



# 数据脱敏与水印

用户指南

天翼云科技有限公司

# 1 用户指南

---

## 1.1 数据脱敏操作指导

### 1.1.1 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,大幅提升了对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。

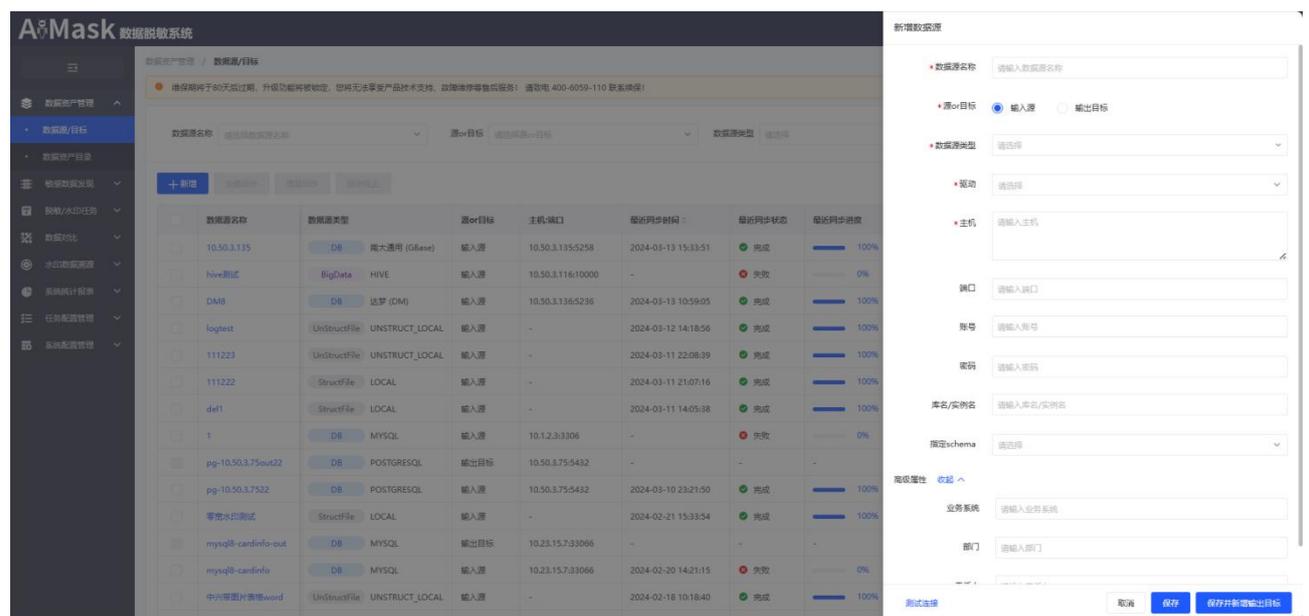
内容修改水印算法：适用于不能影响数据的业务含义并具有较强的业务使用或分析需求的场景，同时隐蔽性要求高。

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。

## 1.1.2 数据资产管理

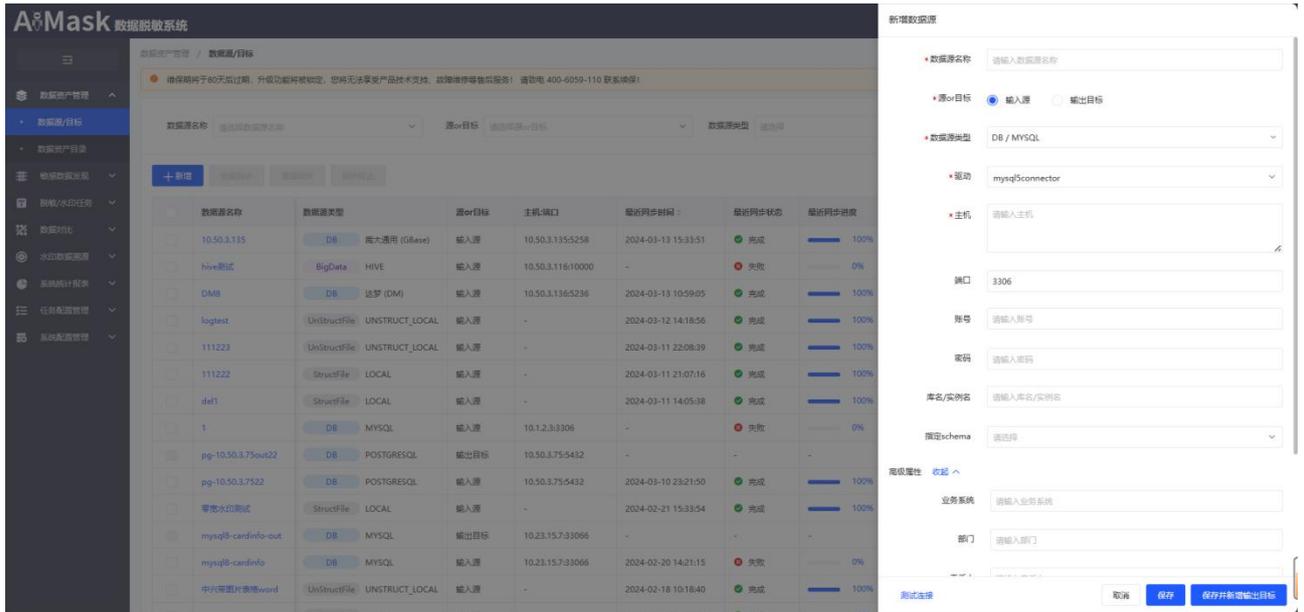
### 新增数据源

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。

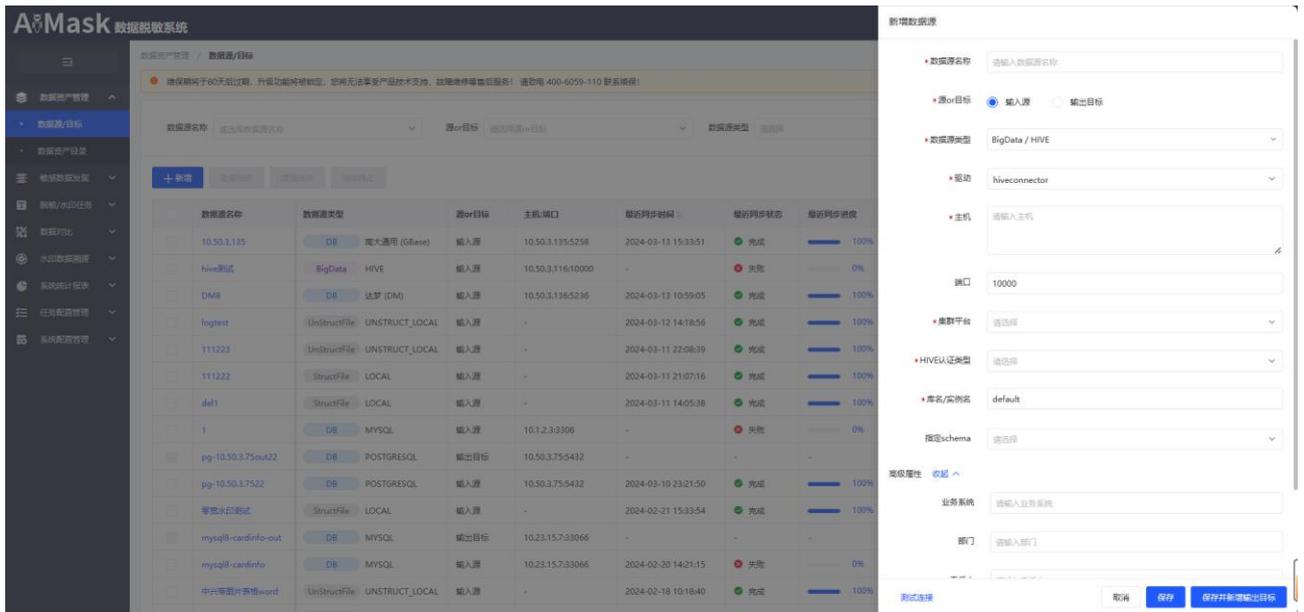


字段前带\*的为必填字段，字段前不带\*的为选填字段。不同的数据库类型，必填项有所差异，请在选择“数据源类型”后根据页面要求仔细核对，例如‘主机’填写内容为数据源的主机地址，一般可填写为 IP 地址。如下图所示。

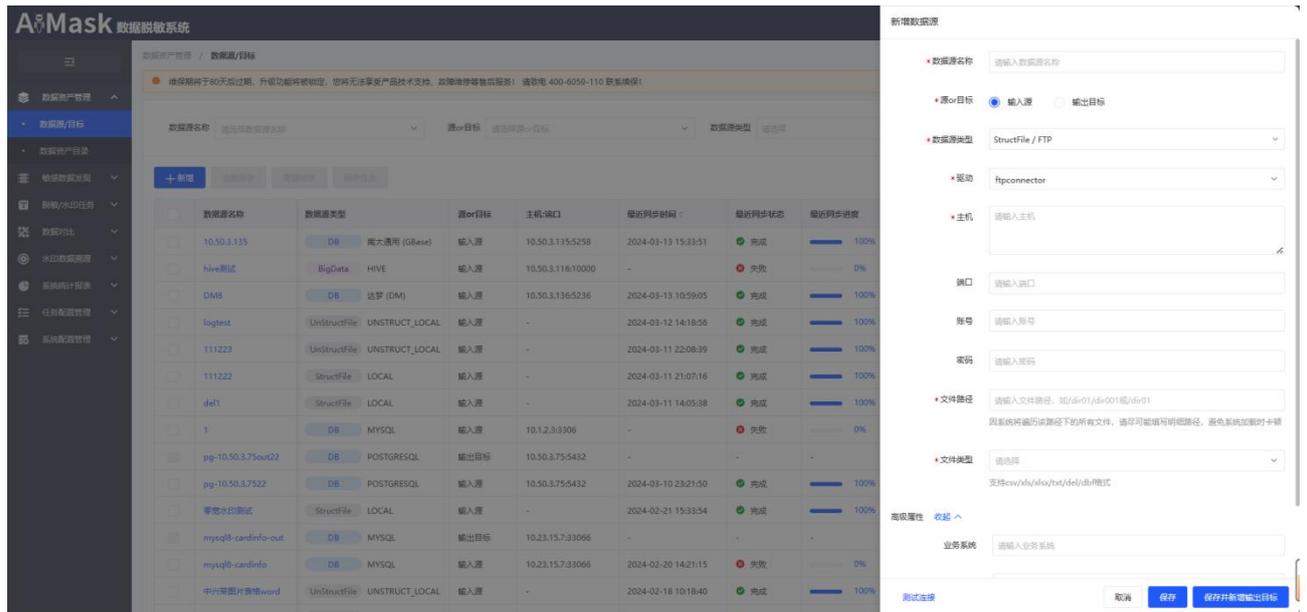
下图示例了 Mysql 类型数据库所需配置项清单。



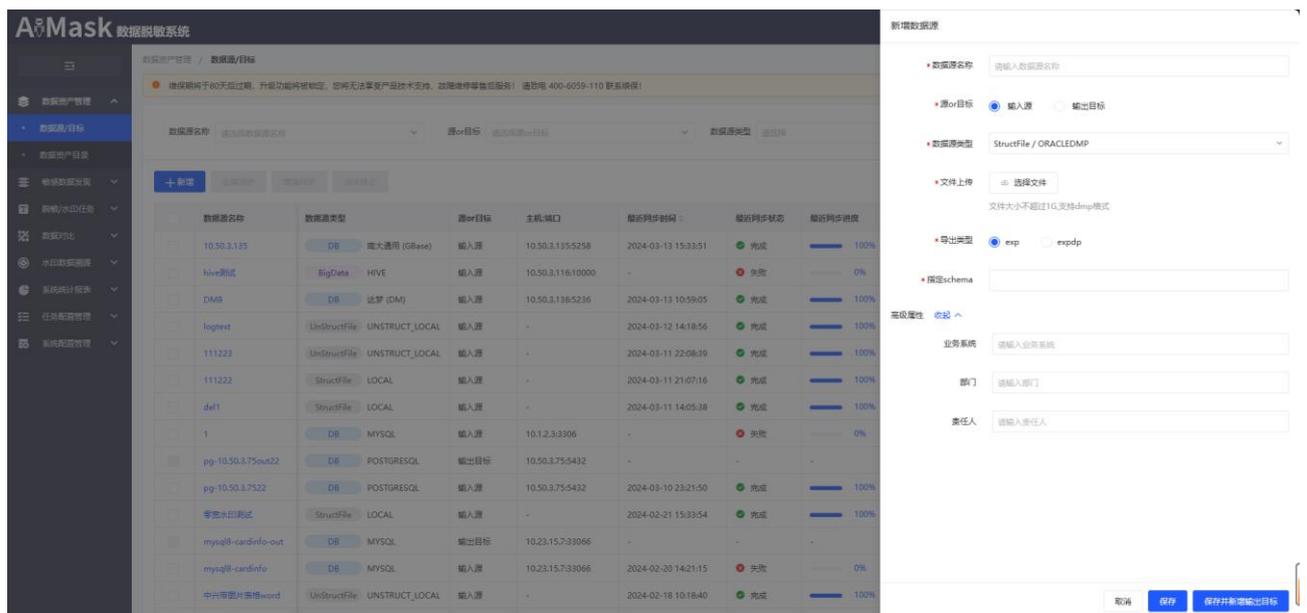
下图示例了 Hive 类型数据库所需配置项清单。



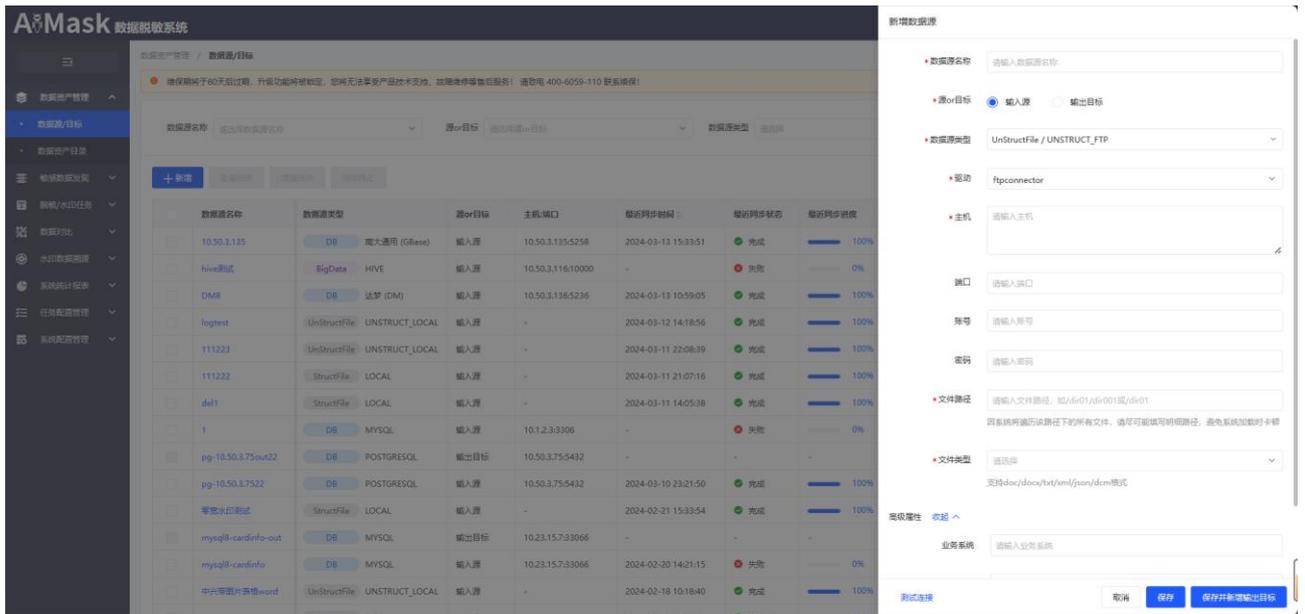
下图示例了结构化 FTP 类型数据所需配置项清单。



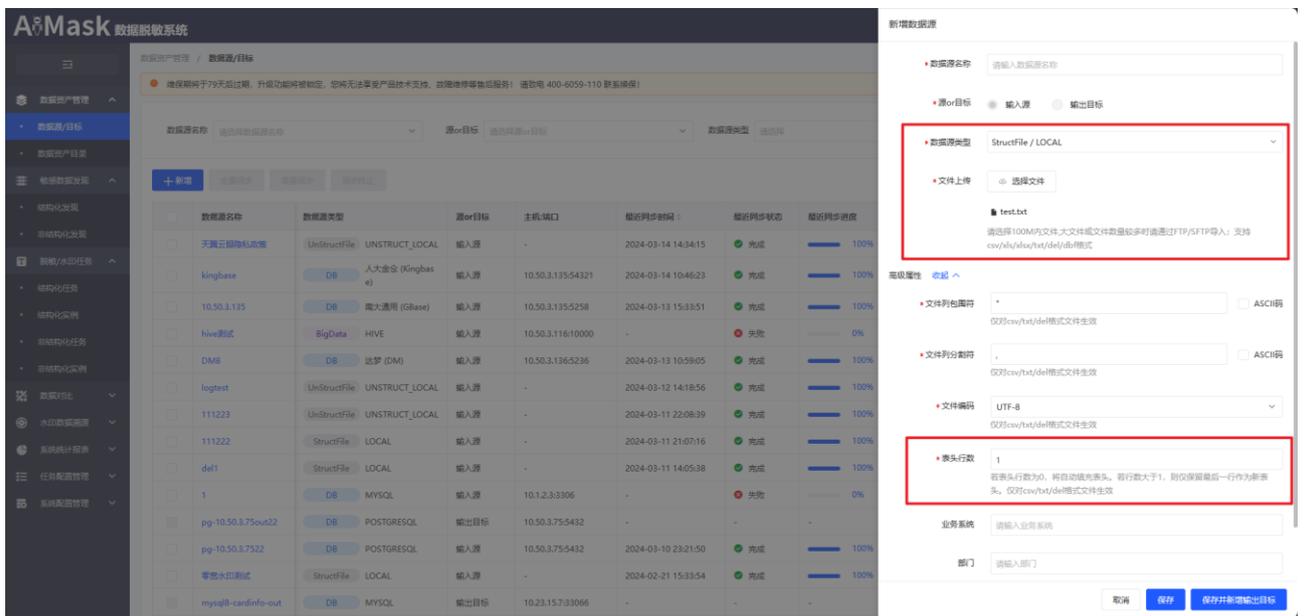
下图示例了 DMP 类型数据所需配置项清单。



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

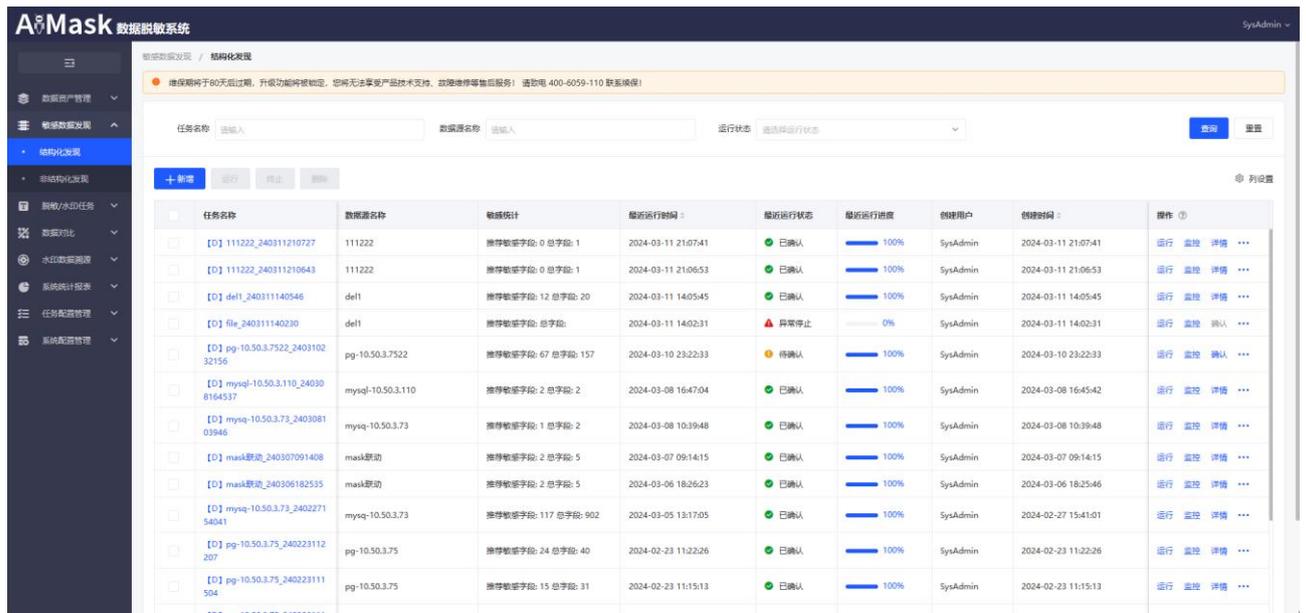
点击<元数据同步>，可以进行同步操作，会获取数据源的数据库、表、字段及账号信息。

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。

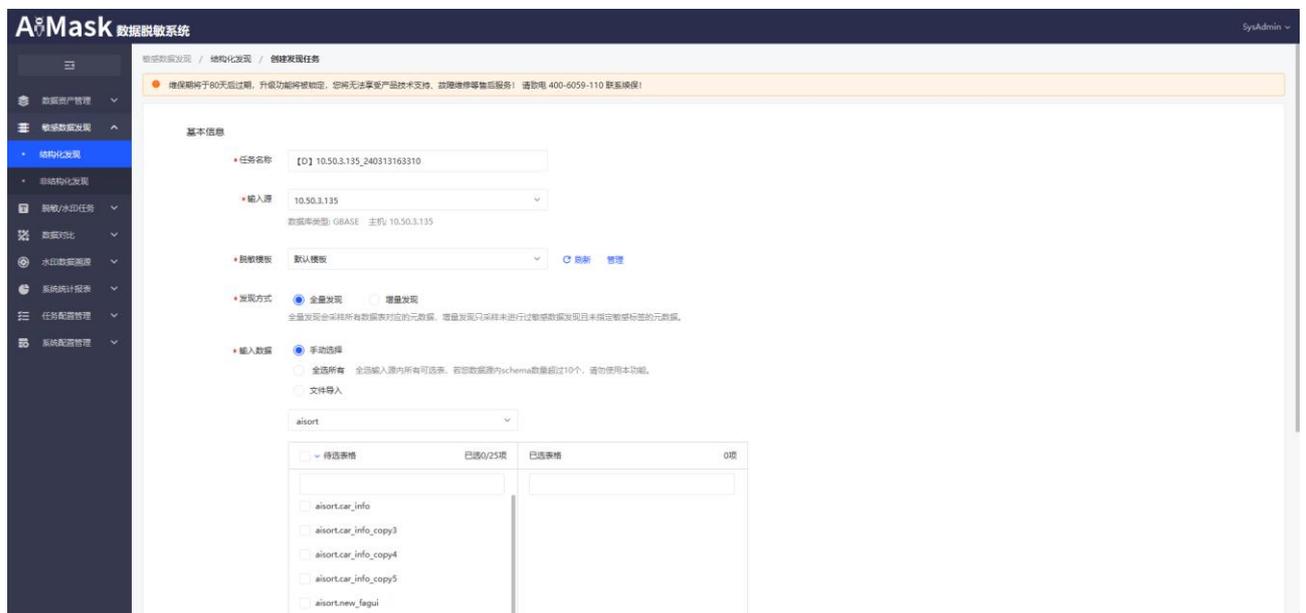
## 1.1.3 敏感数据发现

### 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。



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: 2 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_240308103946	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联动_240307091408	mask联动	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联动_240306182535	mask联动	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



**创建发现任务**

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
- 发现方式:  全量发现  增量发现  
全量发现会扫描所有数据表对应的元数据，增量发现只扫描未进行过敏感数据发现且未指定敏感字段的元数据。
- 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

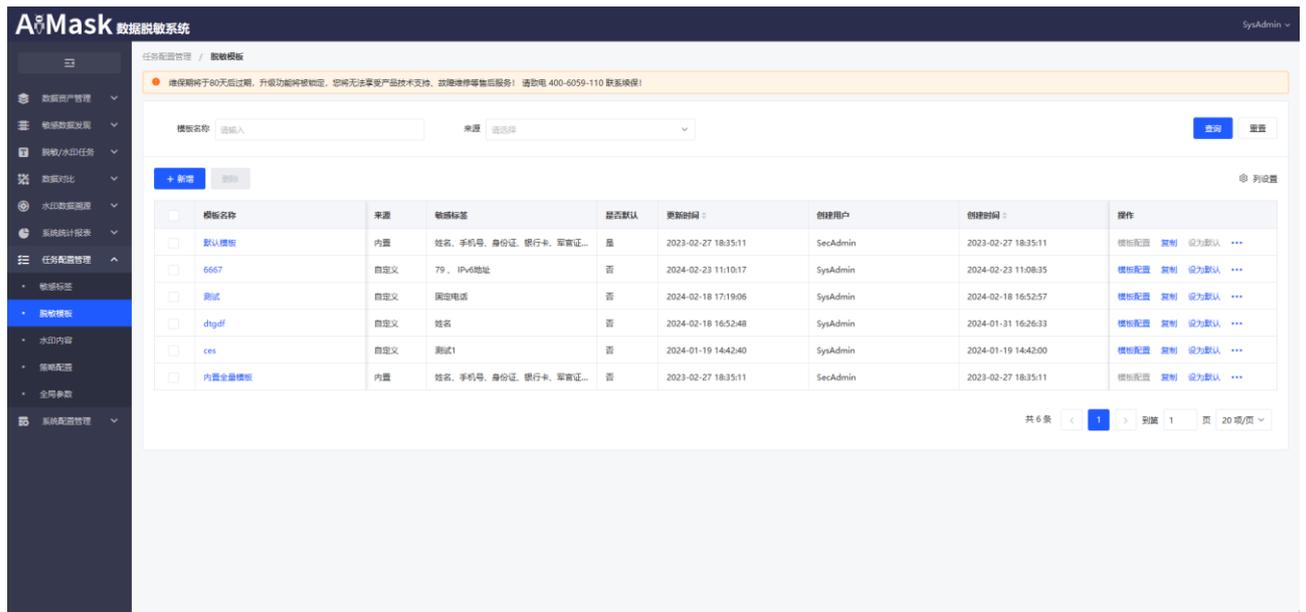
- aisort\_car\_info
- aisort\_car\_info\_copy3
- aisort\_car\_info\_copy4
- aisort\_car\_info\_copy5
- aisort\_new\_fagui

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。

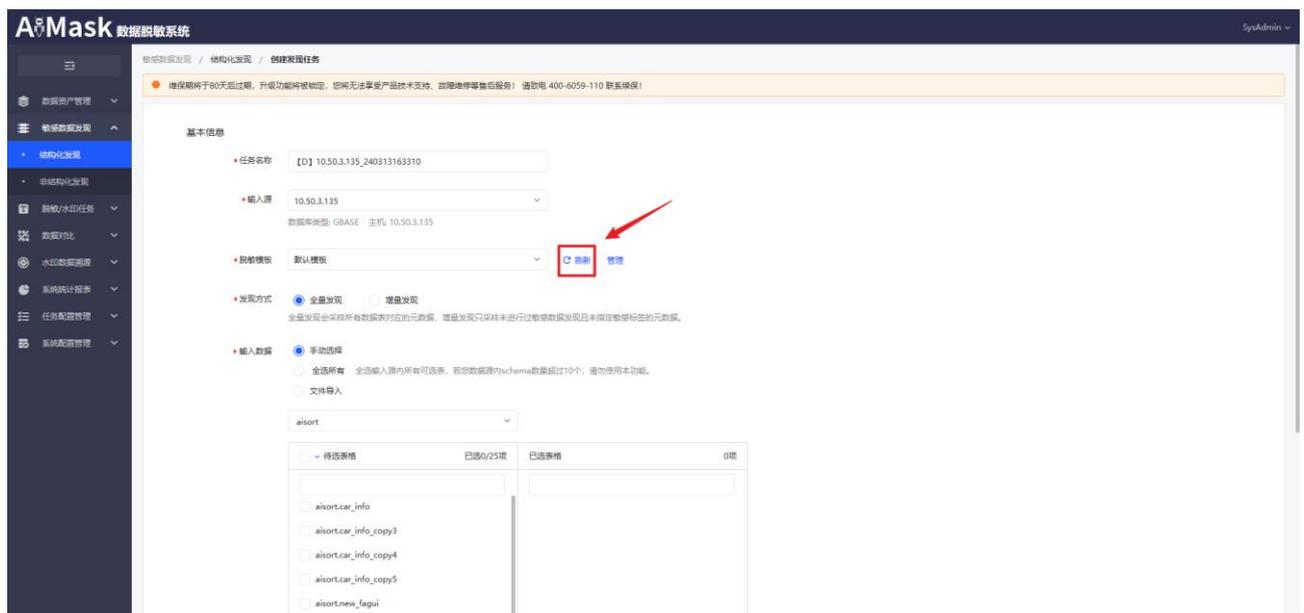
用户可以在“脱敏模板”参数选择进行敏感数据发现的模板，系统内置了默认模板供用户在常规场景下直接使用。如果用户需要自定义或者管理脱敏模板，可点击新增页面中的<管理>，然后系统会弹出一个新的名为‘脱敏模板’的页面供用户管理脱敏模板，如下图所示。



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。



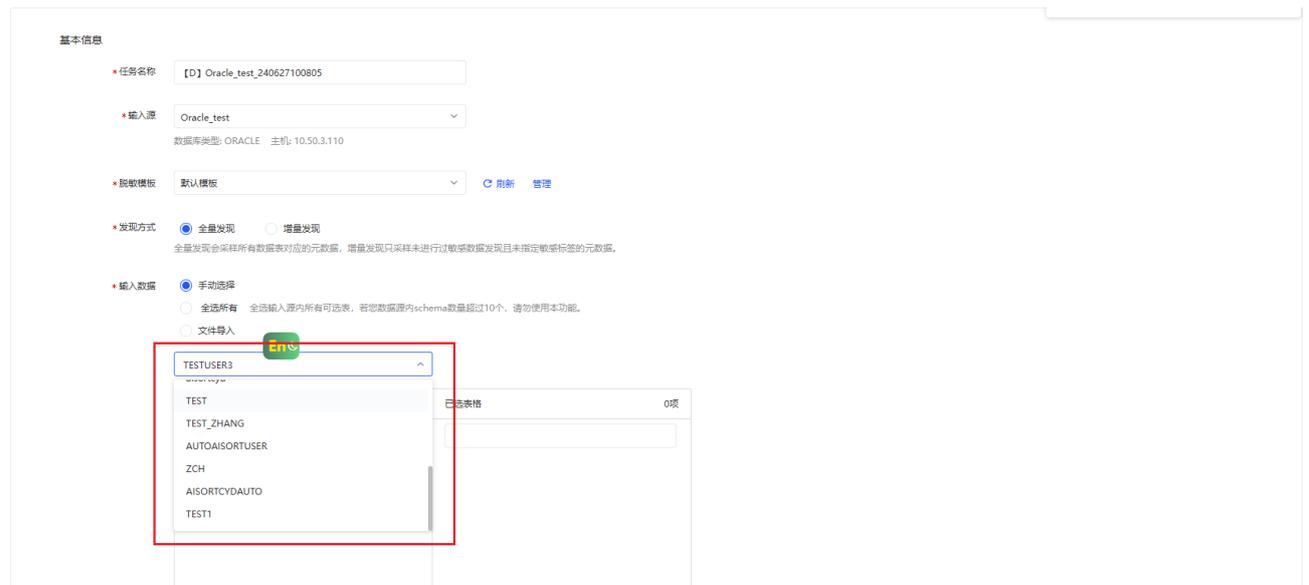
用户配置完毕后点击<保存>, 回到先前配置敏感数据发现任务的页面并点击<刷新>即可在该任务中使用刚保存的自定义脱敏模板, 如下图所示。



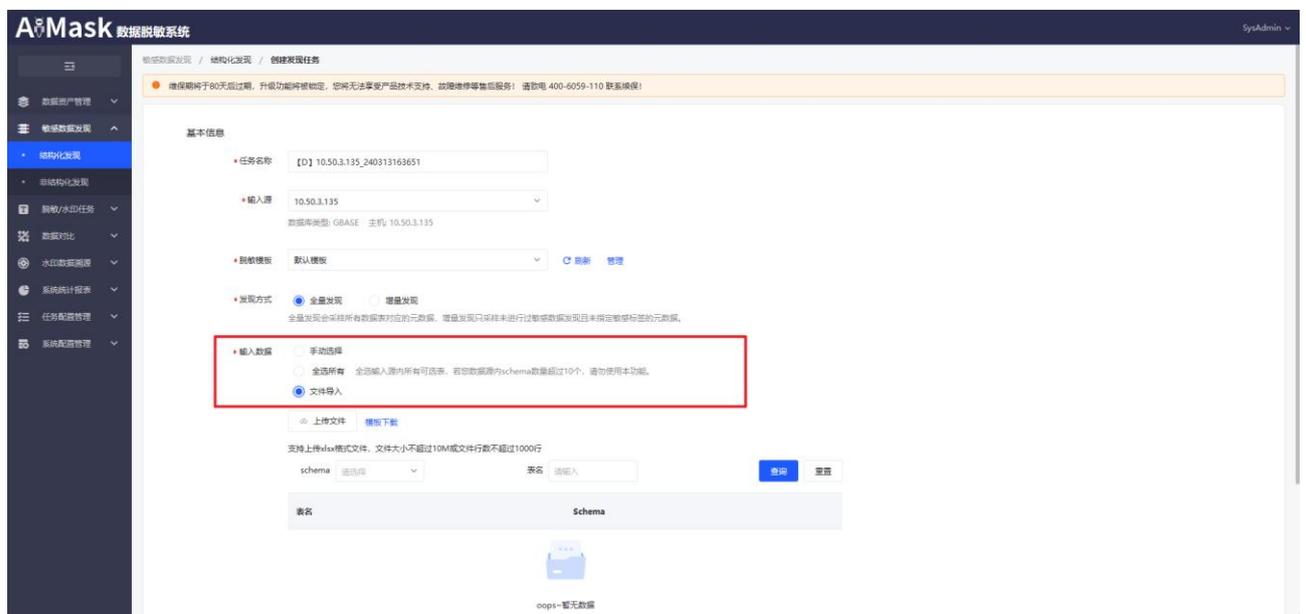
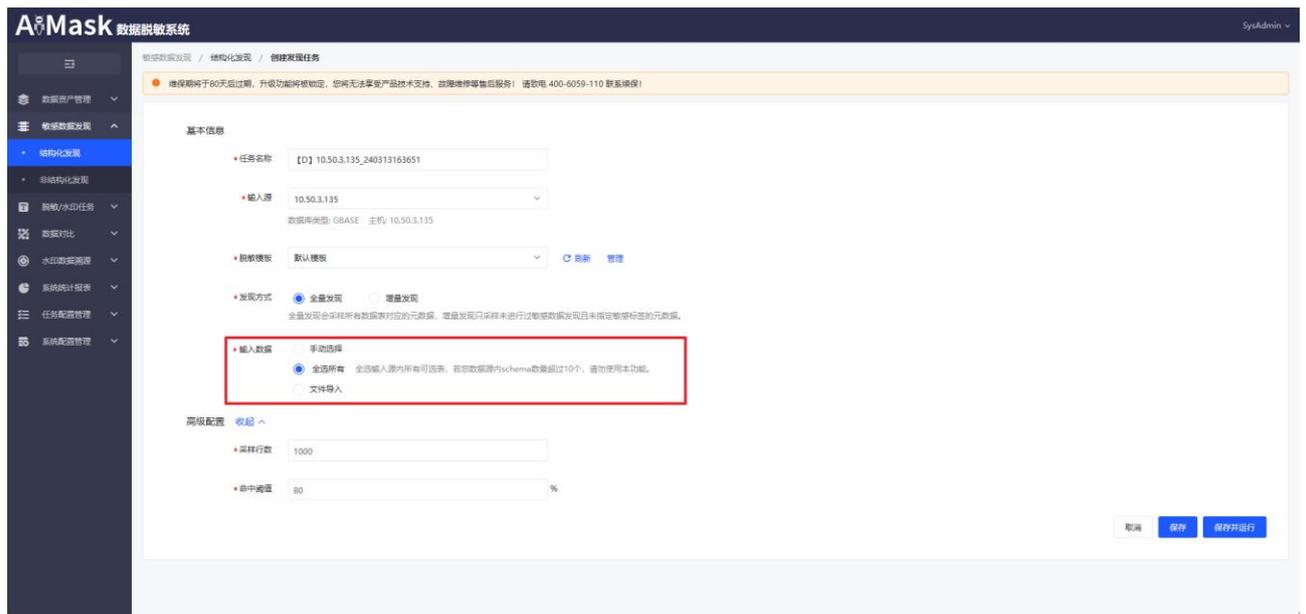
“发现方式”参数包括全量发现和增量发现 2 种。用数据举例, 假设里面一共有 4 个表, 如果我选择的是全量发现, 那他无论之前这个敏感发现任务有没有运行过, 会全部重新对这 4 个表做一次敏感数据发现。如果我选择的是增量发现, 他只会去处理这 4 个表中跟上次这个敏感发现任务运行时相比新增的数据列, 原来老的数据列 AiMask 会认为我已经发现过了, 就不用再去发现一次了。增量和全量相比, 因为处理的数据少很多, 所以会快很多, 那如果我

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。



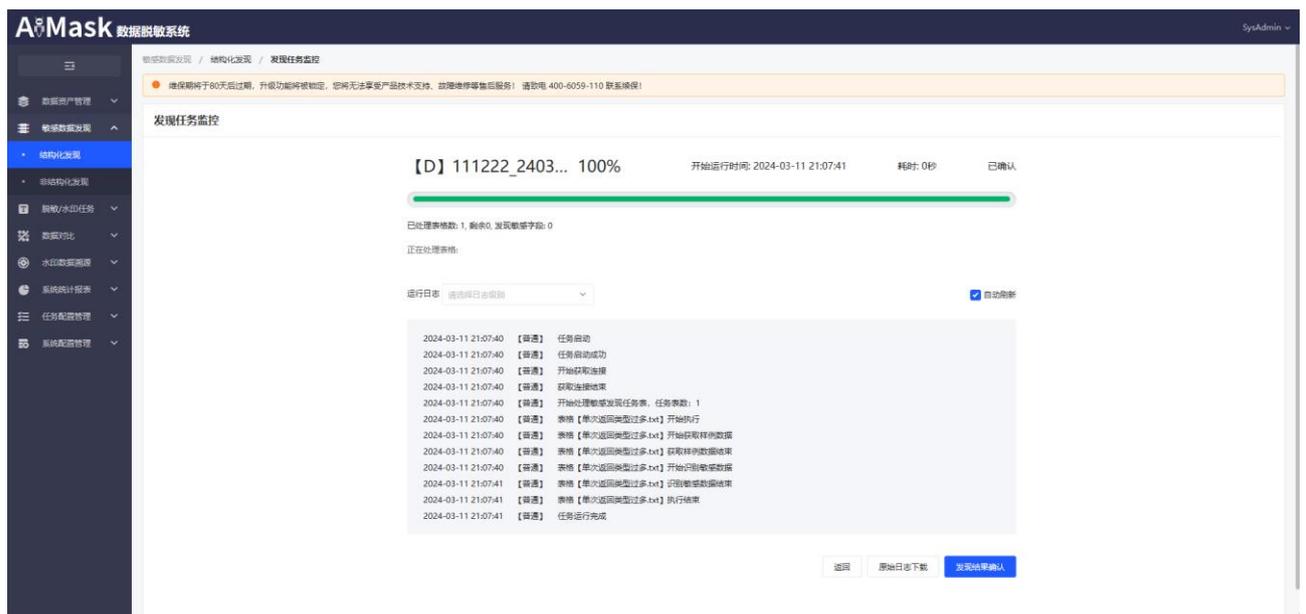
然后勾选他要敏感发现哪些表格的数据，我们提供了全选当页、全选所有、清空已选这些功能（均基于单个 schema）。如果用户是全选所有的，则默认全选输入源内所有 schema 下的全部可选表；若数据源内 schema 数量超过 10 个，请勿使用本功能。如果用户是文件导入的，直接按照我们提供的模板填写信息后上传即可，这个功能是为了方便用户批量操作。用户可以选择直接在新增界面手动选择输入源中的输入数据，或者按照系统要求选择通过文件批量导入。选择<全选所有>或<文件导入>时的界面如下图所示。



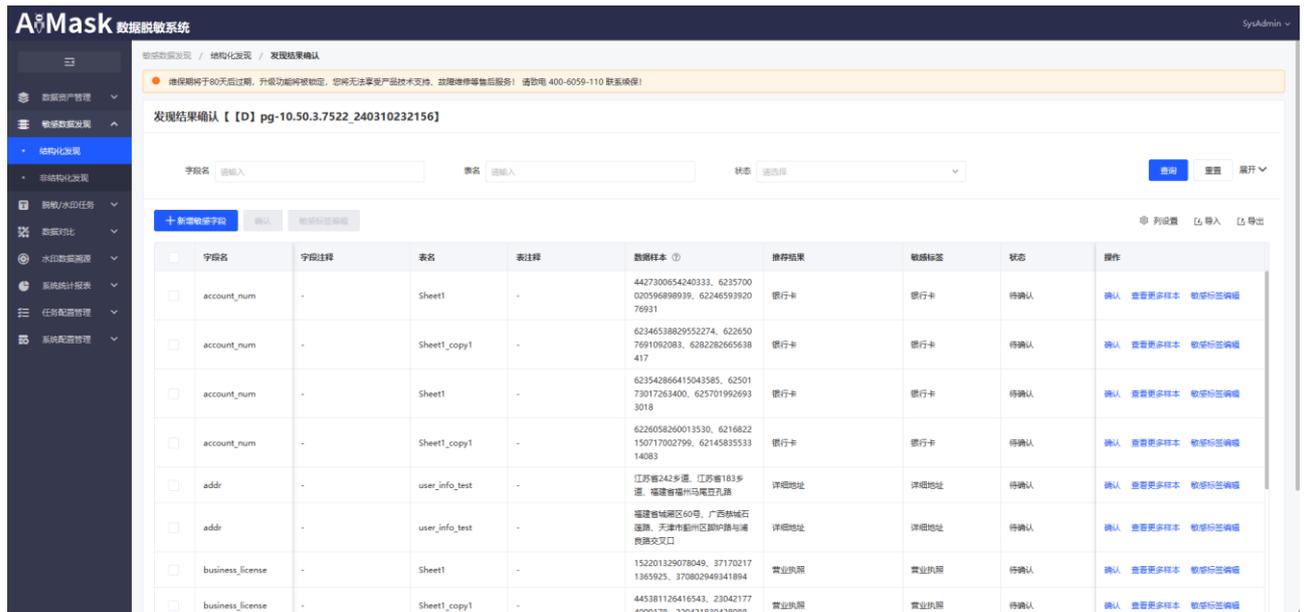
“高级配置”是我们系统默认的一些配置信息，一般是不需要修改的。我分别解释一下采样行数和命中阈值的概念。AiMask 做敏感数据发现时，一般是对数据内容或者数据包含的列名进行解析。如果我们要判断一个列是否是手机号码，一般情况下我们要拿到客户数据库里的一部分数据内容的。具体拿多少就是用采样行数来控制的，我们系统默认推荐的 1000 行是我们在大量客户实践过程中经验所得的一个相对合适的值。命中阈值跟客户的数据质量有关，如果客户数据质量很高，比如说他手机号码这一列的数据大多都是非常正规的手机号码，那我们设置的命中阈值可以高一些，比如 90%或者 95%。如果客户的数据质量比较差，我

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出操作，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，所有已导入字段的的状态列自动变更为“已确认”。

**AiMask 数据脱敏系统**

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

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, 623570020596898939, 6224659392076931	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, 6282282665638417	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, 625701992693018	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, 621458353314083	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, 广西防城港市, 天津市静海区团泊湖与海铁路交叉口	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, 22042183048088	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

**AiMask 数据脱敏系统**

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

导入

导入源 支持EXCEL/CSV/TXT格式文件导入 选择文件 导出

导出为:  xlsx  csv  txt

导入数据标签时请注意:

- 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的任何信息, 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。

取消 导入

导入敏感标签时请注意:

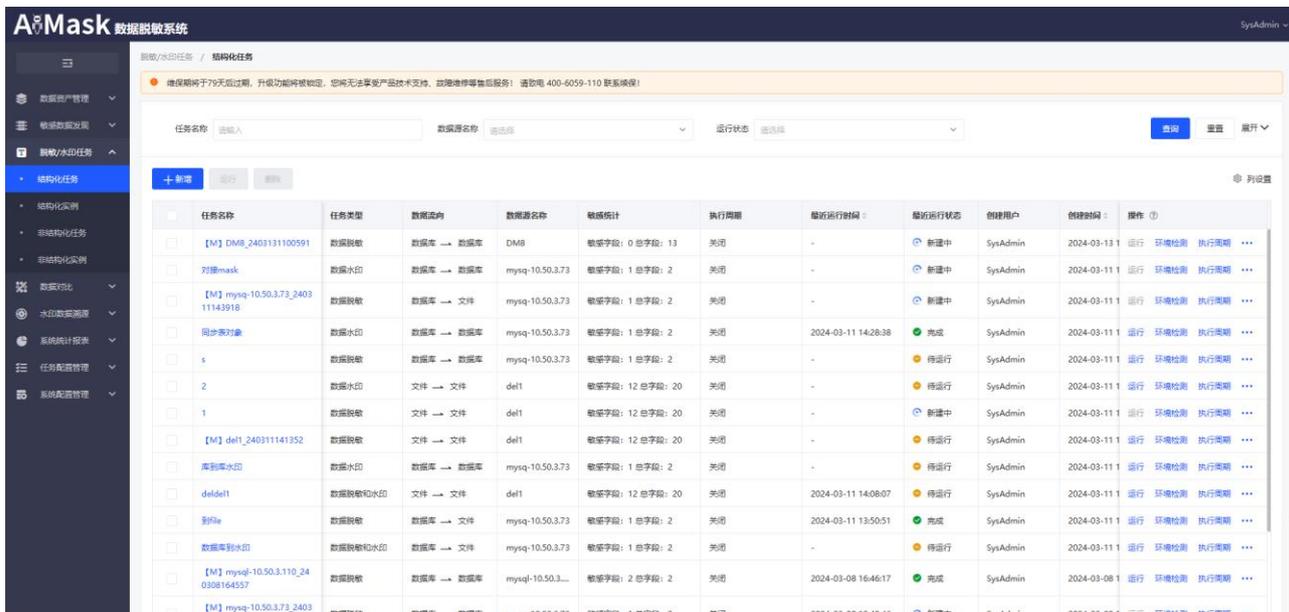
- ◆ 请在系统导出文件基础上修改, 支持 csv/xls/xlsx/txt 格式。
- ◆ 请勿删除或修改 id 列任何信息。若您需要新增敏感字段, 请在表格新增一行内容, 带\*列为必填内容, 对应的 id 信息必须为空。
- ◆ 若您需要编辑的敏感标签未出现在“任务配置管理”下的“敏感标签”页面列表中, 请提前在此页面新增该敏感标签。
- ◆ 若您删除了某些行, 导入本文件后系统不会对这些删除行的数据做任何修改, 保留原始记录。系统自动忽略非必填列数据的修改。



## 1.1.4 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。



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。

**AiMask 数据脱敏系统** SysAdmin

数据/水印任务 / 结构化任务 / 任务新增

● 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，故障维修等服务！ 请致电 400-6059-110 联系维保！

1 基础配置

2 脱敏规则

任务名称: [M] 10.50.3.135\_240314103954

任务类型:  数据脱敏  数据水印

数据流向: 数据库到数据库

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

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。  
 发现结果导入  
 文件导入

aisort

待选表格	已选0/25项	已选表格	0项
<input type="checkbox"/> aisort_car_info			
<input type="checkbox"/> aisort_car_info_copy3			
<input type="checkbox"/> aisort_car_info_copy4			
<input type="checkbox"/> aisort_car_info_copy5			
<input type="checkbox"/> aisort_new_fagui			
<input type="checkbox"/> aisort_new_fagui1			

**AiMask 数据脱敏系统** SysAdmin

数据/水印任务 / 结构化任务 / 任务新增

● 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，故障维修等服务！ 请致电 400-6059-110 联系维保！

1 基础配置

2 脱敏规则

任务名称: [M] 10.50.3.135\_240314103954

任务类型:  数据脱敏  数据水印

数据流向: 数据库到数据库

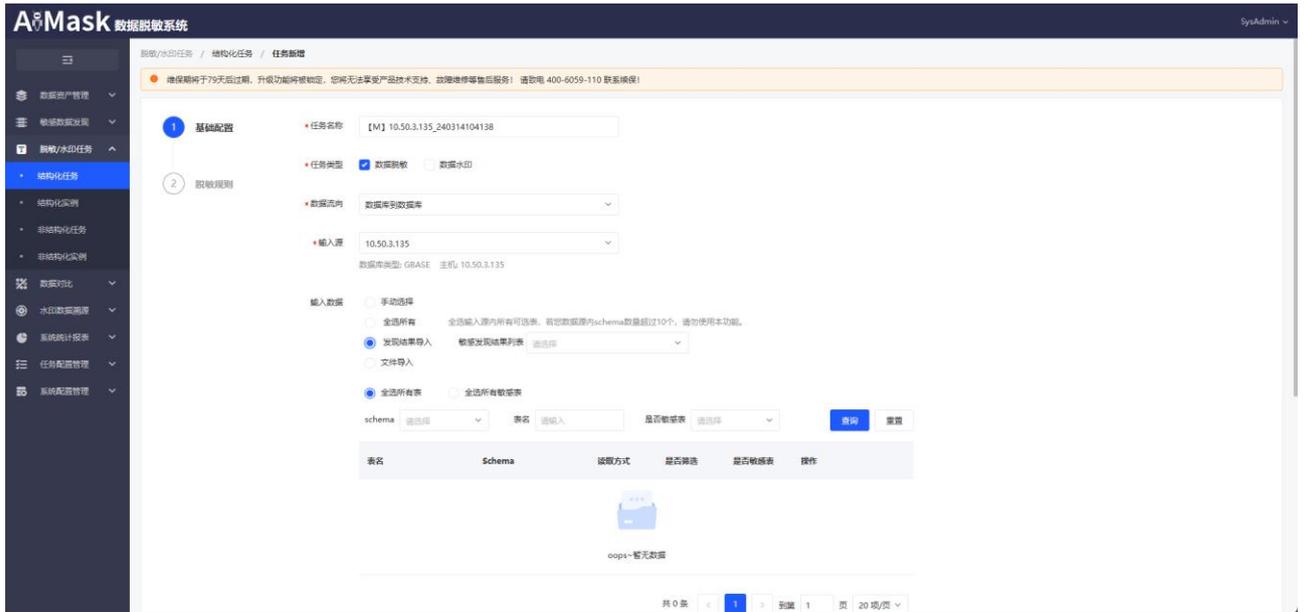
输入源: 10.50.3.135  
数据库类型: GBASE 主机: 10.50.3.135

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。  
 发现结果导入  
 文件导入

输出目标: pg-10.50.3.75out22  
数据库类型: POSTGRESQL 主机: 10.50.3.75  
已选择了不同类型数据库，请确认源和目标数据库所选表的字段类型必须相互兼容，否则脱敏任务将会失败！

Schema映射: 输入源已选schema名: aisort → 输出目标schema名: 请选择  
Schema映射会指定输入源已选schema经处理后输出至输出目标哪个schema的映射关系

表格映射:  继承输入源命名  
 附加前缀 ①  
 附加后缀 ①



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。

**AiMask 数据脱敏系统**

数据/水印任务 / 结构化任务

任务名称: 请输入 数据源名称: 请选择 运行状态: 请选择

+ 新增 运行 删除

任务名称	任务类型	数据流向	数据源名称	敏感估计	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	运行 环境检测 执行策略 ...
刘博mask	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	-	新建中	SysAdmin	2024-03-11	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	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	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	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	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	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	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	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	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	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	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	运行 环境检测 执行策略 ...
[M] mysql-10.50.3.73_240308104046	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	运行 环境检测 执行策略 ...

**AiMask 数据脱敏系统**

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

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 全量配置

共 1 条 < 1 > 到第 1 页 20 页/页

输出目标: pg-10.50.3.75out22

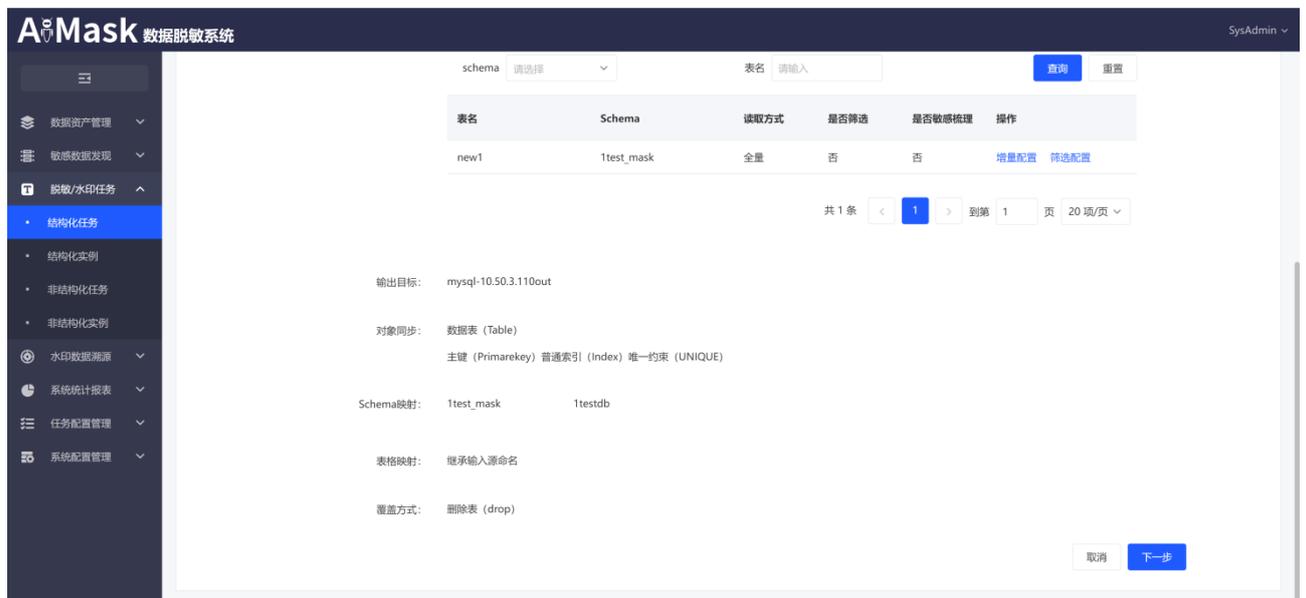
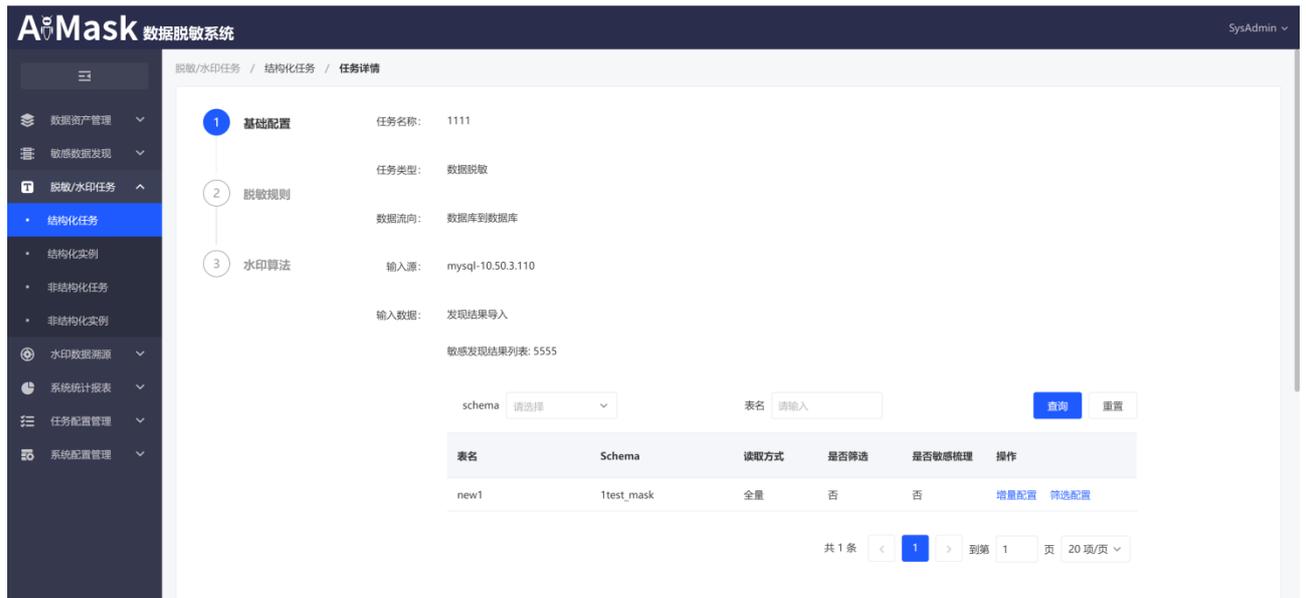
Schema映射: 输入源已选Schema名 AHDB → 输出目标Schema名 hctest

策略映射: 授予输入源命名

覆盖方式: 覆盖数据 (truncate)

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。

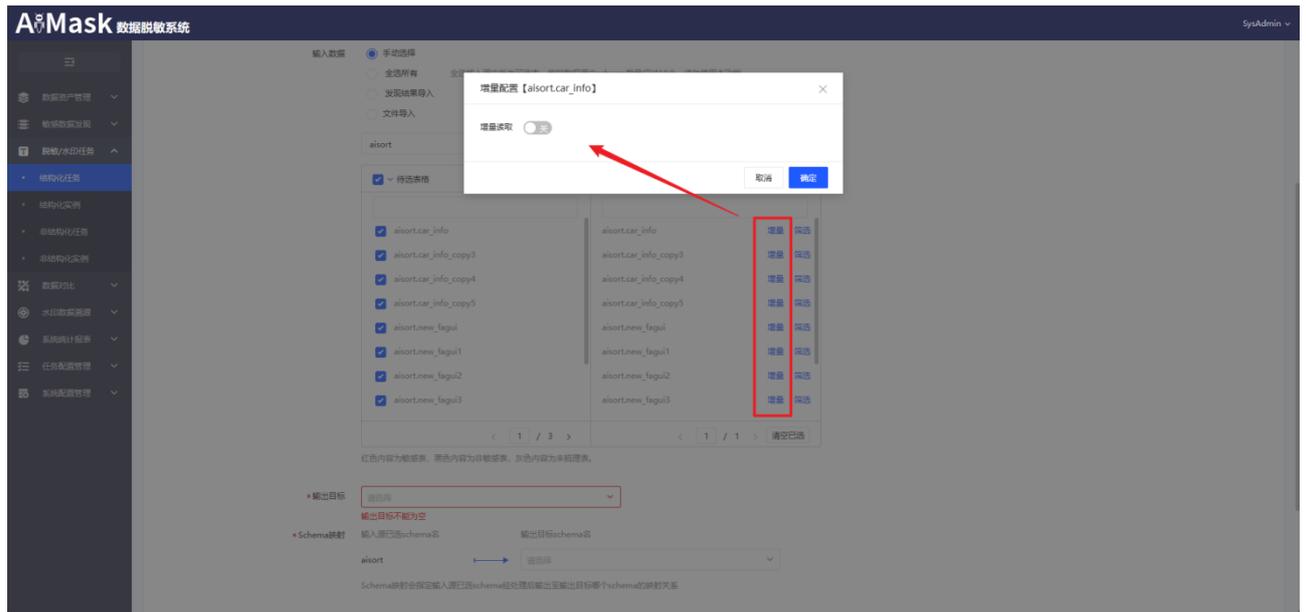


当数据流向为“数据库到数据库”时，用户可以自主设置是否启用“增量读取”，开启后可以以增量方式处理源表更新数据。启用该功能后，可进一步选择增量方式：

- ◆ “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。



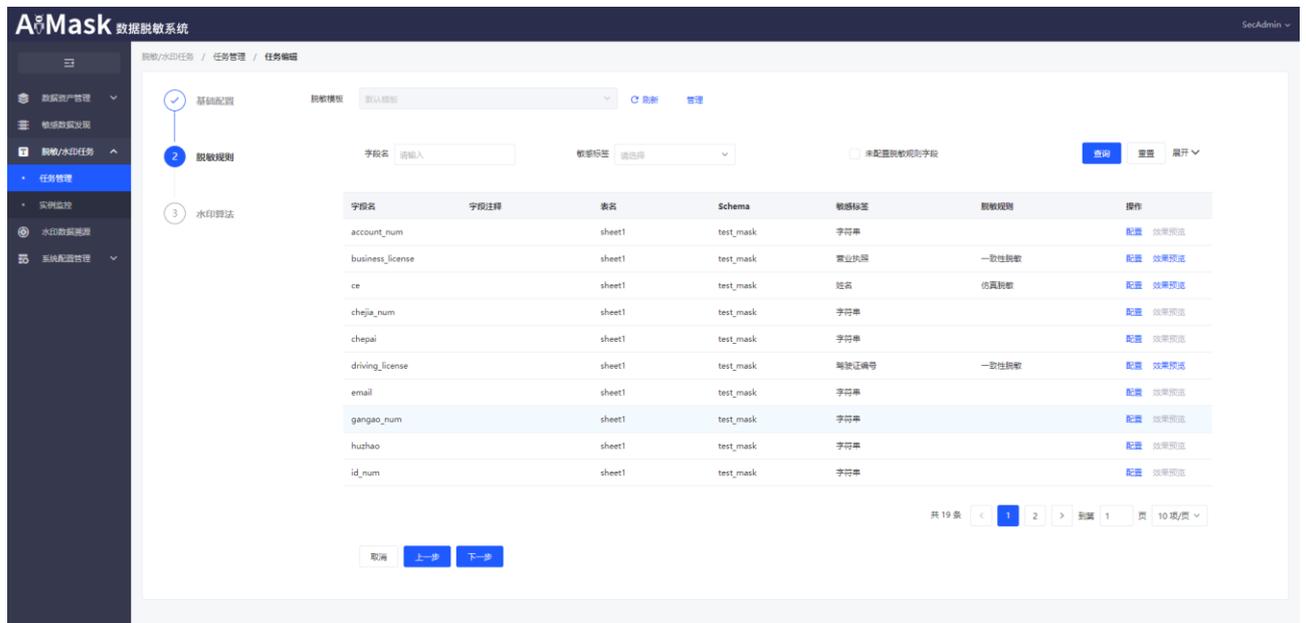
仅有少部分数据库类型支持更新增量。



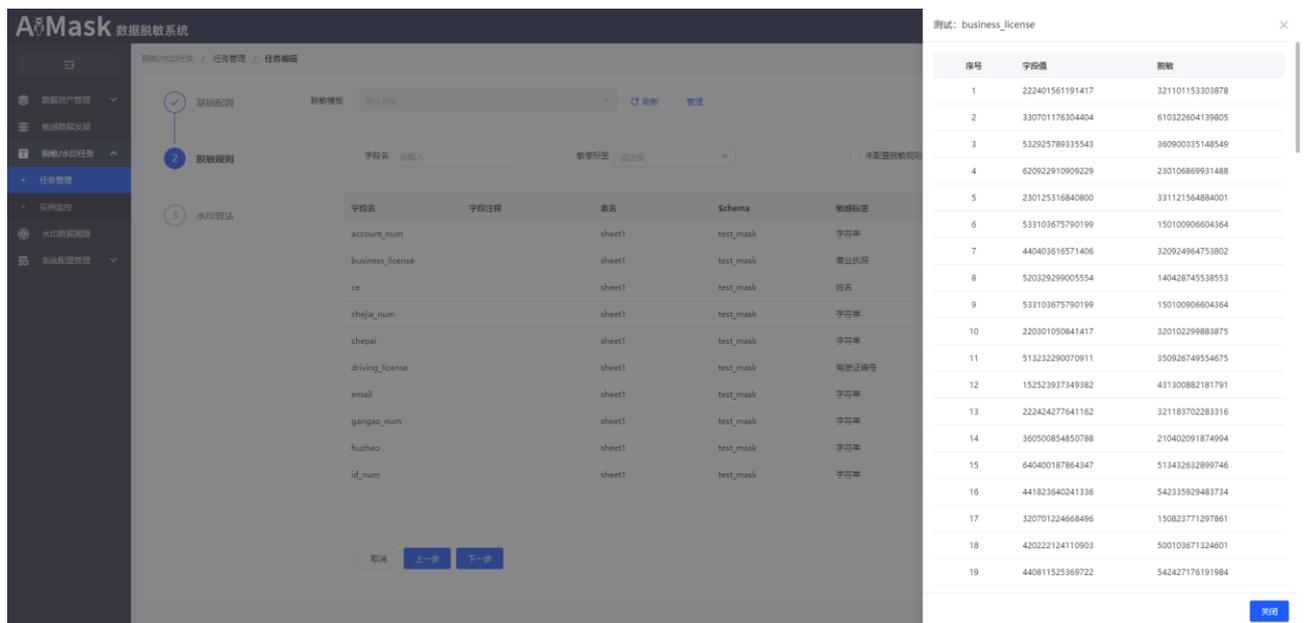
## 脱敏规则

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。

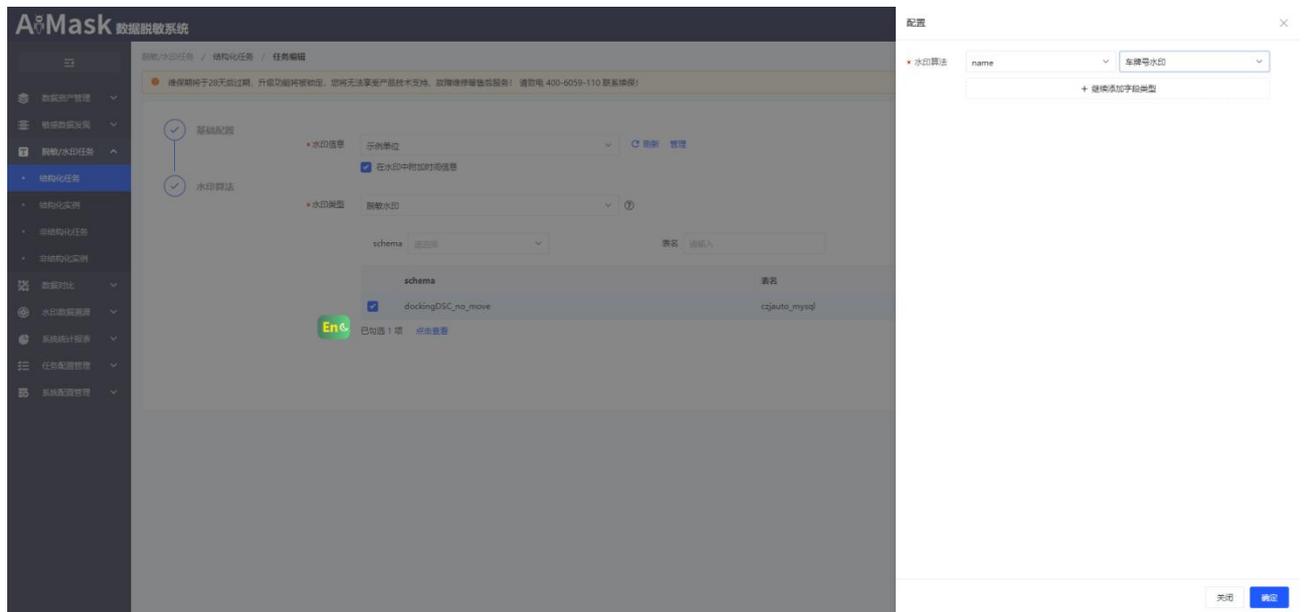


水印算法

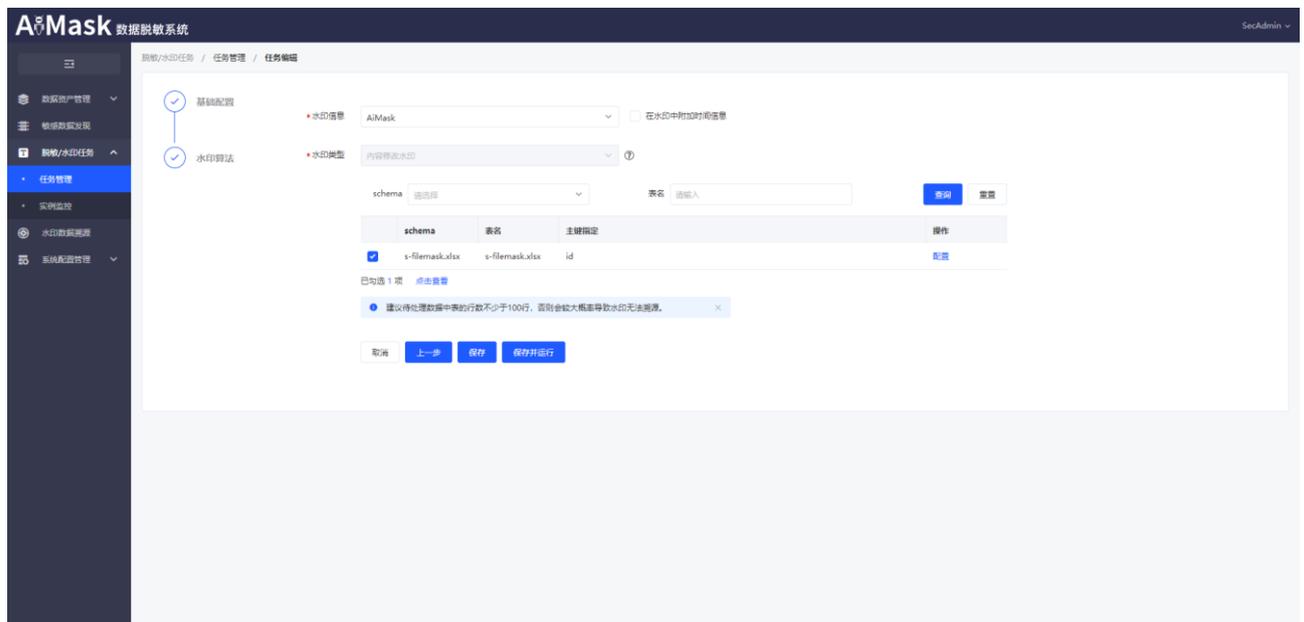
用户若在基础配置的任务类型中同时或单独勾选了“数据水印”，则提供“水印算法”配置页。

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。

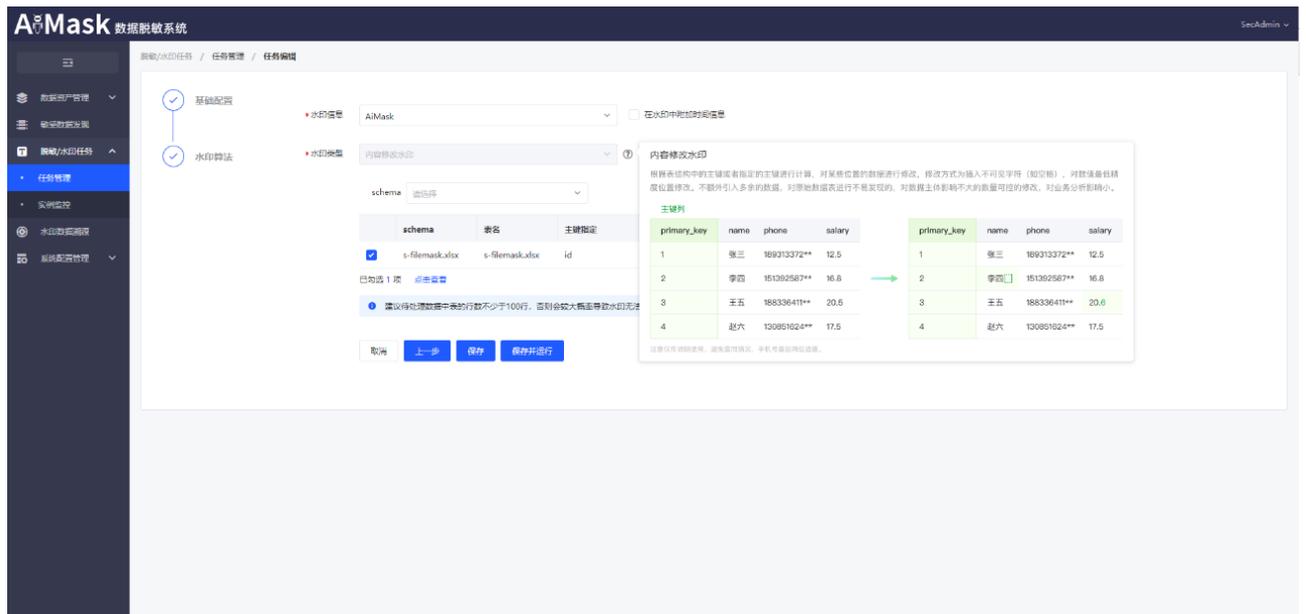
例如脱敏水印 点击配置 进入下面场景 我们选择了车牌号水印类型。然后点击右下角确定



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：

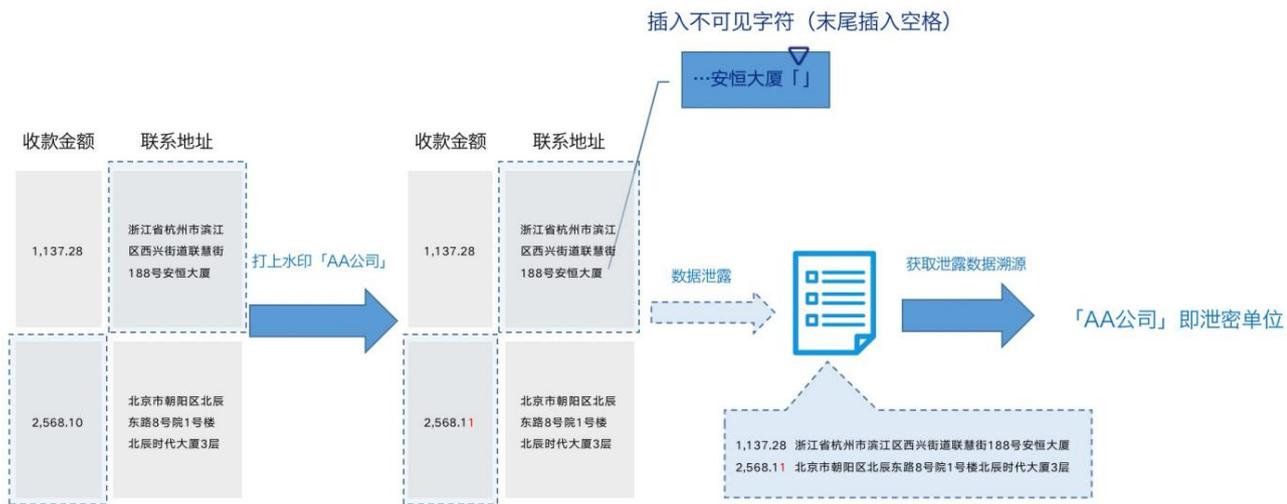
- ◆ 伪行/伪列水印：通过添加配置比例的数据行或数据列来插入水印信息，当数据泄露后，通过提取伪行/伪列水印信息来进行泄漏追溯。



- ◆ 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

● 维保即将于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:  任务名称:  运行状态:

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 ...
DW00000004624 0311140807	deldel1	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 ...
DW00000004624 0311140754	deldel1	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 14:07:54	完成	100%	运行 监控 报告 ...
DM00000004524 0311133050	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 ...
DM00000004324 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 ...
DM00000004224 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 ...
DM00000003724 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 ...
DM00000003624 0306182839	[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 ...
DM00000002624 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

● 维保即将于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

选择日志级别:

- 普通
- 告警
- 错误
- 任务启动
- 任务启动成功
- 错误

2024-03-11 14:28:39 [普通] 任务启动  
2024-03-11 14:28:39 [普通] 任务启动成功  
2024-03-11 14:28:39 [普通] 数据【dockingDSC\_no\_moveczjauto\_mysql】开始处理  
2024-03-11 14:28:39 [普通] 数据【dockingDSC\_no\_moveczjauto\_mysql】已处理数据1条  
2024-03-11 14:28:39 [普通] 数据【dockingDSC\_no\_moveczjauto\_mysql】已完成, 已处理数据1条  
2024-03-11 14:28:39 [普通] 任务运行完成

**A<sup>o</sup>Mask 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

已处理表数: 1, 剩余0 处理速度: 1行/s, 0.02M/s

运行日志 表格明细 对象同步

对象名:  状态:

<input type="checkbox"/>	表名	Schema	状态	错误日志	已处理行数	成功	失败	开始时间	耗时	操作
<input type="checkbox"/>	cjauto_mysql	dockingDSC_no_move	完成	无	1	1	0	2024-03-11 14:28:38	1秒	重跑 停止

共 1 条 < 1 > 到第 1 页 20 页/页

**A<sup>o</sup>Mask 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

已处理表数: 1, 剩余0 处理速度: 1行/s, 0.02M/s

运行日志 表格明细 对象同步

对象名:  状态:

<input type="checkbox"/>	对象名	对象类型	开始时间	耗时	状态	错误日志	操作
<input type="checkbox"/>	dockingDSC_no_move.cjauto_mysql同步对象	表格	2024-03-11 14:28:38	0秒	完成	无	重跑 停止

共 1 条 < 1 > 到第 1 页 20 页/页

点击<报告>将会展示实例对应的各维度分析，如下图所示。

**AiMask 数据脱敏系统** SysAdmin

数据/水印任务 / 结构化实例 / 运行报告

● 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！

## 脱敏任务运行报告

[导出](#)

### 任务运行概述

任务名称: 同步表对象      实例名称: WM0000005224031142837  
 任务创建人: 系统管理员 [SysAdmin]      任务创建时间: 2024-03-11 14:26:25  
 处理表数: 1      处理字段数: 2      敏感字段数: 0  
 开始运行时间: 2024-03-11 14:28:38      结束运行时间: 2024-03-11 14:28:39      运行耗时: 1  
 处理数据量: 1行/0.02M      平均处理速度: 1行/s, 0.02M/s

### 数据处理明细

表名	Schema	处理行数	成功	失败	开始时间	结束时间	耗时	结果
cjauto_mysql	dockingDSC_no_move	1	1	0	2024-03-11 14:28:38	2024-03-11 14:28:39	1秒	完成

共 1 条 < 1 > 到第 1 页 20 页/页

### 对象同步明细

**AiMask 数据脱敏系统** SysAdmin

数据/水印任务 / 结构化实例 / 运行报告

任务名称: 同步表对象      实例名称: WM0000005224031142837  
 任务创建人: 系统管理员 [SysAdmin]      任务创建时间: 2024-03-11 14:26:25  
 处理表数: 1      处理字段数: 2      敏感字段数: 0  
 开始运行时间: 2024-03-11 14:28:38      结束运行时间: 2024-03-11 14:28:39      运行耗时: 1  
 处理数据量: 1行/0.02M      平均处理速度: 1行/s, 0.02M/s

### 数据处理明细

表名	Schema	处理行数	成功	失败	开始时间	结束时间	耗时	结果
cjauto_mysql	dockingDSC_no_move	1	1	0	2024-03-11 14:28:38	2024-03-11 14:28:39	1秒	完成

共 1 条 < 1 > 到第 1 页 20 页/页

### 对象同步明细

对象名	对象类型	开始时间	结束时间	耗时	结果
dockingDSC_no_move.cjauto_mysql同步对象	表 (Table)	2024-03-11 14:28:38	2024-03-11 14:28:38	0秒	完成

共 1 条 < 1 > 到第 1 页 20 页/页

### 任务运行日志

时间	级别	日志
2024-03-11 14:28:38	普通	任务启动
2024-03-11 14:28:38	普通	任务启动成功
2024-03-11 14:28:38	普通	表格【dockingDSC_no_move.cjauto_mysql】开始处理
2024-03-11 14:28:39	普通	表格【dockingDSC_no_move.cjauto_mysql】已处理数据1条
2024-03-11 14:28:39	普通	表格【dockingDSC_no_move.cjauto_mysql】已完成,已处理数据1条
2024-03-11 14:28:39	普通	任务运行完成

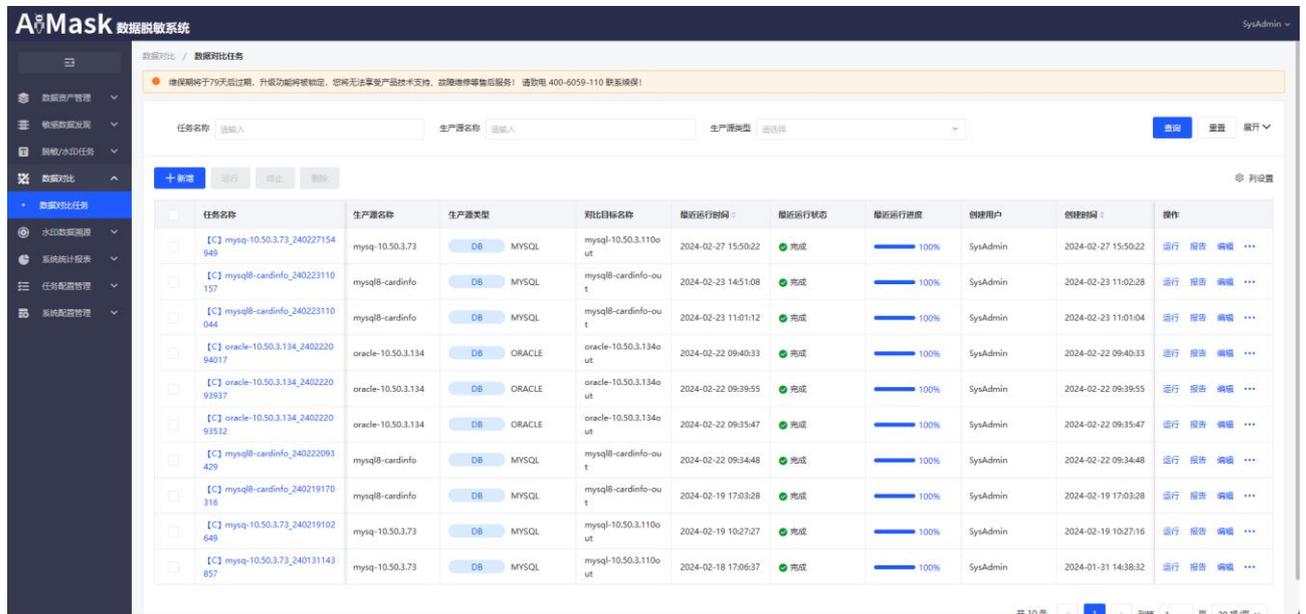
共 6 条 < 1 > 到第 1 页 20 页/页

点击下页面所示的<导出>可以下载 pdf 或 html 格式的任务运行报告。

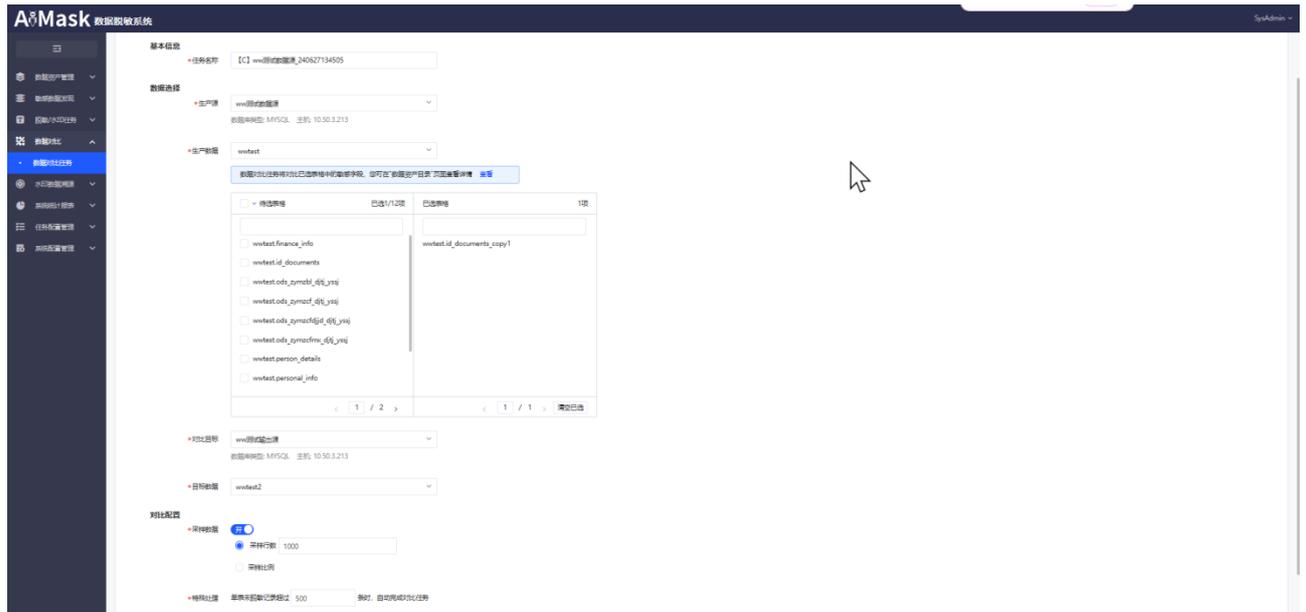


## 1.1.5 数据对比任务

'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。



点击<报告>将会展示任务对应的各维度分析，如下图所示。

## 数据对比任务报告

- 任务运行概述
- 对比结果统计
- 未脱敏字段明细
- 对比结果示例

### 任务运行概述

任务名称: [C] bidui\_240320145757 生产源名称: bidui 生产数据: mysql\_1  
对比目标名称: bidui2 目标数据: mysql\_2  
对比配置: 采样1000行, 由于单表未脱敏记录超过500条, 自动完成对比任务。  
任务创建时间: 2024-03-20 14:58:01 开始运行时间: 2024-03-20 15:02:59 结束运行时间: 2024-03-20 15:02:59  
任务创建人: SysAdmin 运行耗时: 0秒

### 对比结果统计

数据表	未脱敏表	对比记录	未脱敏记录	对比再算表
1 张	0 张	500 行	0 行	0 张

表名	对比记录数	未脱敏记录数	对比结果	日志
sheet1	500	0	已脱敏	-

共 1 条 < 1 > 到第 1 页 20 项/页

### 未脱敏字段明细

schema	表名	未脱敏字段
mysql_1	sheet1	-

共 1 条 < 1 > 到第 1 页 20 项/页

### 对比结果示例

sheet1 表内对比记录共 500 行, 其中未脱敏记录 0 行, 下表仅展示部分数据

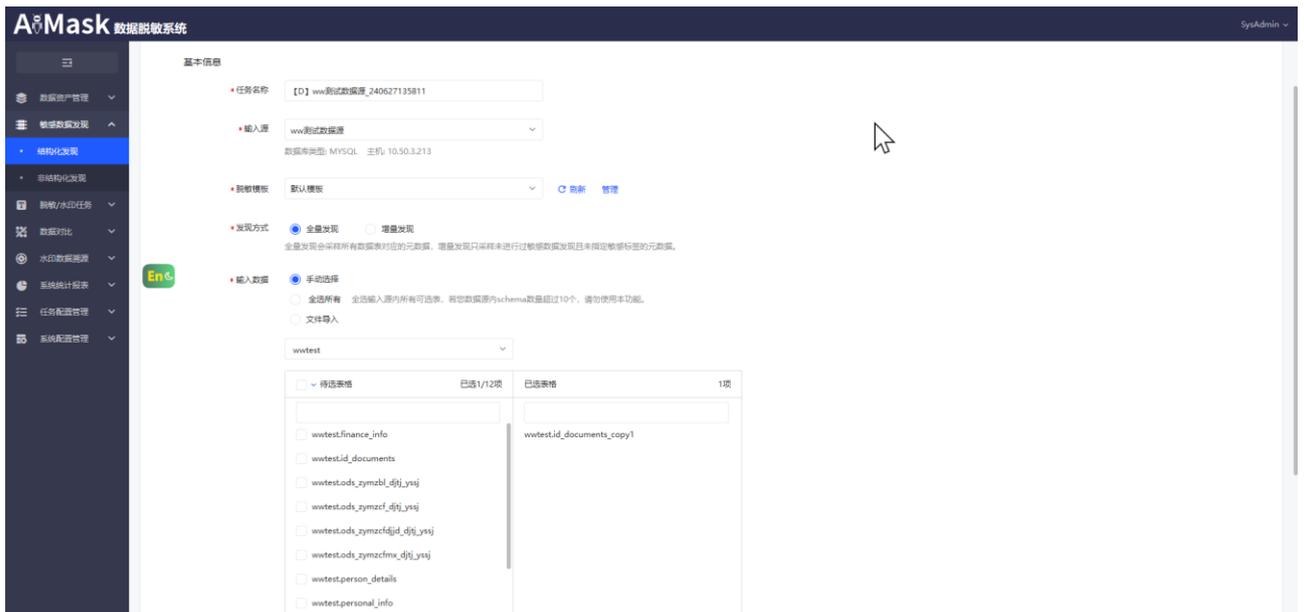
注: 红色灰色代表敏感字段

示例		chepai	meid	gangao_num	nashui_num	chejia_num	huzhao	zuzhijigou_num	business_licens e	mac	junguang_num	ac
1	生产数据	津W50330	C3595912911...	CH1642066	610202216200...	RYK9E99U2D4...	S83928869	344P6074-7	210782207468...	FC-4B-AB-AF-...	沈字第0491222	54
	目标数据	津W50330	C3595912911...	C86514302	220802764434...	RYK9E99U2D4...	S85488606	521N2368-X	510823712097...	FC-4B-AB-AF-...	空字第01192248	62
2	生产数据	渝Z22950	AA951010336...	CU6872943	231181135291...	WRL789600D...	ED1199482	927D4879-X	511322633446...	D9-67-87-94-...	成字第0486518...	62
	目标数据	渝Z22950	AA951010336...	W99324097	331121686403...	WRL789600D...	ED6659874	872P0666-4	130802987080...	D9-67-87-94-...	腾字第2202730...	62
3	生产数据	云N96500	CD883947691...	CT5868091	653024215291...	EUF5G99N1D4...	P91899210	56806046-2	210381269694...	CA-A7-36-05-...	广字第1734322...	62
	目标数据	云N96500	CD883947691...	W81348313	532800766403...	EUF5G99N1D4...	P96559131	150C2392-X	510623755209...	CA-A7-36-05-...	沈字第6070424...	62
4	生产数据	晋W27270	AA893575878...	W21442139	340601118711...	YYM8F848D...	H62119372	327T8833-8	120109348981...	31-29-56-58-F...	参字第0834415...	62
	目标数据	晋W27270	AA893575878...	CT6214256	620721660623...	YYM8F848D...	H67639464	572A8657-4	610625520963...	31-29-56-58-F...	兰字第2370835...	62
5	生产数据	鄂S93300	C2190331190...	W86685760	150105288496...	JAD9S86V2D4...	K40581226	816D5483-9	130706460234...	73-E8-4D-DA-...	参字第86203050	62
	目标数据	鄂S93300	C2190331190...	CZ9507501	431281730000...	JAD9S86V2D4...	K42903122	064P7877-4	411601459449...	73-E8-4D-DA-...	兰字第05854450	62
6	生产数据	赣P22880	DE308080283...	W06847627	510683125178...	SRR4M96K1D...	G12298203	347A4332-4	451024385011...	CE-E1-88-80-6...	北字第5296735...	62
	目标数据	赣P22880	DE308080283...	CR9316929	350424676357...	SRR4M96K1D...	G17758147	522M0454-3	341021536523...	CE-E1-88-80-6...	海字第1752555...	55
7	生产数据	闽A24040	E12358776685...	C82174194	130302206192...	HAN3J8383D4...	S02293130	441F6679-7	540100058712...	D6-01-DE-CE-...	京字第2419998...	94
	目标数据	闽A24040	E12358776685...	CD7620318	411320714202...	HAN3J8383D4...	CD7761361	432C3066-0	370385300633...	D6-01-DE-CE...	浙字第7320010...	62

共 50 条 < 1 2 3 4 5 > 到第 1 页 10 项/页

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：

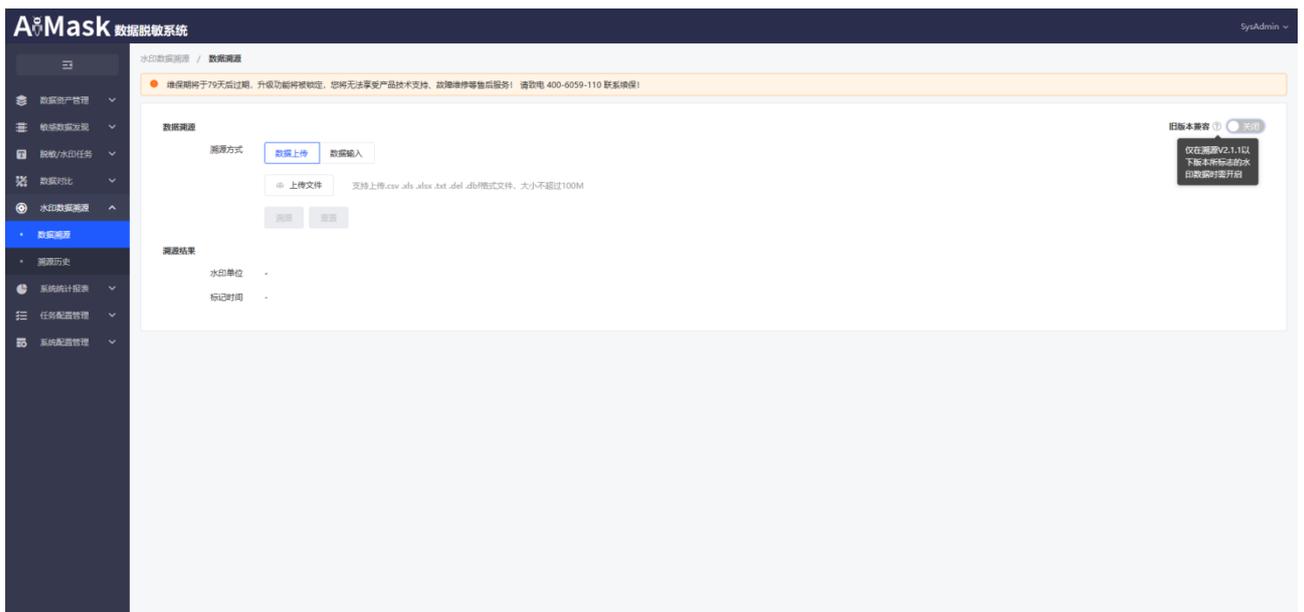
- ◆ 点击报告末尾的<一键创建敏感数据发现任务>，界面将跳转至敏感数据发现的任务创建页，自动生成一个结构化发现任务，并自动填入输入源（对应“生产源”）、输入数据、任务名称等参数；
- ◆ 或点击<一键创建脱敏任务>，界面将跳转至脱敏/水印任务的任务创建页，自动生成一个脱敏任务，并自动填入输入源、输入数据、输出目标（对应“对比目标”）、任务名称等参数。

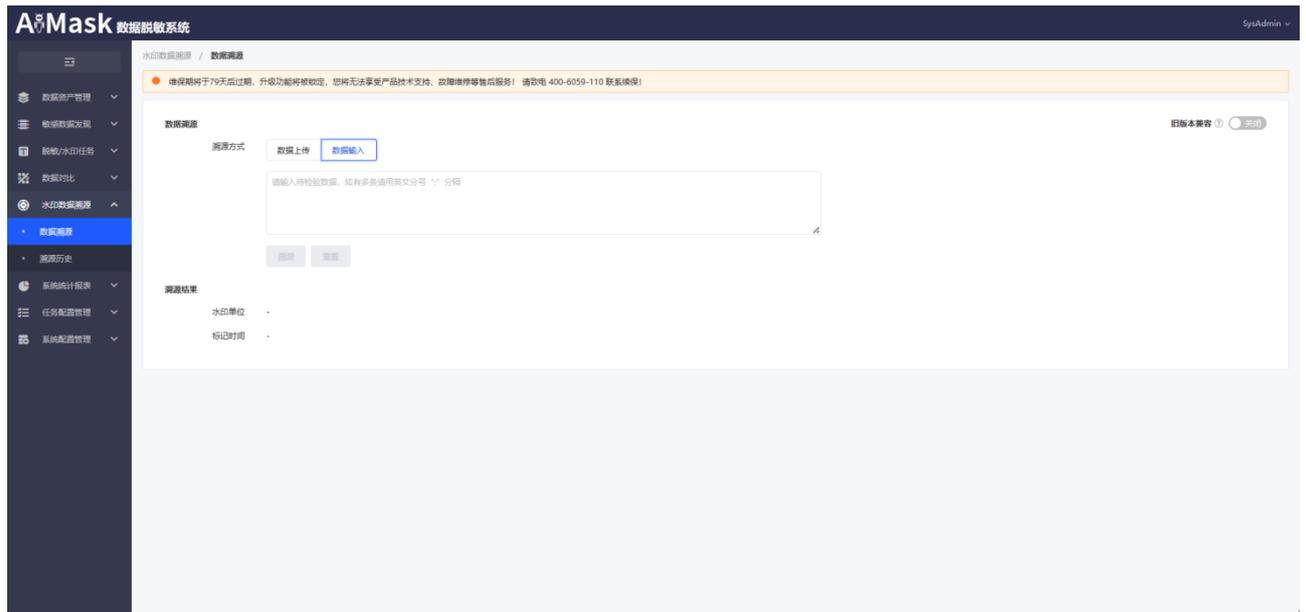


## 1.1.6 水印数据溯源

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）。

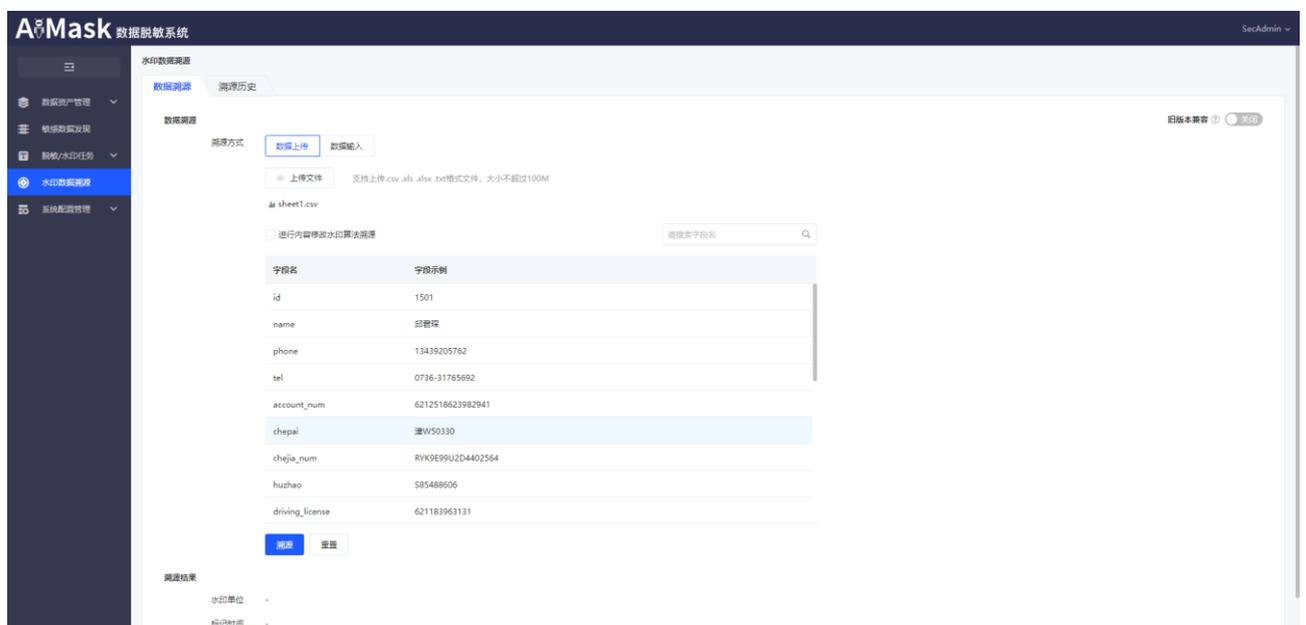
- ◆ 点击<数据输入>后展示的输入框可输入多个数据（长度没有限制），如有多条数据请用英文分号“;”分隔。
- ◆ 点击<数据上传>后可以上传 csv、txt、xls、xlsx、del、dbf 六种文件格式的数据进行溯源。



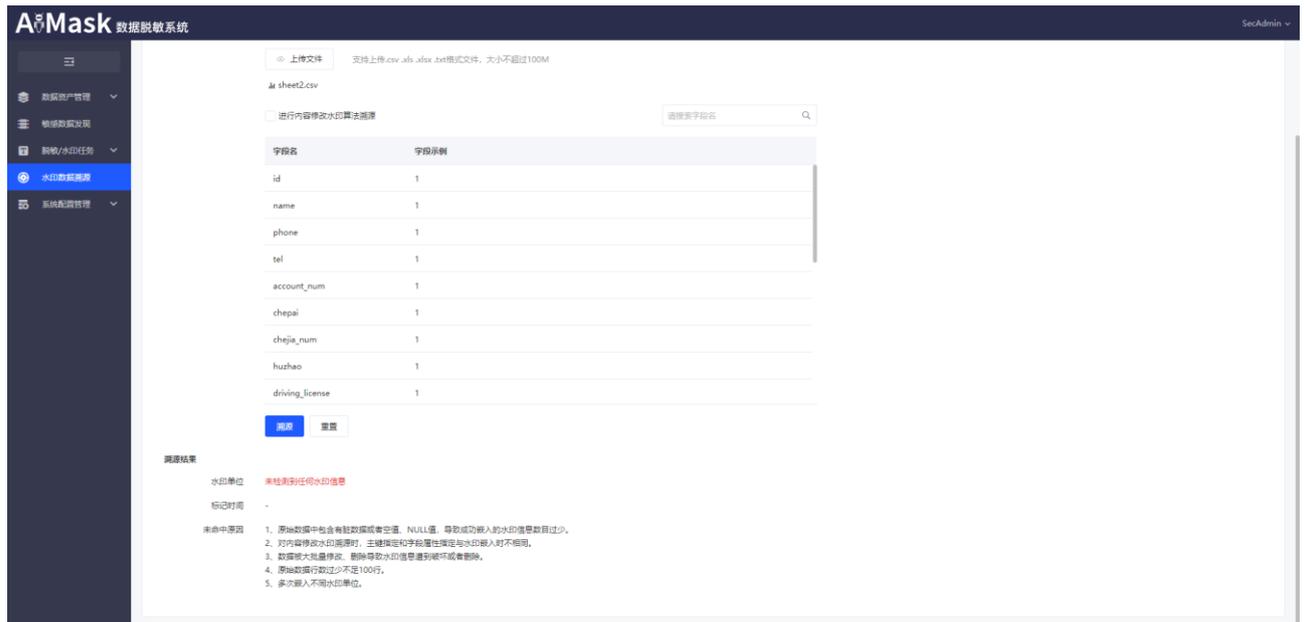


- ◆ 内容修改水印必须指定主键和字符类型，才能溯源到相应结水印单位。如溯源数据为 2.1.1 版本以下打上的水印，需要打开右上角的旧版本兼容开关。
- ◆ 溯源包含水印的数据请尽量避免重复。

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



## 1.1.7 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。

## 1.1.8 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。

**AiMask 数据脱敏系统** SysAdmin

任务配置管理 / 敏感标签

● 维保期将于79天后过期，升级功能即将被锁定，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！

敏感标签 字典管理

敏感标签 请输入 有无发现规则 请选择 有无脱敏规则 请选择 查询 重置 展开

+ 新增 删除

敏感标签	来源	敏感发现规则	脱敏规则数量	更新时间	创建用户	创建时间	操作
ZH+测试	自定义	无	0	2024-03-01 11:26:33	SysAdmin	2024-03-01 11:26:33	发现规则配置 脱敏规则配置 删除
ces	自定义	无	0	2024-02-23 18:05:47	SysAdmin	2024-02-23 18:05:47	发现规则配置 脱敏规则配置 删除
79	自定义	无	2	2024-02-23 11:07:11	SysAdmin	2024-02-23 11:07:11	发现规则配置 脱敏规则配置 删除
sddd	自定义	无	0	2024-01-31 16:25:05	SysAdmin	2024-01-31 16:25:05	发现规则配置 脱敏规则配置 删除
ahaha	自定义	无	0	2024-01-29 19:02:37	SysAdmin	2024-01-29 19:02:37	发现规则配置 脱敏规则配置 删除
测试1	自定义	有	1	2024-01-19 14:36:48	SysAdmin	2024-01-19 14:36:48	发现规则配置 脱敏规则配置 删除
姓名	内置	有	4	2023-02-24 10:57:34	SecAdmin	2023-02-24 10:57:34	发现规则配置 脱敏规则配置 删除
手机号	内置	有	3	2023-02-24 10:57:34	SecAdmin	2023-02-24 10:57:34	发现规则配置 脱敏规则配置 删除
身份证	内置	有	3	2023-02-24 10:57:34	SecAdmin	2023-02-24 10:57:34	发现规则配置 脱敏规则配置 删除
银行卡	内置	有	3	2023-02-24 10:57:34	SecAdmin	2023-02-24 10:57:34	发现规则配置 脱敏规则配置 删除
军官证	内置	有	3	2023-02-24 10:57:34	SecAdmin	2023-02-24 10:57:34	发现规则配置 脱敏规则配置 删除
邮政编码	内置	有	2	2023-02-24 10:57:34	SecAdmin	2023-02-24 10:57:34	发现规则配置 脱敏规则配置 删除

共 55 条 1 2 3 到第 1 页 20 页/页

**AiMask 数据脱敏系统** SysAdmin

任务配置管理 / 敏感标签

● 维保期将于79天后过期，升级功能即将被锁定，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！

敏感标签 字典管理

敏感标签 请输入 有无发现规则 请选择 有无脱敏规则 请选择 查询 重置 展开

+ 新增 删除

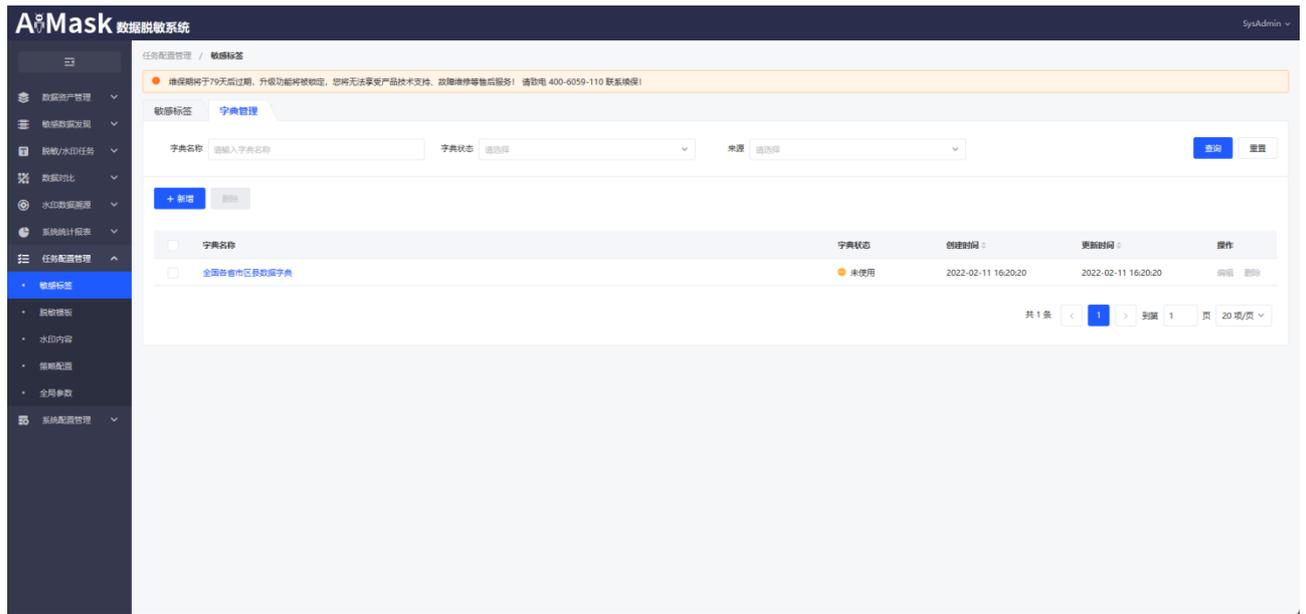
敏感标签	来源	敏感发现规则	脱敏规则数量	更新时间	创建用户	创建时间	操作
ZH+测试	自定义	无	0	2024-03-01 11:26:33	SysAdmin	2024-03-01 11:26:33	发现规则配置 脱敏规则配置 删除
ces	自定义	无	0	2024-02-23 18:05:47	SysAdmin	2024-02-23 18:05:47	发现规则配置 脱敏规则配置 删除
79	自定义	无	2	2024-02-23 11:07:11	SysAdmin	2024-02-23 11:07:11	发现规则配置 脱敏规则配置 删除
sddd	自定义	无	0	2024-01-31 16:25:05	SysAdmin	2024-01-31 16:25:05	发现规则配置 脱敏规则配置 删除
ahaha	自定义	无	0	2024-01-29 19:02:37	SysAdmin	2024-01-29 19:02:37	发现规则配置 脱敏规则配置 删除
测试1	自定义	有	1	2024-01-19 14:36:48	SysAdmin	2024-01-19 14:36:48	发现规则配置 脱敏规则配置 删除
姓名	内置	有	4	2023-02-24 10:57:34	SecAdmin	2023-02-24 10:57:34	发现规则配置 脱敏规则配置 删除
手机号	内置	有	3	2023-02-24 10:57:34	SecAdmin	2023-02-24 10:57:34	发现规则配置 脱敏规则配置 删除
身份证	内置	有	3	2023-02-24 10:57:34	SecAdmin	2023-02-24 10:57:34	发现规则配置 脱敏规则配置 删除
银行卡	内置	有	3	2023-02-24 10:57:34	SecAdmin	2023-02-24 10:57:34	发现规则配置 脱敏规则配置 删除
军官证	内置	有	3	2023-02-24 10:57:34	SecAdmin	2023-02-24 10:57:34	发现规则配置 脱敏规则配置 删除
邮政编码	内置	有	2	2023-02-24 10:57:34	SecAdmin	2023-02-24 10:57:34	发现规则配置 脱敏规则配置 删除

查看内置敏感规则

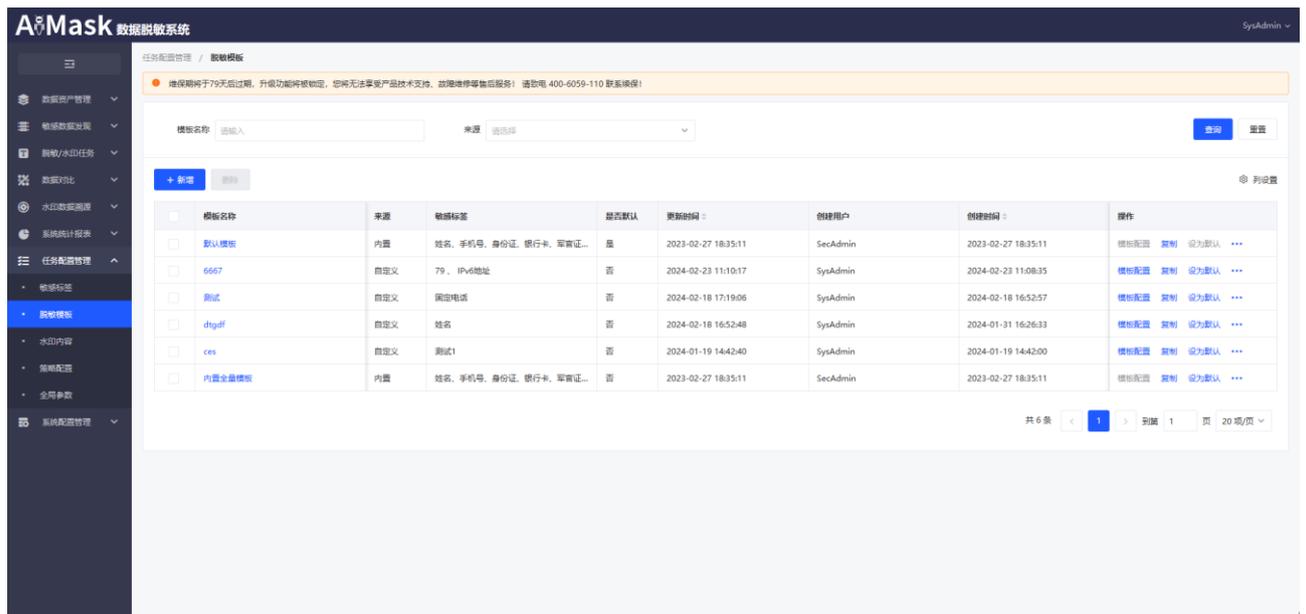
序号	内置敏感规则
7	仿真脱敏(身份证)
8	"出生日期" 遮蔽脱敏(身份证)
9	一致性脱敏(身份证)
10	仿真脱敏(银行卡)
11	"自定义部分" 遮蔽脱敏(银行卡)
12	一致性脱敏(银行卡)
13	仿真脱敏(军官证)
14	"编号" 遮蔽脱敏(军官证)
15	一致性脱敏(军官证)
16	仿真脱敏(邮政编码)
17	一致性脱敏(邮政编码)
18	仿真脱敏(详细地址)
19	"详细信" 遮蔽脱敏(详细地址)
20	仿真脱敏(护照号)
21	一致性脱敏(护照号)

共 55 条 1 2 3 到第 1 页 20 页/页

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。

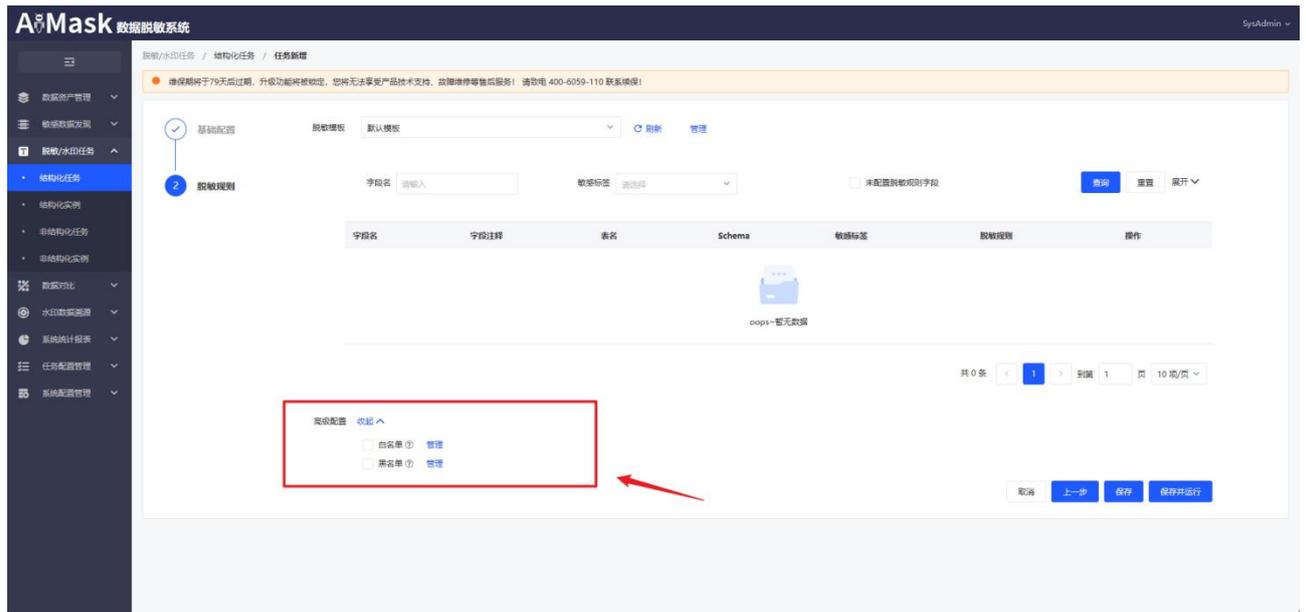
算法名称	说明
可逆算法	根据所选的参数（若有）生成格式类似、符合校验规则（若有）且支持可逆计算的数据。
遮蔽算法	根据所选的参数（若有）对指定区域的数据内容进行替换。
水印算法	根据所选的水印信息，对数据添加水印标记，可实现数据所有权声明和数据泄漏追溯。
取整算法	根据所选的参数（若有）对数据进行取整计算。
分割算法	根据所选的参数（若有）对数据进行分割，保留指定区域的数据。
保留算法	保留原数据不做脱敏。
替换算法	使用固定值替换，注意替换的值应与原数据所在列声明的数据结构保持一致，否则会出现因数据结构不符而造成脱敏失败的情况。
置空算法	使用空值替换，注意需确认若原数据所在列要求为非空，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

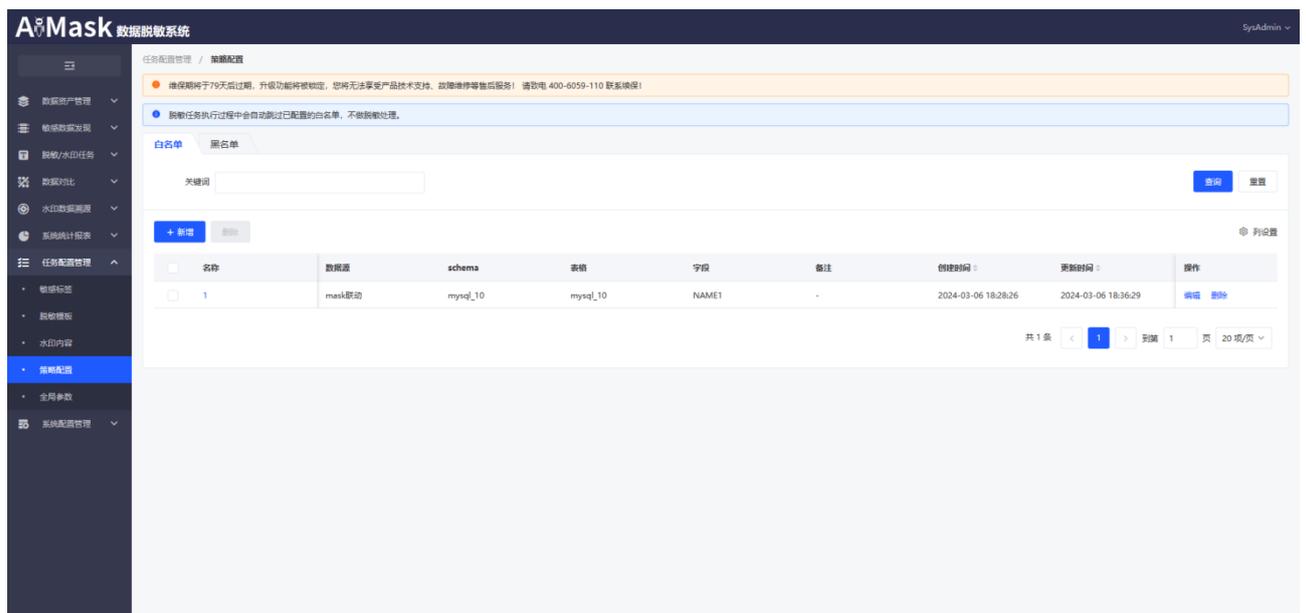
算法名称	说明
法	的表、列中在脱敏后仍保持一致。
计算关系保留算法	指定的列在脱敏后保留减法运算关系。可指定减数或被减数是否保留原值，否则会按照仿真算法重新生成。
乱序关联保留算法	指定的 2 个或多个列（支持至多 5 列）在乱序前后保留对应关系。
身份信息关联保留算法	出生日期/性别/身份证互相之间关系（根据需要）保持一致。

### 策略配置

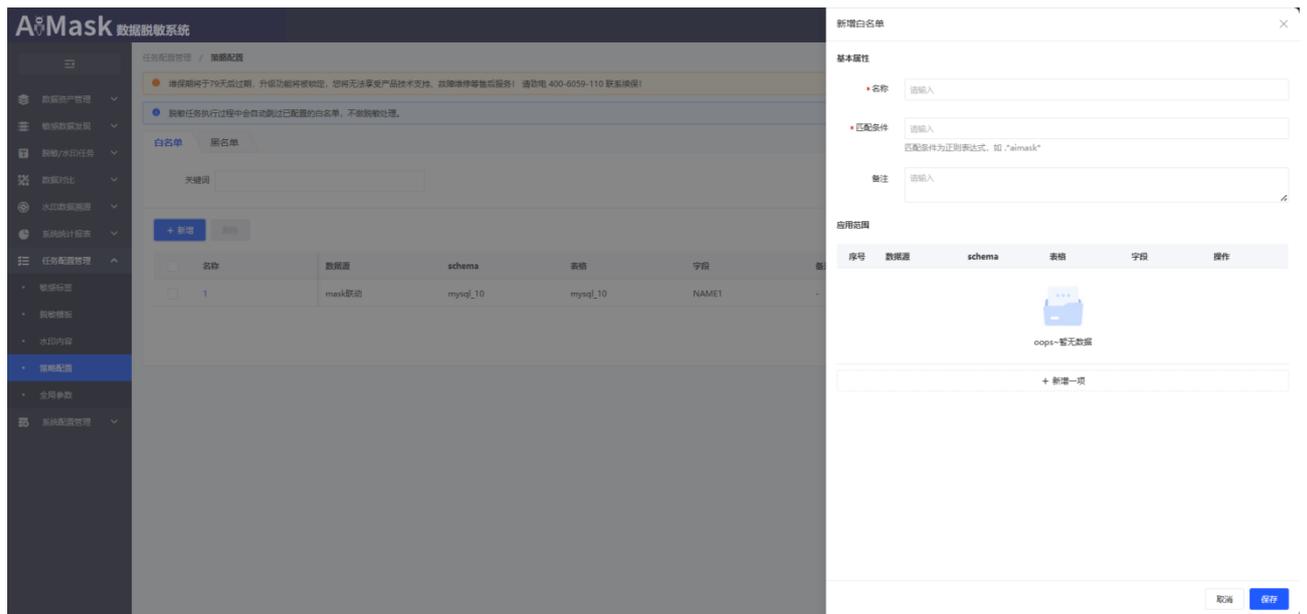
本页面针对结构化脱敏任务提供黑、白名单功能。用户若需使用该功能，可以在配置结构化任务时展开高级配置，勾选对应选项即可。若用户第一次配置结构化任务时未提前新增所需黑白名单，可以点击对应选项右侧的<管理>按钮打开新的策略配置页面。



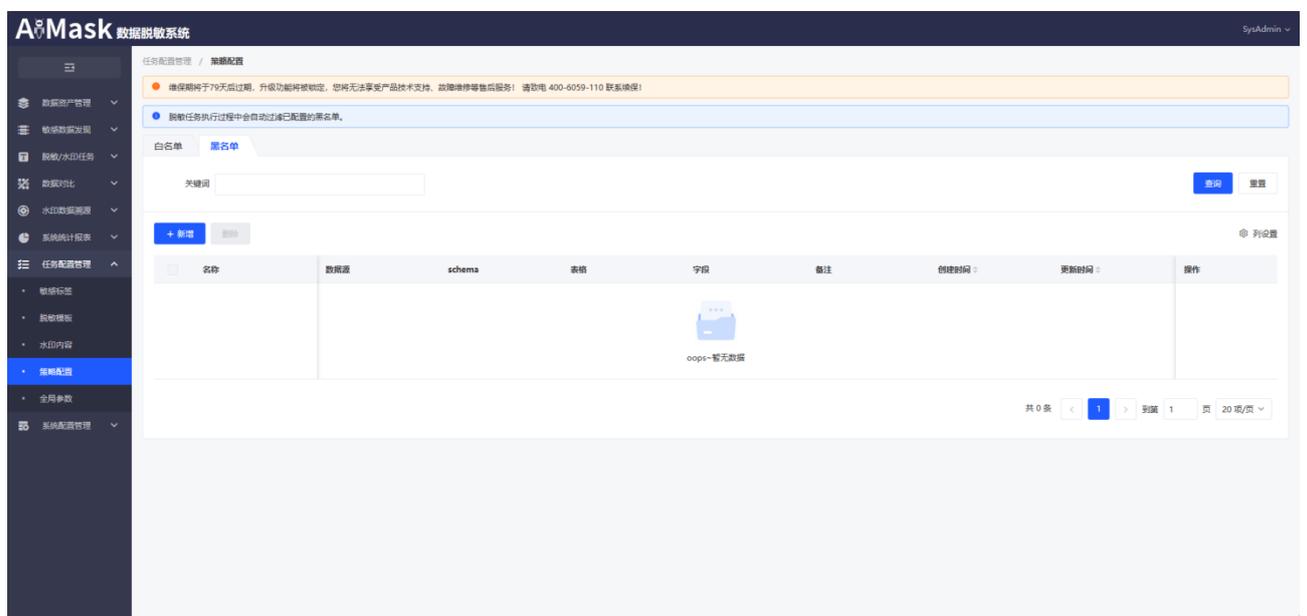
其中，白名单机制的主要目的是为了与未经过脱敏处理的业务系统联调或第三方进行数据对接。所有“白名单”上的数据在脱敏前自动过滤掉。管理页面如下图所示。



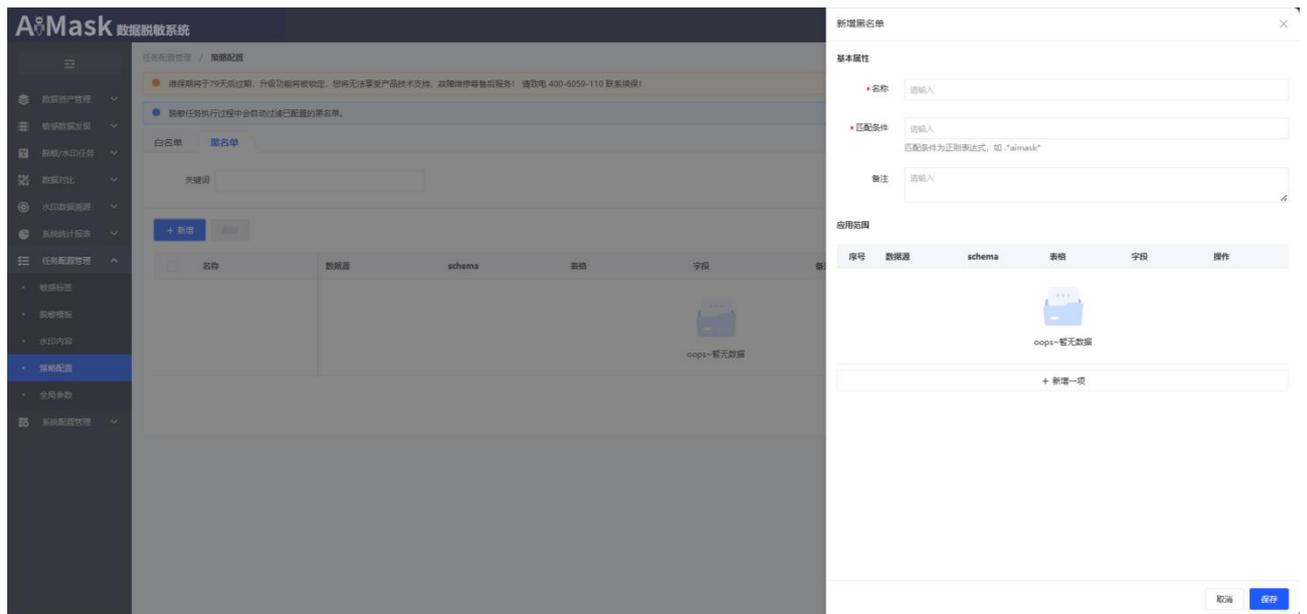
点击<新增>后，用户可以在下页面配置白名单策略的具体信息。



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企业机密数据不被泄漏。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。

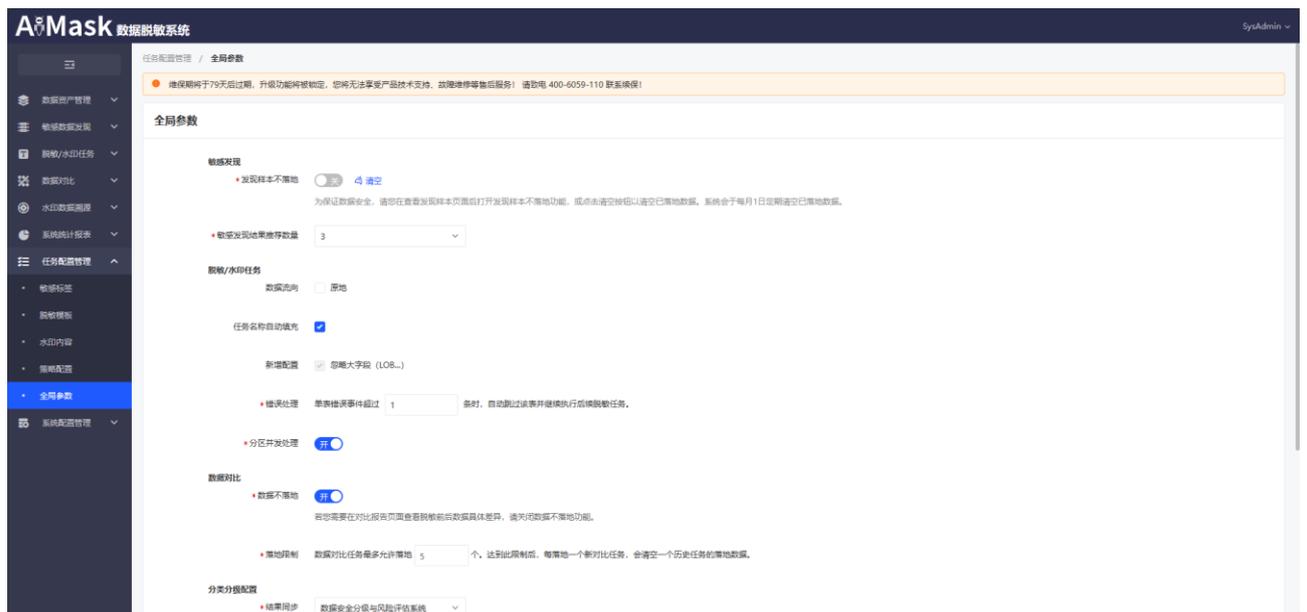


点击<新增>后，用户可以在下页面配置黑名单策略的具体信息。



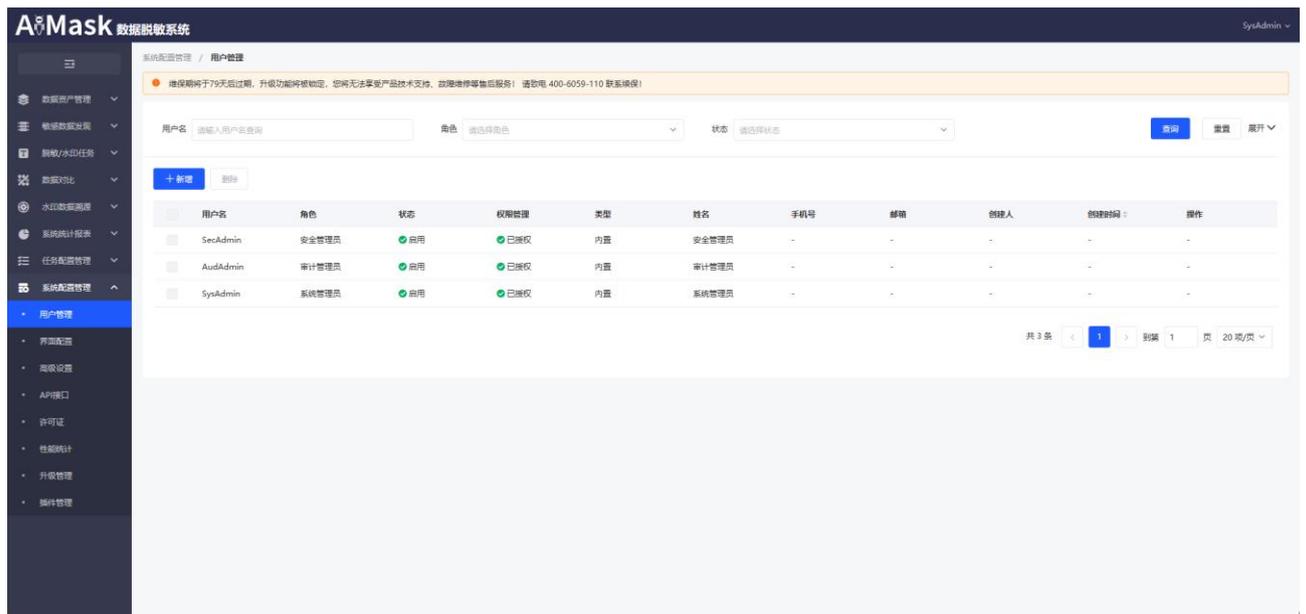
## 全局参数

本页面可配置部分作用于系统全局的全局参数。



## 1.1.9 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减的配置。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>。

