ALIBABA CLOUD

阿里云

内容安全

快速入门

文档版本：20220524

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.入门概述

快速入门手册根据最基础的使用场景，为您梳理业务中的操作流程，方便您能够快速上手对应的功能。

带您快速玩转内容安全

检测场景	支持的功能	使用方式	快速入门
帮助您检测存储在OSS空间的图片、视频和语音文件，是否存在鉴黄、涉政暴恐等风险。 如果存在风险，OSS违规功能会根据您的配置冻结或者删除风险文件。	OSS违规检测包含如下功能： • 增量扫描 对OSS Bucket中新增的图片、视频和语音自动进行违规检测，每当Bucket中有新增内容，将自动触发扫描。 • 存量扫描 对OSS Bucket中的已有图片、视频和语音进行一次性违规检测。	在控制台上配置即可。 该功能无需您开发，只需要在控制台上配置少许页面即可使用。	OSS违规检测
为满足您的个性化需求，内容检测API以接口的方式，帮助您检测存储在公网上（部分支持存储在本地）的图片、视频、文本、音频，是否存在色情、涉政、暴恐、广告、垃圾信息等风险。 支持以同步和异步方式返回检测结果，您可以根据业务特点，选择合适的接口接入使用。	内容检测API包含如下功能： • 内容审核（机审） 包含图片审核、视频审核、文本审核、音频审核、文件审核和网页审核。 • 人工审核 包含图片人工审核、视频人工审核、文本人工审核和语音人工审核。 • 图片OCR识别 包含通用图文OCR、结构化卡证OCR、结构化票据OCR、卡证票据混贴OCR、自动卡证票据分类OCR和自定义模板OCR。 • 人脸识别 包括人脸属性检索、活体翻拍检索、图片敏感人脸识别、视频敏感人脸识别和自定义人脸检索。	通过调用API方式接入使用。 该功能是通过调用接口来进行内容检测，需要您具备一定的编程能力。	内容检测API概览
帮助您检测网站的首页和全站其他网页的图片、文本，是否存在首页篡改、挂马暗链、色情低俗、涉政暴恐等风险。 如果您的网站有疑似违规信息时，站点检测功能会根据您的配置将违规信息发送给您，以便及时对网页内容进行整改。	站点检测包含如下功能： • 首页检测 包含首页篡改、挂马暗链、色情低俗、涉政暴恐。 • 全站检测 包含挂马暗链、色情低俗、涉政暴恐。	在控制台上配置即可。 该功能无需您开发，只需要在控制台上配置少许页面即可使用。	站点检测

2.OSS违规检测

本文以某社交平台为例，该平台的用户每天会上传大量的图片（图片上传后会存储于OSS对象存储服务tmpsample Bucket），为了快速监控该网站新增的图片是否涉及色情、涉政暴恐等，该平台使用内容安全OSS违规检测功能。本文主要介绍如何使用OSS违规检测功能监控该网站的图片。

业务流程

在使用OSS违规检测之前，请先阅读并完成前提条件的内容。快速体验OSS违规检测功能的步骤如下所示：

- 步骤一：授权访问OSS存储空间
- 步骤二：设置增量扫描任务
- 步骤三：查看扫描结果
- 步骤四：查看统计信息

如果您需要更深入地了解OSS违规检测功能，请参见OSS违规检测使用简介。

前提条件

- 已注册阿里云账号。更多信息，请参见阿里云账号注册流程。
- 已开通内容安全产品。更多信息，请参见开通与购买。

步骤一：授权访问OSS存储空间

OSS违规检测只向开通了阿里云对象存储OSS服务的用户提供服务。在使用OSS违规检测前，您需要先开通OSS对象存储服务tmpsample Bucket，并授权内容安全读取该Bucket权限。关于授权的具体操作，请参见授权内容安全访问OSS存储空间。

步骤二：设置增量扫描任务

通过增量扫描设置，您可以对*tmpsample Bucket*中新增的图片、视频自动进行违规检测（每当Bucket中有新增内容，将自动触发扫描），并实时查看近7天的增量扫描结果。

1. 登录内容安全控制台。
2. 在左侧导航栏，选择**设置 > OSS违规检测**。
3. 在**OSS违规检测**页面的**增量扫描**页签，按照实际业务进行如下配置：

i. **选择Bucket。**

从左侧**待选择**框中选择需要检测的Bucket，添加到右侧的**已选择**框中。

然后单击**下一步**。

ii. **设置过滤条件。**

单击已选择的Bucket右侧**设置过滤条件**，按如下说明配置仅扫描*img/test_*前缀的图片文件。关于配置的详细介绍，请参见设置过滤条件。

然后单击**下一步**。

iii. **配置检测场景。**

展开**推荐配置**下拉框，选择**社交行业推荐配置**，并关闭视频和语音的检测场景。

关于参数说明的具体介绍，请参见配置检测场景。

然后单击**下一步**。

iv. **配置扫描范围。**

扫描范围根据您配置的检测场景显示。例如，您在**场景配置**页面只勾选**图片**，那么**扫描配置**页面就只显示图片的配置信息。

设置**每日图片扫描上线**为100张，开启**检测无后缀文件**。

然后单击**下一步**。

v. **配置冻结范围**。

- 开启图片**需要自动冻结**开关。

  设置**涉黄**、**涉政**、**图文违规**、**不良场景**选项按结论冻结。

- 选择**修改权限**冻结方式。

  将您Bucket中public权限的违规文件设置为private访问权限。

然后单击**下一步**。

4. 选中**我已经同意OSS违规检测服务条款**，并单击**保存**。

   OSS违规检测功能会根据您的配置为您预估出扫描费用的上限，您可以根据实际业务选择按量付费或者资源包抵扣方式。

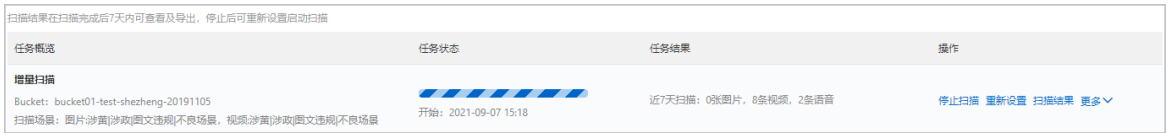
然后单击**确定**。增量任务设置成功后，系统会自动跳转到**OSS违规检测 > 增量扫描任务**页面。

增量扫描配置保存后即时生效。系统会按照增量扫描配置，自动对已选择Bucket中新增的图片、视频进行违规检测。

步骤三：查看扫描结果

OSS违规检测服务为您提供查看扫描结果的功能，当您完成增量扫描任务后，您可以随时在内容安全控制台查看扫描结果，并根据扫描结果执行自助审核。

1. 登录内容安全控制台。
2. 在左侧导航栏，选择**OSS违规检测 > 增量扫描**。



3. 在**增量扫描**页面，单击操作列**扫描结果**。

   在**增量扫描结果**页面，查看最近7天的增量扫描结果和处理违规内容。

如果扫描结果不符合您的业务需要，您可以对扫描结果进行自助审核，自助审核包含如下操作：

- **违规并删除**

  通过单击**违规并删除**，可将图片或视频从内容安全控制台和OSS Bucket中一并删除。支持单选或者多选。

- **正常并忽略**

  通过单击**正常并忽略**，则忽略该检测结果。忽略后该图片或视频将不再在控制台展示，并不影响存储在OSS Bucket中的图片或视频。支持单选或者多选。

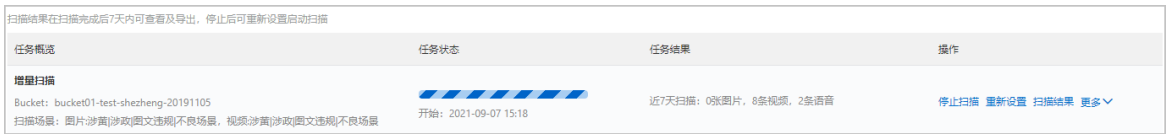
- **正常并解冻**

  若您设置了自动冻结功能，则还可以在选中图片或视频后单击**正常并解冻**，将已冻结的图片或视频解冻。

步骤四：查看统计信息

OSS违规检测服务为您提供数据统计功能，当您完成增量扫描任务后，您可以随时在内容安全控制台查看数据统计信息。您可以通过监控一段时间的统计数据，根据网站内容的违规情况，对网站加强管控。

1. 登录内容安全控制台。
2. 在左侧导航栏，选择**OSS违规检测 > 增量扫描**。



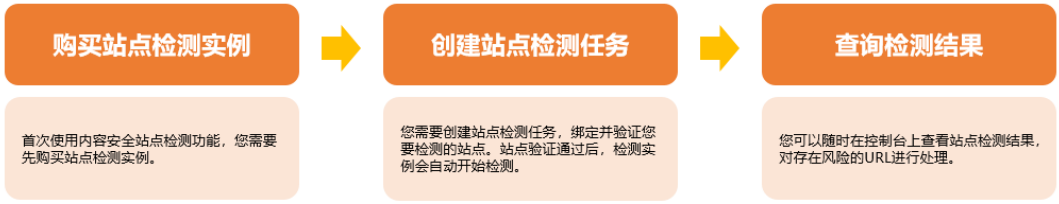
3. 在**增量扫描**页面，选择**更多 > 数据统计**。

4. 在**OSS违规检测调用量**页面的**图片**页签，查看最近7天增量扫描的统计信息。

3.站点检测

本文以某网站为例。为了能够快速监控该网站在内容安全方面可能存在的风险（例如，首页篡改、挂马暗链、色情低俗、涉政暴恐等），并反馈违规内容的具体地址，帮助您查看和修复，该网站使用内容安全站点检测功能。本文主要介绍如何使用站点检测功能监控网站内容。

业务流程



在使用站点检测之前，请先阅读并完成前提条件的内容。快速体检站点检测功能的步骤如下所示：

- 步骤一：购买站点检测实例
- 步骤二：创建站点检测任务
- 步骤三：查看检测结果

如果您需要更深入地了解站点检测功能，请参见站点检测使用简介。

前提条件

已开通内容安全产品。更多信息，请参见开通与购买。

步骤一：购买站点检测实例

首次使用内容安全站点检测功能，您需要先购买站点检测实例。

1. 登录内容安全控制台。
2. 在左侧导航栏，选择**设置 > 站点检测**。
3. 在**实例管理**页签，单击**购买实例**。
4. 在**站点检测（包年）**页面，根据实际需求配置相关参数，单击**立即购买**，按照页面指引完成订单。

步骤二：创建站点检测任务

完成购买站点检测实例后，您需要创建站点检测任务，绑定并验证您要检测的站点。站点验证通过后，检测实例会自动开始检测。

1. 在**实例管理**页签，选择**未绑定**状态的实例，单击右侧操作列的**绑定站点**。
2. 在**绑定站点**对话框中，设置**协议**、**域名**、**默认首页地址**、**首页检测间隔**和**全站检测频率**，并单击**下一步**。

   关于绑定站点的具体介绍，请参见绑定站点页面说明表。
3. 在**验证站点**对话框，选择**阿里云账户验证**，然后单击**立即验证**。

   关于验证站点的具体介绍，请参见创建站点检测任务。

步骤三：查看检测结果

您可以前往控制台查看站点检测结果，对存在风险的URL进行处理。如果您对检测的结果有异议，可以将问题反馈给我们。

1. 在左侧导航栏，选择**站点检测 > 首页监测**或者**站点检测 > 全站监测**。
2. 单击存在风险的URL，查看并确认风险。
   - 消除风险后，单击**已处理**，完成处理。
   - 如果您对结果有异议，您可以单击**纠错**或**问题反馈**，通过表单将问题反馈给我们。在确认问题后，我们将在算法层面进行优化改进。