



# 昂楷数据脱敏系统 用户手册

文档版本：V3.0

发布日期：2019/9/12

---

## 版权声明

昂楷数据脱敏系统“ AAS-DM”是由深圳昂楷科技有限公司自主研发的专用信息安全产品，其版权受中华人民共和国版权法保护。

深圳昂楷科技有限公司拥有本手册的全部版权，未经本公司许可，任何单位及个人不得对本手册中的任何部分进行转印、影印或复印。

## 信息反馈

深圳昂楷科技有限公司尽最大的努力保证本手册的准确性和完整性。如您在使用过程中发现任何问题，希望及时将情况反馈给我们以完善产品，我们将非常感谢您的支持。

# 目录

1、产品概述.....	5
1.1 产品特点.....	5
1.1.1 敏感数据自动发现.....	5
1.1.2 丰富的脱敏算法.....	5
1.1.3 全面兼容性.....	5
1.1.4 脱敏有效性验证.....	6
1.1.5 敏感数据字典管理.....	6
1.1.6 产品部署方式.....	6
1.2 主要功能.....	6
1.2.1 监控墙.....	7
1.2.2 日志管理.....	7
1.2.3 用户管理.....	7
1.2.4 数据源管理.....	7
1.2.5 静态脱敏.....	7
1.2.6 动态脱敏.....	7
1.2.7 脱敏配置管理.....	7
1.2.8 脱敏报告管理.....	8
1.2.9 互联服务器.....	8
1.2.10 告警管理.....	8
1.3 产品价值.....	8
1.3.1 隐私信息全面防护.....	8
1.3.2 数据合理化管理.....	8
1.3.3 满足合规性要求.....	8
1.4 产品描述.....	8
1.5 兼容性.....	9
1.5.1 浏览器兼容.....	9
1.5.2 屏幕分辨率兼容.....	9

---

1.6	系统中常用名词解释 .....	9
<b>2、</b>	<b>软件使用介绍 .....</b>	<b>10</b>
2.1	系统登录 .....	10
2.2	许可证 .....	10
2.3	系统管理平台 .....	11
2.3.1	监控墙 .....	11
2.3.2	日志管理 .....	12
2.3.3	告警管理 .....	13
2.3.4	用户管理 .....	14
2.3.5	系统管理 .....	16
2.4	操作管理平台 .....	17
2.4.1	数据源管理 .....	17
2.4.2	静态脱敏 .....	22
2.4.3	动态脱敏 .....	27
2.4.4	脱敏配置管理 .....	33
2.5	审计管理平台 .....	39
2.5.1	审核数据源 .....	39
2.5.2	审核脱敏任务 .....	39
2.5.3	查看脱敏配置 .....	40
<b>3、</b>	<b>常见问题及解决方案 F&amp;Q .....</b>	<b>40</b>
3.1	修改数据源库中内容 .....	40
3.2	现场环境中源数据库中的数据表、字段内容多 .....	40
3.3	不同数据库类型相互脱敏 .....	41
3.4	HIVE 数据库脱敏注意事项 .....	41
3.5	数据字典导入 .....	41
3.6	加密算法使用注意事项 .....	41
3.7	可逆和可逆还原使用方式介绍 .....	41

## 1、产品概述

昂楷数据脱敏系统致力于防止敏感信息的泄露，又可保证数据的“正确性”而开发设计。可针对不同的敏感类型、不同的应用场景使用不同的脱敏规则，使敏感数据呈现不同的脱敏效果。

昂楷数据脱敏系统简称为：“AAS-DM”。

“AAS-DM”是指对某些敏感信息通过脱敏规则进行数据的变形，实现敏感隐私数据的可靠保护。在涉及客户安全数据或者一些商业性敏感数据时，在不违反系统规则条件下，对真实数据进行改造并提供测试使用，如对身份证号、手机号、银行卡号等个人信息进行数据脱敏。

“AAS-DM”脱敏系统具有流程化、自动化和作业复用等特点。作为轻部署的一体化设备，它拥有强大的功能、能够达到开箱即用的优势，能够极大减轻工作人员的工作强度以及项目周期。

### 1.1 产品特点

#### 1.1.1 敏感数据自动发现

脱敏系统可通过预定义敏感数据特征库，在任务执行过程中通过智能匹配识别敏感数据，最大限度的实现脱敏工作自动化，简化以手工方式按字段名定义敏感数据的繁琐工作。

#### 1.1.2 丰富的脱敏算法

数据脱敏系统内置同义替换、数据覆盖、随机替换、加密、映射、可逆脱敏等多种脱敏规则，内置二十多种中国本地隐私数据类型的漂白规则和算法，可根据实际的应用场景灵活使用。

#### 1.1.3 全面兼容性

支持 Oracle、SQL Server、MySQL、DB2、MongoDB、Hive、PostgreSQL 等国内、外主流数据库。

### 1.1.4 脱敏有效性验证

在脱敏产品实际使用过程中，因生产环境中数据或数据结构频繁发生变化，会出现脱敏失效的情况，昂楷数据脱敏系统可对脱敏后的数据进行验证，确保脱敏工作准确有效的执行。

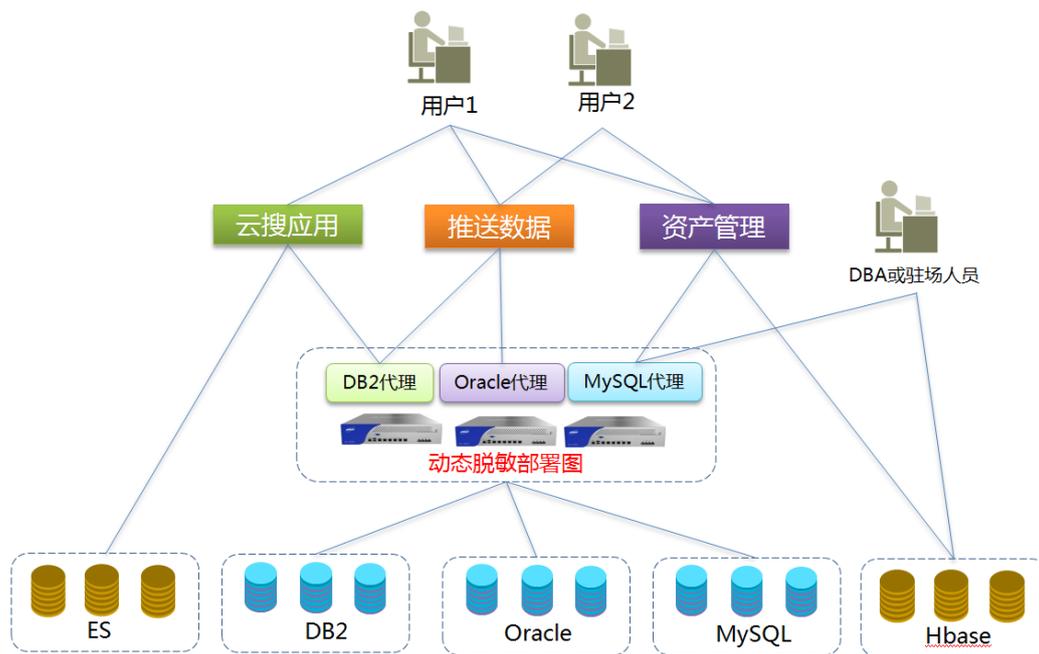
### 1.1.5 敏感数据字典管理

实现敏感数据类型的统一管理，保证组织内的敏感数据具有相同脱敏策略，保证系统的脱敏结果一致性。

### 1.1.6 产品部署方式

昂楷数据脱敏系统部署时采用旁路部署-逻辑串联的方式，不会在生产环境中安装任何客户端或者插件，对现有系统架构无需任何调整。所有配置、操作均通过浏览器访问实现完善的用户权限管理机制，可以兼容多个业务系统的脱敏。

系统部署为单机部署，部署构架图如下图所示。



## 1.2 主要功能

本产品涵盖以下几大功能。包括监控墙、日志管理、用户管理、数据源管理、静态脱敏、动态脱敏、脱敏配置管理和脱敏报告管理。

### 1.2.1 监控墙

展示“AAS-DM”系统使用时 CPU、内存以及磁盘使用率的实时信息；最近七天已完成的作业信息。

### 1.2.2 日志管理

记录用户操作“AAS-DM”系统时产生的操作日志信息以及系统运行时产生的系统日志信息。系统日志和操作日志均可导出为 Excel 或 PDF 进行本地保存。

### 1.2.3 用户管理

用于对系统登录账号的新增、修改、删除和配置用户使用权限的管理。

### 1.2.4 数据源管理

- 数据源配置：对需要参与脱敏的源、目标数据库进行添加、修改、删除、查询等操作。
- 自动发现：自动发现即为对表中数据自动进行敏感类型的匹配，该模块是对自动发现任务的添加和修改、删除、搜索等操作。

### 1.2.5 静态脱敏

对数据库脱敏、文件脱敏任务进行管理。

- 数据库脱敏：对数据库进行脱敏任务的配置，脱敏到数据库或文件
- 文件脱敏：对文件进行脱敏任务的配置，脱敏到数据库或文件

### 1.2.6 动态脱敏

对动态脱敏任务进行配置与管理。

- 代理用户配置：对连接代理数据库的用户名进行添加、修改和删除等操作
- 应用层账号配置：第三方业务系统的登录账号即为应用层账号，该模块是对应用层账号进行添加、修改和删除等操作
- 代理数据库配置：对需要动态脱敏处理的数据源进行代理配置的添加、修改、删除和启用禁用等操作

### 1.2.7 脱敏配置管理

对在数据进行脱敏时使用的脱敏规则、敏感类型、数据字典的管理。

## 1.2.8 脱敏报告管理

可对执行成功的数据库脱敏和文件脱敏任务进行任务报告的生成与导出。

## 1.2.9 互联服务器

对脱敏系统的告警发件邮箱以及告警 syslog 接收服务器进行管理。

## 1.2.10 告警管理

- 告警配置：配置需要进行告警的内容和告警接收端
- 告警信息：对系统产生的告警信息进行查看和导出操作

# 1.3 产品价值

## 1.3.1 隐私信息全面防护

随着信息化的快速发展，信息化建设已经渗透到日常工作的各个方面，在线业务的增长，对虚拟化技术基础设施的要求在逐步提升，企业把越来越多的业务数据都迁移到数据库环境中，虽然降低了企业运营成本，但是在数据传输、数据共享和使用过程中，其安全性无法保证，企业将面临着数据泄露或丢失的高危风险以及数据防护的巨大挑战。

数据脱敏系统可以在数据共享过程中，通过对数据的脱敏处理，使数据在数据库端访问更加安全，防止核心重要数据的泄露，保护企业的核心资产，避免遭受损失。

## 1.3.2 数据合理化管理

利用人工智能技术，强化数据脱敏技术和算法，在保证脱敏随机性的同时还能保留数据的原始特性，参照完整性和业务应用的逻辑性，以便满足开发和测试等场景下的数据使用需求。

## 1.3.3 满足合规性要求

脱敏系统的使用，在使企业能更好的保护自身数据安全外，还能有效帮助客户满足等保、ISO/IEC27001、国际标准（PCI-DSS、HIPPA）和行业监管要求中对敏感数据的使用要求。

# 1.4 产品描述

本产品支持软硬件部署。采用 B/S 的多用户架构，无需在数据库中安装任

何插件, 只要扫描的数据库网络可达。

本产品的运行环境为: CentOS7.4 64 位系统、MySQL 5.7、JDK1.8、Tomcat8 的环境下。

本产品的硬件至少要满足如下条件:

CPU: Intel(R) Celeron(R) CPU G1620 @ 2.70GHz 及以上性能

内存: 至少 4G 内存

磁盘: 至少 500G

## 1.5 兼容性

### 1.5.1 浏览器兼容

本系统采用 B/S 架构, 支持 IE10 及以上浏览器、Chrome 浏览器、Firefox 火狐浏览器。为了保证良好的浏览效果, 推荐使用 Chrome 浏览器。

### 1.5.2 屏幕分辨率兼容

本系统可保证屏幕水平方向像素范围为 1366 到 1920 时的显示效果, 推荐使用屏幕分辨率 1920 x 1080 (即 1080p) 以获得最佳显示体验。

## 1.6 系统中常用名词解释

- 自动发现中的样本数匹配率: 样本数即将会用来进行匹配敏感类型的样本数, 匹配率即这些样本数中匹配了敏感类型的数据量达到或超过的样本数的百分比
- 脱敏规则中的脱敏方式:
  1. 随机: 在符合敏感类型的范围内随机脱敏。
  2. 覆盖: 分区中的数据将会被输入的数据覆盖, 即遮蔽。
  3. 替换: 敏感数据将会被输入的数据替换掉。
  4. 映射: 再符合敏感类型的范围内随机脱敏, 但是两张表中同样的数据脱敏后还是一样的。
  5. SHA256 算法: 采用 SHA 加密算法, 脱敏后的数据为加密的字符串。
- 应用层账号: 第三方应用系统的登录用户名
- 账号自定义: 对需要特殊脱敏处理的代理用户或应用层账号配置单独的脱敏规则

- 黑白名单：对表中数据不需要脱敏处理或不能查看的数据进行黑白名单的配置

## 2、软件使用介绍

### 2.1 系统登录

在 PC 上打开 chrome 浏览器，输入数据脱敏系统的访问地址（如：<https://192.168.0.1:8543>，系统内置三个账号，操作员：ruleadmin、审核员：auditadmin、管理员：admin，默认密码：3edc\$RFV）。登录界面输入用户名、密码和验证码，点击登录后直接跳转到相应的页面。



图 2.1 登录界面

注意：一个账号的密码连续输错 5 次，该账号会被锁定 5 分钟，5 分钟后自动解锁。

### 2.2 许可证

未上传证书时，登录系统，首先会出现上传证书界面。需上传证书后方可正常使用该系统。

**操作：**在上传证书的页面，点击[导入证书] > 点击[请选择文件]>[开始上传] > [确认]

点击确认按钮后页面提示“证书上传成功，请重新登录系统”，输入正确的账号密码点击登录即可登录成功。



登录成功后，点击[软件授权]，授权界面展示证书的有效期和设备的机器码信息，具体信息如下：



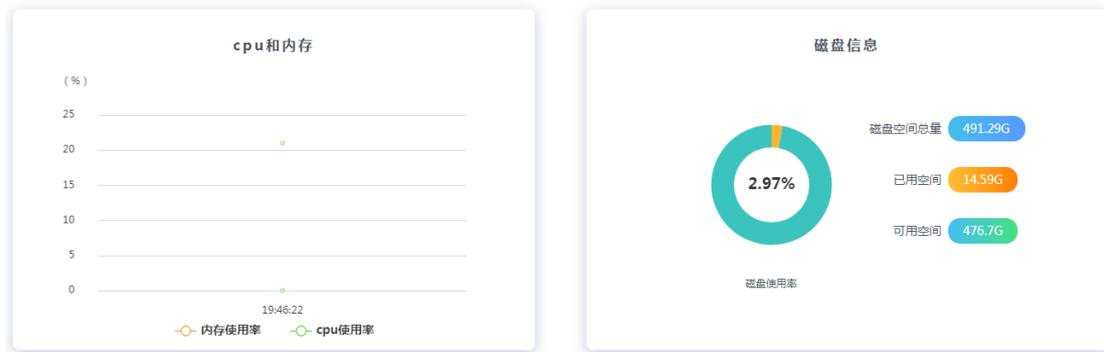
注意：若证书已过期，可成功登录，只能查看历史数据而无法执行任何添加和开始等操作，只有在重新上传有效的证书后系统才可正常使用。

## 2.3 系统管理平台

使用 admin 账号进行登录时，可登录系统管理平台。对以下几个模块的内容进行管理：日志管理、告警管理、用户管理、系统管理、软件授权。

### 2.3.1 监控墙

监控墙主要展示了“ AAS-DM ”系统在使用时硬件设施的 CPU、内存以及磁盘使用率的实时信息；以及最近七天完成的作业数据记录列表。如下图所示。



### 最近完成的作业分析

作业类型	任务名	执行时间	消耗时间	执行效率	数据量
自动发现	62mysql	2019-08-01 17:55:02	1S	10张/s	9个
数据库脱敏	测试2	2019-07-31 14:36:22	145S	1381行/S	200000 行
数据库脱敏	测试	2019-07-31 14:11:42	311S	644行/S	200001 行
自动发现	oracle_test1	2019-07-29 09:57:22	2S	7张/s	16个
自动发现	SQLServer_newlogin2	2019-07-26 09:51:32	4S	10张/s	38个

## 2.3.2 日志管理

日志管理包含了操作日志和系统日志两种，其中操作日志记录所有登录用户在系统上的所有操作行为，而系统日志则是记录了系统在运行中出现的所有系统问题。

操作日志支持按时间段查询，输入用户名和动作的关键字模糊查询及点击模块的下拉筛选查询。操作日志页面展示如下图所示。



时间 从 2019-07-29 17:58 到 请选择日期 请输入搜索内容 查询 导出文件

用户名	登录IP	操作时间	模块	动作	结果	描述
admin	172.18.1.6	2019-08-01 17:50	系统日志	查询系统日志	成功	
ruleadmin	172.21.1.132	2019-08-01 17:48	敏感数据扫描	查询敏感数据扫描任务	成功	

系统日志支持按时间段查询和关键字的模糊查询。系统日志页面展示如下图所示。



时间 从 2019-07-29 17:56 到 请选择日期 请输入搜索内容 查询 导出文件

时间	模块	级别	结果	描述
2019-08-01 16:28	互联服务器	信息	Syslog告警成功;	告警等级：一般告警；任务类型:自动发现；任务名:mysql_read；描述信息：自动发现任务失败，请到操作日志或系统日志查看具体错误信息!

操作日志和系统日志支持导出为 Excel 或者 PDF 文件，方便用户本地保存查看。



注意：按时间段进行查询时，可设置最大的时间段为 30 天；一次最大导出量为 10 万条。

### 2.3.3 告警管理

告警管理模块分为告警配置和告警信息两部分。告警配置中可选择告警方式、接收人和配置告警条件。

**操作：** [告警管理] > [告警配置] > 填写相关信息 > [保存]按钮。

告警方式： 邮件  syslog

邮件接收人：

Syslog接收人：

告警配置：

CPU使用率超过阈值	严重告警： 阈值 <input type="text" value="≥90"/>	<input checked="" type="checkbox"/> ON 启用
	一般告警： 阈值 <input type="text" value="≥90"/>	<input checked="" type="checkbox"/> ON 启用
内存使用率超过阈值	严重告警： 阈值 <input type="text" value="≥80"/>	<input type="checkbox"/> OFF 禁用
	一般告警： 阈值 <input type="text" value="≥80"/>	<input type="checkbox"/> OFF 禁用
系统异常	告警等级： <input type="text" value="严重告警"/>	<input checked="" type="checkbox"/> ON 启用
任务失败	告警等级： <input type="text" value="一般告警"/>	<input checked="" type="checkbox"/> ON 启用

告警方式有邮件告警和 Syslog 告警两种方式，可任选其中一种也可同时选择两种告警方式。选择告警接收人时需先在互联服务器中添加 Syslog 服务器或邮件服务器。



告警信息模块可查看已产生的告警信息。点击下拉图标可查看告警内容的详细信息，如图所示：



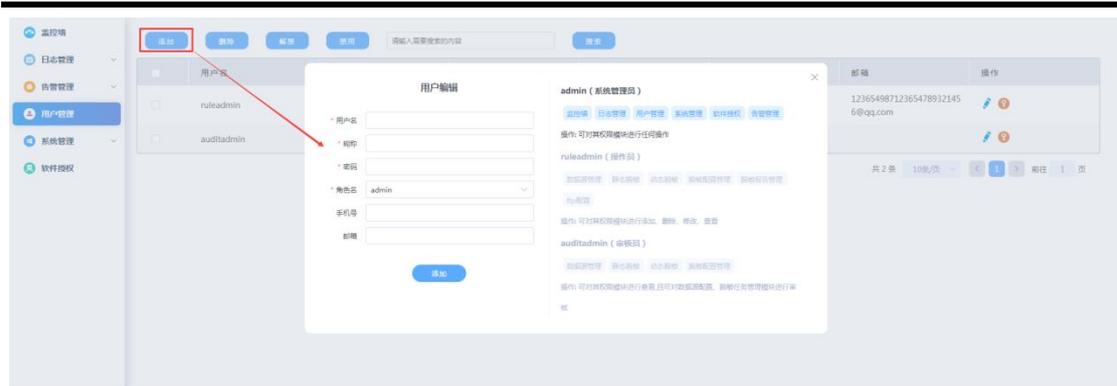
点击导出文件按钮，可选择导出选中或导出所有，点击确定即可导出告警信息的 Excel 文件。



## 2.3.4 用户管理

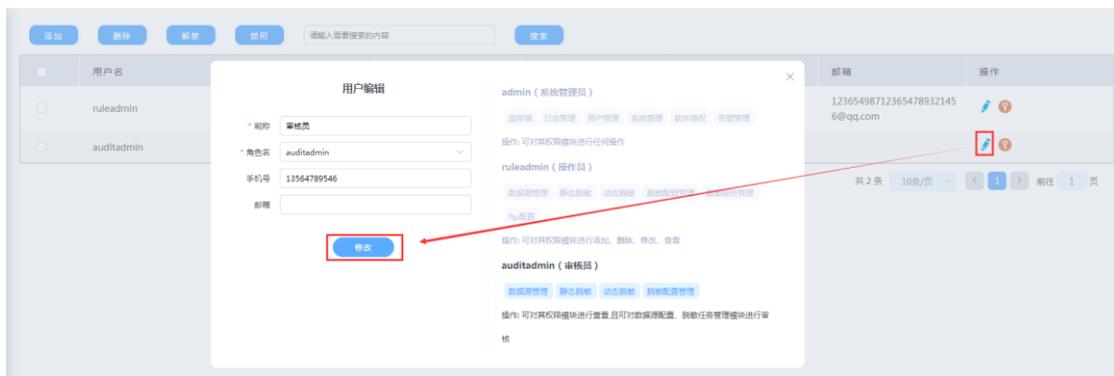
该功能用于对系统用户进行添加、修改、删除、重置密码、用户禁用、用户解禁操作。

添加用户的操作如下图所示。



提示：用户在没有填写邮箱信息的状态下，在邮件告警配置时不会展示出该用户

修改用户的操作如下图所示。



对创建的用户进行密码的初始化、用户解禁、用户禁用等操作。如下图所示。



重置密码按钮，重置后密码为12345678



被禁用的用户无法登录，只有解禁后才可登录

提示：重置后的密码为：12345678，用户在使用重置后的密码第一次登陆时系统会强制性要求修改密码

## 2.3.5 系统管理

### 系统设置

系统设置模块又分为单机模式和集群模式。

单机模式时系统设置可对系统进行升级、回退、重启、关机、恢复出厂设置、管理口配置操作，如下图所示。



单机模式也可转为集群模式，如下图所示：

集群模式时可对系统进行升级、回退操作，可对集群设备进行管理，增加或减少设备，如下图所示：

集群模式也可转为单机模式，如下图所示：

### 互联服务器

互联服务器主要包含邮件服务器和 Syslog 服务器的配置。

**操作：** [系统管理] > [互联服务器] > 在邮件服务器模块填写相关的配置信息 > [测试] > [保存]按钮。点击[测试]按钮之后提示“测试成功”，测试邮箱收到一封测试邮件，则表明发件邮箱可用。



系统发件人可以填写 QQ 邮箱、网易邮箱、139 邮箱以及企业邮箱等，测试邮箱填写可接收邮件的邮箱地址即可。

提示：邮件主机需根据不同的邮箱进行更改。如 QQ 邮箱的邮件主机为 smtp.qq.com，端口号 465 或 587，密码为对应的服务授权码。其他邮箱密码为自定义的邮箱密码。

本系统可以添加备用邮箱服务器，当主邮箱不能发送邮件或者出现故障时就会启用备用邮箱作为发件邮箱。



其配置操作和主邮箱配置一致，此处不再复述。

Syslog 服务器配置，点击添加按钮即可添加 syslog 服务。不可添加 IP 和端口重复的配置。点击网络测试可验证 syslog 服务是否可用。

**操作：** [系统管理] > [互联服务器] > Syslog[添加]按钮 > 填写服务器地址和端口 > [网络测试] > [保存]按钮。



## 2.4 操作管理平台

使用 ruleadmin 账号进行登录时，可登录到操作管理平台界面。可以用于对数据源、静态脱敏任务、动态脱敏任务、脱敏配置以及脱敏报告进行管理。

### 2.4.1 数据源管理

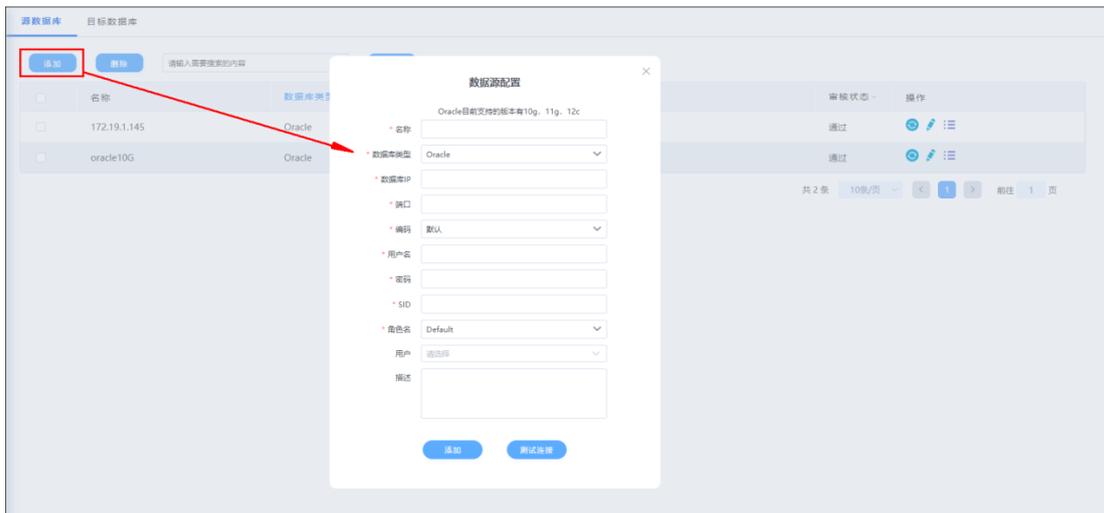
#### 数据源配置

昂楷数据脱敏系统支持以数据库中的表做为源来进行脱敏，数据源管理模块中主要是对源数据库、目标数据库进行管理，包括添加、修改、删除、更新、查询操作。

数据源模块支持添加包含 Oracle、MySQL、SQLServer、DB2、Hive、MongoDB、PostgreSQL 等数据库类型。

名称	数据库类型	ip地址	编码	描述	审核状态	操作
MOngoDB	MongoDB	172.19.1.210 : 27017			未审核	
DB203	DB2	172.19.1.203 : 50000			通过	
pgsql	postgresql	172.19.1.219 : 5432			通过	
200mysql	MySQL	172.19.1.200 : 3306			通过	
hive200	Hive	172.19.1.200 : 10000			通过	
Oracle12c	Oracle	172.19.1.211 : 1521			通过	
SQLServer2017	SQL Server	172.19.1.211 : 1433			通过	

**操作：** 在[数据源配置]>[添加]按钮后，进入添加源数据库配置界面、填写源数据库的名称、选择数据库类型，填写数据库 IP、端口、用户名、密码等信息。（\*号的内容为必填项）



填写完成后点击[测试连接]，测试数据库是否可连接成功，测试成功后点击[添加]按钮，保存数据库配置。

数据源配置在进行测试连接时若测试连接时出错，会给出对应的错误信息。根据对应的提示信息能够快速定位到测试连接出错原因。

在数据源列表的操作栏中点击[更新]按钮，可更新源数据库的数据、表结构。



更新数据库

**注意：** 在更新数据源信息时，若数据库中的表、字段信息太多时，更新时间可能较长。

点击数据源详情按钮，可查看更新后数据库详细信息，包含数据库的配置信息、表信息。



数据源详情。

## 自动发现

**功能介绍：**添加的数据源信息，需经过审核员审核后才能进行自动发现。

**操作：**[数据源管理] > [自动发现] > [添加]按钮。

具体添写内容如下所示：

**第一步：**选择需要进行自动发现的数据源信息。



**第二步：**若选择的数据源在添加时没有进行更新操作则需更新数据源，进行源数据库中的表，字段信息采集。



第三步：选择需要进行自动发现的表，设置样本数、匹配率，以及执行方式的选择。然后点击保存按钮进行保存。

## 添加自动发现

选择全部

表名	描述
<input checked="" type="checkbox"/> 选择当前页	
<input checked="" type="checkbox"/> ACCOUNT	开户许可证号
<input checked="" type="checkbox"/> ADDRESS	
<input checked="" type="checkbox"/> Business_license	
<input checked="" type="checkbox"/> CERTIFICATE	
<input checked="" type="checkbox"/> CREDITCARD	
<input checked="" type="checkbox"/> CVV	
<input checked="" type="checkbox"/> DATETIME	日期
<input checked="" type="checkbox"/> EMAIL	

共 31 条
 <
1
2
3
4
>

\* 样本数

匹配率

执行方式

注意：样本数为在源数据库中取样的数据量。匹配率为样本数中

添加成功后点击开始按钮，自动发现任务状态转换为“等待中”，此时系统会判定是否有其他任务在执行，若没有其它正在执行的任务，15秒内任务状态转换为进行中。

在修改自动发现页面可对所选表名、样本数、匹配率和执行方式进行修改。

## 修改自动发现

名称

选择全部

表名	描述
<input checked="" type="checkbox"/> 选择当前页	
<input checked="" type="checkbox"/> ACCOUNT	开户许可证号
<input checked="" type="checkbox"/> ADDRESS	
<input checked="" type="checkbox"/> Business_license	
<input checked="" type="checkbox"/> CERTIFICATE	
<input checked="" type="checkbox"/> CREDITCARD	
<input checked="" type="checkbox"/> CVV	
<input checked="" type="checkbox"/> DATETIME	日期
<input checked="" type="checkbox"/> EMAIL	

共 31 条 < 1 2 3 4 >

\* 样本数

匹配率

执行方式

选择任意一条或多条自动发现任务，点击[删除]按钮，可删除自动发现任务。

The screenshot shows a table of tasks with columns for '名称', 'ip地址', '审核状态', '任务状态', '任务进度', and '操作'. A red box highlights the '删除' button in the top toolbar. A red arrow points from this button to a modal dialog box that says '自动任务将会被删除 确认要删除吗' with '确认' and '取消' buttons.

名称	ip地址	审核状态	任务状态	任务进度	操作
Hive	172.19.1.203 : 10000	通过	未启动	0%	
145	172.19.1.145 : 51511	通过	未启动	0%	
62mysql	172.23.1.62 : 3306	通过	已完成		

自动发现任务执行完成后点击结果按钮，可查看自动发现后的结果信息。



结果中展示出源数据库中对应表、字段名和匹配的字段敏感类型和匹配率。如下图所示。

表名	字段名	匹配结果
ACCOUNT	ACCOUNTNUM	开户许可证号(100.0%)
ADDRESS	ADDRESS	地址(93.0%)
Business_license	Business_num	营业执照(100.0%)
CERTIFICATE	CERTIFICATE_NUM	医师资格证书编号(100.0%)
CREDITCARD	CREDITCARD_NUM	银行卡号(100.0%)
EMAIL	EMAILADDR	电子邮箱(100.0%)

### 2.4.2 静态脱敏

静态脱敏管理模块用于管理在脱敏系统中所有静态脱敏任务的管理，添加、修改、删除、实时查看各类任务。静态脱敏任务包括数据库脱敏和文件脱敏。

#### 数据库脱敏

**操作：** [静态脱敏] > [数据库脱敏] > [添加] 按钮后在弹出页面中根据提示信息填写就可创建一个数据库脱敏任务。

库到库的脱敏添加步骤：

第一步：源、目标数据库的选择



第二步：脱敏表、字段的选择

添加脱敏作业



第三步：脱敏范围的确认

添加脱敏作业



第四步：敏感类型的确认

添加脱敏作业



点击[完成]按钮后会自动创建一条目标类型为数据库的未审核状态下的脱敏作业信息在数据库脱敏列表中显示。

在未执行脱敏任务时，点击样本对比，会显示目标库中随机的 10 条数据进

行脱敏后效果。

对选择的脱敏规则进行确认。



样本对比

样本比较详情

表名  
ACCOUNT

源数据	目标数据
J5086786043407	L6622554861108
L7836268512103	Z7348991213079
J6824689029076	L4288903950060
Z5915039244610	L7830024164936
L2464550052070	J4521051876377
L1996456974304	J7550116122179
Z7585199727735	L4635288134229
L7727791137109	Z1684949591512
J6638563320266	L2520135240049
Z3063957126659	L6592484234469

若需要执行数据库脱敏任务，任务必须经过审核员审核通过。若没有审核通过则无法开始执行数据库脱敏任务。



数据库脱敏任务可以被修改，修改成功后，脱敏作业审核状态变为“未审核”，任务状态变为“未启动”状态。



数据库脱敏作业详情展示了任务配置以及需要进行脱敏的库名、表名、配置的脱敏规则、耗时、脱敏量等。

### 脱敏作业详情

**任务名称** 测试2

**所用模板**

**目标类型** 数据库

**开始时间** 2019/07/31 14:36:22

**结束时间** 2019/07/31 14:38:46

**运行时长** 2分钟24.726秒

**源数据库名称** mysql\_read

**数据库类型** MySQL

**IP地址** 172.19.1.200:3306

**数据库名** test

**目标数据库名称** 203mysql

**数据库类型** MySQL

**IP地址** 172.19.1.203:3306

**数据库名** testtab

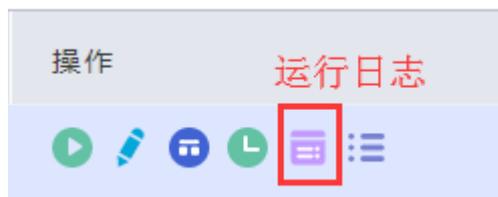
请输入需要搜索的内容 搜索

源数据库表名	目标数据库表名	字段名	敏感类型	所使用的规则	耗时	脱敏量
EMAIL	EMAIL	EMAILADDR		默认映射_电子邮箱	0.133秒	100000
ADDRESS	ADDRESS	ADDRESS		地址替换	0.134秒	100000

共 2 条 10条/页 1 前往 1 页

**注意：**若是在进行脱敏作业时查看，脱敏耗时和脱敏量为当前的实时数据。

可点击运行日志，查看整个脱敏过程中的日志，任务进行中会实时展现作业运行日志。



## 文件脱敏

昂楷数据脱敏系统现支持 Excel、CSV、TXT、DUMP、DEL 这几种格式的文件脱敏，可将文件中的敏感数据脱敏至对应文件中；也可将脱敏后的数据插入到目标数据库。

添加文件脱敏任务步骤如下所示：

第一步：点击添加按钮，选择脱敏文件上传及选择脱敏目标类型。



第二步：对上传文件的格式进行选择匹配。

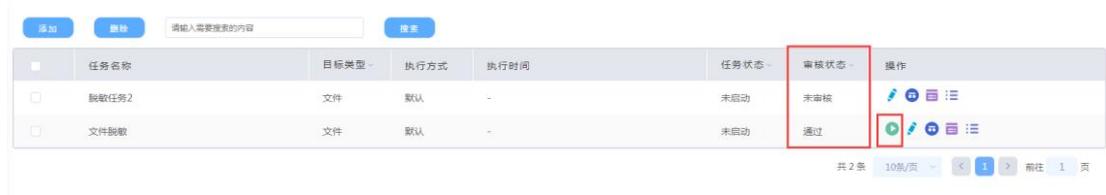


第三步：选择脱敏模板与执行方式，点击添加。一个文件脱敏任务添加成功。执行方式包括默认和定时执行。默认方式时是用户手动点击开始执行文件脱敏任务，定时执行可设置文件脱敏任务自动执行时间，无需用户手动点击开始任务。



**注意：未审核通过的定时执行任务将不会生效。**

添加成功后，未审核任务没有开始按钮。需等待审核员审核通过后，才可点击开始按钮，开始执行文件脱敏任务。



点击修改按钮，可修改除任务名之外的配置信息，修改完成后该任务审核状态变为未审核。需再次审核通过后才可执行脱敏任务。

点击样本对比，可查看按脱敏模板进行脱敏后的效果展示。

点击下载按钮可下载文件脱敏成功的文件压缩包。只有脱敏到文件才会有该文件生成并支持下载。

注意：若脱敏的行数大于 65536 行时，会分成几个 Excel 文件进行存储。因此下载执行脱敏到 Excel 文件中的任务文件时会出现几个 Excel 文件

点击运行日志按钮，可查看文件脱敏的运行日志。

点击详情按钮可查看文件脱敏任务配置详情。



### 2.4.3 动态脱敏

动态脱敏模块用于管理在脱敏系统中所有动态脱敏任务的配置，包括对代理用户、应用层账号以及代理数据库进行配置。

#### 代理用户配置

管理代理用户的信息，可对代理用户进行增加、删除、修改、查询操作。添加的代理用户可被应用于代理数据库配置中。

点击添加按钮，输入代理用户名和密码，点击保存即可添加成功。不可添加重复的代理用户名。



添加的代理用户可被删除，若已应用到代理数据库中，则代理数据库中的代理用户也会对应被删除。

点击修改按钮，可修改代理用户密码。修改后需使用新密码重新登录代理数据库。

### 应用层账号配置

管理应用层账号的配置信息，可对应用层账号进行增加、删除、修改、查询操作。选择某一个应用层账号可在引用任务列表查看代理数据库配置中应用该应用层用户的代理服务名及状态。



点击添加按钮，可添加应用层账号。添加应用层账号有两种方式：手动添加和文件导入。添加的应用层账号可被应用于代理数据库配置中。

文件导入步骤如下：

点击添加按钮，添加方式选择文件导入，点击[选择文件]上传一个包含应用层账号的文件，导入方式选择清除原表，点击[保存]。



注意：导入方式选择清除原表会删除之前添加的所有应用层账号，若选择直接插入则会保留之前添加的应用层账号；导入的文件格式要求每行一个账号；本系统在导入应用层账号时会自动去重。

手动添加步骤如下：

点击添加按钮，添加方式选择手动添加，账号输入框中输入账号信息即可。可点击新增账号的[添加]按钮，可同时新增多个应用层账号。如下图所示。



点击修改按钮，可修改应用层账号名。若已应用到代理数据库中，则代理数据库中的应用层账号也会对应被修改。

点击删除按钮，可删除一个或多个应用层账号。若已应用到代理数据库中，则代理数据库中的应用层账号也会对应被删除。

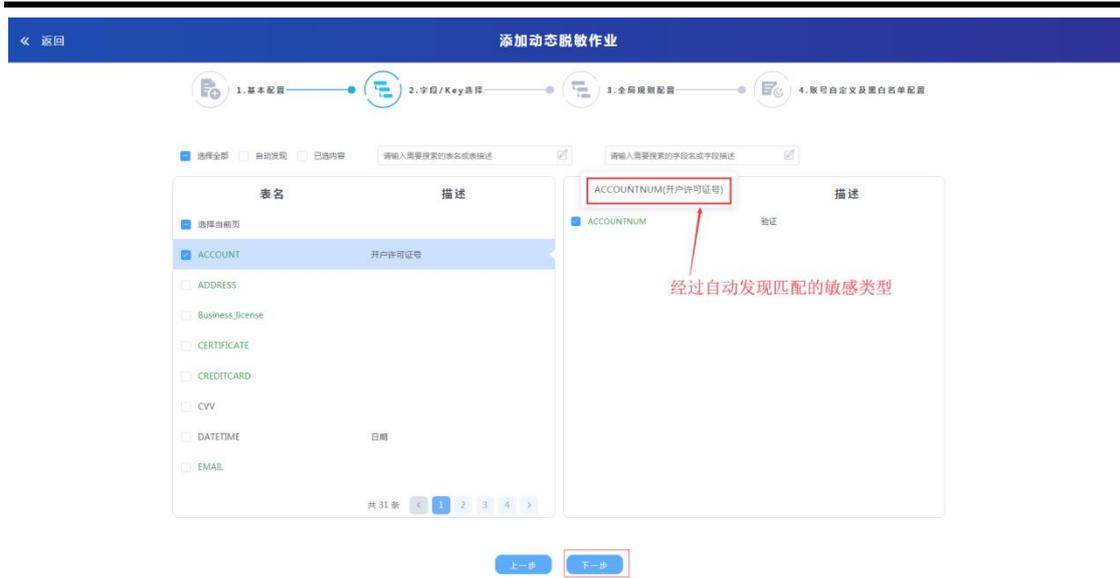
### 代理数据库配置

点击添加按钮后在弹出页面中根据提示信息填写就可创建一个动态脱敏任务。具体操作步骤如图：

第一步：在添加动态脱敏作业页面填写各项信息，点击[下一步]。



第二步：选择需要脱敏的表、字段，点击[下一步]。



第三步：选择该动态脱敏作业的全局脱敏规则，全部脱敏规则为必选项，点击[下一步]。



第四步：在该页面可自定义账号和黑白名单，账号自定义和黑白名单非必填项。

① 加自定义用户，点击[账户自定义]：



在弹出的账号自定义界面，选择账号类型，可选择代理用户或者应用层账号，选择账号和规则名称点击[添加]后点击[完成]按钮，则完成账号自定义的添加。点击[删除]也可删除自定义的账号。



② 添加黑白名单，点击[黑白名单]



在弹出的黑白名单添加页面，可添加黑白名单关键词。在黑名单关键词输入框中输入指定的关键词点击[添加]，即可添加一个黑名单。白名单添加方式和黑名单一样。点击[确定]即可对该表字段成功添加自定义的黑白名单。



第五步：完成账号自定义和黑白名单关键字的添加之后，点击[完成]即成功添加一个未审核的动态脱敏任务。

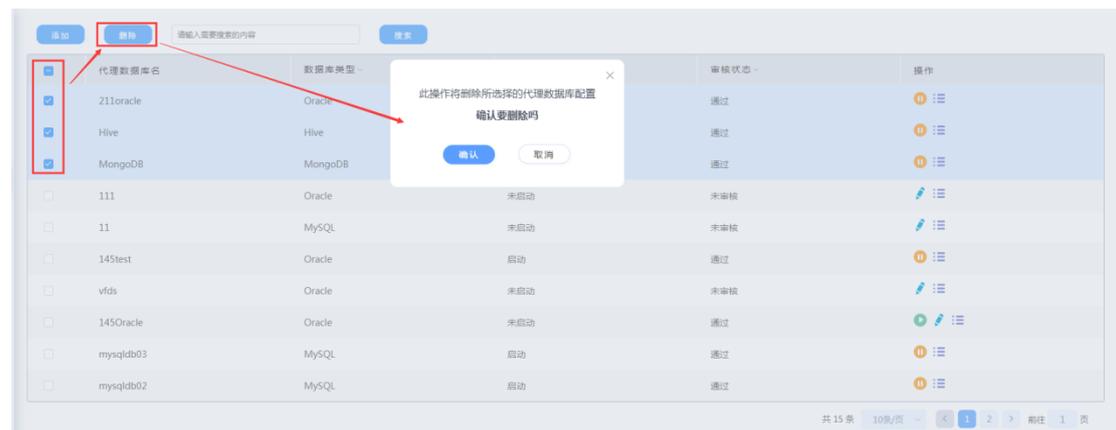


**注意：**如果同一个关键词同时出现在黑名单和白名单中则以黑名单为准。

添加成功后等待该任务被审核员审核通过之后，即可开启成功。开启的任务不可删除且不可修改，只可停止或者查看动态脱敏任务详情。



选择点击删除按钮，可删除一个或多个动态脱敏作业。已启动的动态脱敏作业不可删除。



点击修改按钮即可跳转到动态脱敏任务修改页面，在一个任务修改完成后其审核状态会变更为未审核。需重新审核才可启动任务。

点击详情按钮查看该脱敏任务的配置信息。

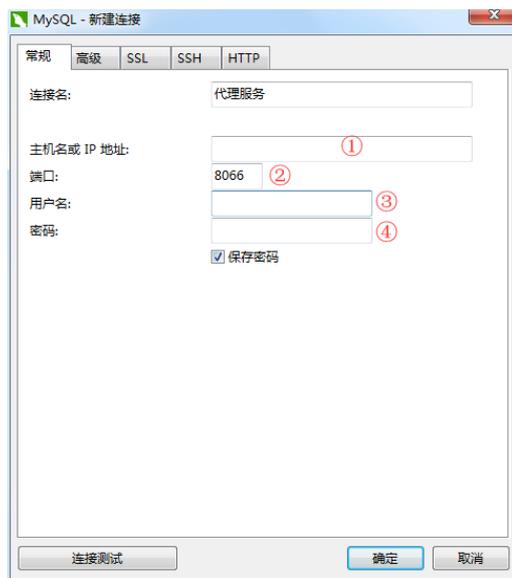


## 代理数据库的连接

在代理数据库配置页面添加了代理数据库并启用后，使用工具连接代理数据库（推荐 Navicat）。

打开 Navicat 工具，新建一个连接，类型选择 MySQL，连接的配置信息如下：

- ①主机名或 IP 地址：脱敏设备的 IP 地址
- ②端口：端口为设备的代理服务端口，固定为 8066
- ③用户名：用户名是添加代理数据库配置的第一步中选择的代理用户
- ④密码：添加代理用户时设置的密码



## 2.4.4 脱敏配置管理

此模块提供的是脱敏规则、敏感类型、数据字典的配置管理。

### 脱敏规则

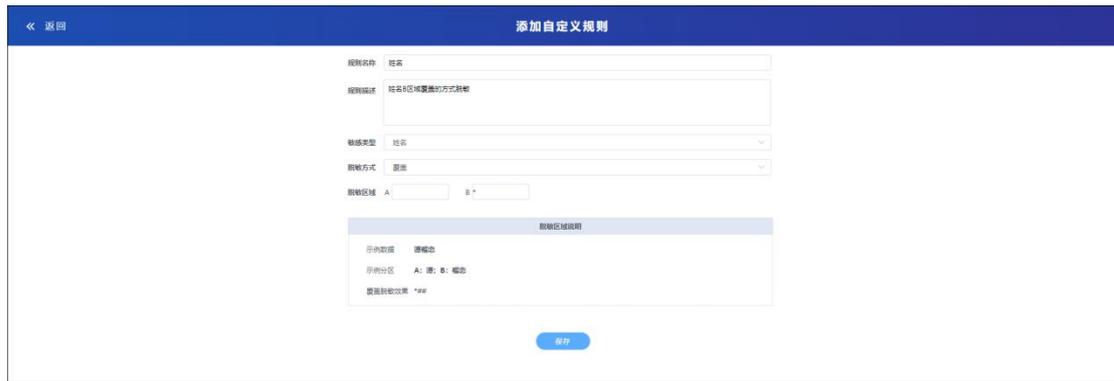
该功能主要用于用户对敏感数据的脱敏方式自主选择，可对需要脱敏的数据区域进行覆盖、随机、替换、映射、加密的配置。

系统内置常用的脱敏规则，方便用户在未创建自定义脱敏规则的情况下直接使用进行数据的脱敏。

用户可点击添加按钮，添加自定义规则。其中，脱敏方式为以下几种

- 随机：在符合敏感类型的范围内随机脱敏。
- 覆盖：分区中的数据将会被输入的数据覆盖，即遮蔽。
- 替换：敏感数据将会被输入的数据替换掉。
- 映射：再符合敏感类型的范围内随机脱敏，但是两张表中同样的数据脱敏后还是一样的。

➤ SHA256 算法：采用 SHA 加密算法，脱敏后的数据为加密的字符串。



点击修改按钮，可修改自定义规则。敏感类型不可修改。

点击删除按钮，可删除系统自带的脱敏规则和自定义的脱敏规则。

点击详情按钮，可查看脱敏规则配置信息。

### 脱敏模板

该功能对敏感信息进行脱敏处理的模版管理。在列表中选中一条模板信息点击查看全部按钮，可查看出该模板下已添加的所有规则信息。

模板名称	详细信息	创建时间	操作
默认_可逆还原	默认可逆还原_IP地址;默认可逆还原_电子邮箱;默认可逆还原_港澳通行证	2018-11-23 11:38	查看全部
默认_可逆	默认可逆_IP地址;默认可逆_电子邮箱;默认可逆_港澳通行证	2018-11-23 11:38	查看全部
默认_SHA256	默认SHA256_IP地址;默认SHA256_电子邮箱;默认SHA256_港澳通行证	2018-11-23 11:38	查看全部
默认_映射	默认映射_IP地址;默认映射_电子邮箱;默认映射_港澳通行证	2018-11-23 11:38	查看全部
默认_随机	默认随机_IP地址;默认随机_电子邮箱;默认随机_港澳通行证	2018-11-23 11:38	查看全部

点击添加按钮，在模板配置框中可选择需要的敏感类型和脱敏规则。已选择的会显示在模板内容中，点击保存即可添加成功。

### 添加模板

\* 模板名称

模板配置

- 手机号 ✓
- 座机号
- 公积金
- 道路运输经营许可证号
- 军密认证号
- 车牌号
- 医师执业证书编号

脱敏规则

- 默认可逆还原\_手机号 ✓
- 默认可逆\_手机号
- 默认SHA256\_手机号
- 默认映射\_手机号
- 默认随机\_手机号

模板内容

敏感类型	脱敏规则	
手机号	默认可逆还原_手机号	删除

点击修改按钮，可修改模板内容，删除已选择的脱敏规则或新增脱敏规则。  
 点击删除按钮，可删除自定义的脱敏模板和内置脱敏模板。

## 敏感类型

敏感类型是用于对需要脱敏的数据进行分类，比如姓名、手机号、地址等就是不同的敏感类型。

系统自带了 28 种数据敏感类型，并且还可支持手动添加自定义敏感类型。  
 添加敏感类型步骤如下图所示。

第一步：填写类型名和需要配置的敏感数据示例信息。

添加敏感类型

\* 类型名

\* 示例

\* 分区方式  指定分隔符  数据分段

\* 分隔符

第二步：划分敏感数据的区域

敏感数据区域的划分方式有两种，一种是按指定分隔符来划分，分割符前面的字段为一个区域，分割符后面的字符为一个区域；一种是按数据分段来划分，每个分区都可以选择按字符长度或以某字符结尾进行分区。

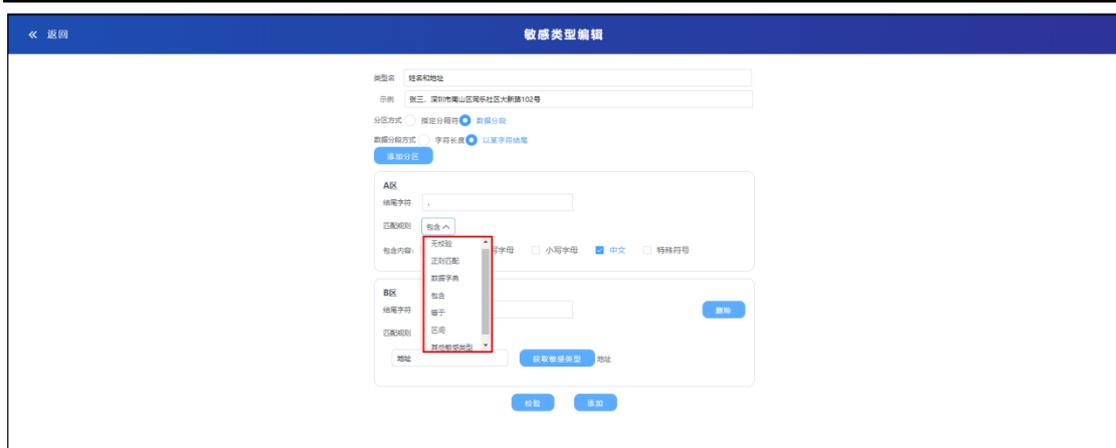


### 第三步设置每个区域的匹配规则

系统带有七种匹配规则无校验、正则匹配、数据字典、包含、等于、区间、其他敏感类型。每个分区的数据都符合分区的校验规则时才能添加敏感类型。

- 无校验：该分区的数据不进行校验，任何字符串都可以匹配
- 正则匹配：该分区的数据必须符合输入的正则表达式，才能校验成功
- 数据字典：该分区中的数据必须包含在数据字典中才能校验成功，数据字典可在脱敏配置管理中的数据字典页面进行配置
- 包含：该分区的数据必须包含所选择的字符例如数据、中文、英文等字符才能校验成功
- 等于：该分区的数据必须包含在输入的关键字中才能校验成功
- 区间：该分区的数据只能是整数、浮点数或日期格式的数据，且必须包含在输入的区间中才能校验成功

其他敏感类型，该分区的数据必须符合其他敏感类型的校验规则才能校验成功

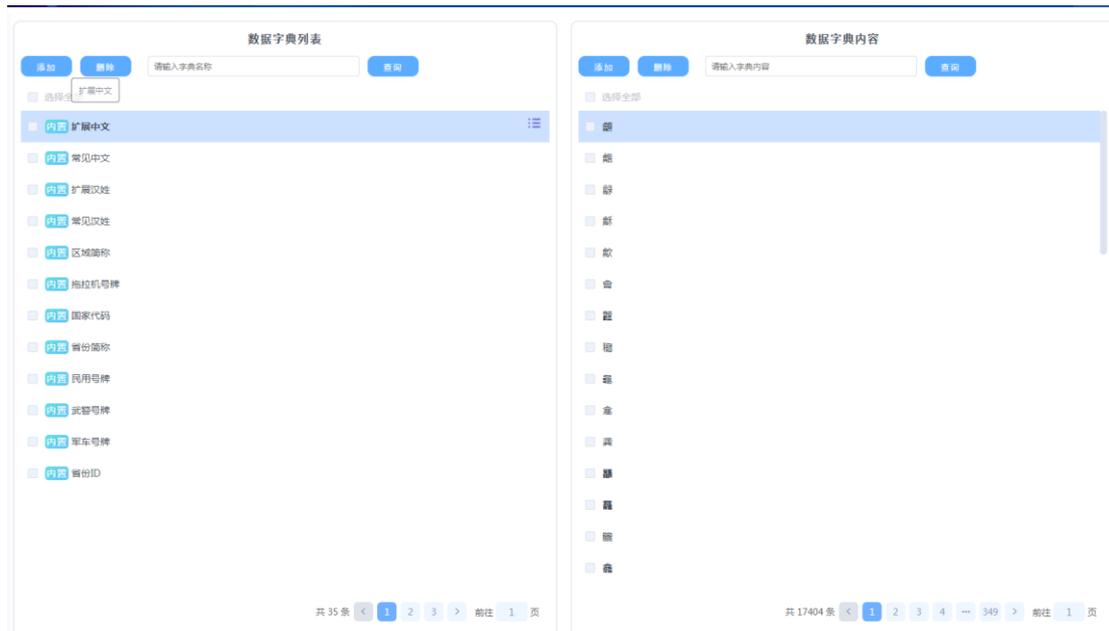


修改、删除、查看的方式和数据规则的操作方式一致，请参照数据规则的操作方式。

## 数据字典

数据字典即敏感类型分区数据的集合

数据字典模块系统内置 35 个字典信息，也可点击添加按钮，自定义数据字典信息。



系统自带的字典信息展示

添加自定义数据字典步骤如下：

第一步：点击添加数据字典，输入字典名称和描述，点击添加。



第二步：点击添加数据字典内容，可选择文件添加和手动添加两种方式。添加、修改、删除、查看详情方式与应用层账号模块操作一致，可参考应用层账号配置操作。



## 黑白名单

当对数据库执行脱敏操作时，如果对某个字段配置了黑名单，且该字段的某条记录的值在黑名单中，则将该条记录所在的整行数据进行过滤，不记录在目标中。

当对数据库执行脱敏操作时，如果对某个字段配置了白名单，且该字段的某条记录的值在白名单中，则将该条记录所在的整行数据不执行脱敏操作，直接将原值记录在目标中。

当黑白名单冲突时以黑名单为主。

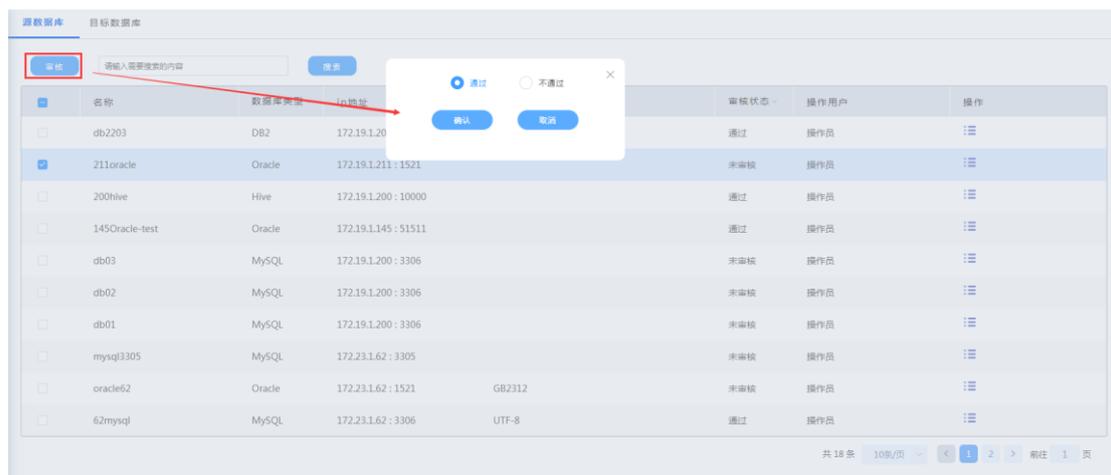


## 2.5 审计管理平台

使用 auditadmin 账号进行登录时，可登录到审计管理平台界面。该平台可对数据源、静态脱敏任务、动态脱敏任务、脱敏配置进行查看与审核。

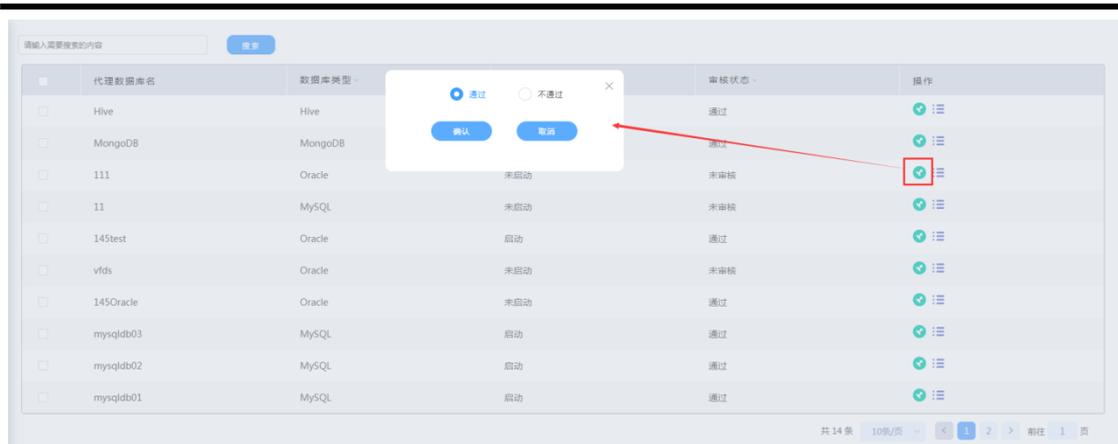
### 2.5.1 审核数据源

审核员可查看源数据库和目标数据库的详细配置信息，审核源、目标数据库的配置信息是否正确，操作员可修改源、目标数据库的信息后再次提交到审核员进行审核。



### 2.5.2 审核脱敏任务

审核员通过查看脱敏任务的详细信息，确认操作员配置的脱敏任务是否正确。若是合法即可审核通过，若是不合法可审核不通过，操作员可修改脱敏作业信息后再次提交到审核员进行审核。



## 2.5.3 查看脱敏配置

审核员可对操作员添加的脱敏规则，敏感类型，数据字典的配置信息进行查看。



## 3、常见问题及解决方案 F&Q

### 3.1 修改数据源库中内容

源数据库执行了字段、表等信息的修改后，请在源数据库列表页面执行更新操作。

**注意：**1，源数据库处于审核通过的状态、进行修改后审核状态将会变为未审核的状态，并且源数据库信息已应用于未运行的任务中，则该任务状态转换为未审核。2，若修改了源数据信息，并且源数据库信息已应用于正在运行的任务中，则会停止任务，任务状态转换为未审核状态。

### 3.2 现场环境中源数据库中的数据表、字段内容多

在使用系统之前请先确定现场环境，库中一共有多少张表，表中字段有多少，若库中的表和字段，大于 40 张表，字段总数大于 200 个字段时，在执行任何任务时需要等待。可在执行任务时打开 F12 进行查看，处理的进度信息。（实验室中测试的是 30 张表，字段总数在

260 个字段左右，能够很快更新出数据源信息在 6 秒钟左右，超过实验室中测试的值时，在执行操作时时间会相对加长，请耐心等待。）

### 3.3 不同数据库类型相互脱敏

执行不同数据库类型之间的脱敏时，请注意源表中的字段类型在目标数据库中字段类型是否支持，不支持时请先手动在目标数据库中建表。

### 3.4 Hive 数据库脱敏注意事项

hive 表名和字段名不支持中文和特殊字符，脱敏到 hive 数据库时不要选择这些类型的表，大写的表会自动转换成小写

脱敏到 hive 数据库时若后台日志中显示下图中的内容，则需要把 hive 数据库的 HDFS 修改为主机名，并在设备的/etc/hosts 文件中配置 IP+主机名进行解析。

出现的具体信息内容如下：`movefrom: hdfs://IP:端口/hive/...`

### 3.5 数据字典导入

导入数据字典的文件要一行数据一行，不要多列一行。

### 3.6 加密算法使用注意事项

使用加密算法时确保目标库建表时字段长度够长，否则插入数据时会报数据超出字段长度的错误

### 3.7 可逆和可逆还原使用方式介绍

使用可逆还原的脱敏方式时，只可以使用经过可逆脱敏方式脱敏后的数据作为源数据，若使用其他脱敏方式脱敏后的数据作为源数据，会出现乱码情况。