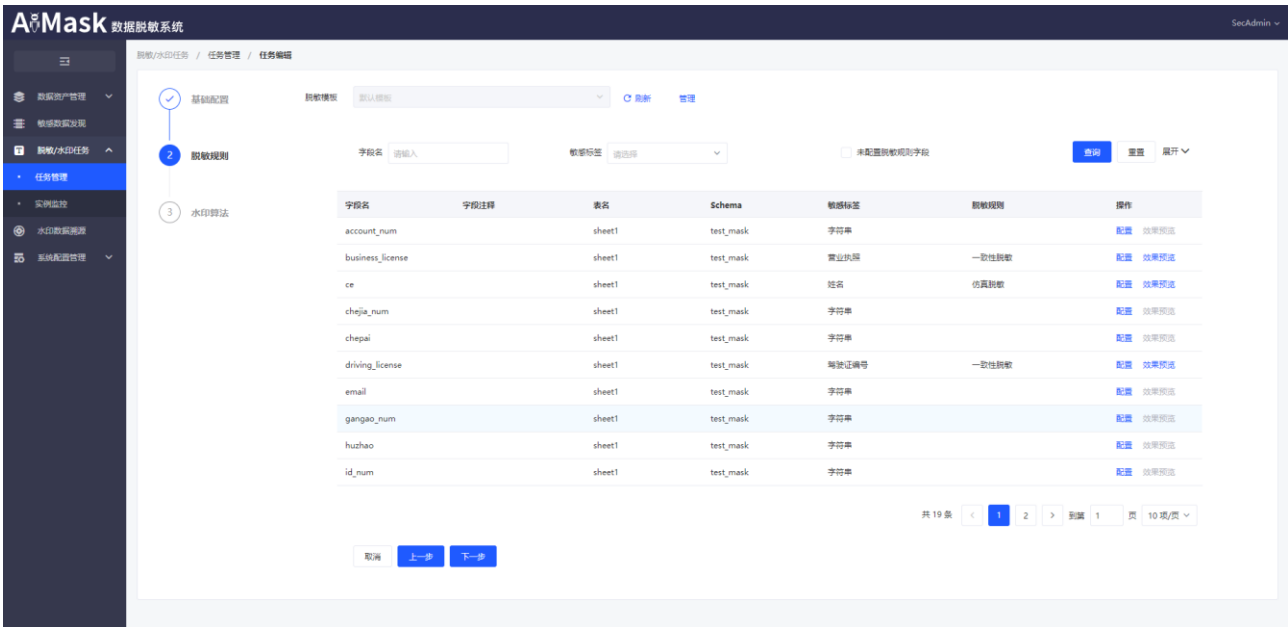
仅有少部分数据库类型支持更新增量。



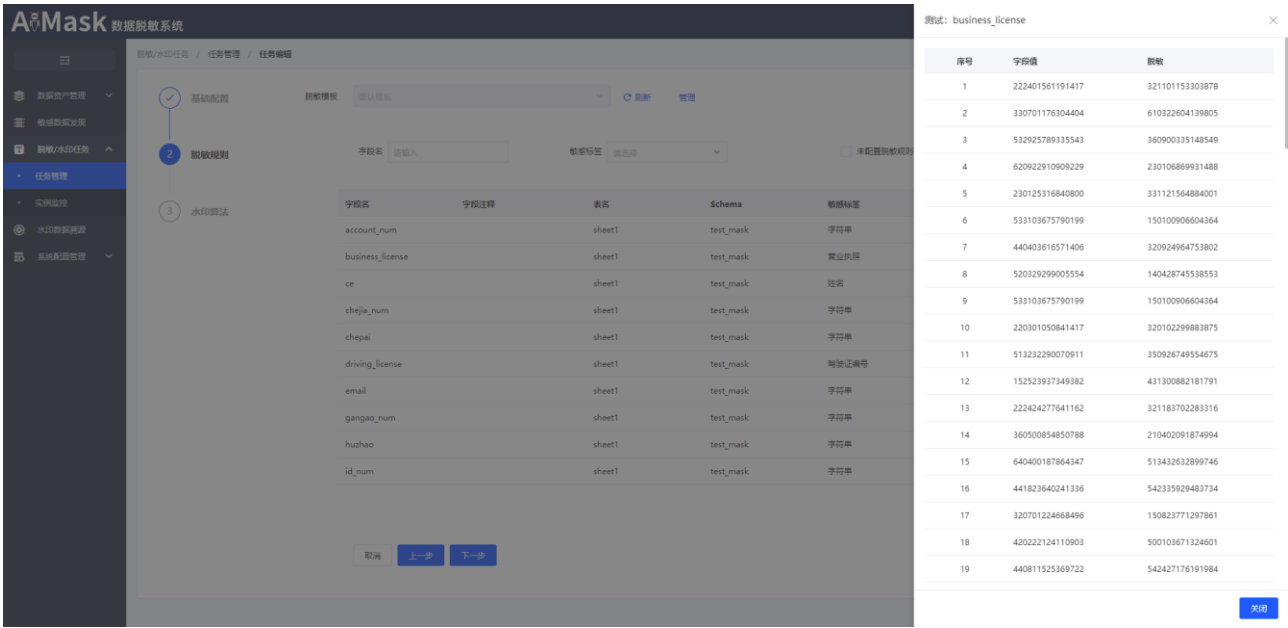
脱敏规则

完成基础配置后，点击<下一步>，用户若勾选的任务类型是“数据脱敏”，则进入“脱敏规则”配置页。配置数据脱敏任务时，用户可以选择手动配置所需字段对应的脱敏规则，或者一键选择 AiMask 提供的脱敏模板。如用户希望创建自己的脱敏模板，可点击管理按钮，进入对

应配置页面配置自定义脱敏模板。<保存>后点击<刷新>，即可在下拉框中选择新创建的自定义模板。如下图所示。点击未配置脱敏规则字段可以方便用户快速在众多字段中快速找到没有配置脱敏规则的字段，点击配置则可以手动切换脱敏规则。



AiMask 提供<效果预览>按钮，提供用户在运行任务前预知部分脱敏效果。

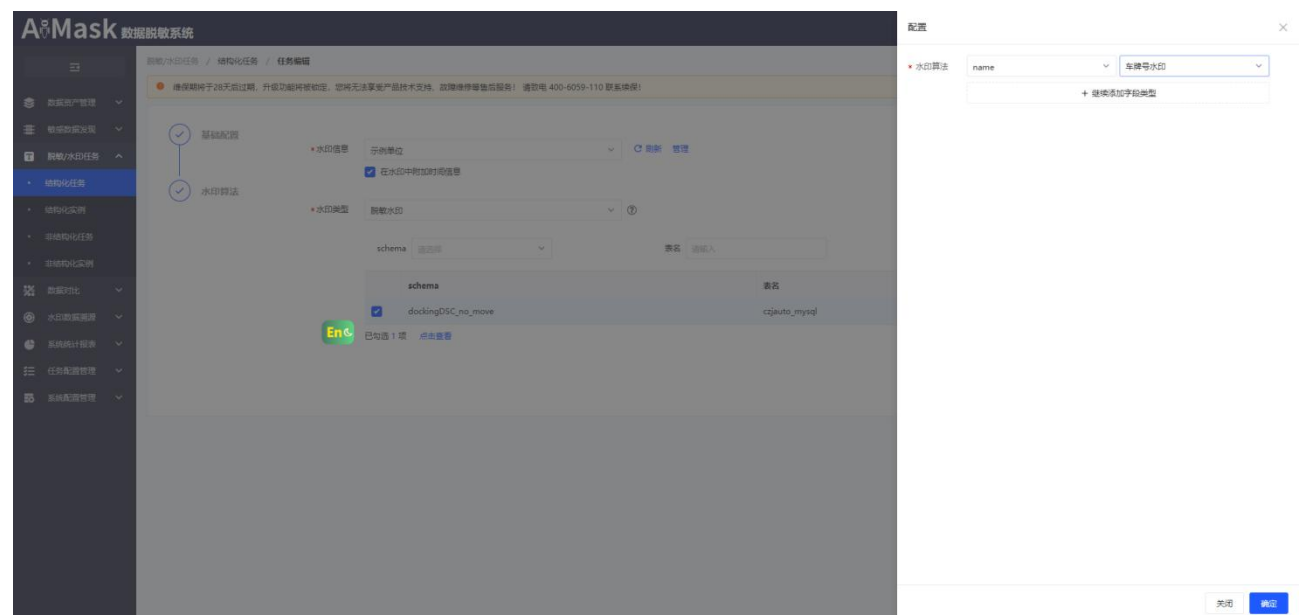


水印算法

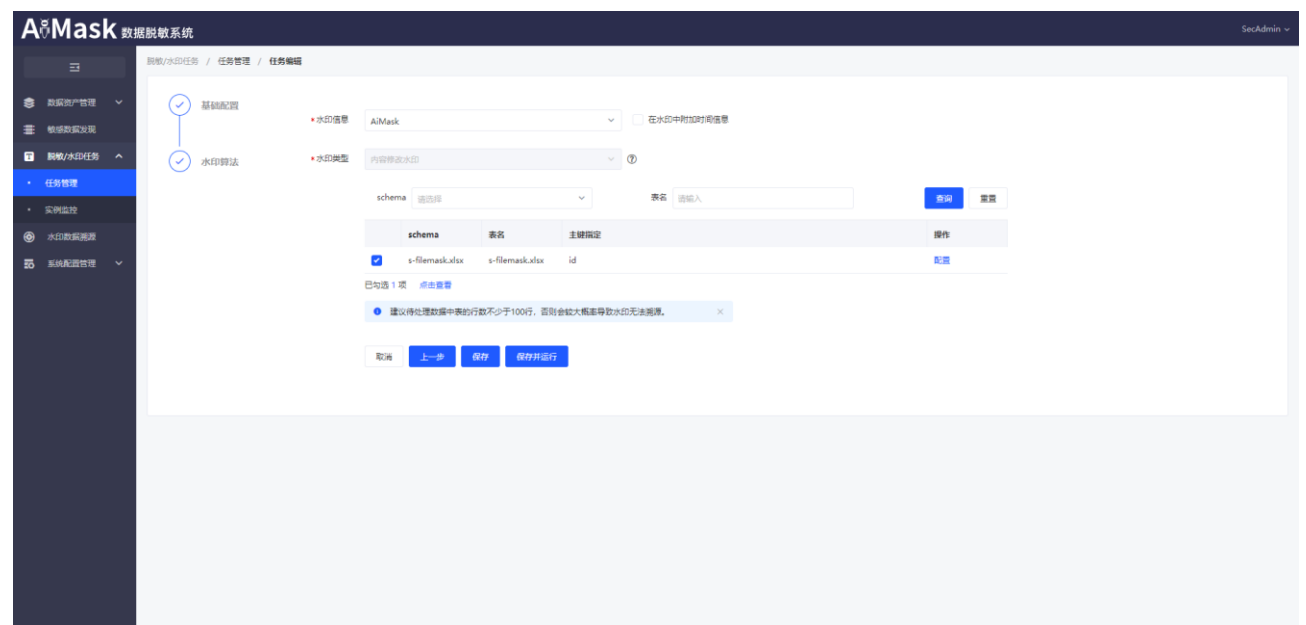
用户若为基础配置的任务类型中同时或单独勾选了“数据水印”，则提供“水印算法”配置页。

配置数据水印任务时，用户需要基于使用场景选择所需的数据水印类型。

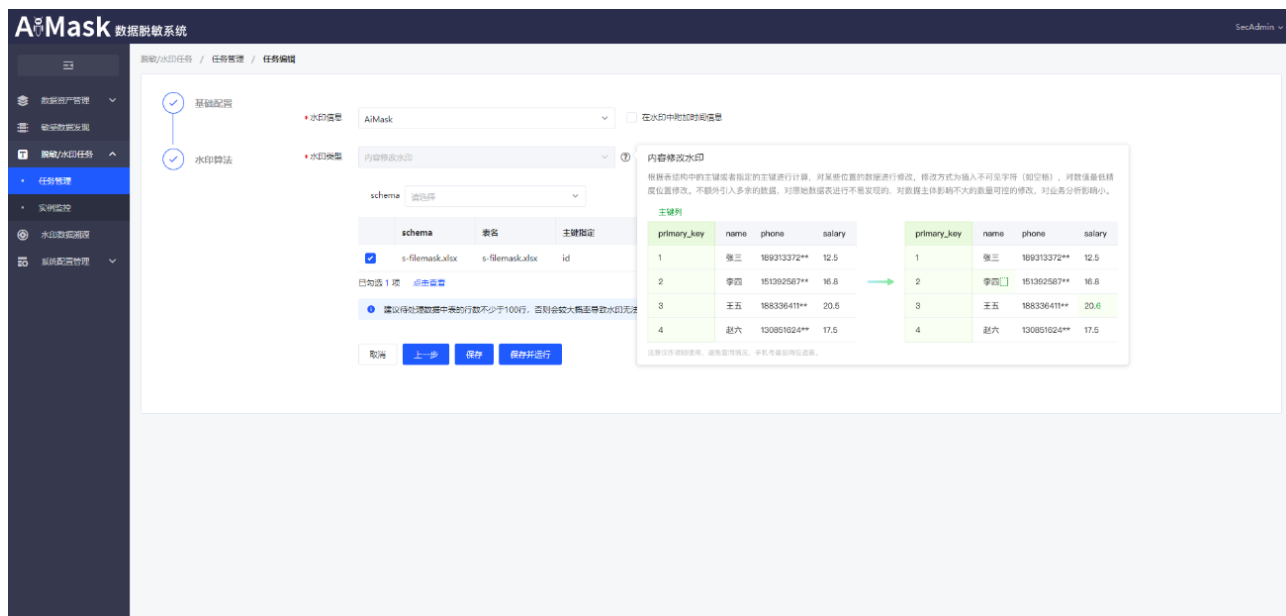
例如脱敏水印 点击配置 进入下面场景 我们选择了车牌号水印类型。然后点击右下角确定



再点击保存运行，水印任务即可启动。

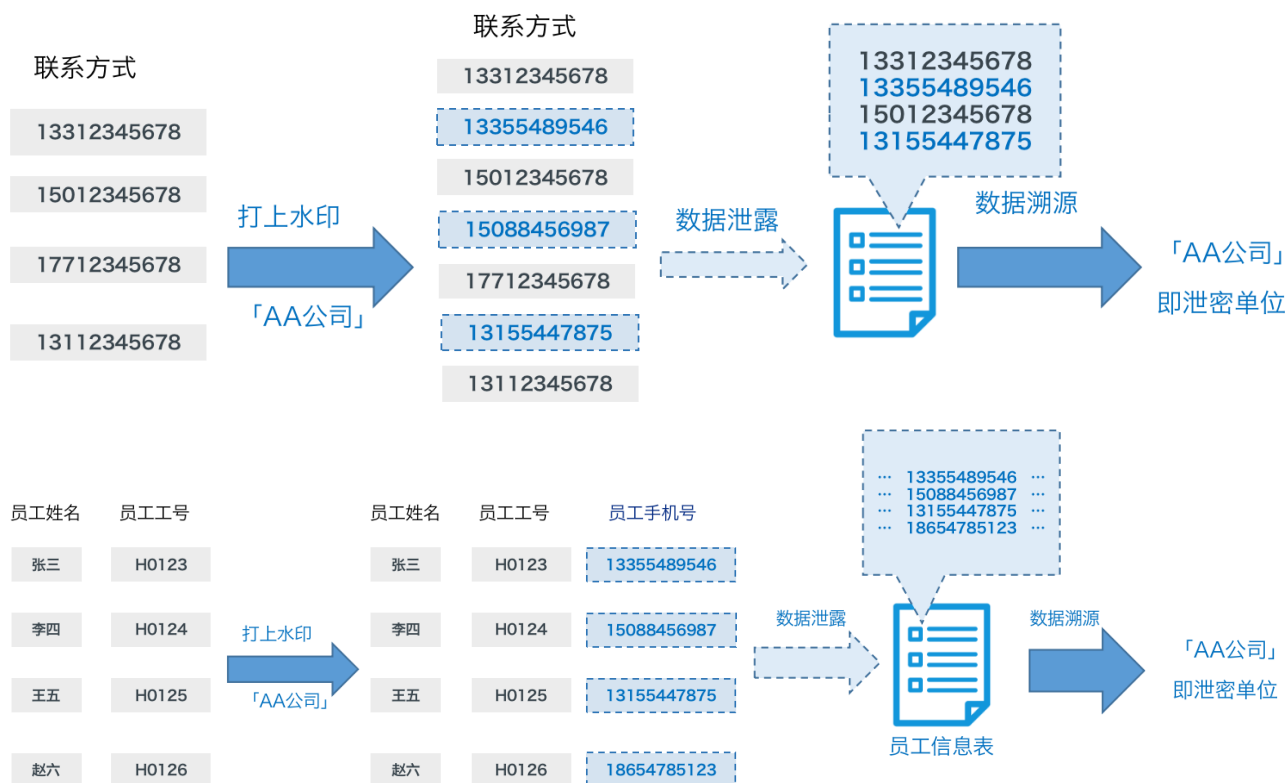


页面会展示选定水印类型的介绍，如下图所示。



AiMask 支持多种数据水印格式，包括并不限于伪行水印、伪列水印、脱敏水印、内容修改水印、零宽字符水印：

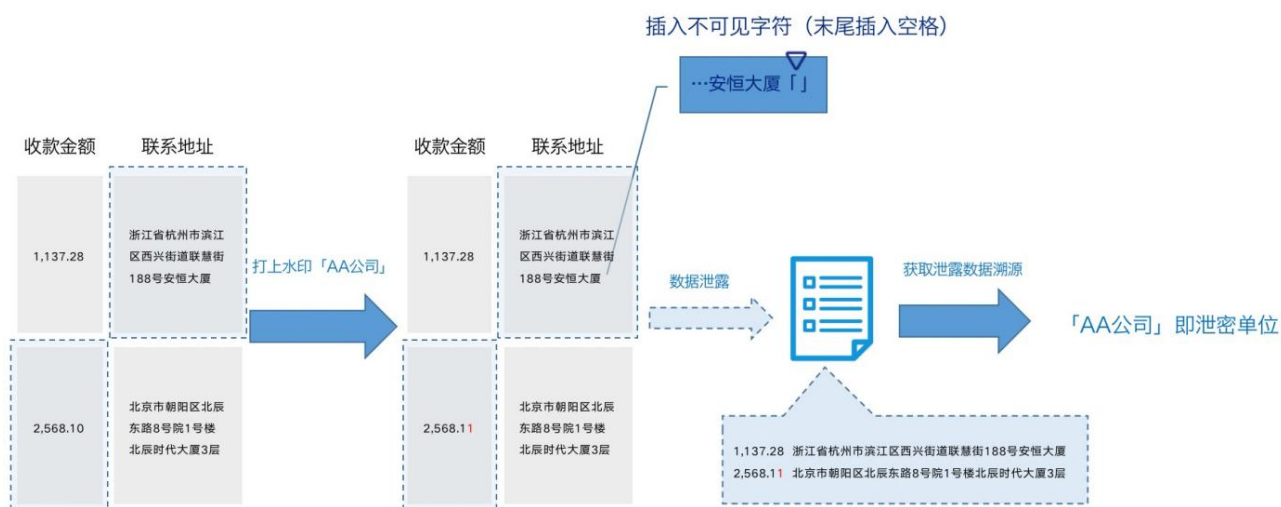
- ◆ 伪行/伪列水印：通过添加配置比例的数据行或数据列来插入水印信息，当数据泄露后，通过提取伪行/伪列水印信息来进行泄漏追溯。



- ◆ 脱敏水印：支持基于数据本身特征脱敏水印方式，不影响业务逻辑的同时能有效的避免了水印数据被绕开。



- ◆ 内容修改水印算法: 适用于不能影响数据的业务含义并具有较强的业务使用或分析需求的场景，同时隐蔽性要求高。



- ◆ 零宽字符水印算法: 可视宽度为零、溯源耗时短、无需主键列，适用于在文件数据内容中添加水印的场景，隐秘性和鲁棒性强。



结构化实例

本页面展示所有结构化实例的详细数据，包括已被删除的结构化任务对应的实例，如下图所示。

AiMask 数据脱敏系统 SysAdmin

数据/水印任务 / 结构化实例

● 维保期将于79天后过期，升级功能将被锁定，您将无法享受产品技术支持、故障维修等售后服务！ 请致电 400-6059-110 联系维保！

实例名称: 请输入 任务名称: 请输入 运行状态: 请选择 [按钮: 查询] [按钮: 重置] [按钮: 展开]

[按钮: 重置] [按钮: 暂停] [按钮: 运行] [按钮: 终止] [按钮: 列设置]

	实例名称	任务名称	任务类型	数据流向	运行耗时	结果统计	错误/告警	最近运行时间	最近运行状态	最近运行速度	操作
☐	WM0000005224 0311142837	同步表对象	数据水印	数据库 → 数据库	1秒	成功: 1 失败: 0	错误: 0 告警: 0	2024-03-11 14:28:38	完成	100%	运行 监控 报告 ...
☐	WM0000005224 0311142748	同步表对象	数据水印	数据库 → 数据库	2秒	成功: 1 失败: 0	错误: 0 告警: 0	2024-03-11 14:27:49	完成	100%	运行 监控 报告 ...
☐	WM0000005224 0311142705	同步表对象	数据水印	数据库 → 数据库	3秒	成功: 1 失败: 0	错误: 0 告警: 0	2024-03-11 14:27:05	完成	100%	运行 监控 报告 ...
☐	DW0000004624 0311140807	delid1	数据脱敏和水印	文件 → 文件	1秒	成功: 1 失败: 0	错误: 0 告警: 0	2024-03-11 14:08:07	完成	100%	运行 监控 报告 ...
☐	DW0000004624 0311140734	delid1	数据脱敏和水印	文件 → 文件	2秒	成功: 1 失败: 0	错误: 0 告警: 0	2024-03-11 14:07:34	完成	100%	运行 监控 报告 ...
☐	DM0000004524 0311130590	到File	数据脱敏	数据库 → 文件	3秒	成功: 1 失败: 0	错误: 0 告警: 0	2024-03-11 13:50:51	完成	100%	运行 监控 报告 ...
☐	DM0000004324 0308164616	[M] mysql-10.50.3.110_24 0308164557	数据脱敏	数据库 → 数据库	1秒	成功: 1 失败: 0	错误: 0 告警: 0	2024-03-08 16:46:17	完成	100%	运行 监控 报告 ...
☐	DM0000004224 0308104045	[M] mysql-10.50.3.73_2403 08104026	数据脱敏	数据库 → 数据库	2秒	成功: 1 失败: 0	错误: 0 告警: 0	2024-03-08 10:40:46	完成	100%	运行 监控 报告 ...
☐	DM0000003724 0306183743	[M] mask联动_测试台名单 扫描	数据脱敏	数据库 → 文件	1秒	成功: 1 失败: 0	错误: 0 告警: 0	2024-03-06 18:37:44	完成	100%	运行 监控 报告 ...
☐	DM0000003624 0306182859	[M] mask联动_240306182 656	数据脱敏	数据库 → 文件	5秒	成功: 1 失败: 0	错误: 0 告警: 0	2024-03-06 18:29:01	完成	100%	运行 监控 报告 ...
☐	DM0000002624 0304130019	[M] oracle-10.50.3.134_24 0222093811	数据脱敏	数据库 → 数据库	4秒	成功: 0 失败: 10	错误: 10 告警: 0	2024-03-04 13:00:20	完成	100%	运行 监控 报告 ...

点击<监控>将从运行日志、表格明细、对象同步三个角度对实例运行情况进行展示，如下

图所示。

AiMask 数据脱敏系统 SysAdmin

数据/水印任务 / 结构化实例 / 监控详情

● 维保期将于79天后过期，升级功能将被锁定，您将无法享受产品技术支持、故障维修等售后服务！ 请致电 400-6059-110 联系维保！

任务实例监控

同步表对象【WM00... 100%】 开始运行时间: 2024-03-11 14:28:38 耗时: 1秒 完成

已处理数据量: 1, 剩余: 0 处理速度: 1行/s, 0.02M/s

[按钮: 运行日志] [按钮: 表格明细] [按钮: 对象同步]

选择日志级别: [下拉菜单: 普通, 告警, 错误]

任务启动 任务启动成功

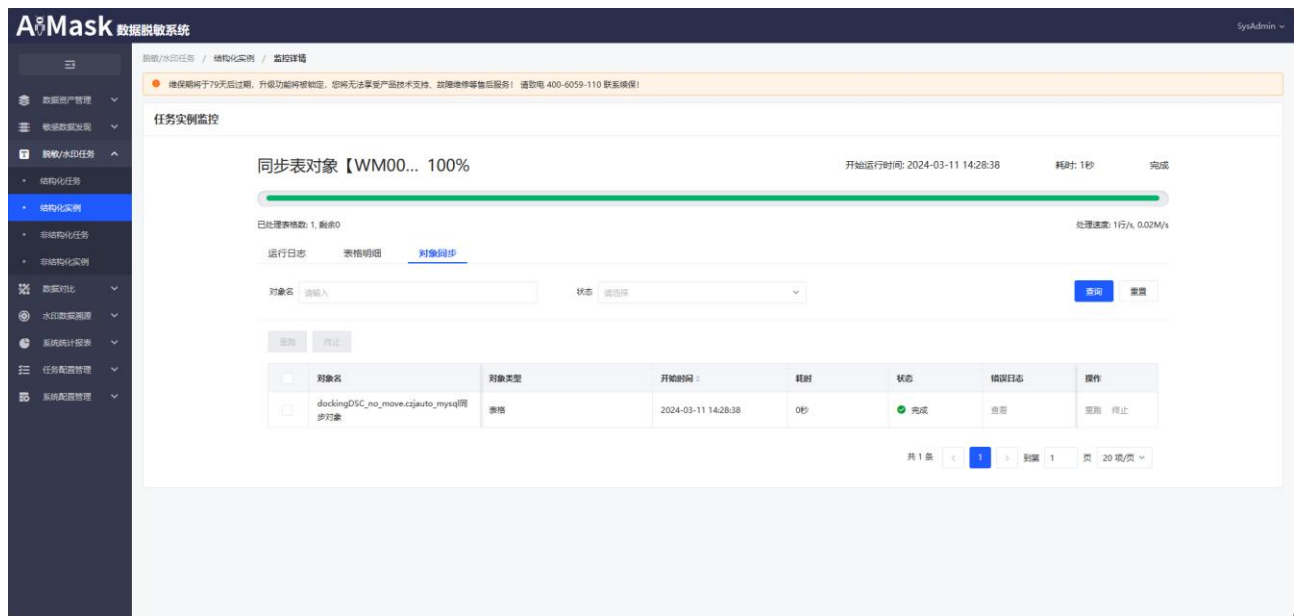
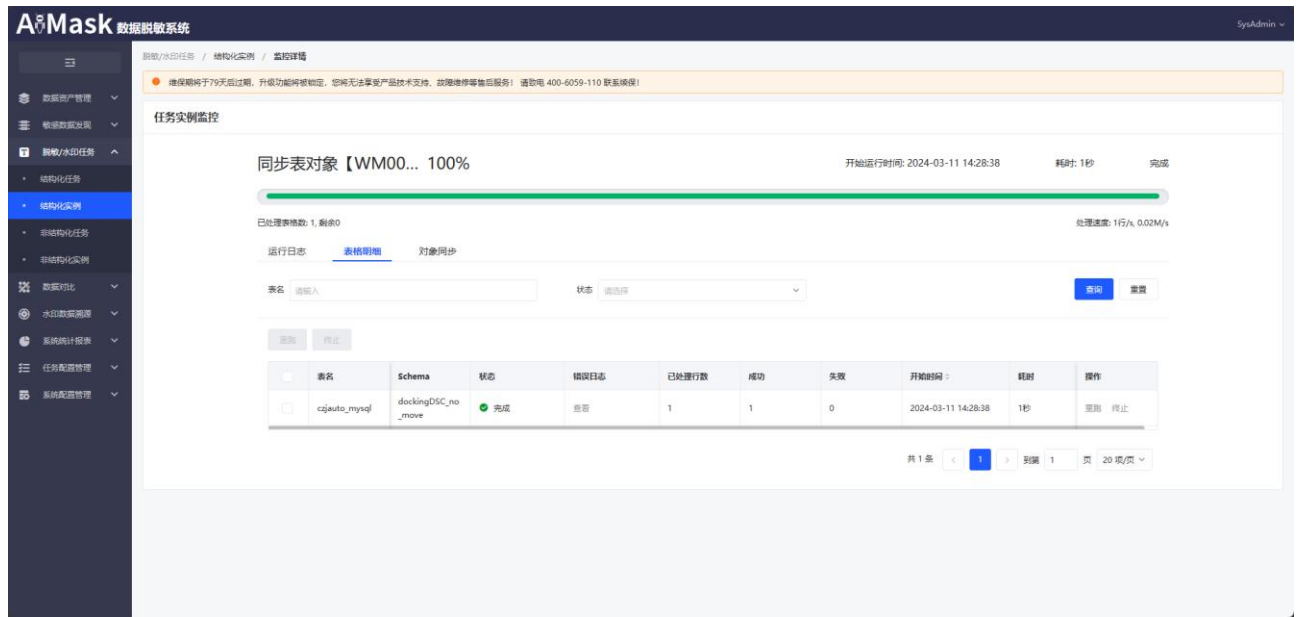
数据【dockingDSC_no_moveccjauto_mysql】开始处理

数据【dockingDSC_no_moveccjauto_mysql】已处理数据1条

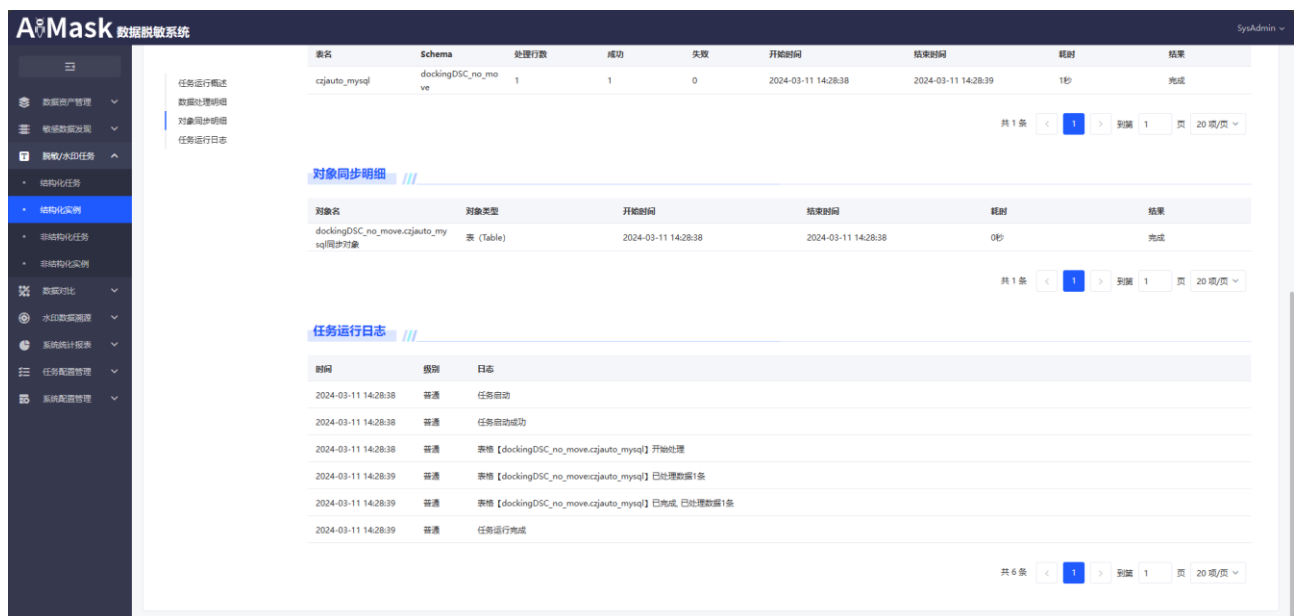
数据【dockingDSC_no_moveccjauto_mysql】已完成, 已处理数据1条

任务运行完成

[按钮: 返回] [按钮: 清除日志下载]



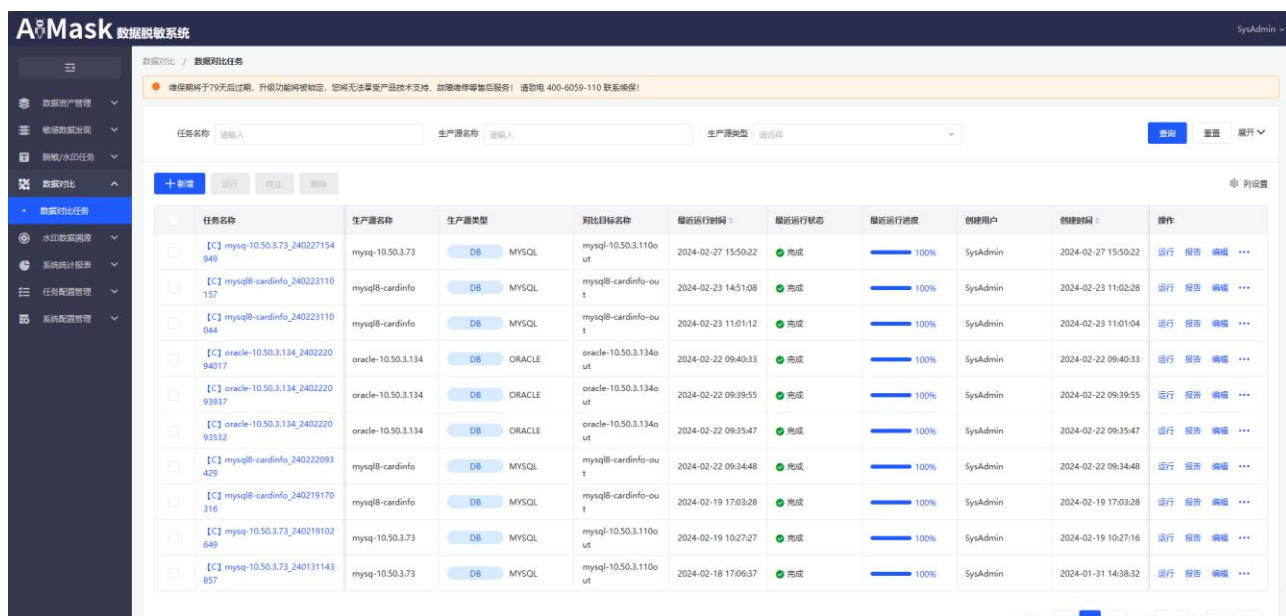
点击<报告>将会展示实例对应的各维度分析，如下图所示。





4.1.5 数据对比任务

‘数据对比任务’页面用于比较生产源和对比目标，轻松发现目标数据是否脱敏；也可用于验证数据脱敏处理效果，即对比脱敏前后的数据差异。



用户可以通过<新增>增加数据对比任务。用户可以在本页面填写任务名称，选择生产源及生产数据、对比目标及目标数据，并设置采样、特殊处理等对比配置。目前，系统仅支持对比同构数据库中的敏感字段，自动跳过非敏感字段，且表名及表所含字段名完全相同才算匹配上；不支持对比主键经过脱敏的数据。

天翼云

数据脱敏系统

任务名称

wwi测试数据源_340627134505

数据源选择

生产源

wwi测试数据源

数据库类型: MySQL 主机: 10.50.3.213

生产数据

wwitest

数据源/表选择: 已选1/12项 点击查看选择 重置

wwitest.financa_info

wwitest.id_documents

wwitest.odt_zymochi_dtl_yinq

wwitest.odt_zymochi_dtl_yinq

wwitest.odt_zymochi_dtl_yinq

wwitest.odt_zymochi_dtl_yinq

wwitest.person_details

wwitest.personal_info

wwitest.id_documents_copy1

1 / 2

1 / 1

添加已选

对比目标

wwi测试数据源

数据库类型: MySQL 主机: 10.50.3.213

目标数据

wwitest2

对比配置

采样数据

采样数据

采样比例

采样比例

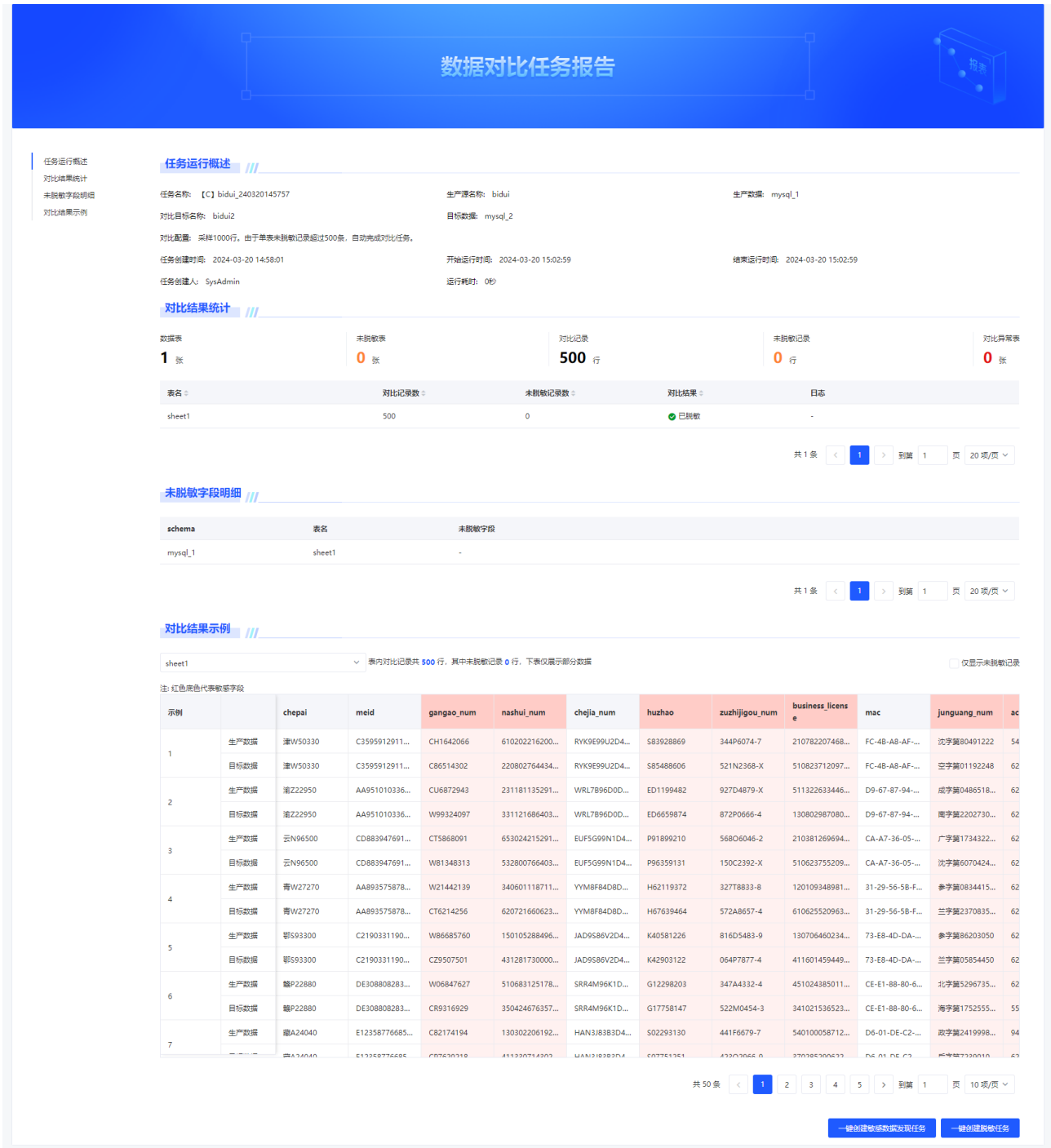
1000

采样比例

500

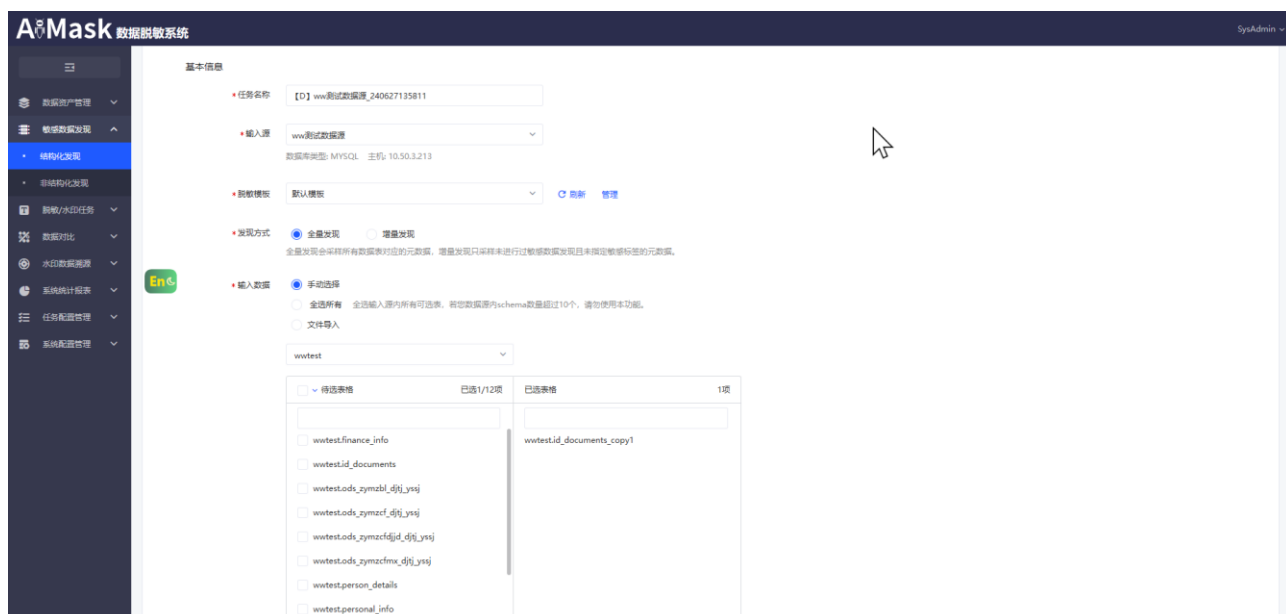
秒时, 自动完成对比任务

点击<报告>将会展示任务对应的各维度分析，如下图所示。



用户可以根据对比结果，一键处理未脱敏数据，具体操作路径如下：

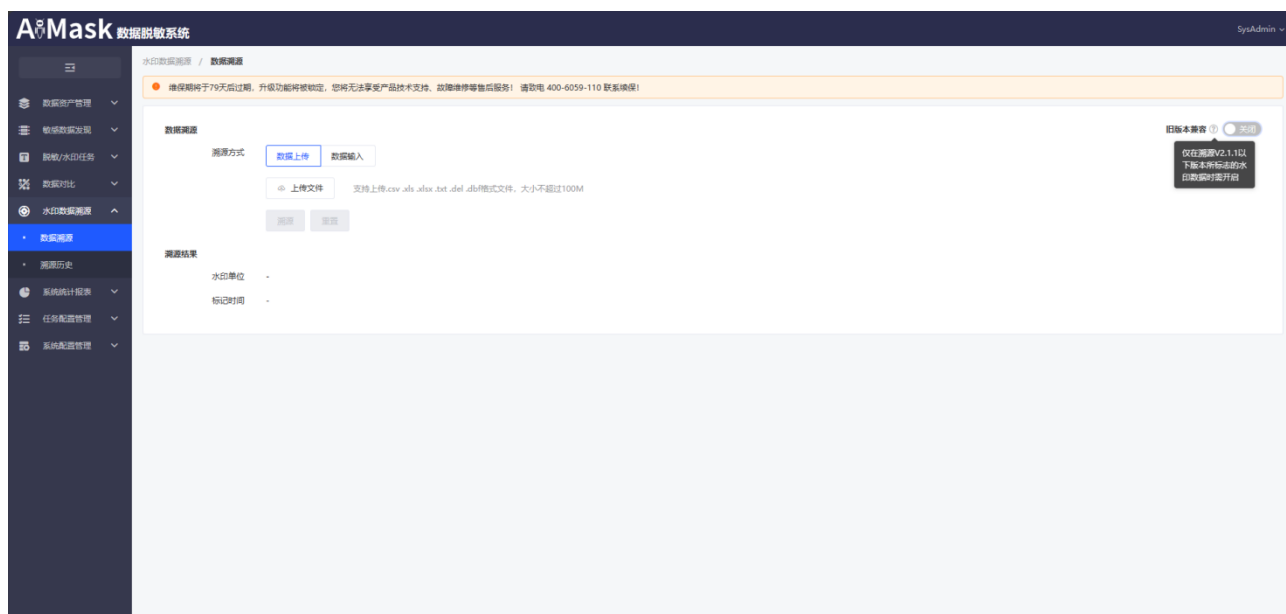
- ◆ 点击报告末尾的<一键创建敏感数据发现任务>，界面将跳转至敏感数据发现的任务创建页，自动生成一个结构化发现任务，并自动填入输入源（对应“生产源”）、输入数据、任务名称等参数；
- ◆ 或点击<一键创建脱敏任务>，界面将跳转至脱敏/水印任务的任务创建页，自动生成一个脱敏任务，并自动填入输入源、输入数据、输出目标（对应“对比目标”）、任务名称等参数。

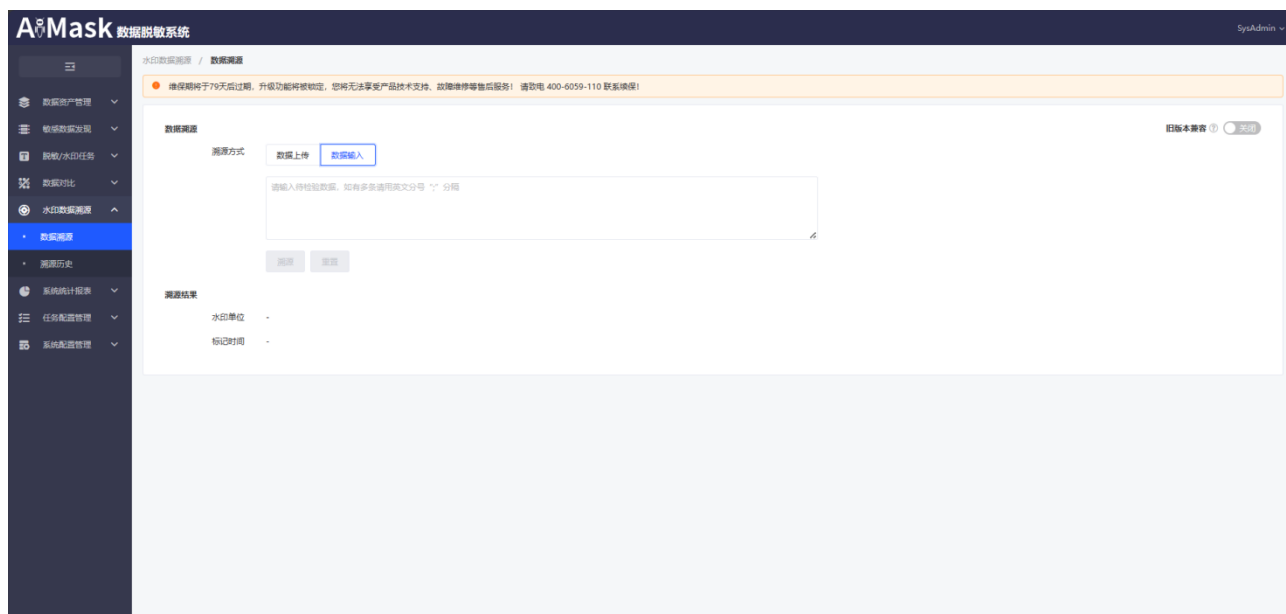


4.1.6 水印数据溯源

用户可以用此功能来定位水印处理后数据的单位信息（即对应的水印名称）。

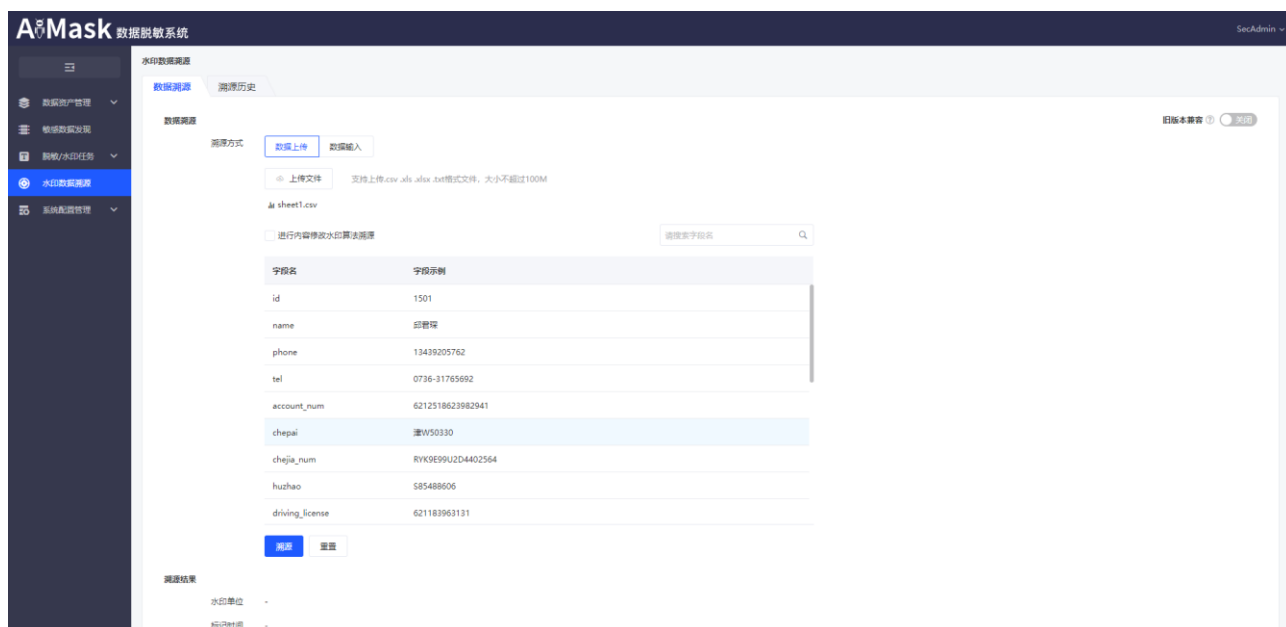
- ◆ 点击<数据输入>后展示的输入框可输入多个数据（长度没有限制），如有多条数据请用英文分号“;”分隔。
- ◆ 点击<数据上传>后可以上传 csv、txt、xls、xlsx、del、dbf 六种文件格式的数据进行溯源。



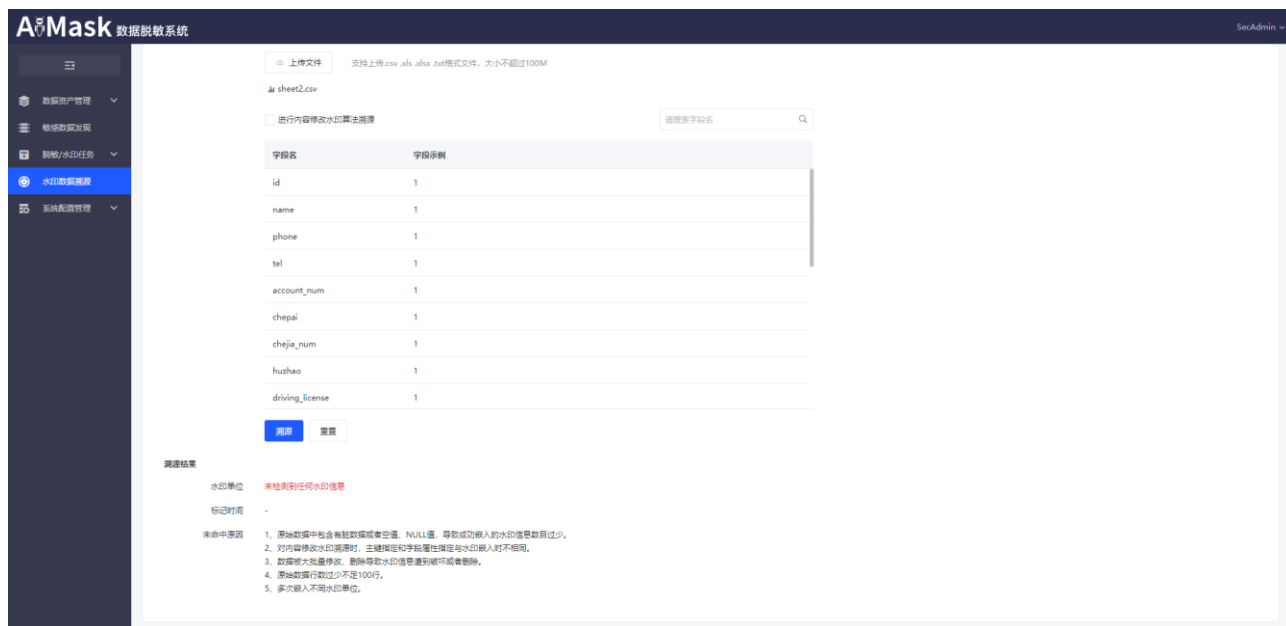


- ◆ 内容修改水印必须指定主键和字符类型，才能溯源到相应结水印单位。如溯源数据为 2.1.1 版本以下打上的水印，需要打开右上角的旧版本兼容开关。
- ◆ 溯源包含水印的数据请尽量避免重复。

点击<溯源>按钮，系统将会对输入数据进行溯源计算，将结果展示在下方溯源结果当中，如图所示：



若未溯源到任何结果，则会系统提示未命中信息。如图



4.1.7 系统统计报表

“数据资产分析”页面的功能是让用户知道自己有哪些资产，哪些资产更加敏感，哪些资产经常做数据脱敏等。刚进入本页面时，页面会展示系统自动更新过的数据。AiMask 会在每天凌晨自动更新数据。但是用户也可以选择点击手动更新按钮来获得当前时间最新的数据。左边展示了最近更新的时间。左侧可以切换数据源或者查看所有数据源粒度的数据。数据资产分析针对的统计对象仅包含非文件的数据源，不包括数据目标。本页信息一共展示了任务运行概述、数据源类型分布、数据源和 schema 粒度的敏感程度排名 top10、敏感标签词云以及资产统计清单这些数据分析内容。

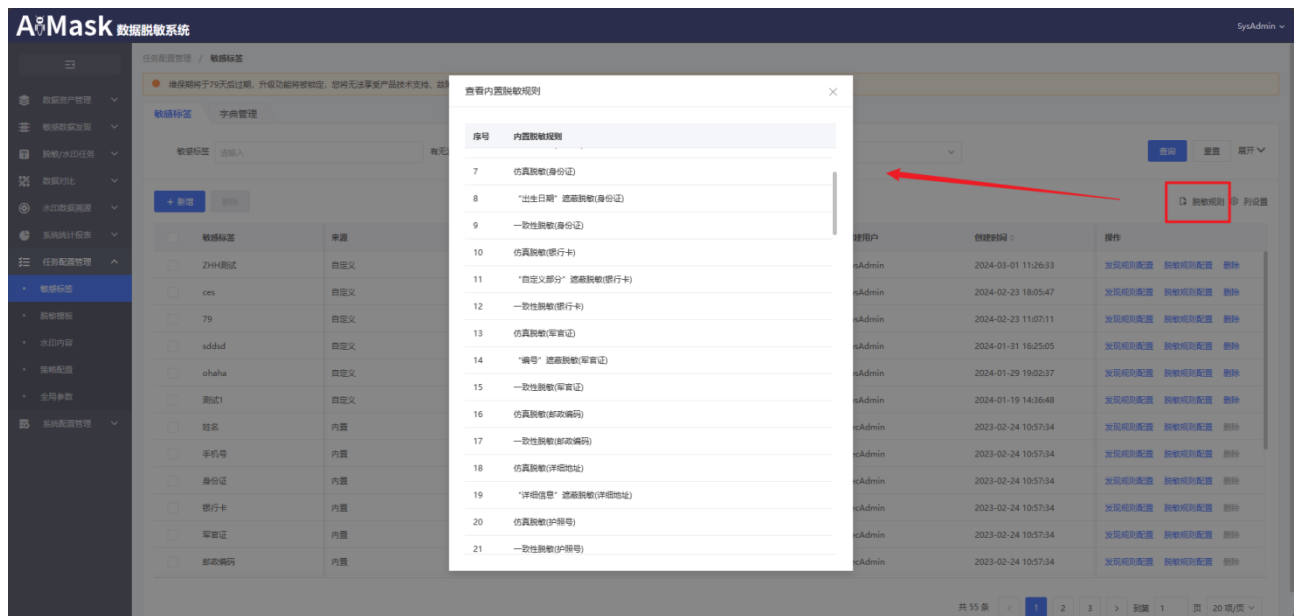
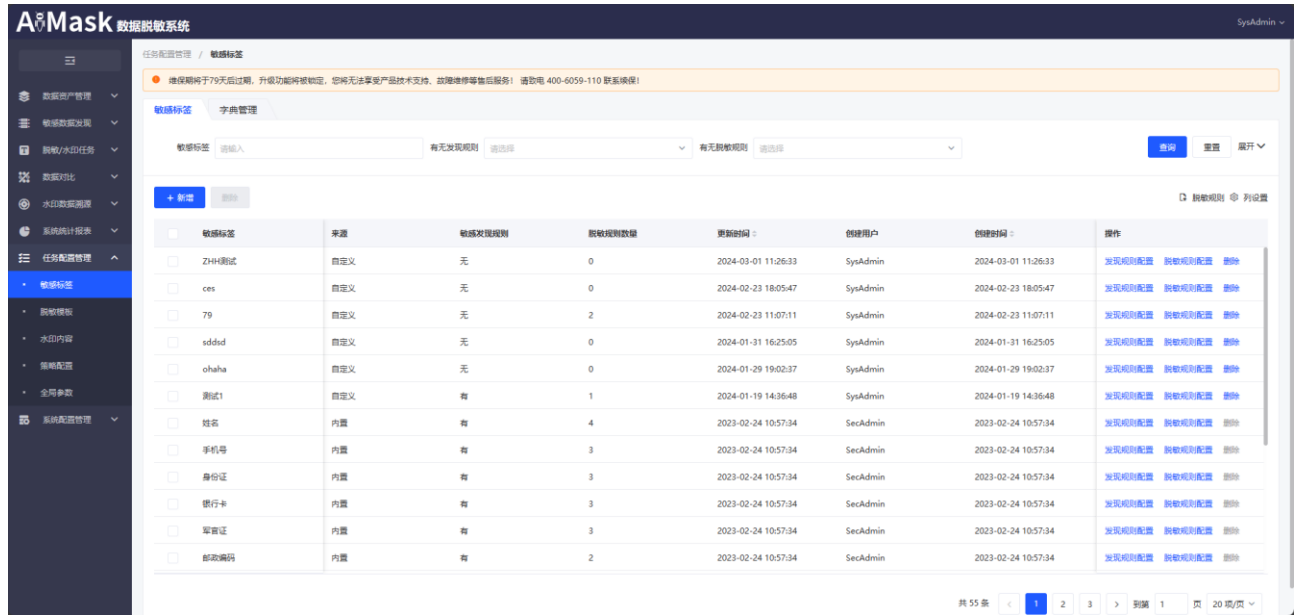
“脱敏任务分析”页面的功能是让用户知道脱敏任务运行是否正常，整体任务运行质量如何（比如错误数的变化），任务调度分配是否合理等。AiMask 会实时更新本页面的大部分数据，其中任务属性是针对任务进行统计，任务运行是针对实例进行统计，而且我们只对结构化任务的数据进行分析。左侧可针对任务运行时间（即实例运行时间）进行搜索，包括最近 7 天（默认）、最近 15 天、最近 30 天和自定义。这个页面分析的维度非常多。

此外，2 个页面都带“导出”功能，点击后可导出页面内容，支持 PDF 和 HTML 两种文件格式。

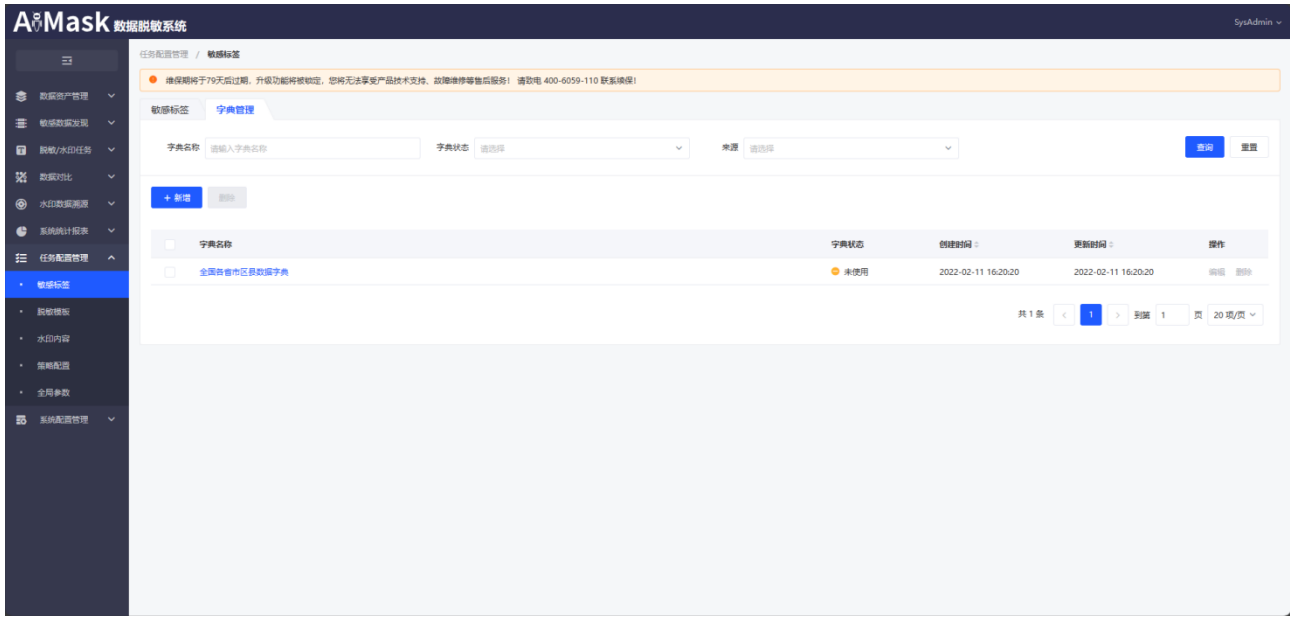
4.1.8 任务配置管理

敏感标签

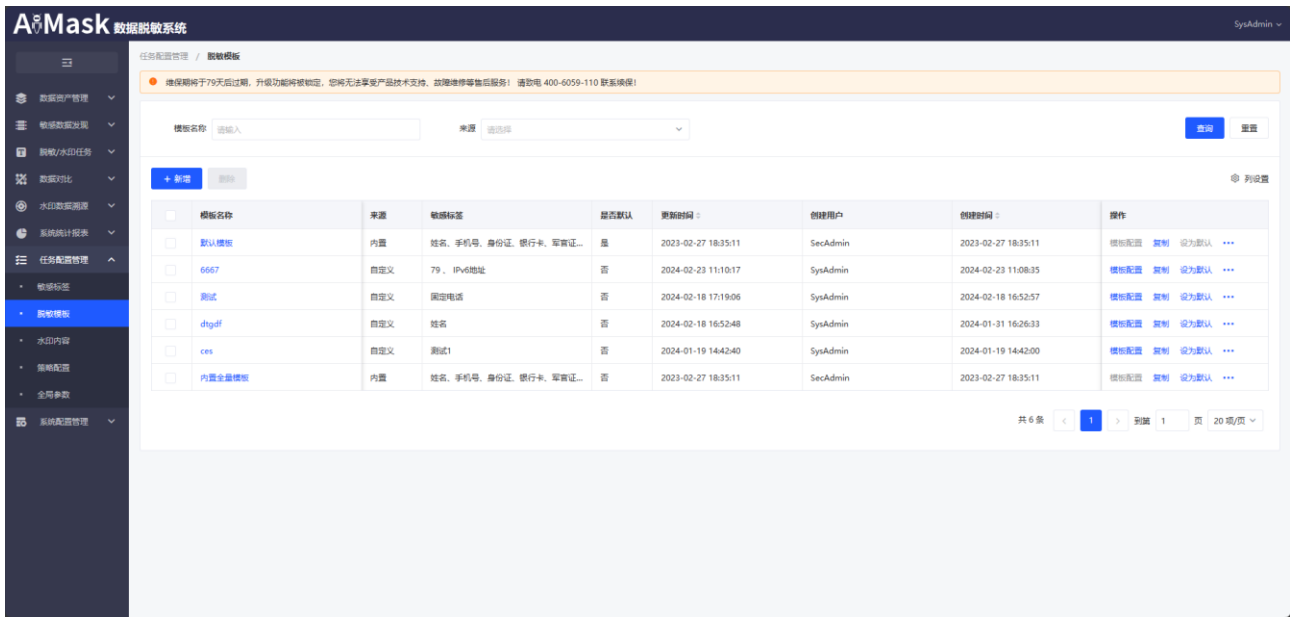
本页面用以配置敏感标签对应的敏感发现规则及脱敏规则。点击<脱敏规则>，一键速览内置脱敏规则。



字典管理



脱敏模板



部分算法及说明如下：

算法名称	说明
仿真算法	根据所选的参数（若有）生成格式、语义等属性与原数据相同且符合校验规则（若有）的仿真数据。

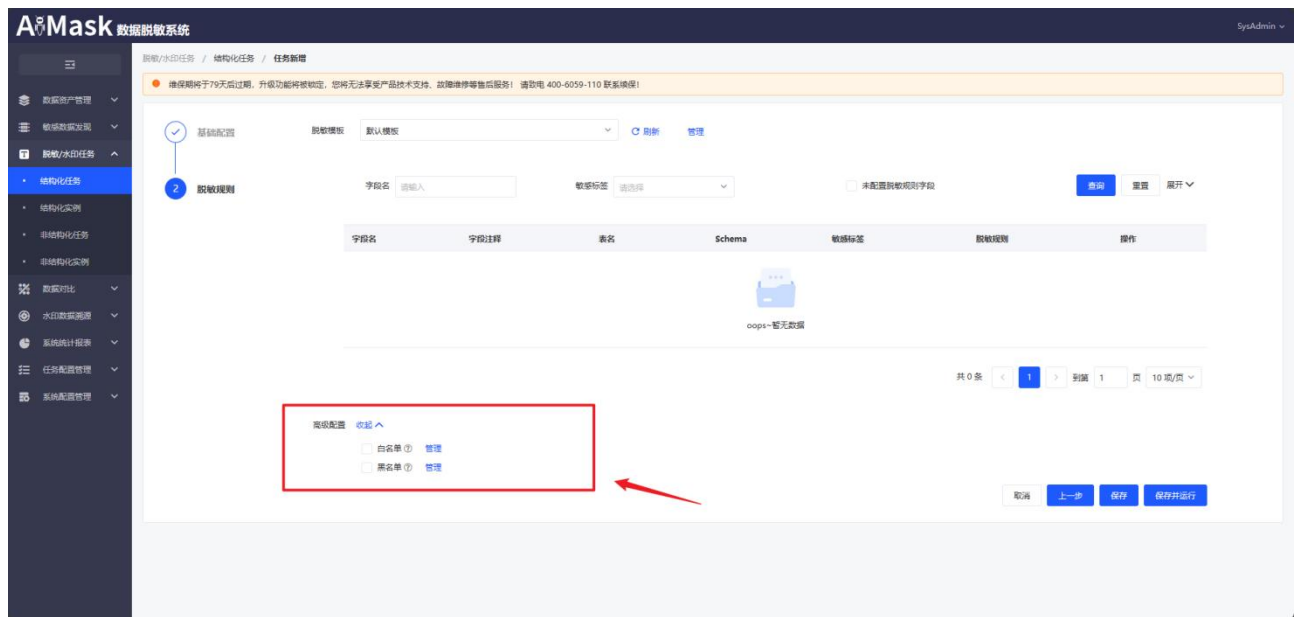
算法名称	说明
可逆算法	根据所选的参数（若有）生成格式类似、符合校验规则（若有）且支持可逆计算的数据。
遮蔽算法	根据所选的参数（若有）对指定区域的数据内容进行替换。
水印算法	根据所选的水印信息，对数据添加水印标记，可实现数据所有权声明和数据泄漏追溯。
取整算法	根据所选的参数（若有）对数据进行取整计算。
分割算法	根据所选的参数（若有）对数据进行分割，保留指定区域的数据。
保留算法	保留原数据不做脱敏。
替换算法	使用固定值替换，注意替换的值应与原数据所在列声明的数据结构保持一致，否则会出现因数据结构不符而造成脱敏失败的情况。
置空算法	使用空值替换，注意需确认若原数据所在列要求为非空，则使用本算法会出现失败的情况。
乱序算法	打乱数据元组的顺序。
哈希算法	对数据进行哈希编码，支持输入盐值。
密码学算法	根据所选参数指定的密码学加密算法对数据进行加密。支持的加密算法有：RSA/AES/SM2/SM4。
编码算法	根据所选参数指定的编码方式对数据进行编码。支持的编码方式有：UTF-8/GBK。
平均算法	对数据进行均值填充。

算法名称	说明
区间算法	根据步长对数据进行规整。
保留频次算法	对类别属性的数据（取值数目有限）进行编码，去除字段内容含义，仅保留类别区分性。适用于机器学习等数据分析中需保留标签区分性的场景。
标准化算法	对数值类型的数据进行标准化缩放，使得数据均值归为 0，方差归为 1。本算法脱敏后的数据基本保留数据分布类型，可后续用于常见的分类、聚类等数据分析任务。注意本算法不适合用于大量数据（建议行数为 $2w$ 以下），因平台在该种情况下会批量计算，由此造成结果不准确。
归一化算法	对数值类型的数据进行归一化缩放，将数据线性缩放至 $[0,1]$ 区间。本算法脱敏后的数据可限定数据范围，保留数据相对大小，剔除量纲影响，可根据分析模型和分析需求选用本算法。注意本算法不适合用于大量数据（建议行数为 $2w$ 以下），因平台在该种情况下会批量计算，由此造成结果不准确。
分布重建算法	根据指定的直方图的数量对原数据分布进行估计和采样重建。本算法可以使得脱敏后的数据保留原数据的高阶统计特征，适用于对数据质量有较高要求的分析场景。注意本算法不适合用于过少量数据（建议行数为 100 以上），因在此种情况下会造成分布估计不准确。
添加噪声算法	根据指定的均值和标准差，对原数据添加加性高斯噪声。本算法可以使得（在参数合理配置情况下）脱敏后的数据仍满足常见信号估计和趋势分析的噪声假设，适用于序列数据，可后续用于回归拟合和预测任务。
一致性算	根据所选的参数（若有）生成格式类似、符合校验规则（若有）且支持相同的数据在指定

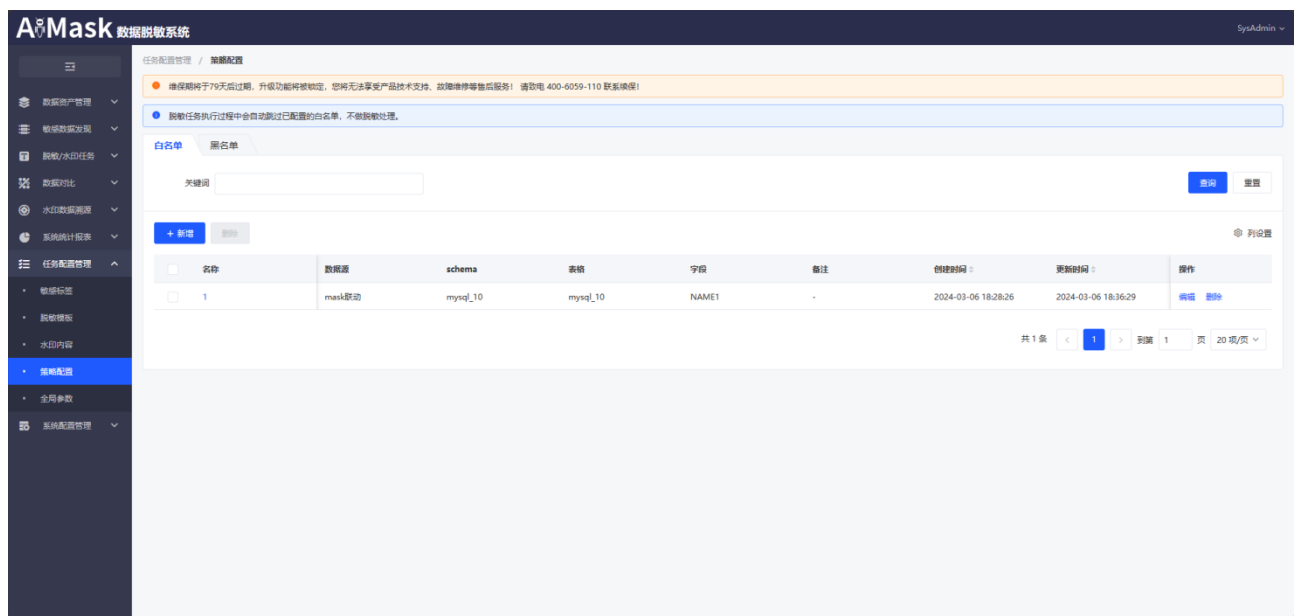
算法名称	说明
法	的表、列中在脱敏后仍保持一致。
计 算 关 系 保留算法	指定的列在脱敏后保留减法运算关系。可指定减数或被减数是否保留原值，否则会按照仿真算法重新生成。
乱 序 关 联 保留算法	指定的 2 个或多个列（支持至多 5 列）在乱序前后保留对应关系。
身 份 信 息 关 联 保 留 算法	出生日期/性别/身份证互相之间关系（根据需要）保持一致。

策略配置

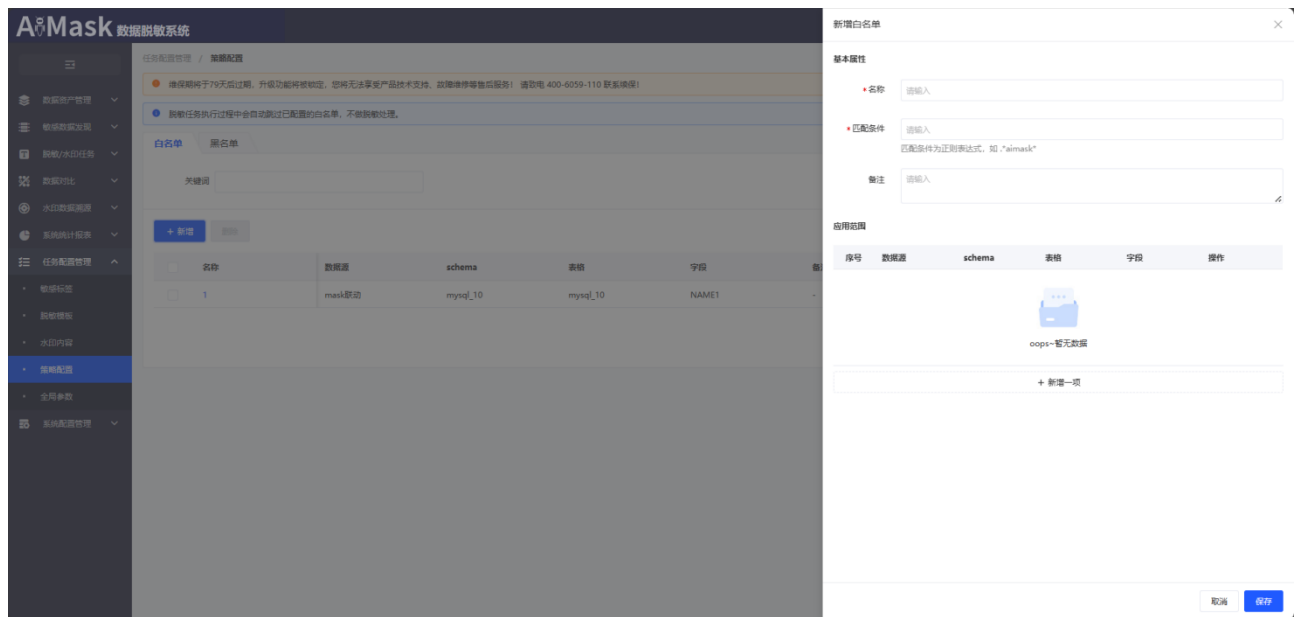
本页面针对结构化脱敏任务提供黑、白名单功能。用户若需使用该功能，可以在配置结构化任务时展开高级配置，勾选对应选项即可。若用户第一次配置结构化任务时未提前新增所需黑白名单，可以点击对应选项右侧的<管理>按钮打开新的策略配置页面。



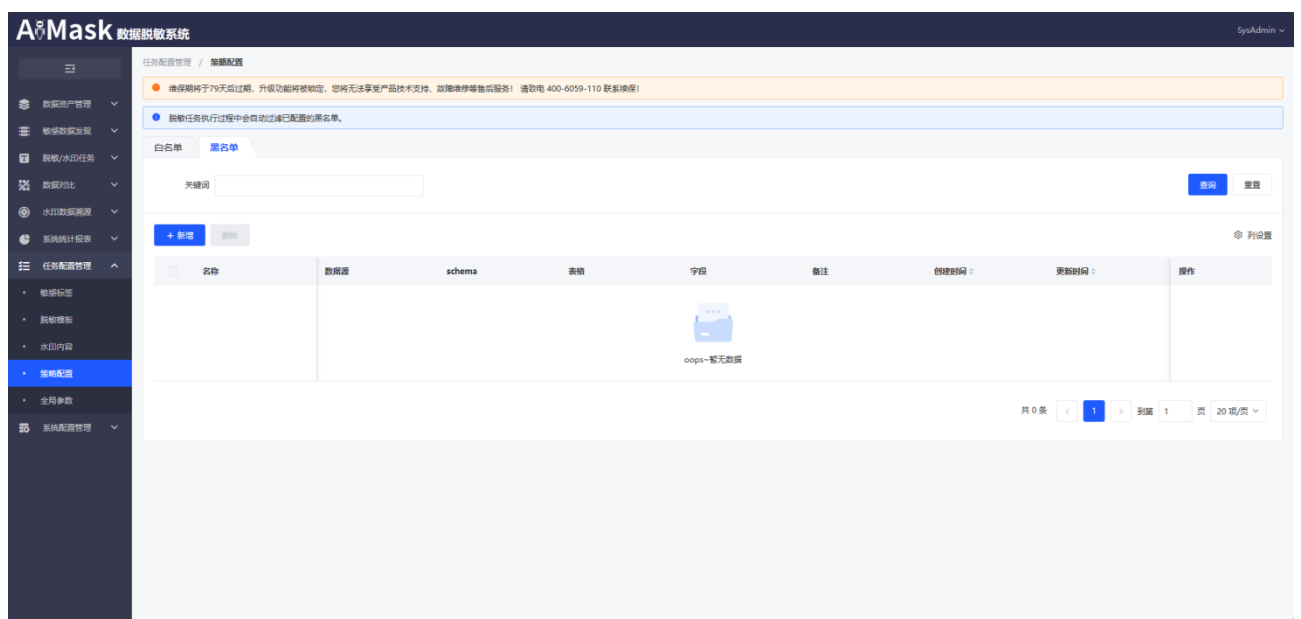
其中，白名单机制的主要目的是为了与未经过脱敏处理的业务系统联调或第三方进行数据对接。所有“白名单”上的数据在脱敏前自动过滤掉。管理页面如下图所示。



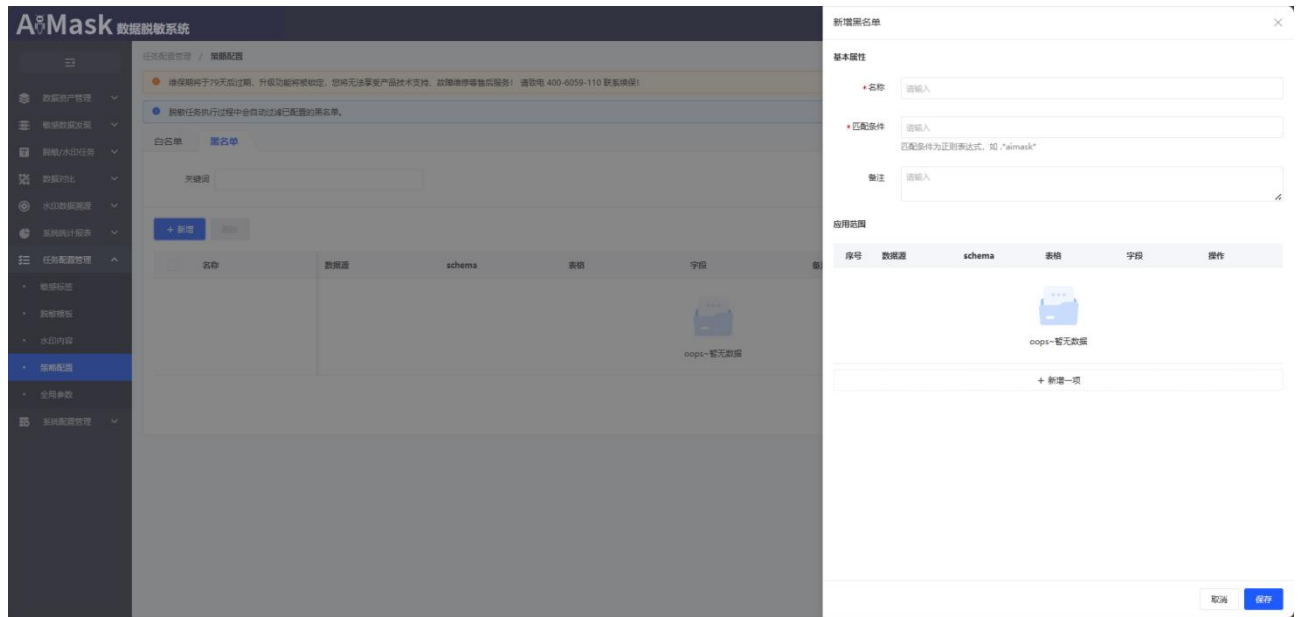
点击<新增>后，用户可以在下页面配置白名单策略的具体信息。



黑名单模块提供为企业“核心数据”设置黑名单功能。黑名单机制的主要目的是为了保护环境机密数据不被泄漏。所有“黑名单”上的数据在脱敏前自动过滤掉。管理页面如下图所示。

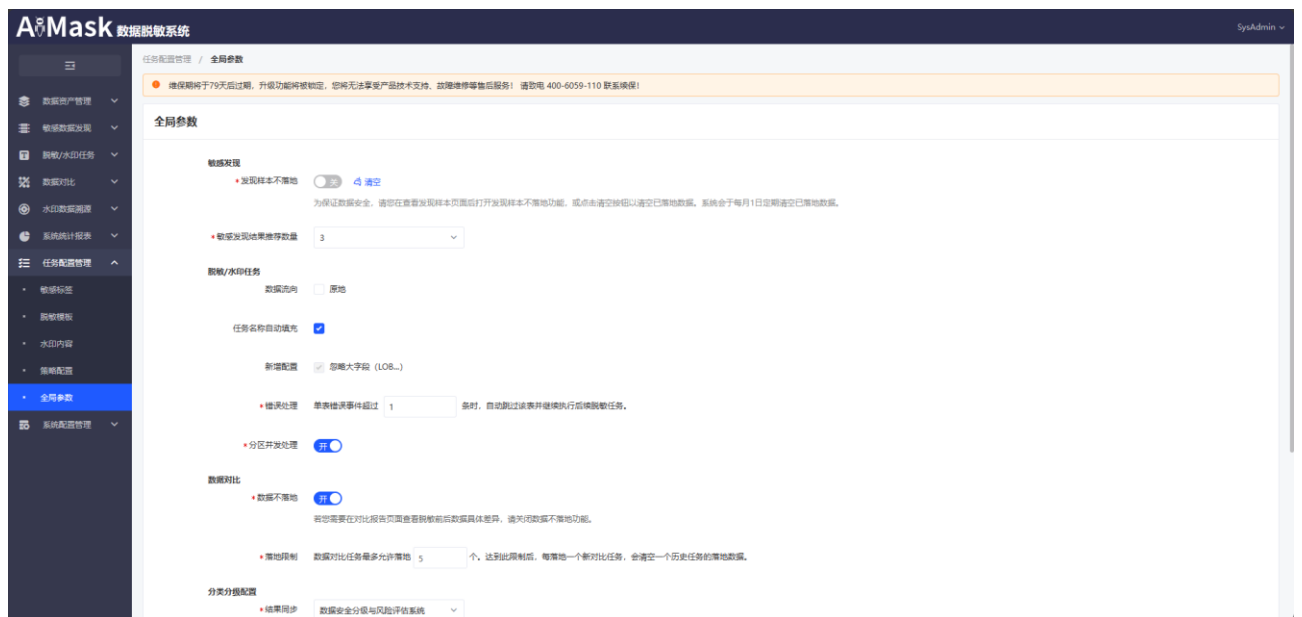


点击<新增>后，用户可以在下页面配置黑名单策略的具体信息。



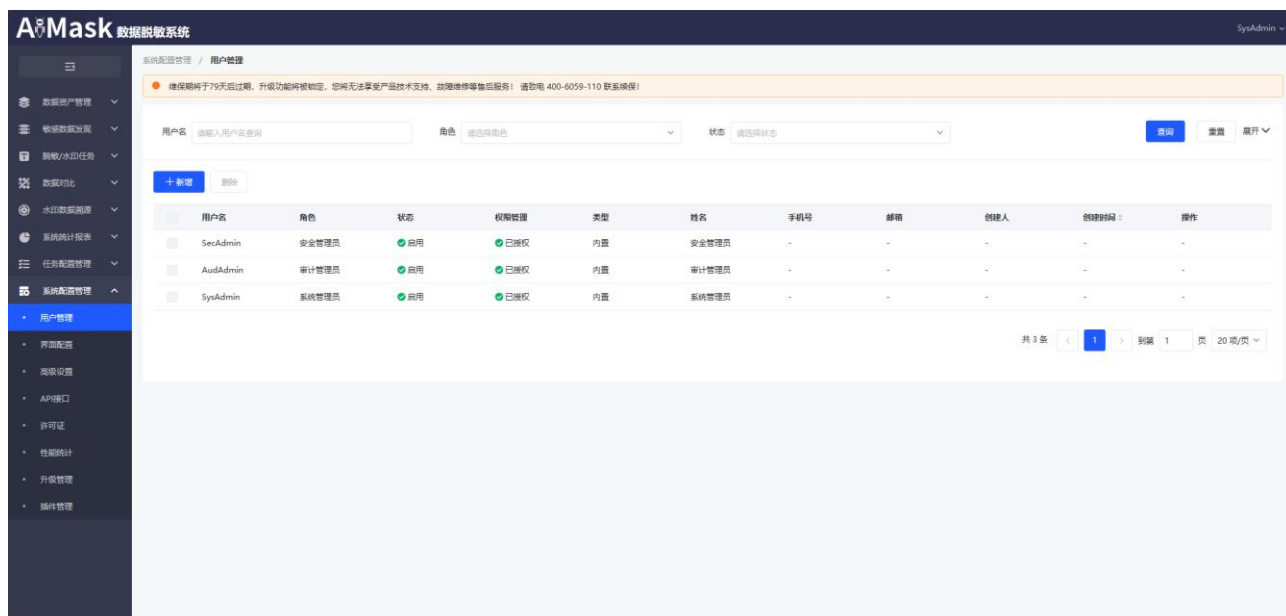
全局参数

本页面可配置部分作用于系统全局的全局参数。



4.1.9 系统配置管理

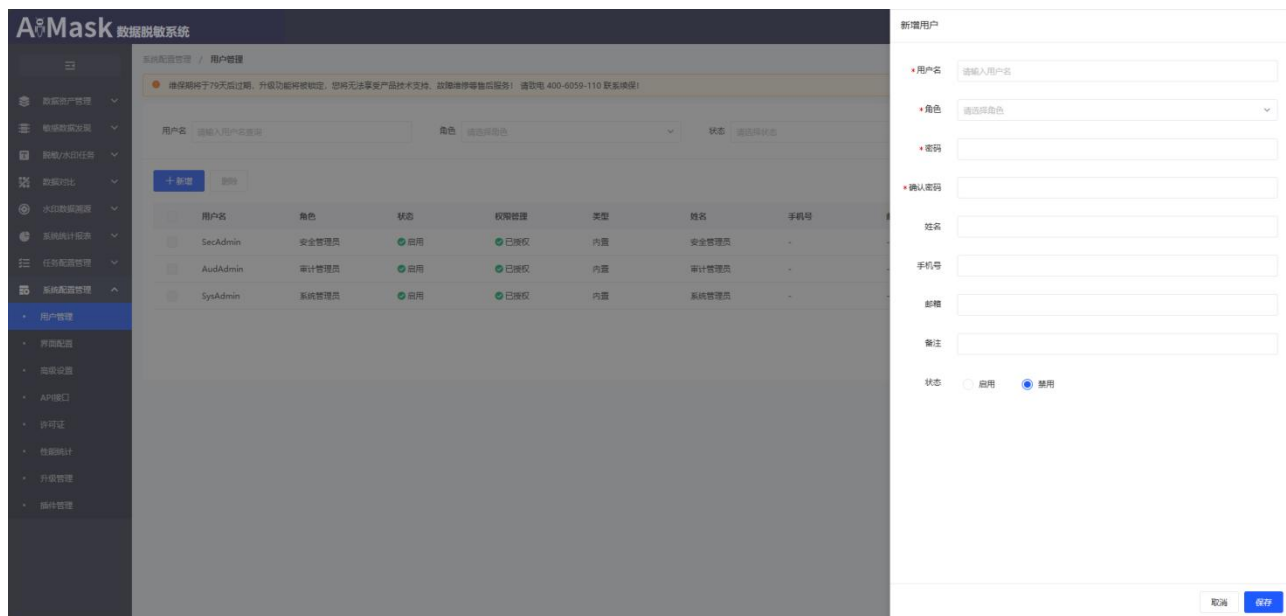
‘系统配置管理>用户管理’页面主要用于用户增改的配置。列表内置 3 个管理员用户。



新增用户操作步骤如下：

步骤1. 在 ‘系统配置管理>用户管理’ 标签页点击<新增>。

步骤2. 在右侧抽屉界面中填写用户信息，带*的字段为必填字段。完成填写后点击<保存>。



4.2 数据分类分级操作指导

4.2.1 概述

数据安全分级是一款智能化、流程化、开放化的数据分类分级工具。

产品内置各行业分类分级标准及框架，并以此为依据，结合已有的上万条数据分类分级规则与行业专有数据分类分级 Ai 模型，帮助您快速有效完成数据分类工作，以支持后续精细化的数据安全风险分析及安全管控。

主要功能

数据资产梳理：产品可通过网络嗅探技术，自动化发现网络环境中存在的数据库系统，同时也可通过对接安恒数据库审计系统获取数据库清单，以对于数据库资产台账进行校对及补充。

数据资产目录：产品可通过数据库扫描技术，自动化采集数据库基础元数据信息，同时结合后续数据分类分级结果，以库表列的形式，展现完整地数据资产目录，以供用户进行数据资产的查询。

结构化数据分类分级：结构化数据的分类分级是产品核心能力。产品通过流程与算法相结合的方式，以达到做最少的人工操作，实现最好的分类分级结果的目标。用户无须编写复杂的分类分级规则，仅需完成少量数据的分类分级打标工作。

非结构化数据分类分级：产品支持对于非结构化文件进行自动化分类分级扫描，基于关键词的 NER 识别，准确确定文件所对应分类和级别，同时可查看文件中所包含的常见敏感信息。

分类分级模板管理：产品支持通过页面方式对于分类分级框架及对应规则进行管理，用户可自定义上传分类分级模版，并支持以系统名称、库名、表名、表注释、字段名、字段注释、字段内容等各种维度进行条件组合的方式维护分类分级规则。

分类分级模型训练：产品基于有监督学习模型训练的能力，用户在一定量手动打标的基础上，即可通过模型管理模块，对于该部分数据进行自定义模型训练。通过训练得到的模型，可有效对于剩余未分类分级数据进行预测，从而大幅降低手工打标的工作量。

资产分类分级报告：产品以可视化图表的方式，将数据资产分类分级的各项统计指标进行展示，帮助用户清晰直观地了解数据资产整体分布、分类分级梳理、敏感数据占比、数据分类分布等情况。

数据库账户梳理：产品支持对于数据库账号状态及权限进行梳理评估，并对于账号新增、权限变更删除等情况进行分类展示，有效检测账号的违规授权和恶意提权。

数据库风险检测：产品提供数据库漏洞检查、配置基线检查、弱口令检查等手段进行数据资产安全评估，通过安全现状评估，能有效发现当前数据库系统的安全问题，对数据库的安全状况进行持续化监控，保持数据库的安全健康状态。

资产分类分级大屏：产品提供数据安全分类分级监管大屏展示，内容包含数据源的分类分级情况、漏洞检查、配置基线检查、弱口令等风险评估情况，以统一展示相关工作成果。

操作流程

系统主要业务流程如下：

系统初始化：首先，登录系统管理员，修改密码，完成运维和安全性配置；其次，登陆安全管理员，执行数据资产管理等操作。

添加数据源：新增需要分类分级的数据源，并完成数据源同步；

执行分类分级任务：为数据源新增分类分级任务，跟随流程向导完成分类分级任务；

获取分类分级结果：查看和导出分类分级结果，提供分类分级报告、报表、可视化大屏功能；

优化分类分级配置：优化分类分级的规则、模型、策略等配置；首次使用可以直接使用内置模板、内置模型和默认参数。

4.2.2 数据资产管理

本模块主要用于数据源管理。用户可以在本模块手动添加或导入数据源，或者利用**主动嗅探资产**和**被动发现资产**功能，待系统自动发现数据源后，再对数据源执行添加操作。

A8Sort+ 数据安全分级与风险评估系统										
维护期将于06天后过期，升级功能将被锁定，您将无法享受产品技术支持、故障维修等售后服务！请致电 400-6059-110 联系续保！										
数据源管理 / 数据源管理										
数据源管理 主动嗅探资产 被动发现资产										
数据源名称 请输入数据源名称 主机 请输入主机 分类分级状态 请选择分类分级状态 查询 重置 展开										
+ 新增 删除 同步 终止 列设置 导入 导出										
☐	数据源名称	数据源类型	主机:端口	业务系统	部门	责任人	创建用户	创建时间	最近同步时间	操作
☐	test4	DB MySQL	10.50.3.126:13306				SecAdmin	2024-10-30 19:29:58	2024-10-30 19:29:58	同步 编辑 终止 ...
☐	test3	DB MySQL	10.50.3.126:13306				SecAdmin	2024-10-30 19:13:02	2024-10-30 19:13:02	同步 编辑 终止 ...
☐	test2	DB MySQL	10.50.3.126:13306				SecAdmin	2024-10-30 19:09:27	2024-10-30 19:09:27	同步 编辑 终止 ...
☐	冷舍	StructFile METADATA					SecAdmin	2024-10-30 16:35:28	2024-10-30 16:35:28	同步 编辑 终止 ...
☐	111111	StructFile LOCAL					SecAdmin	2024-10-29 19:33:50	2024-10-29 19:33:50	同步 编辑 终止 ...
☐	test	StructFile METADATA					SecAdmin	2024-10-29 16:32:51	2024-10-29 16:32:51	同步 编辑 终止 ...
☐	SZH	StructFile METADATA					SecAdmin	2024-10-29 09:23:21	2024-10-29 09:23:21	同步 编辑 终止 ...
☐	政府演示数据源	DB MySQL	10.50.3.126:13306				SecAdmin	2024-10-28 09:32:09	2024-10-28 09:32:09	同步 编辑 终止 ...
☐	医疗演示数据源	DB MySQL	10.50.3.126:13306				SecAdmin	2024-10-28 09:30:53	2024-10-28 09:30:53	同步 编辑 终止 ...
☐	教育演示数据源	DB MySQL	10.50.3.126:13306				SecAdmin	2024-10-28 09:28:48	2024-10-28 09:28:48	同步 编辑 终止 ...
☐	金融演示数据源	DB MySQL	10.50.3.126:13306				SecAdmin	2024-10-25 23:06:54	2024-10-25 23:06:54	同步 编辑 终止 ...

4.2.2.1 数据源管理

数据源列表的展示字段包括：数据源名称、数据源类型、主机:端口、业务系统、部门、责任人、创建用户、创建时间、最近同步时间、最近分类分级时间、发现结果、同步状态、同步标识、分类分级状态、同步进度、操作。由于页面宽度的限制，部分字段需要拉动页面下方水平滚动条进行查看。其中：

- 同步状态：分为待同步、已同步、同步中、失败四种；
- 同步标识：分为已更新、建议重配两种；同步状态为待同步或失败的数据源，其同步标识为空；
- 分类分级状态：分为待执行、执行中、暂停、完成、失败五种，执行分类分级任务见“分类分级管理>结构化数据”页面；
- 发现结果：显示最近一次分类分级中发现的新增字段数和减少字段数。

用户可以点击列表右上方<列设置>，来控制每列的显示/隐藏状态。

列表默认按照创建时间降序排序。用户可以点击表头的<创建时间>和<最近同步时间>，使表格切换为按该列降序或升序排序。

用户可以通过输入数据源名称、主机、分类分级状态、数据源类型、同步标识来查询数据源。

☐	数据源名称	数据源类型	主机/端口	业务系统	部门	责任人	创建用户	创建时间	最近同步时间	同步状态	分类分级状态	操作
☐	greenplum-10.50.3.113	DB GREENPLUM	10.50.3.113:5432	系统测试a*4	部门测试a*4	责任人a*4	test15aA1	2021-08-25 19:07:28	2021-08-25 19:08:55	已同步	完成	...
☐	oracle-10.50.3.110	DB ORACLE	10.50.3.110:1521	系统测试a*3	部门测试a*3	责任人a*3	test15aA1	2021-08-25 19:07:28	2021-08-25 19:08:55	已同步	完成	...
☐	clickhouse-10.50.3.113	DB CLICKHOUSE	10.50.3.113:8123	系统测试a*14	部门测试a*14	责任人a*14	test15aA1	2021-08-25 19:07:28	2021-08-25 19:08:55	已同步	完成	...
☐	db2-10.50.3.120	DB DB2	10.50.3.120:50010	系统测试a*6	部门测试a*6	责任人a*6	test15aA1	2021-08-25 19:07:28	2021-08-25 19:08:55	已同步	完成	...
☐	elasticsearch-10.50.3.136	DB ELASTICSEARCH	10.50.3.136:9200	系统测试a*9	部门测试a*9	责任人a*9	test15aA1	2021-08-25 19:07:28	2021-08-25 19:08:55	已同步	完成	...
☐	postgresql-10.50.3.135	DB POSTGRESQL	10.50.3.135:1921	系统测试a*8	部门测试a*8	责任人a*8	test15aA1	2021-08-25 19:07:28	2021-08-25 19:08:55	已同步	完成	...
☐	sqlserver-10.50.3.115	DB MSSQL	10.50.3.115:1433	系统测试a*5	部门测试a*5	责任人a*5	test15aA1	2021-08-25 19:07:28	2021-08-25 19:08:55	已同步	完成	...
☐	sybase-10.50.3.145	DB SYBASE	10.50.3.145:5003	系统测试a*12	部门测试a*12	责任人a*12	test15aA1	2021-08-25 19:07:28	2021-08-25 19:08:55	已同步	完成	...
☐	mongodb-10.50.3.138	DB MONGODB	10.50.3.138:27017	系统测试a*10	部门测试a*10	责任人a*10	test15aA1	2021-08-25 19:07:28	2021-08-25 19:08:55	已同步	完成	...
☐	dm7-10.50.3.136	DB DM7	10.50.3.136:5236	系统测试a*16	部门测试a*16	责任人a*16	test15aA1	2021-08-25 19:07:28	2021-08-25 19:08:55	已同步	完成	...
☐	mysql-10.50.3.110	DB MYSQL	10.50.3.110:3377	系统测试a*1	部门测试a*1	责任人a*1	test15aA1	2021-08-25 19:07:28	2021-08-25 19:08:55	已同步	完成	...
☐	KINGBASE-10.50.3.135	DB KINGBASE8	10.50.3.135:54321	系统测试a*13	部门测试a*13	责任人a*13	test15aA1	2021-08-25 19:07:28	2021-08-25 19:08:55	已同步	完成	...
☐	odps-service-10.50.3.135	DB ODPS	http://service.odps.aliyun.com/api	系统测试a*7	部门测试a*7	责任人a*7	test15aA1	2021-08-25 19:07:28	2021-08-25 19:08:55	已同步	完成	...
☐	gbase8a-10.50.3.135	DB GBASE8A	10.50.3.135:5258	系统测试a*15	部门测试a*15	责任人a*15	test15aA1	2021-08-25 19:07:28	2021-08-25 19:08:55	已同步	完成	...

4.2.2.2 主动嗅探资产

主动嗅探资产功能可以针对指定主机进行扫描，从而发现数据库资产。扫描任务列表的展示字段包括：任务名称、IP 段、周期、服务数、状态、创建时间、最新发现时间。其中：

- 服务数表示任务扫描出的端口数量；
- 状态分为待扫描、扫描中、失败、已完成四种；
- 任务列表默认按照更新时间降序排列。

用户可以通过输入任务名称、状态来查询现有任务。

数据源管理

主动嗅探资产

被动发现资产

任务名称

请输入任务名称

状态

请选择

查询

重置

+ 新增

立即扫描

任务报表

导出

☐	任务名称	IP段	周期	服务数	状态	创建时间	更新时间	操作
☐	> 2	1	手动执行		待执行	2023-12-13 18:57:39	2023-12-13 18:57:39	编辑 删除
☐	> 1	10.50.3.110,10.50.3.113	手动执行	14	完成	2023-12-12 19:08:32	2023-12-12 19:28:56	编辑 删除
☐	> 测试	10.50.3.1	手动执行	1	完成	2023-12-12 19:05:51	2023-12-12 19:43:16	编辑 删除
☐	> test_ip2	10.50.3.110-130	手动执行	64	完成	2023-10-17 18:36:25	2023-12-12 19:19:32	编辑 删除
☐	> test_ip	10.50.3.110	手动执行	1	完成	2023-10-17 18:35:13	2023-10-26 15:16:36	编辑 删除

共 5 条

< 1 >

到第 1 页

20 项/页

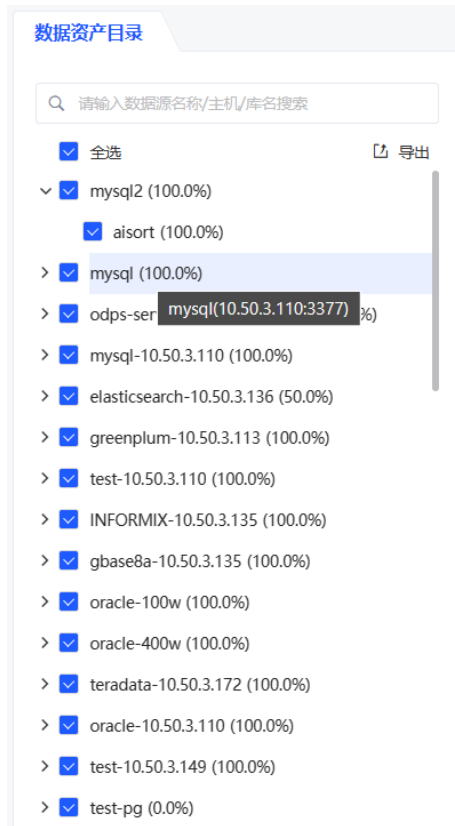
4.2.2.3 数据资产目录

数据源列表默认显示全部数据源及其分类分级梳理率。点击数据源名称，可以展开下属数据库。用户可以在顶部搜索框输入数据源名称、主机或库名，搜索对应的数据源或数据库。

点击本区域右上角的<导出>，可以导出选中数据源或数据库的全部表格和字段资产信

息。

勾选特定数据源或数据库，用户可以在右侧数据表列表区域，实时预览该数据源或数据库的资产信息。



4.2.3 分类分级任务

4.2.3.4 结构化数据

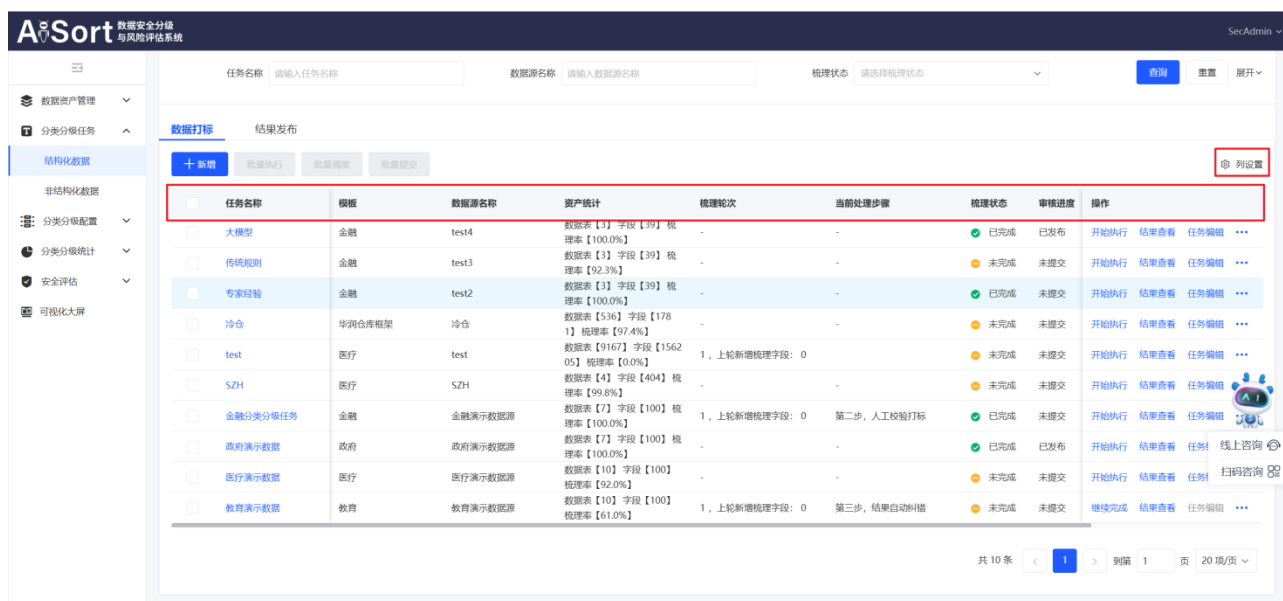
结构化数据的分类分级是产品核心能力。在本模块，用户可以为数据源管理模块中汇总的数据源绑定分类分级任务，批量调度任务的执行周期和执行次数，查看、确认和导出分类分级结果。

该模块包括数据打标以及结果发布两个子模块，其中共性功能包括：

◆ 每个模块的列表展示了分类分级任务的各种状态和配置信息，展示字段包括：任务名称、模版、数据源名称、梳理状态、审核进度、审核员、最近审核时间等。用户可以点击列表右上角<列设置>，来控制每列的显示/隐藏状态。其中：

- 资产统计：该分类分级任务涉及的表数、字段数和梳理率；
- 梳理轮次：该分类分级任务的已梳理轮次（总轮数取决于数据量，以及任务打标方式中设置的每轮打标数量）、上轮新增梳理字段（“梳理字段”指字段的分类分级结果经过了人工确认）；

- 执行周期：分类分级任务支持按照特定周期自动执行，默认为手动执行、不限制执行次数，用户可以通过点击操作列<任务编辑>或任务列表上方的<批量调度>按钮进行设置；
- 执行状态：分为待执行、执行中、完成、失败；
- 审核模式：
 - 1) 审核验收模式开启（仅审核员审核）：未提交、审核中、审核打回、待发布、发布中、已发布；
 - 2) 审核验收模式开启（审核员审核+验收员审核）：未提交、审核中、审核打回、验收中、验收打回、待发布、发布中、已发布；
 - 3) 审核验收模式关闭：未提交、待发布、发布中、已发布。
- 流程优化：在 V2.1R24C30 中对当前任务流程进行了优化，仅依靠<数据达标>及<结果发布>两个主要功能即可完成分类分级任务的主线流程，该部分的操作详情可见 3.1.2 数据打标及 3.1.5 结果发布两个部分的详细说明；其余部分的功能入口可通过登录系统管理员账号进行选择配置。



任务名称	模板	数据源名称	资产统计	梳理次数	当前处理步骤	梳理状态	审核进度	操作
大模型	金融	tes4	数据表【3】字段【39】梳理率【100.0%】	-	-	已完成	已发布	开始执行 结束查看 任务编辑 ...
传统规则	金融	tes3	数据表【3】字段【39】梳理率【92.3%】	-	-	未完成	未提交	开始执行 结束查看 任务编辑 ...
专家经验	金融	tes2	数据表【3】字段【39】梳理率【100.0%】	-	-	已完成	未提交	开始执行 结束查看 任务编辑 ...
冷仓	华润仓库框架	冷仓	数据表【536】字段【1781】梳理率【97.4%】	-	-	未完成	未提交	开始执行 结束查看 任务编辑 ...
test	医疗	test	数据表【9167】字段【156205】梳理率【0.0%】	1, 上轮新增梳理字段: 0	-	未完成	未提交	开始执行 结束查看 任务编辑 ...
SZH	医疗	SZH	数据表【4】字段【404】梳理率【99.8%】	-	-	未完成	未提交	开始执行 结束查看 任务编辑 ...
金融分类分级任务	金融	金融演示数据源	数据表【7】字段【100】梳理率【100.0%】	1, 上轮新增梳理字段: 0	第二步, 人工校验打标	已完成	未提交	开始执行 结束查看 任务编辑 ...
政府演示数据	政府	政府演示数据源	数据表【7】字段【100】梳理率【100.0%】	-	-	已完成	已发布	开始执行 结束查看 任务编辑 ...
医疗演示数据	医疗	医疗演示数据源	数据表【101】字段【100】梳理率【92.0%】	-	-	未完成	未提交	开始执行 结束查看 任务编辑 ...
教育演示数据	教育	教育演示数据源	数据表【101】字段【100】梳理率【61.0%】	1, 上轮新增梳理字段: 0	第三步, 结果自动排错	未完成	未提交	继续完成 结束查看 任务编辑 ...

- ◆ 用户可以通过输入任务名称、数据源名称、梳理状态、模板、数据源类型、执行状态、敏感数据识别执行状态、创建人、创建时间、审核进度、打标员、审核员、验收员、发布员来查询现有任务。

AiSort 数据安全分级与风险评估系统

分类分级任务 / 结构化数据

● 维保期将于85天后过期，升级功能将被锁定，您将无法享受产品技术支持，故障维修等售后服务！请致电 400-6059-110 联系续保！

任务名称

请输入任务名称

数据源名称

请输入数据源名称

梳理状态

请选择梳理状态

白列

重置

收起

模板

请选择模板

数据类型

请选择

执行状态

请选择执行状态

敏感数据识别执行状态

请选择执行状态

创建人

请选择创建人

创建时间

审核进度

请选择审核进度

打标员

请选择打标员

审核员

请选择审核员

验收员

请选择验收员

发布员

请选择发布员

◆ 点击任务名称，可以快速查看任务配置。

AiSort 数据安全分级与风险评估系统

分类分级任务 / 结构化数据

● 维保期将于85天后过期，升级功能将被锁定，您将无法享受产品技术支持，故障维修等售后服务！请致电 400-6059-110 联系续保！

任务名称

请输入任务名称

数据源名称

请输入数据源名称

梳理状态

请选择梳理状态

数据打标 **结果发布**

+ 新增 **数据执行** **数据清理** **数据回收**

任务名称	模板	数据源名称	资产统计	梳理轮次	当前处理步骤	梳理状态	审核进度	操作
☐ 大模型	金融	test4	数据表【3】字段【39】梳理率【100.0%】	-	-	已完成	已发布	开始执行 结果查看 任务编辑 ...
☐ 传统规则	金融	test3	数据表【3】字段【39】梳理率【92.3%】	-	-	未完成	未提交	开始执行 结果查看 任务编辑 ...
☐ 专家经验	金融	test2	数据表【3】字段【39】梳理率【100.0%】	-	-	已完成	未提交	开始执行 结果查看 任务编辑 ...
☐ 冷合	华润仓库框架	冷合	数据表【536】字段【1781】梳理率【97.4%】	-	-	未完成	未提交	开始执行 结果查看 任务编辑 ...
☐ test	医疗	test	数据表【9167】字段【1562】梳理率【0.0%】	1, 上轮新增梳理字段: 0	-	未完成	未提交	开始执行 结果查看 任务编辑 ...
☐ SZH	医疗	SZH	数据表【7】字段【100】梳理率【99.0%】	-	-	未完成	未提交	开始执行 结果查看 任务编辑 ...
☐ 金融分类分级任务	金融	金融演示数据源	数据表【7】字段【100】梳理率【100.0%】	-	-	已完成	未提交	开始执行 结果查看 任务编辑 ...
☐ 政府演示数据	政府	政府演示数据源	数据表【10】字段【100】梳理率【100.0%】	-	-	已完成	未提交	开始执行 结果查看 任务编辑 ...
☐ 医疗演示数据	医疗	医疗演示数据源	数据表【10】字段【100】梳理率【92.0%】	-	-	未完成	未提交	开始执行 结果查看 任务编辑 ...
☐ 教育演示数据	教育	教育演示数据源	数据表【10】字段【100】梳理率【61.0%】	1, 上轮新增梳理字段: 0	-	未完成	未提交	开始执行 结果查看 任务编辑 ...

分类分级任务配置

基础信息

- *任务名称 大模型
- *数据源 test4
- *行业模板 金融
- *抽样策略 每张表抽取 100 条数据
- *执行策略 ☐ 分类分级匹配规则
- ☒ 分类分级模型分析
 - ☐ 内置小模型分析（置信度） 90 %
 - ☒ “假脑”大模型分析 ①
- ☐ 分类分级模型分析（置信度） 90 %
- 辅助识别 请选择
- *打标方式 ☒ 一次处理模式

关闭

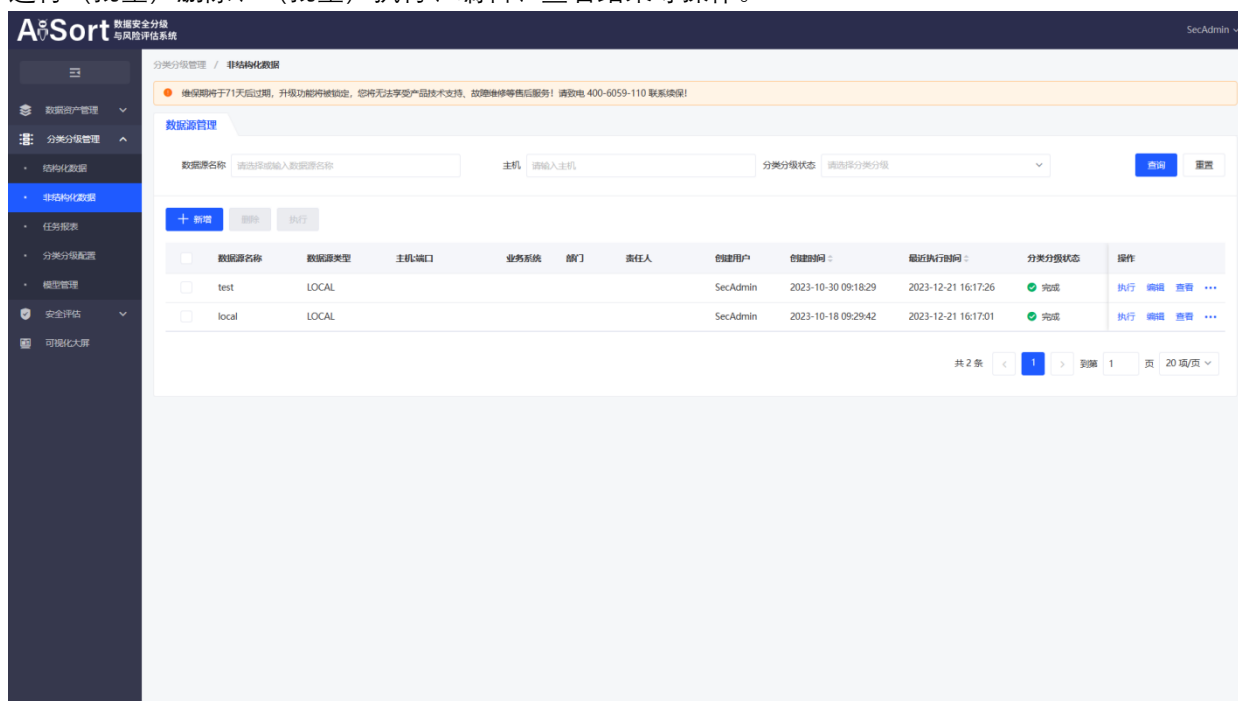
注：该模块默认仅启动数据打标、结果发布功能页面，若要启用任务分配、结果审核以及结果验收功能页面，需登录 SysAdmin，进入“系统管理-安全设置”页面，点击<审核验收模式>开关并在右侧下拉框选择“审核员审核+验收员审核”，启动任务分配、结果审核与结果验收功能；下拉框中选择“仅审核员审核”，仅启动任务分配、结果审核功能。



4.2.3.5 非结构化数据

针对非结构化数据，分类分级任务列表的展示字段包括：数据源名称、数据源类型、主机窗口、业务系统、部门、责任人、创建用户、创建时间、最近更新时间、分类分级状态。

用户可以通过输入数据源名称、主机、分类分级状态来查询现有任务，并且可以对任务进行（批量）删除、（批量）执行、编辑、查看结果等操作。



4.2.4 分类分级配置

4.2.4.6 分类分级模板

系统内置多个行业模板。点击当前模板名，展开模板列表。模板列表中，置顶模板为默认模板，其余模板的排列顺序基于原内置顺序和自定义导入顺序。

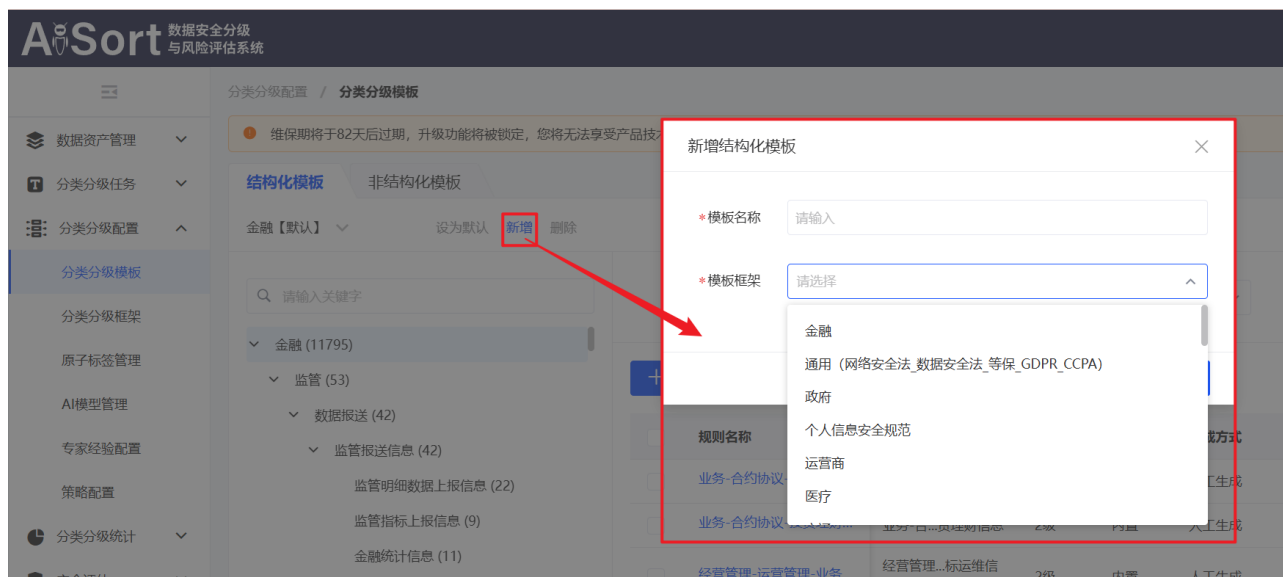
◆ 设置默认模板

选定目标模板后，点击<设为默认>，可以将当前模板设为默认模板。



◆ 新增模板

点击<新增>，填写模板名称（要求不能与已有模板名重复）并选择模板框架。可通过分类分级框架页面新增分类分级框架（详见 4.2 分类分级框架）。



◆ 删除模板

点击<…>图标后选择<删除>，可以删除指定模板。无法删除的情况有：（1）该模板为默认模板；（2）该模板为系统内置模板；（3）该模板已经被分类分级任务所引用。



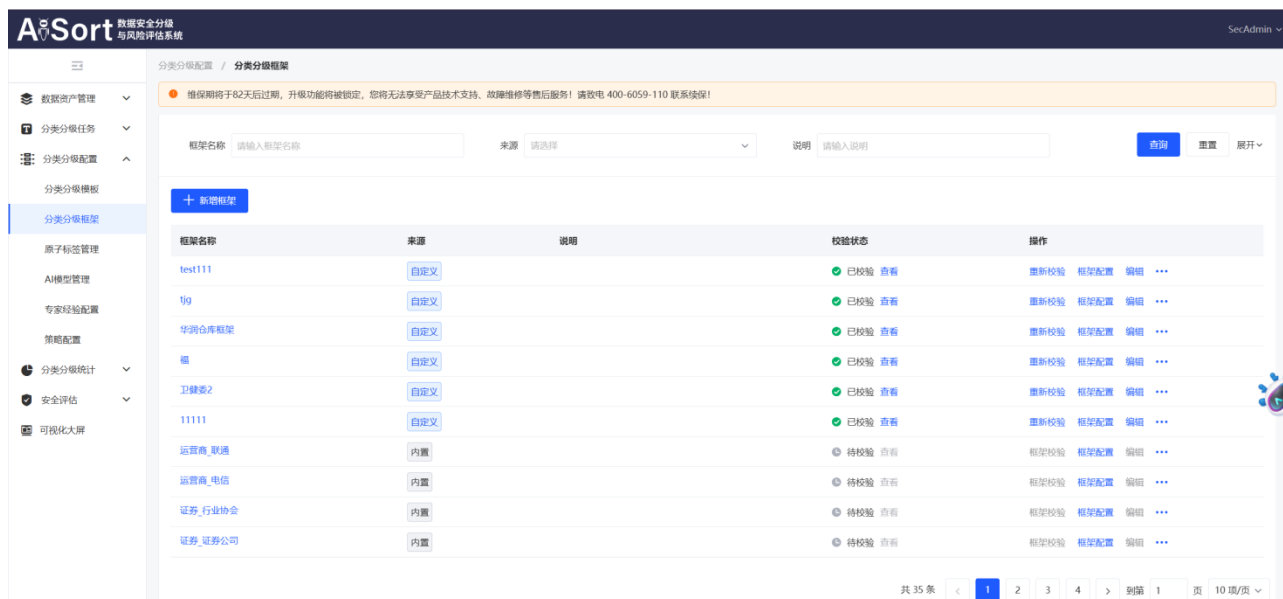
4.2.4.7 分类分级框架

◆ 框架列表

框架列表的展示字段包括：框架名称、来源、说明、校验状态、操作。其中：

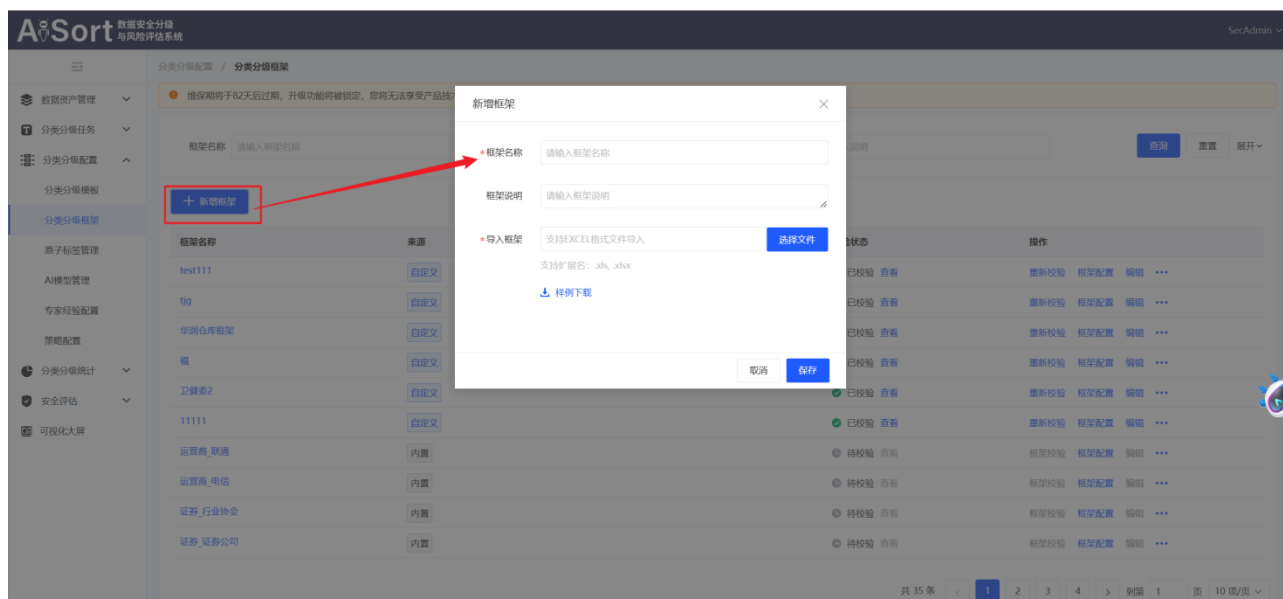
- 校验状态：分为待校验、已校验、校验中、校验失败四种；

用户可以通过输入框架名称、来源、说明及校验状态来查询已有的框架。



◆ 新增框架

点击<新增框架>，填写框架名称（要求不能与已有框架名重复），并通过导入框架分类分级框架页面新增分类分级框架。



点击<样例下载>，可以获取模板填写样例，模板包含以下字段：一级分类、二级分类、三级分类、四级分类、分级、重要数据定级、个人金融信息类别。其中，一级分类和分级为必填项。

4	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	四级子类	类别	数据级别	数据示例	分类说明	重要数据定级	金融个人信息类别									
2																
3																
4																
5																
6																
7																
8																
9																
10																
11																
12																
13																
14																
15																
16																
17																
18																
19																
20																
21																
22																
23																
24																
25																
26																
27																
28																
29																
30																

填写说明：

1. 黄色底色列为必填项，其余列可选填。
2. “类别”与“一级子类、二级子类、三级子类、四级子类”两者填写一即可，不存在时以“类别”为准，不包含特殊字符，例如：()!@#%&*~+~+[]{}|<=>?等。
3. 样例：客户、个人、个人自然信息、个人联系信息、客户-个人-个人自然信息-个人联系信息、3、姓名、分类说明、重要数据、C3

数据示例填写规范：

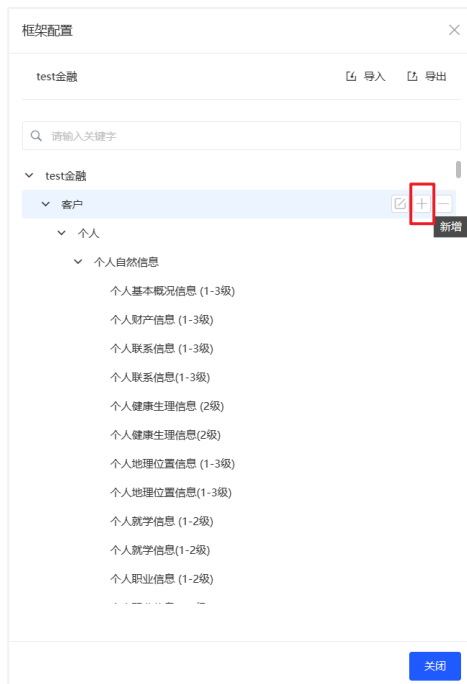
1. 不包含特殊字符，例如：()!@#%&*~+~+[]{}|<=>?等。
2. 不同分类中不能包含相同或含义相似的数据示例，如两个分类中确实存在相似数据，请通过加限定词的方式进行区分，如“姓名”，可区分成“客户姓名”和“员工姓名”。
3. 同一分类下避免出现相似含义的字段示例，如“客户编号”和“客户编码”，写一个即可。
4. 数据示例中如包含英文单词，请确保单词含义相对单一，如“id”含义明确，但“code”的含义较多，不建议放在字段示例中。
5. 每个分类下的数据示例不宜多过，一般不超过30个。

文件上传后，系统将自动对内容进行逐行校验；若校验失败，用户可以下载带有错误信息提示的文件。错误文件在用户原上传文件内容的基础上，新增“错误原因”一列，用于记录对应行的数据未通过校验的原因。根据提示，调整文件后重新上传符合要求的文件。

◆ 新增分类

在模板框架区域点击<配置框架>，进入框架配置页面。点击右侧<新增>按钮，可以为该分类新增子类。需要填写的信息包括：分类名称、分级、数据示例。其中，分类名称、分级为必填项，上级分类默认为点击<新增>按钮时所在的分类；当用户在策略配置中启用重要数据识别和金融个人信息分类时，此处可以填写重要数据定级和金融个人信息类别（见 4.7 策略配置）；填写数据示例，可以为分类分级预测提供依据。

<新增>功能对所有模板下的全部分类开放。新增子类支持新增、编辑、删除操作。



新增分类

*分类名称

上级分类

监管-数据报送-监管报送信息-监管明细数据上报信息

*分级

请选择

重要数据定级

请选择

金融个人信息类别

请选择

数据示例

取消

确认

◆ 编辑、删除分类

在框架配置页面点击<编辑>或<删除>按钮，可以编辑或删除特定分类。

点击<删除>，存在二次确认环节，删除分类将一并删除该分类下所有分类分级规则。若用户确认删除，系统将判断是否允许删除，系统无法删除已经在分类分级结果中被引用的分类。

注意：<删除>功能与<编辑>功能仅对非系统内置的分类开放。

点击<新增框架>，填写框架名称（要求不能与已有框架名重复），并通过导入框架分类分级框架页面新增分类分级

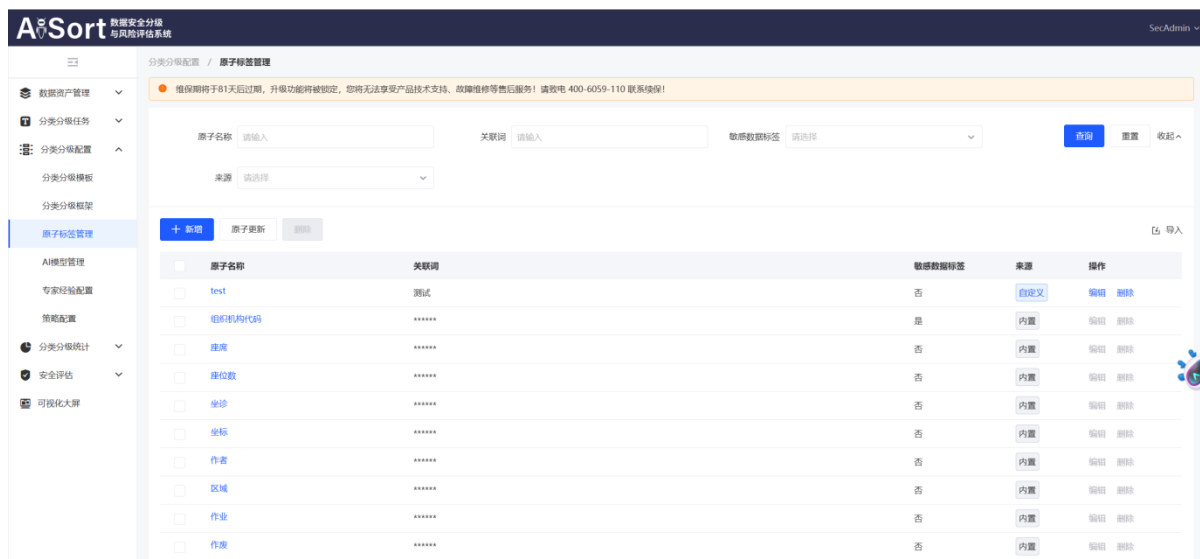
4.2.4.8 原子标签管理

◆ 原子标签列表

原子标签列表的展示字段包括：原子名称、关联词、敏感数据标签、来源、操作。其中：

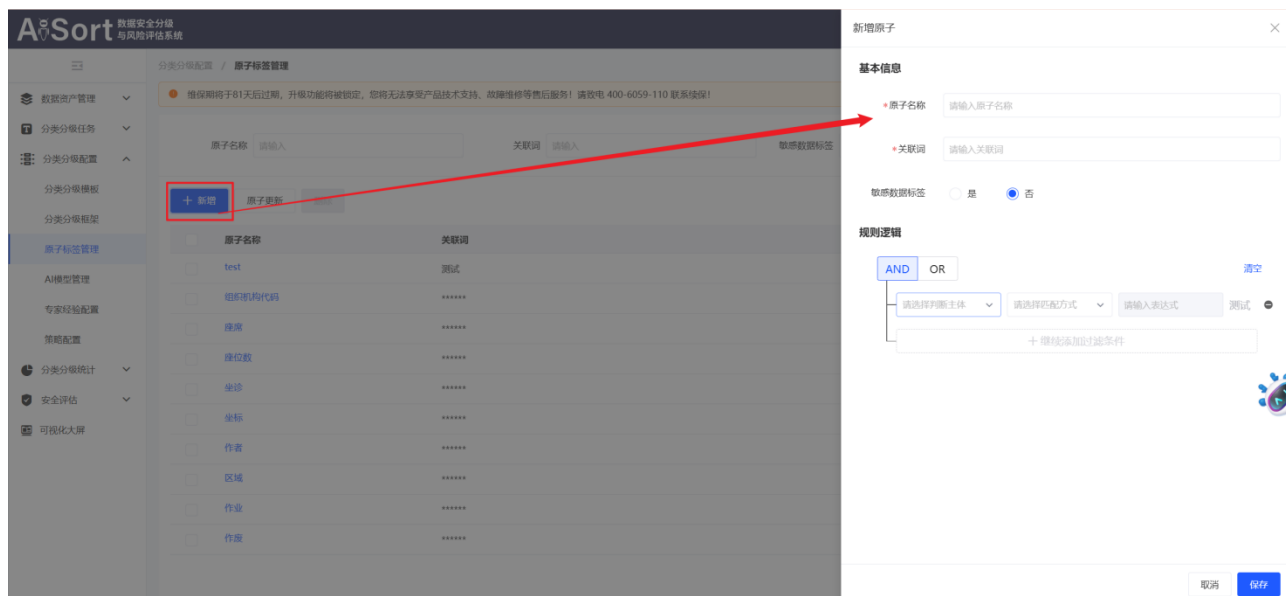
- 仅自定义的原子标签可进行编辑与删除等操作；

用户可以通过输入原子名称、关联词、敏感数据标签、来源来查询已有的原子标签。

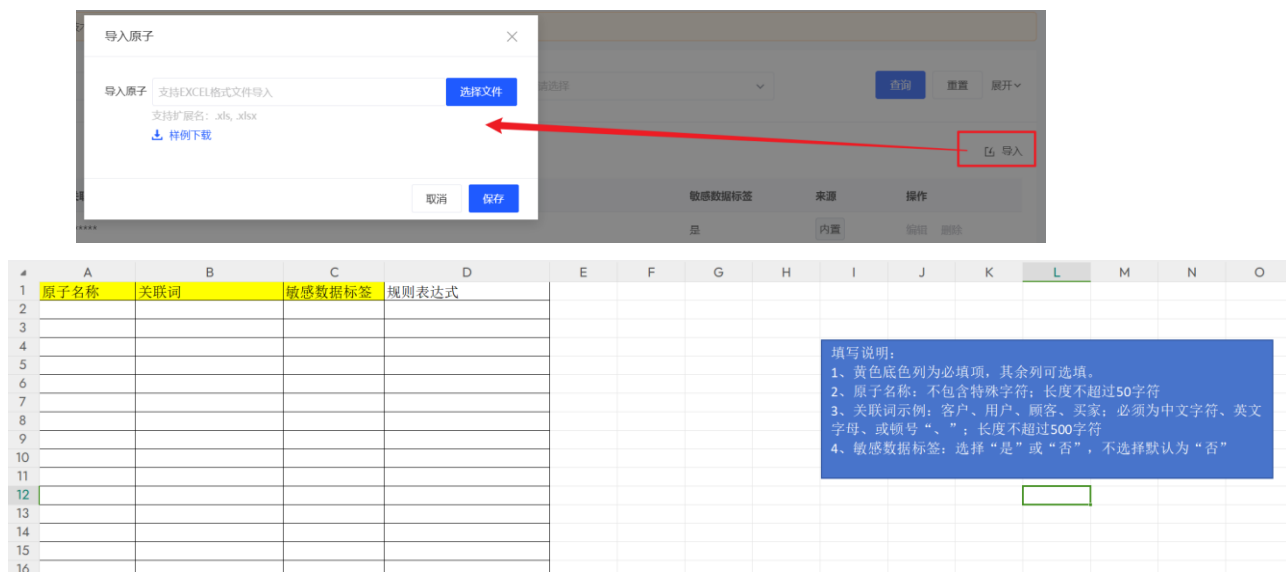


◆ 新增原子

点击<新增>，填写原子名称（要求不能与已有原子名称重复），并填写关联词、敏感数据数据标签、规则逻辑等信息来新增原子。



点击<导入>可批量导入原子，点击<样例下载>，可以获取模板填写样例。

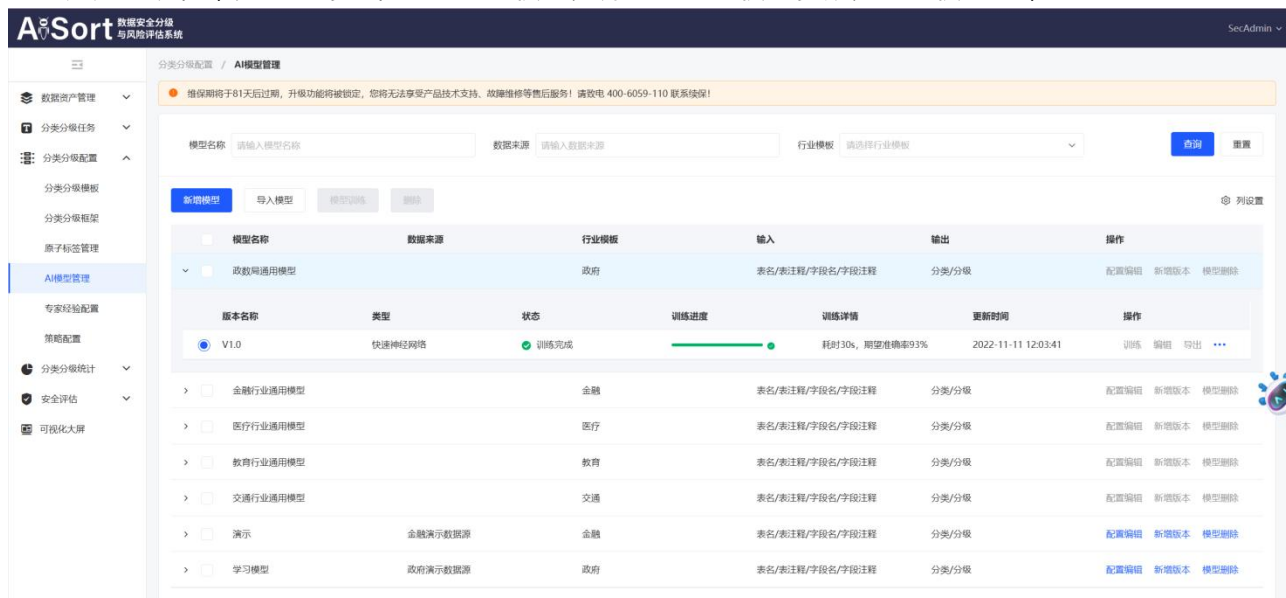


文件上传后，系统将自动对内容进行逐行校验；若校验失败，用户可以下载带有错误信息提示的文件。错误文件在用户原上传文件内容的基础上，新增“错误原因”一列，用于记录对应行的数据未通过校验的原因。根据提示，调整文件后重新上传符合要求的文件。

4.2.4.9 AI 模型管理

模型列表的展示字段包括：模型名称、数据来源、行业模板、输入、输出。用户可以通过输入模型名称、数据来源、行业模板来查询现有模型。

用户可以在本页面训练和管理自定义模型，并通过调整模型参数，进行模型版本迭代。



4.2.5 分类分级统计

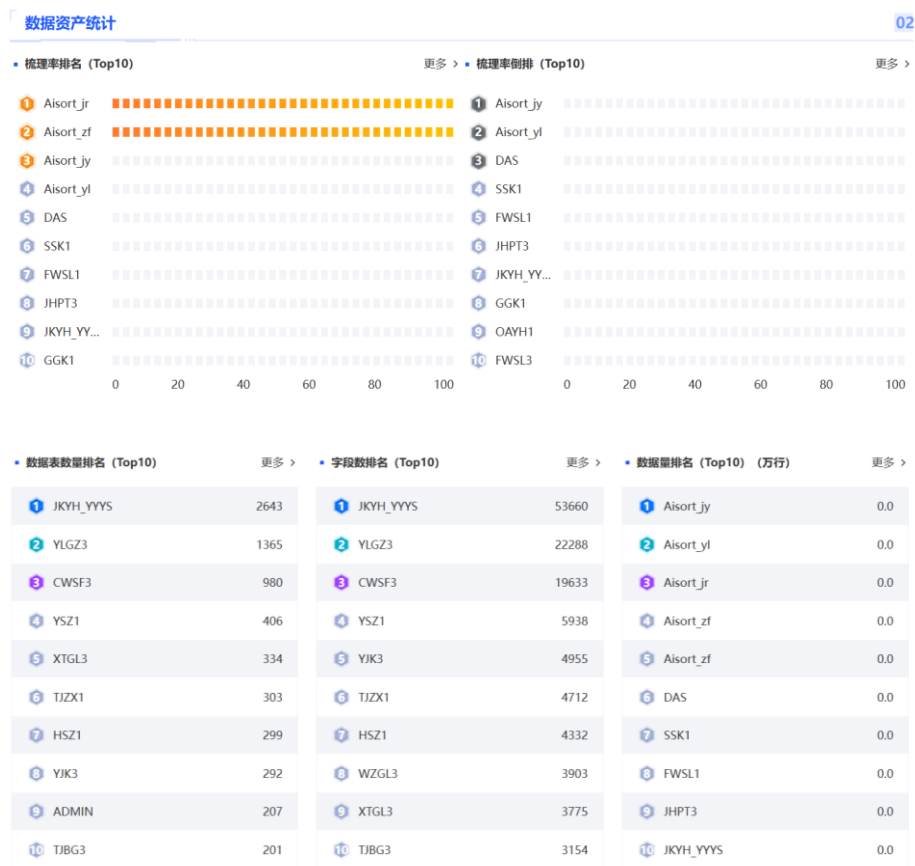
本模块页面左侧为数据源列表，展示的最细颗粒度可达数据库级；未同步的数据源置灰，无法勾选。每日零时系统自动刷新统计数值，用户也可以点击报告顶部的<手动刷新>按钮，获取实时最新数据。

右侧报告区根据列表选择范围动态变化，展示内容如下：

- **数据资产概况**：数据源数、数据库数、Schema 数、数据表数、字段数、未梳理字段数、梳理率、数据量（万行，数据源下能统计到的所有表的行数） 、近 6 个月资产/梳理新增趋势与分类分级梳理统计；



- **数据资产统计**：按 Schema 进行统计，展示梳理率正排 TOP10、梳理率倒排 TOP10、数据表数量正排 TOP10、字段数正排 TOP10、数据量正排 TOP10； 点击 <更多>，跳转分类分级报表页的数据资产视角查看 Schema 详情（详见 5.2 分类分级报表）；



- **分类分级统计**：选择要查看的行业模板，下方分类分级模板和数据分类分布联动显示对应内容。

■ **分类分级模板**：包含数据分类概述和分类详情统计

- ◆ **数据分类概述**：以树状图形式展示模板完整框架。点击蓝色节点可展开或折叠下级分类，节点显示白色时表示该分类无下级子类，可通过鼠标滚轮、双指缩放（触控板）、单击拖动画布调整显示范围；
- ◆ **分类详情统计**：用户在数据分类概述中单击选中特定分类后，本区域联动显示属于当前分类的数据表数、字段数、数据量、敏感字段数；点击<更多>，跳转分类分级报表页的**分类分级视角**查看分类详情。

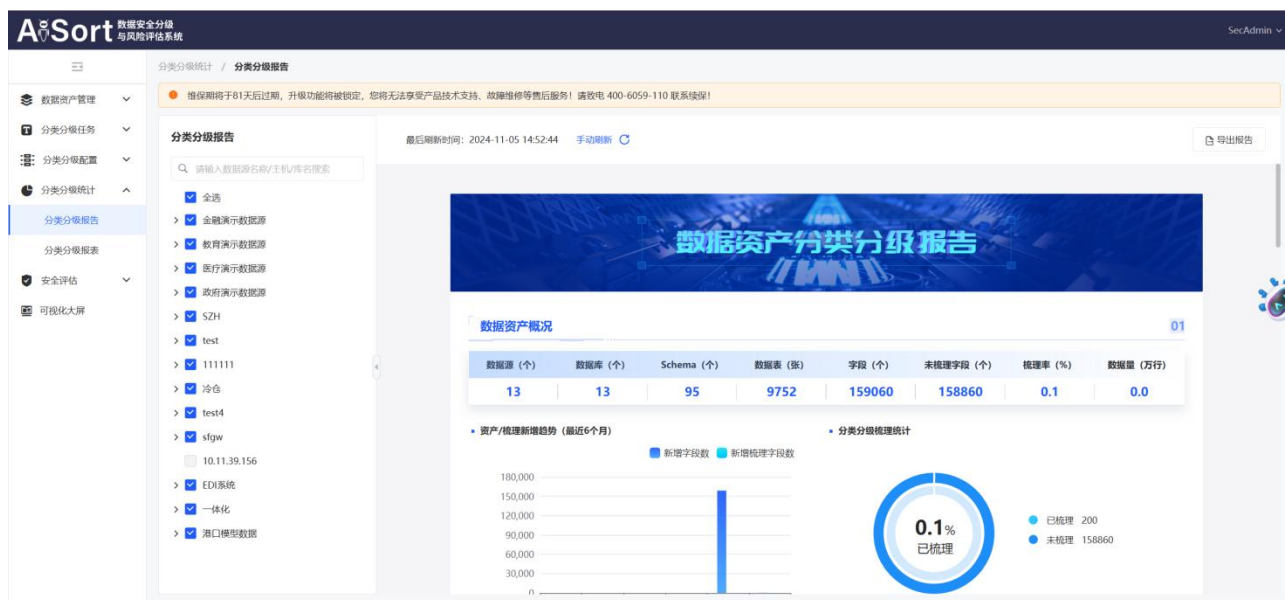


■ **数据分类分布**：

- ◆ 按字段数排序可视化展示数据分类分布。只统计到模板下的二级分类（如客户-个人），每个层级显示的分类不超过 6 个，若超过 6 个则显示字段数最多的 5 个分类，同层级内的其他分类合并统计为“其他”。不同的一级分类以不同色相区分；一级分类内部，不同的二级分类以色彩深浅区分；
- ◆ 按 Schema 进行统计，展示敏感字段数正排 TOP10、敏感数据数据量正排 TOP10、不同分级字段数正排 TOP10（支持多选不同分级）。



- 导出报告：点击<导出报告>，可以选择以 pdf 或 html 格式导出当前分类分级报告。



4.2.6 安全评估

可视化图表直观展示了数据源中存在的漏洞、配置基线、弱口令风险数，并分别指出风险数最多的前 5 项数据源。点击数据源名称，可以跳转至对应的风险检测详情界面。

风险数量统计基于所有已完成风险扫描任务的数据源，不包括扫描失败的任务。当未成功扫描到特定类型的风险时，对应图表区域显示“暂无数据”。

点击本区域右上角<图表展开>，可以展开或折叠上述可视化图表。



4.2.7 可视化大屏

点击左侧导航栏<可视化大屏>，选择<分类分级大屏>，跳转至分级分类大屏页面。分级分类大屏集中展示了数据分类分级结果及安全评估概况，每五分钟系统自动刷新统计数值。具体展示模块如下：

- 重点数据源（安全风险）：数据源风险综合评分 TOP6、数据源弱口令风险数 TOP5、数据源漏洞风险数 TOP5、数据源配置基线风险数 TOP5；
- 重点数据源（敏感数据）：数据源敏感评分 TOP6、数据源敏感字段数 TOP5。
- 数据分类分级结果概览：数据源统计、重点数据源的分类分级和风险评估结果、数据分级分布、数据分类分布。



6.1 产品类

- **数据安全专区支持哪些计费方式？**

目前数据安全专区的产品套餐（包括数据分类分级、数据脱敏与水印）为包周期，包年 8.33 折，服务套餐（数据分类分级服务、数据安全咨询规划服务）为一次性计收计费，注意不支持退订。

- **数据安全专区是否支持试用？**

产品套餐（包括数据分类分级、数据脱敏与水印）支持试用 7-30 天，服务套餐（数据分类分级服务、数据安全咨询规划服务）不支持试用。

- **购买数据安全产品是否还要额外购买云主机、弹性 IP？**

数据分类分级、数据脱敏与水印不需要额外购买云主机和弹性 IP，开通成功后点击管理即可使用产品。

- **数据脱敏可以应用于哪些场景？**

- **数据共享交换：灵活策略**

为了确保数据安全，数据提供方需要按照需求方的实际用途对数据进行脱敏处理并添加水印信息。在数据开放时，提供方需要筛选出可开放字段，对不可开放字段进行脱敏处理。可通过数据脱敏与水印系统基于用户业务情况制定不同的脱敏策略，灵活遮蔽、清洗各类敏感数据。

- **开发测试环境：规避风险**

在业务系统开发测试过程中，为确保数据质量存在直接引用生产数据的情况。可通过数据脱敏与水印系统一致性脱敏算法，将生产数据经过脱敏处理后导入开发测试环境，在保证数据的“真实性”的同时规避生产数据泄露的风险。

- **共享数据泄漏：溯源追责**

当数据被共享给第三方后，第三方有责任对数据进行有效保护。可通过数据脱敏与水印系统将三方单位信息、共享时间等信息以数据水印的方式嵌入到共享数据中。一旦发生数据泄漏，用户只需提供少量的泄漏数据样本就可以通过嵌入的水印信息溯源泄漏源头，明确责任方并采取相应的追责措施。

- **数据分析计算：隐私保护**

数据脱敏与水印系统支持保留计算特征的脱敏方式，在剔除数据敏感性的同时最大程度保留 AI 建模相关数据的可用性。可通过数据脱敏与水印系统对个人识别信息进行脱敏处理，可完成数据去标识化且不影响统计分析结果，确保个人隐私安全。

- **数据脱敏适配数据的类型有哪些？**

数据脱敏与水印系统适配于目前市场中常见的数据类型，可高效满足目前市场中不同行业对不同数据类型脱敏或水印溯源的要求。此外数据脱敏与水印系统不仅支持 xls、xlsx、csv、dmp 等结构化数据文件的脱敏及水印处理，还支持 word、txt、xml、json 等半/非结构化文

件的脱敏处理。

数据类型	支持的数据
传统型数据库	MYSQL、GOLDENDB、TIDB、MYCAT、DBLE、DRDS、TELEDB、TDSQL_FOR_M、TELEPG、MsSQL(SQLServer)、MARIADB、INFORMIX、SYBASE、DB2、CHCHEDB
非关系型数据库	ElasticSearch、MongoDB、ClickHouse
云数据库+大数据组件	HIVE、星环(INCEPTOR)、IMPALA、GREENPLUM、HBASE、ODPS(Maxcompute)、
国产数据库	南大通用(GBASE)、达梦(DM)、人大金仓(KINGBASE)、高斯 DB-T(Gauss100)、高斯 D (UXDB)
StructFile	FTP (支持 csv/xls/xlsx/txt/del/dbf 格式)、SFTP(同 FTP)、LOCAL(同 FTP)、ORACLEDMP、

- 数据分类分级系统支持的数据类型

数据类型	支持的数据
传统型数据库	Oracle、DB2、MySQL、MariaDB、SQLServer (MsSQL)、PostgreSQL、Informix、Sy
非关系型数据库	ElasticSearch、MongoDB、ClickHouse
云数据库+大数据组件	ODPS(MaxCompute)、GaussDB、Teradata、Greenplum、Hive、Hbase、CDH、TDH、
国产数据库	高斯 DB-T (GAUSS100)、高斯 DB-A (GAUSS200)、达梦 (DM)、南大通用 (GBA 优炫数据库 (UXDB)
文件传输协议	FTP、SFTP、Local (本地上传)

数据类型	支持的数据
------	-------

- 数据脱敏系统支持哪些水印算法？

数据脱敏系统支持多种数据水印算法，包括并不限于伪行水印、伪列水印、脱敏水印、内容修改水印、零宽水印。

- 数据脱敏系统国产化适配能力如何？

支持的国产数据库：高斯 DB-T (GAUSS100)、高斯 DB-A (GAUSS200)、达梦 (DM)、南大通用 (GBASE)、人大金仓 (KINGBASE)、OceanBase、TiDB、优炫数据库 (UXDB)。

6.2 产品使用类

- 数据脱敏是否支持原地脱敏？

数据脱敏支持原地脱敏功能，请您先打开“系统配置管理->全局配置”的“数据流向”开关。然后在‘脱敏/水印任务→结构化任务’下新增任务中的“数据流向”配置项下，就会出现相关选项。

- 数据脱敏是否支持非结构化脱敏？

数据脱敏支持非结构化脱敏（包括 XML、JSON、TXT、WORD、DICOM 等），具体请先登录 SysAdmin 用户，查看‘脱敏/水印任务→非结构化任务’。

- 数据脱敏数据是否会落地？

数据脱敏系统保证敏感数据全程不落地，所有敏感数据全部在内存中处理，避免产品本身成为数据泄漏的一个风险点。

实现原理：代码用 Java 语言实现，脱敏步骤在 JVM 内存中进行，脱敏结束后，方法自动出栈，堆内存中的数据对象会被自动清理，不会落地到磁盘。

- 数据脱敏支持哪些脱敏算法？

支持 24 大类 200 多种系统内置脱敏算法，适用各种业务场景主要算法支持情况如下：

- 置空：直接将待脱敏的信息以填充空字符或者删除的形式抹除。
- 乱序：在结构化数据（例如数据库）中颇为常用，对于待脱敏的列，不对列的内容进行修改，仅对数据的顺序进行随机打乱。
- 遮蔽：保留数据一些位置上的信息，对于敏感位置的信息使用指定的字符进行替换。
- 分割：保留数据一些位置上的信息，对于敏感位置的信息进行删除。
- 替换：使用固定值或字典映射表对敏感数据进行替换。
- 取整：对数值类型和日期时间类型的数据进行取整操作。
- 哈希：将哈希（或哈希加盐）编码后的数据作为脱敏结果输出。
- 仿真（保留业务含义）：考虑到业务含义，则生成的数据需符合校验规则，主要包括长度、取值范围以及校验规则和校验位的计算等。
- 密码学：根据所选参数指定的密码学加密算法对数据进行加密。支持的加密算法有：RSA/AES/SM2/SM4。
- 编码：根据所选参数指定的编码方式对数据进行编码。
- 保留类别频次特征：主要指的是反映事物类别的数据类型，具有有限个无序的值，或枚举类型，脱敏后各个类型出现的频次可保持不变。
- 保留数值统计特征：支持分布重建、平均、区间、标准化、归一化、添加噪声、一致性等算法：对数值型数据在脱敏后可保留数据中的高阶统计特征及数据分布特征，仍满足趋势分析要求。
- 关联关系保留：支持计算关系保留、乱序关联保留、身份信息关联保留等算法：对数据在脱敏后可保留其原始的字段间关联关系。
- 保留原数据：对指定的列的数据不做脱敏处理。

溯源算法：对数据本身添加仿真水印，做到精准溯源。

- 数据脱敏支持哪些水印算法？

支持多种数据水印算法，包括并不限于伪行水印、伪列水印、脱敏水印、内容修改水印、零宽水印。

- 数据脱敏是否支持异构脱敏？

暂不支持大部分异构脱敏自动生成，仅支持少量兼容性极强的异构脱敏，例如 MySQL、RDS。

- 数据脱敏是否支持 PDF 脱敏？

暂不支持 PDF 脱敏，但支持非结构化脱敏（包括 XML、JSON、TXT、WORD、DICOM 等）。