ALIBABA CLOUD

阿里云

敏感数据保护

常见问题

文档版本：20200911

阿里云

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。

2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。

3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。

4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。

5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。

6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
危险	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	危险 重置操作将丢失用户配置数据。
警告	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	警告 重启操作将导致业务中断，恢复业务时间约十分钟。
注意	用于警示信息、补充说明等，是用户必须了解的内容。	注意 权重设置为0，该服务器不会再接受新请求。
说明	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击**设置**> **网络**> **设置网络类型**。
粗体	表示按键、菜单、页面名称等UI元素。	在**结果确认**页面，单击**确定**。
`Courier字体`	命令或代码。	执行 `cd /d C:/window` 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	`bae log list --instanceid` *Instance_ID*
[] 或者 [a|b]	表示可选项，至多选择一个。	`ipconfig [-all|-t]`
{} 或者 {a|b}	表示必选项，至多选择一个。	`switch {active|stand}`

目录

1.常见问题总览

本文档介绍了敏感数据保护的各类常见问题和解决方法。

功能相关

数据安全

数据授权

数据扫描和识别

数据脱敏

支持列表

SDDP支持识别的敏感数据

SDDP支持解析的非结构化文件类型

SDDP内置的异常检测规则

SDDP内置的安全审计规则

SDDP支持的数据脱敏算法

2.数据安全

本文介绍数据安全相关的常见问题。

SDDP是否会保存您的数据和文件？

敏感数据保护（SDDP）不会保存您的数据和文件。在您授权访问数据源后，SDDP会对数据进行扫描，并仅将扫描的分析结果展示在SDDP控制台界面，供您使用。

如何对SDDP的操作记录进行审计？

SDDP的所有操作都会通过API形式记录在操作审计服务（ActionTrail）中。

SDDP已经与阿里云操作审计服务集成。开通操作审计服务后，您可以在操作审计服务中查看有关SDDP的所有操作记录，供安全审查使用。关于如何开通操作审计的详细内容，请参见操作审计服务的计费说明。

3.数据授权

您授权允许SDDP访问MaxCompute、RDS和对象存储OSS数据时，可能会出现授权失败的情况。您可以参考以下原因排查数据授权问题。

RDS连接授权失败有哪些原因？

- RDS数据库账号或密码输入错误。
- 您自行删除了RDS访问白名单中SDDP自动添加的服务器地址。
- 部署在经典网络中的阿里云数据产品外网地址未放行流量的访问控制，导致网络不通。

MaxCompute连接授权失败有哪些原因？

- MaxCompute项目名称输入错误。
- MaxCompute项目中添加SDDP账号失败。

相关文档

数据资产授权

4.数据扫描和识别

SDDP支持扫描的数据源有哪些？

SDDP支持对结构化数据源和非结构化数据源进行扫描。

支持的时间	数据源类型
2019年7月	阿里云RDS：MySQL类型（结构化数据）
	阿里云MaxCompute项目（结构化数据）
	阿里云OSS对象存储文件（非结构化数据）

数据源授权完成后需要多长时间完成扫描？

敏感数据保护（SDDP）完成数据源授权后，会在2小时内启动扫描。扫描时长将由您所需扫描的数据量决定。当存在大量数据表时（例如：表数量超过10000张），或者OSS文件总量特别大（例如：OSS总量超过PB）时，扫描周期会相应延长。在SDDP扫描数据的过程中，已经完成扫描的阶段性结果，会在SDDP控制台概览页面展现。详细内容请参见控制台概览。

SDDP对于非结构化数据源（OSS）的扫描机制是怎样的？

SDDP对非结构化数据源中存储的内容进行扫描，根据扫描结果判断是否为敏感数据。

- **首次扫描**：完成授权后，SDDP会对授权的OSS存储桶（Bucket）中的文件进行全量扫描。
- **增量扫描**：如果OSS文件有新增或修改时，SDDP会扫描该新增或修改的文件。

是否支持对已扫描过的OSS文件重新扫描？

如果文件没有修改，SDDP不会对已扫描过的文件重新扫描；如果文件已被修改，SDDP会在4-8小时内对该文件重新扫描。

SDDP后续将会上线手动扫描功能，支持对指定的OSS存储桶执行手动扫描任务。

SDDP对于结构化数据源（RDS/MaxCompute）的扫描机制是怎样的？

SDDP扫描数据库类型和数据表类型数据源中的字段名称和字段值，同时根据字段名称和值综合判断该数据是否为敏感数据。例如：年龄数据。如果只通过字段值无法判断数据是否敏感，SDDP会结合数据源列中的字段名称和对应的数值来综合判断。

- **首次扫描**：完成授权后，SDDP会扫描整个数据库/数据项目中所有的表。
- **增量扫描**：当有新增数据库/数据项目表时，SDDP会对新增表进行扫描；如果现有数据表结构（列）发生变化，SDDP也会对该表进行扫描。

SDDP是否会登录到数据库内获取数据？

已获取授权的情况下，SDDP会登录到数据库内以数据采样的方式对数据进行敏感识别，SDDP不会保存您MaxCompute项目/RDS数据库中的数据。

目前存在哪些触发重新扫描的场景？

目前，SDDP会在以下场景中自动触发对已授权数据源中的数据进行重新扫描。

重新扫描的场景	扫描逻辑	计费影响
数据源首次完成授权接入。	扫描该数据源中的所有数据。	对该数据源中的所有数据收取全量扫描费用。
数据源完成授权接入并已进行过扫描后，数据源发生了变化。	在MaxCompute/RDS数据表结构发生变化后（仅指数据表的列有新增或删减），会触发自动扫描并扫描有变化的列；数据表的行发生变化不会触发自动扫描。	对该数据源中的所有数据收取全量扫描费用。
	在OSS文件新增和修改后会触发自动扫描。 说明 OSS Bucket中的文件仅被删除时不会触发自动扫描。	仅对该新增或修改的文件收取扫描费用。
敏感数据识别规则的配置发生了变化（包括新增、开启、关闭或删除规则）。	会对所有已授权的数据源中的全部数据进行自动扫描。	对所有已授权的数据源收取全量扫描费用。

是否存在不触发扫描的场景？

如果单个OSS文件大小在200MB以上，SDDP不会对其进行扫描。目前，SDDP只扫描单个文件大小在200MB以下的OSS文件。

说明 压缩包视为一个单独的文件。因此，压缩包中所有文件加起来大小在200MB以上，不会触发扫描。

5.数据脱敏

静态脱敏是否对原始数据有影响？

没有影响。静态脱敏功能只会对数据进行读取、脱敏后保存到您选择的目标位置，不会对源数据进行改动。

6.支持识别的敏感数据类型

敏感数据保护服务可识别的敏感数据包括敏感图片信息、个人敏感信息、企业敏感信息等七类。本文档介绍敏感数据保护支持识别的数据类型详情。

敏感分类	数据类型
敏感图片信息	身份证图片
	护照图片
个人敏感信息	身份证
	银行卡
	姓名
	手机号
	邮箱
	护照号
	港澳通行证
	车牌号
	电话号码
	军官证
	性别
	种族
	中国香港身份证
	繁体姓名
	英文姓名
	车辆识别代码
企业敏感信息	营业执照号码
	税务登记证号码
	组织机构代码
	统一社会信用代码
	PEM证书

敏感分类	数据类型
密钥敏感信息	KEY私钥
	AccessKeyId
	AccessKeySecret
	哈希密码
设备敏感信息	IP地址
	MAC地址
	JDBC连接串
	IPv6地址
	IMEI
	MEID
位置敏感信息	省份
	城市
	GPS位置
	地址
通用敏感信息	日期

7.SDDP支持解析的非结构化文件类型

本文介绍敏感数据保护（SDDP）支持解析的非结构化文件类型。

序号	文件类型	序号	文件类型
1	C/C++源代码	85	Tokyo Cabinet数据库文件
2	Lua源代码	86	X3D (Extensible 3D) model xml文件
3	Javascript源代码	87	XML文档
4	VRML虚拟现实建模语言代码	88	XML sitemap文件
5	BCPL源代码	89	DBF数据库文件
6	配置文件Windows Initialization	90	PGP文件
7	Java源代码	91	FTP会话文件
8	BAT文件	92	二进制文件
9	Objective-C源代码	93	EML邮件文件
10	Pascal源代码	94	Visio文档
11	Perl源代码	95	Iwork文档
12	Python源代码	96	WPD文档
13	Ruby源代码	97	WPS文档
14	TCL源代码	98	MSWorks文档
15	Java JCE KeyStore文件	99	Office文档
16	Java KeyStore文件	100	XPS文档
17	Shell脚本	101	邮件文档
18	HTML文件	102	Vcf名片文件
19	GO代码文件	103	Microsoft Reader文档
20	Datax配置文件	104	Excel文档
21	IIS配置文件	105	Outlook文件
22	Tomcat配置文件	106	Word文档
23	MaxCompute配置文件	107	PDF文档

序号	文件类型	序号	文件类型
24	OpenVPN配置文件	108	FDF文档
25	OSS配置文件	109	lotus多字节字符集文件
26	Tomcat application配置文件	110	Lotus WordPro文件
27	Tomcat users配置文件	111	PowerPoint文档
28	Weblogic配置文件	112	SVG图片xml文件
29	SSH配置文件	113	CAD文档
30	DICOM医学成像数据	114	MP4视频文件
31	HDF文件	115	MPEG视频文件
32	GIF图像文件	116	音视频媒体文件
33	JP2图像文件	117	Mp4a视频文件
34	JPEG图像文件	118	3GPP视频文件
35	JPM图像文件	119	GnucCash财务xml文件
36	JPX图像文件	120	3GPP2视频文件
37	JXR图像文件	121	H264视频文件
38	PCX图像文件	122	MJ2视频文件
39	PNG图像文件	123	MP2视频文件
40	TIFF图像文件	124	MP2T视频文件
41	Photoshop图像文件	125	SSH公钥
42	DJVU文档	126	MP4V视频文件
43	ICON图像文件	127	TEXT文件
44	BIOS ogo图像文件	128	MPEG4视频文件
45	BMP图像文件	129	MPV视频文件
46	佳能CR2图像文件	130	Quicktime视频文件
47	佳能CR图像文件	131	DVB视频文件
48	Cartesian Perceptual Compression	132	FLC视频文件
49	DPX图像文件	133	FLI视频文件

序号	文件类型	序号	文件类型
50	EPS图像文件	134	JNG视频文件
51	OpenEXR图像文件	135	M4V视频文件
52	GEM图像文件	136	MNG视频文件
53	MacOS icon图像文件	137	ASF视频文件
54	Windows帮助全文搜索引	138	SG视频文件
55	Windows安装信息	139	ACC音频文件
56	NIFF图像文件	140	SSH私钥
57	Olympus ORF图像文件	141	m4a音频文件
58	Paint.NET图像文件	142	DOS可执行文件
59	Windows帮助文档	143	空文件
60	Polar Monitor Bitmap图像文件	144	COM可执行文件
61	Windows预编译文件	145	ELF可执行文件
62	Greymap图像文件	146	Object文件
63	Pixmap图像文件	147	Tcpdump捕获文件
64	Quicktime图像文件	148	7z压缩文件
65	TGA图像文件	149	Bzip2压缩文件
66	WindowsINF文件	150	Cabinet压缩文件
67	X3F原始图像文件	151	普通压缩文件
68	XPM图像文件	152	LHa压缩文件
69	XWD X Window Dump图像文件	153	LRZIP压缩文件
70	BPG图像文件	154	LZ4压缩文件
71	MongoDB数据库文件	155	Lzip压缩文件
72	MySQL数据库文件	156	Unix压缩文件
73	Oracle数据库文件	157	LZMA压缩文件
74	Postgres数据库文件	158	Gzip压缩文件
75	SQLite3数据库文件	159	XZ压缩文件

序号	文件类型	序号	文件类型
76	Redis数据库文件	160	Zstandard压缩文件
77	SQLServer数据库文件	161	Zstandard字典文件
78	Berkeley数据库文件	162	Zip压缩文件
79	DBase数据库文件	163	Zlib压缩文件
80	GNU dbm或ndbm数据库文件	164	Qpress压缩文件
81	Access数据库文件	165	Snappy压缩文件
82	可扩展存储引擎数据库文件	166	RAR压缩文件
83	MSVC程序数据库文件	167	tar文件
84	Windows应用程序兼容性数据库	168	-

8.SDDP支持的数据脱敏算法

本文介绍敏感数据保护（SDDP）系统支持的脱敏算法。

算法分类	分类描述	算法描述	输入参数	适用类型和典型场景
哈希脱敏	不可逆算法。 适用于密码或需要通过对比进行敏感数据确认的场景。 支持常见的哈希算法，并支持偏移量（加盐值）配置。	MD5	Salt值	• 敏感类型：密钥类 • 适用场景：数据存储
		SHA-1	Salt值	
		SHA-256	Salt值	
		HMAC	Salt值	

遮盖脱敏（*和#）	不可逆算法。
	适用于前端展示或敏感数据分享的场景。
	通过使用特殊字符（*或者#），对部分文字进行遮盖实现敏感数据的脱敏。
		保留前n后m	n、m	• 敏感类型：个人敏感
				• 适用场景：
				  ○ 数据使用
				  ○ 数据分享

		保留自x至y	x、y	
		遮盖前n后m	n、m	
		遮盖自x至y	x、y	
		特殊字符前遮盖（针对首次出现该字符）	“@”、”&”、“.”	
		特殊字符后遮盖（针对首次出现该字符）	“@”、”&”、“.”	

替换脱敏	部分可逆算法。
	适用于证件号等构成规则固定的字段脱敏。
		身份证映射替换	行政区划随机码表	• 敏感类型：
				  ○ 个人敏感
				  ○ 企业敏感

		身份证随机替换	行政区划随机码表	
		军官证随机替换	种类编码随机码表	
		护照随机替换	用途字段随机码	
		港澳通行证随机替换	用途字段随机码	
		银行卡随机替换	Bin码随机码表	
		座机号码随机替换	行政区划随机码表	

算法分类（支持新增）	分类描述	算法描述	输入参数	适用类型和典型场景
替换脱敏	使用替换码表进行映射替换（可逆），或使用随机区间进行随机替换（不可逆），实现字段整体或者部分内容的脱敏。系统预置多套码表可供选择，并支持用户自定义替换算法。	手机号随机替换	网号	• 适用场景： ◦ 数据存储 ◦ 数据分享 ◦ 设备敏感
		统一信用码随机替换	登记部门随机码表、类别码随机码表、行政区划随机码表	
		通用表格映射替换	大写字母映射码 、小写字母映射码、数字映射码、特殊映射码	
		通用表格随机替换	大写字母随机码 、小写字母随机码、数字随机码 、特殊随机码	
变换脱敏	部分可逆算法。 适用于对敏感数据集进行分析和统计类场景。 提供对数字或日期等进行取整操作（不可逆）和对文字进行位移操作（可逆）两类变换脱敏算法。	数字取整，保留小数点前第N位	N	• 敏感类型：通用敏感 • 适用场景： ◦ 数据存储 ◦ 数据使用
		日期取整	日期取整保留到	
		字符位移	整体循环位移Bit数、向左/向右	
加密脱敏	可逆算法。 适用于对需要回源的字段进行加密的场景。 支持常见的对称加密算法。	DES算法	加密密钥	• 敏感类型： ◦ 个人敏感 ◦ 企业敏感 • 适用场景：数据存储
		3DES算法	加密密钥	
		AES算法	加密密钥	
洗牌脱敏	不可逆算法。 适用于结构化数据列级别的数据脱敏场景。 在源数据表抽取数据并确认数值范围后，对该字段（在范围内）进行列级别的打散重排和随机选择，实现混淆脱敏。	随机洗牌	打散重排/随机选择	• 敏感类型： ◦ 设备敏感 ◦ 位置敏感 • 适用场景：数据存储

9.SDDP内置的安全审计规则

本文介绍敏感数据保护（SDDP）安全审计功能支持的内置审计规则。

敏感数据保护的内置审计规则适用的数据源为RDS。审计规则详情请参见下表。

攻击审计场景	审计规则
数据库漏洞攻击	数据库漏洞利用
数据库注入	基于DBMS_PIPE.RECEIVE_MESSAGE的时间盲注
	基于PG_SLEEP的时间盲注
	基于SLEEP的时间盲注
	基于WAITFOR DELAY的时间盲注
	基于DELAY的时间盲注
	基于GENERATE_SERIES的时间盲注
	基于BENCHMARK的时间盲注
	基于EXTRACTVALUE的报错注入
	基于DBMS_UTILITY.SQLID_TO_SQLHASH的报错注入
	基于UTL_INADDR.GET_HOST_ADDRESS的报错注入
	基于CTXSYS.DRITHSX.SN的报错注入
	探测PLUGINS信息
	探测表字段个数
	MAKE_SET/ELT注入探测
	条件判断注入探测
	引号型注入探测
	Char注入探测
	Concat注入探测
	CaseWhen注入探测
	异常操作UDF
	创建角色
	删除角色

攻击审计场景	审计规则
数据库高危操作	删除用户账号
	删除、修改、创建Tablespace
	收回用户权限的行为
	探测系统版本
	探测数据库用户及密码
	探测数据库名称
	探测登录用户
	探测数据库版本
	使用load_file函数读取系统文件信息
	使用 `outfile` / `dumpfile` 函数往系统写入文件
	删除数据库
	删除表
	清空表
	创建账号
	授权
	修改密码

10.SDDP内置的异常检测规则

本文介绍敏感数据保护（SDDP）异常检测功能支持的内置异常检测规则。

模型类别	模型名称	异常描述	适用数据源
数据流转异常	异常地理位置下载敏感数据	来自异常地理位置的数据下载可能是由于账号访问权限被外部攻击者获取，导致数据泄露。	OSS、RDS、MaxCompute
	异常终端下载敏感数据	来自异常终端的数据下载可能是由于账号访问权限被外部攻击者获取，或者员工使用非工作终端进行数据下载。	OSS
	异常时间下载敏感数据	来自异常时间的数据下载可能是由于账号访问权限被外部攻击者获取，或者员工在非正常工作时间内进行数据下载。	OSS、RDS、MaxCompute
	初次下载敏感数据	账号首次下载敏感数据可能是由于账号被错误分配敏感数据下载权限，导致敏感数据外泄。	OSS、RDS、MaxCompute
	敏感数据下载量异常	敏感数据下载量异常可能是由于账号访问权限被外部攻击者获取，或者员工恶意备份敏感数据。	OSS、RDS、MaxCompute
	下载非常用敏感表	账号下载非常用敏感表可能是由于账号被错误分配敏感数据下载权限，导致敏感数据外泄。	RDS、MaxCompute
	日志输出量异常降低	日志输出量异常，可能存在产品日志输出故障，导致产品上的数据操作异常事件无法有效识别。	OSS、RDS、MaxCompute
	文件下载量异常	账号数据下载量异常可能是由于账号访问权限被外部攻击者获取，或者员工恶意备份敏感数据。	OSS
	数据下载量异常	账号数据下载量异常可能是由于账号访问权限被外部攻击者获取，或者员工恶意备份敏感数据。	RDS、MaxCompute
	登录时间异常	来自异常时间的鉴权记录可能是由于账号访问权限被外部攻击者获取，或者员工在非工作时间访问数据。	OSS、RDS、MaxCompute
	登录使用终端异常	来自异常终端鉴权记录可能是由于账号访问权限被外部攻击者获取，或者员工在使用非办公终端访问数据。	OSS、RDS、MaxCompute
	登录地址异常	来自异常地理位置的鉴权记录可能是由于账号访问权限被外部攻击者获取，导致数据泄露。	OSS、RDS、MaxCompute

模型类别	模型名称	异常描述	适用数据源
权限使用异常	下载非常用Bucket内敏感文档	账号下载非常用Bucket内敏感文档可能是由于账号被错误分配敏感数据下载权限，导致敏感数据外泄。	OSS
	MaxCompute敏感项目未设置保护	包含敏感数据的项目未设置Protection标识（详见项目空间的数据保护），则该项目无法实现数据流出保护。	MaxCompute
	MaxCompute敏感项目未设置标签安全	包含敏感数据的项目未设置Label Security标识（详见列级别访问控制），则无法控制用户对敏感数据的访问。	MaxCompute
	OSS敏感Bucket被设置为可公开访问	包含敏感数据的Bucket被设置为公开，则外部人员都可以通过接口访问到敏感数据。	OSS
	权限闲置期限超过阈值	闲置权限的存在违反权限最小化授权原则，长期闲置的权限如果被外部攻击者获取更不容易被发现。	OSS、RDS、MaxCompute
	多次访问不存在的文件	出现多次访问不存在的文件可能是存在外部攻击尝试。	OSS
	多次访问没有权限的文件	出现多次访问没有权限的文件可能是存在外部攻击尝试。	OSS
	多次尝试访问未成功	出现多次尝试访问未成功可能是存在外部攻击尝试。	OSS、RDS、MaxCompute
数据操作异常	MaxCompute打标结果低于识别值	恶意调低数据打标结果可能导致恶意绕过权限控制、数据安全保护措施不能有效覆盖等风险。	MaxCompute
	后台变更敏感数据字段	与后台订正数据相比，通过应用实现数据更正，存在更大的恶意风险。	MaxCompute