MaaS 模型即服务

安全与合规

文档版本 01

发布日期 2026-09-17

华为云计算技术有限公司

商标声明


和其他华为商标均为华为技术有限公司的商标。
本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为云计算技术有限公司

地址： 贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

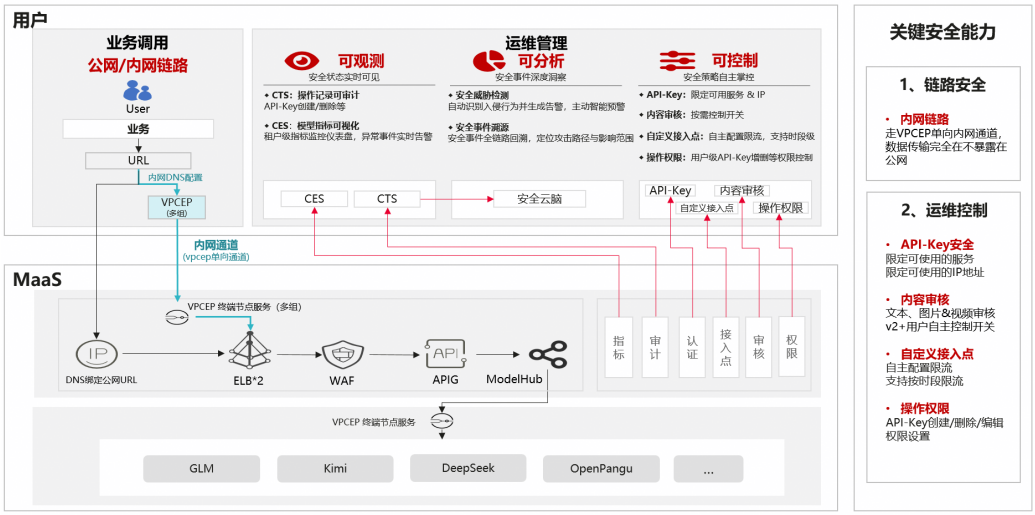
网址： https://www.huaweicloud.com/

目 录

1 安全总览

华为云MaaS围绕“**可观测 · 可分析 · 可控制**”构建推理态安全能力，使您对业务与数据安全有清晰可见度与全面掌控力。

图 1-1 MaaS 安全总览



可观测

- **云审计 CTS**

  模型服务开通、API Key创建/删除/编辑、自定义接入点的创建/删除/编辑等操作事件都会上报到CTS，可用于支撑安全分析、合规审计、资源跟踪和问题定位等。详情可参考**操作审计与模型监控**。

- **云监控 CES**

  MaaS对接了云监控服务（CES），监控在线服务和对应模型负载，如RPM/TPM、Token用量、模型调用成功率/失败率等监控指标，且支持指标仪表盘可视化。监控指标详情请参见**监控指标说明**。

- **模型调用统计**

  统计预置服务、自定义接入点在指定时间段内的调用数据，如总调用Tokens数、输入Tokens数、输出Tokens数、端到端时延等信息，并以分钟为最小时间粒度展示数据趋势，帮助您了解服务的使用情况和性能变化，从而更有效地进行模型评估、问题定位、故障排除和性能优化。统计详情请前往**调用统计**页面查看。

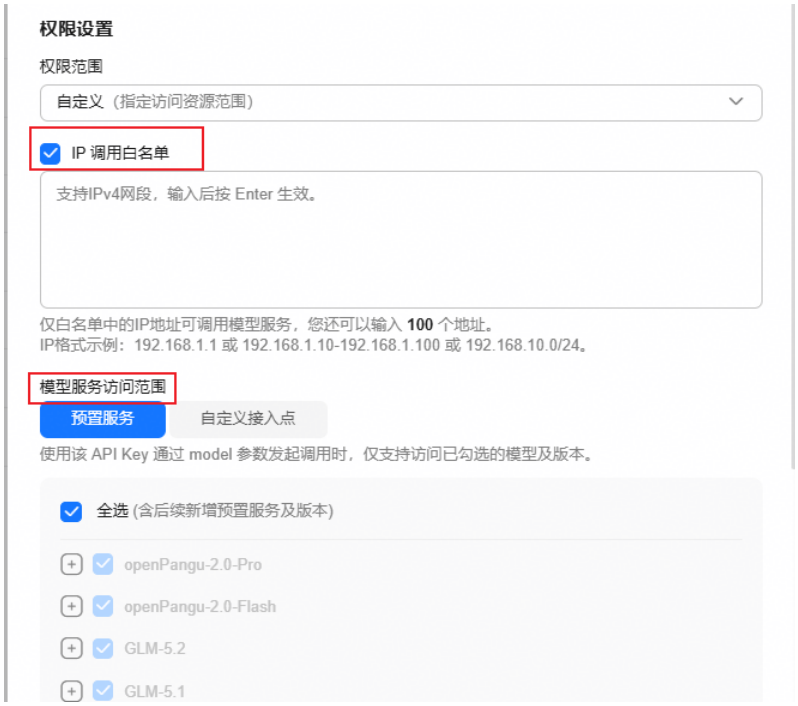
可分析

- **安全威胁检测**
  主动识别安全威胁与入侵行为并生成告警。
- **安全事件溯源**
  汇聚安全日志，展示威胁态势，支持安全事件全链路回溯与定位。
- **异常事件告警**
  基于SMN对流控、异常请求配置短信、邮件、语音等告警。

可控制

- **API Key**：
  - 支持限定IP调用白名单和模型服务访问范围。在创建API Key时，支持用户指定访问资源范围，如下图所示：

    图 1-2 API Key 权限设置

  - 定期轮换：建议用户定期创建并使用新的API Key，删除旧API Key。
  - 明文仅创建时展示一次，平台不留存，丢失不可恢复；建议立即导入凭据托管（CSMS/DEW）。
  - 建议按应用/环境隔离凭据边界，不同应用、不同环境使用相互独立的API Key。
- **内容审核**：
  输入/输出文本/图片/视频审核，其中文本审核支持敏感词审核，以及涉政、涉敏、暴恐、违禁、辱骂、广告等语义审核。客户等级V2及以上用户可自主控制开关，如下图所示：

图 1-3 内容安全护栏开关

- **自定义接入点**：

  MaaS支持自定义接入点功能，通过创建独立的调用入口，允许用户设置限流规则，并基于自定义接入点名称实现费用的精准统计，帮助用户高效管理推理服务资源，优化使用成本。在创建自定义接入点时可自主配置限流，支持按时段精细流控。如下图所示：

图 1-4 限流配置

- **操作权限**：

  用户级API Key增删改权限细粒度控制，避免无权限的子用户删除/编辑API Key，具体操作请参考**管理子用户的API Key操作权限**。

链路安全

- **公网链路**：

  公网访问支持HTTPS/TLS加密传输，且模型自身支持防御提示词攻击、指令注入；MaaS服务内置WAF防火墙、公网高防、漏洞扫描、态势感知，提供入侵检测和安全防御能力。

- **内网链路**：

  通过VPCEP单向内网通道访问，数据不暴露在公网。具体操作请参考**内网访问MaaS**。

- **合规认证**：

  严格遵循国内外数据安全法规，当前已获得国内外权威认证（云平台获等保四级认证，具备网络安全等级保护）。详情请参考**认证证书**。

2 模型安全

面向模型推理的输入与输出，提供内容合规审核和风险拦截能力，覆盖自营预置模型与云商店三方模型两类场景。

MaaS 提供的风控能力

MaaS除了本身具有算力底座安全优势外，还通过各种风控能力保障模型安全。

表 2-1 MaaS 具备的风控能力

风控能力	说明
模型原生安全	模型本身具备的内置安全能力，由模型厂商提供，默认生效，无需配置。
内容安全护栏	MaaS在推理链路额外提供输入/输出有害内容拦截，支持**文本、图片与视频审核**。文本审核覆盖敏感词以及涉政、涉敏、暴恐、违禁、辱骂、广告等语义审核。
安全护栏开关控制	**注意** 开启内容安全护栏会引入额外时延；关闭后可能无法拦截部分违规内容。 客户等级V2及以上支持自主关闭内容安全护栏，关闭后仅依赖模型原生安全。具体操作如下： ● 进入**在线推理**预置服务界面，在操作列选择“更多 -> 关闭内容安全护栏”。 ● 或者在创建/编辑自定义接入点时勾选“内容安全护栏”。
风险识别	当模型检测到输入或输出内容可能包含敏感信息时，模型会返回HTTP 403状态码及ModelArts.81011错误码，便于客户端识别和处理。

风控检测方式

MaaS内容安全护栏会针对推理链路中的输入和输出环节进行检测，流程如下：

- **输入检测**：用户请求到达MaaS网关后，用户输入内容先送至内容安全护栏检测，命中风险规则时请求不继续调用模型。
- **输出检测**：输入检测通过后请求正常调用模型，模型返回内容再次送检，命中规则时模型输出内容被相应处理。
- **处置策略**：命中风控时默认为**拦截**（返回ModelArts.81011错误码），未命中则正常返回。

预置模型与三方模型的风控策略

MaaS当前提供**自营预置模型**和**云商店三方模型**。对于不同来源的模型有不同的风控策略。

- **自营预置模型**
  - 推理链路**默认开启**内容安全护栏，对输入与输出双向检测。
  - 命中风险内容时请求被拦截，返回ModelArts.81011。
  - V2及以上客户可自主关闭，关闭后使用模型原生安全能力。
- **云商店三方模型**

  针对云商店提供的三方模型API服务，遵循“**平台兜底、责任共担**”原则：
  - **责任界定**：三方模型由云商店第三方商家提供，签约主体为“您与第三方商家”。华为云不对三方模型的性能、可用性、合规性承担担保责任。
  - **平台侧防护**：MaaS服务内置WAF防火墙、公网高防、漏洞扫描、态势感知，对平台链路上三方模型API的输入/输出进行底线防护。
  - **客户义务**：使用三方模型前应进行独立安全评估，遵守《云商店通用商品用户协议》、《云商店用户隐私声明》及三方服务支持条款。
  - **信息共享**：首次调用三方模型API时授权华为云创建云商店服务关联角色，用于查询商品信息与调用服务；华为云可能与三方商家共享调用信息以提供技术支持。

安全防护能力

MaaS在推理链路上构建了分层安全防护体系，按能力维度归为以下三类：

- **网络安全防护层**

  面向公网调用链路，在MaaS网关侧提供入侵防护与流量清洗能力，拦截恶意请求与攻击流量：
  - WAF防火墙：防护注入、爬虫、恶意Payload，关注大报文/超长上下文特征。
  - 公网高防 + Anti-DDoS：ELB高防，防护DDoS攻击。
  - APIG网关：统一入口，集中TLS、限流限并发、访问控制与观测。
- **安全检测与响应层**

  面向平台整体安全态势，提供威胁检测、漏洞管理与事件溯源能力：

  - 漏洞扫描与态势感知：主动识别安全威胁与入侵行为并生成告警。
  - 安全云脑：汇聚安全日志，展示威胁态势，支持安全事件全链路回溯与定位。
- **安全审计与监控层**

  面向操作行为与运行指标，提供审计留痕、指标监控与异常告警能力，详情请参见**操作审计与模型监控**。

风控免责说明

关于内容审核与风控，有以下事项需要您知悉：

- **模型原生安全由模型厂商提供**：MaaS平台上的模型（含开源模型、三方模型）的原生安全能力由模型厂商负责，华为云不对模型原生安全的合规性、准确性、完整性做出担保。对于开源模型，华为云不对其原生安全能力承担责任。
- **内容安全护栏不能保证100%过滤与合规**：MaaS提供的内容安全护栏是基于现有技术手段的辅助风险识别与拦截能力，受内容风控技术客观局限性影响，无法保证对所有违规内容均能识别和拦截。您应结合自身业务场景对输入与输出内容进行独立的内容合规审核，不应将平台内容安全护栏作为唯一的合规保障手段。
- **三方模型免责**：三方模型及输出不代表华为云观点，华为云不对其合规性、安全性、真实性、准确性、完整性做出担保，由此引发的损害由您与三方商家之间解决。
- **关闭护栏的风险**：关闭内容安全护栏后，因模型输出违规内容引发的法律责任与第三方索赔由您自行承担，建议在生产环境谨慎关闭并配套自建内容审核机制。

详情请参见**《MaaS模型即服务服务声明》**。

常见问题

Q1：关闭护栏后会有什么影响?

关闭后将仅使用模型原生安全能力，可能无法拦截部分违规内容，因模型输出违规内容引发的法律责任由您自行承担。

Q2：命中风控拦截后如何处理?

命中拦截时模型会返回HTTP 403状态码及ModelArts.81011错误码，您可更改请求内容（如Prompt）后重试。

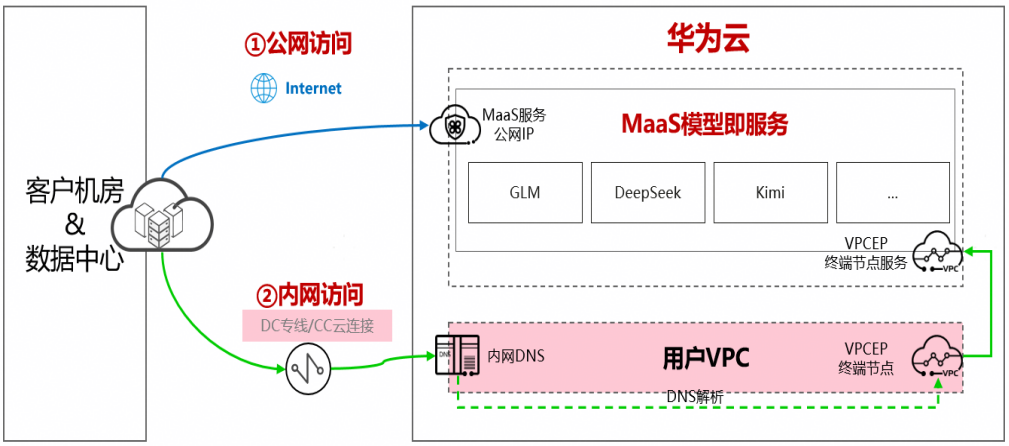
3 内网访问 MaaS

面向对数据出公网敏感的场景，MaaS提供VPCEP内网通道，使调用流量在内网闭环、不经过公网。

VPCEP 内网通道原理

下图展示了公网访问MaaS与内网访问MaaS的不同路径与处理方式。

图 3-1 公网访问与内网访问示意图



在您的VPC内创建接口终端节点（VPCEP），与MaaS服务建立**单向**私网连接。VPC内资源通过终端节点访问MaaS，流量全程留在华为云内网，不经过公网。MaaS无法通过该连接反向访问您的VPC资源。

- **单向访问**：网络ACL策略只允许您单向访问MaaS，MaaS无法反向访问您的VPC。
- **双路接入**：MaaS对外提供双路接入，需创建两个VPCEP，内网DNS解析到这两个VPCEP上以实现高可用。
- **支持地域**：西南-贵阳一。
- **终端节点不具备公网访问能力**，也无法绑定EIP。

操作步骤

通过VPC内网访问MaaS的操作步骤如下：

步骤1 创建VPC终端节点。

须知

- VPC所在的区域和MaaS业务部署的区域一致。
- 为保障高可靠，需创建两个VPCEP。MaaS会提供两个VPCEP ServiceID，基于这两个ServiceID各创建一个VPCEP。
- 终端节点按购买时长计费。

1. 进入**VPC终端节点**控制台，单击“购买终端节点”，进入“购买终端节点”页面。
2. 请参考**购买终端节点**进行配置。配置时服务名称选择MaaS提供的VPCEP ServiceID（ServiceID需**提交申请**获取）。如下图所示：

3. 创建成功后记录VPCEP IP，如下图所示：

4. 重复**步骤1**~**步骤2**创建第二个VPCEP。

步骤2 配置内网DNS解析。

1. 进入**DNS控制台**，单击“内网域名”进入“内网域名”页面。
2. 请参考**创建内网域名指导**创建内网域名（例如“api.modelarts-maas.com”），创建成功后，列表页会展示您刚创建的域名。

注意

- 一定要勾选“子域名递归解析代理”，否则在添加解析前会中断服务。
- VPC关联时，所选“区域”和“VPC”需与**创建VPCEP**时一致。

图 3-2 创建访问 MaaS 服务的内网域名

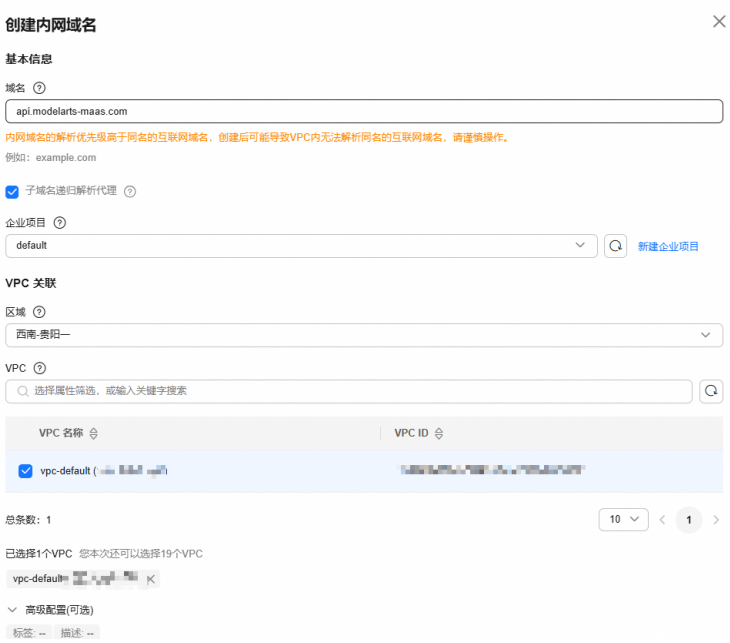


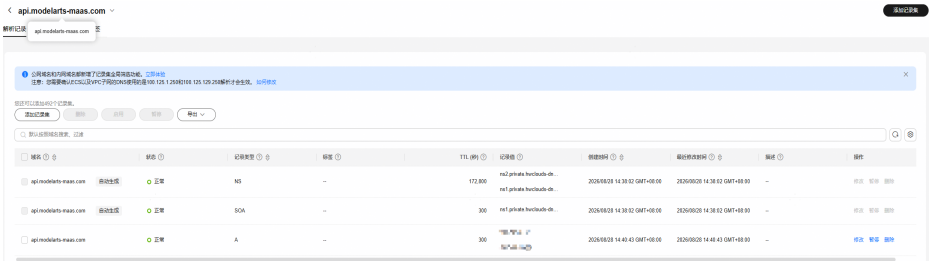
图 3-3 内网域名列表

3. 单击域名名称，默认进入解析记录页，单击“添加记录集”弹出“添加记录集”弹窗，按下表进行配置，单击“确定”，配置成功后，记录集列表会显示您刚添加的记录集。

表 3-1 添加记录集参数配置

参数	取值
记录类型	A – 将域名指向IPv4地址
TTL (秒)	默认300秒
记录值	填写**步骤1**创建的两个VPCEP的IP地址。

图 3-4 解析记录列表



步骤3 调用MaaS模型API。

1. 登录业务部署的节点。
2. 执行如下命令验证内网连通性。
   ```
   ping api.modelarts-maas.com
   ```
3. 执行如下代码调用MaaS模型API。如果请求成功响应则表示内网链路已打通。
   ```
   curl -X POST "https://api.modelarts-maas.com/v2/chat/completions" \
     -H "Content-Type: application/json" \
     -H "Authorization: Bearer $MAAS_API_KEY" \
     -d '{
       "model": "glm-5.2",
       "messages": [
         {"role": "system", "content": "You are a helpful assistant."},
         {"role": "user", "content": "介绍下你自己"}
       ]
     }'
   ```

----结束

4 数据处理与隐私承诺

什么是数据?

用户在MaaS中的数据包含用户发布的**AI数据**以及用户提供的**个人数据**。

AI数据包括但不限于文本、图形、文章、照片、图像、插图、代码、AI算法、AI模型等。

个人数据包括：

- 用户注册时提供的昵称、头像、邮箱、手机号。
- 用户注册时提供的企业信息。
- 用户申请白名单时提供的华为云账号ID、邮箱、企业信息。

数据处理

MaaS对用户发布的AI数据做统一保存管理：文件类数据存放在MaaS官方OBS桶内；用户提供的个人数据保存在数据库中，其中敏感信息（如手机号、邮箱等）在数据库中进行加密处理。

MaaS提供了数据与隐私保护声明，主要对以下方面进行保障，具体承诺请参见**《MaaS模型即服务服务声明》**。

表 4-1 数据处理与隐私承诺项

承诺项	说明
数据不用于训练	在未获得您明确授权情况下，不使用您的内容（含提示词、模型输出）用于模型训练、优化等。
不缓存/不存储/不留存	用户输入数据不缓存、不存储、不留存；推理交互数据仅做处理，不进行存储，服务完成后立即删除。
数据处理与跨境策略	在中华人民共和国境内处理您的内容，服务完成后立即删除。
不向第三方披露	除提供服务外，在未获得用户明确授权的前提下，不向第三方授权使用及披露用户的内容。

承诺项	说明
合规审核权	华为云有权采取人工或技术方式对您的内容进行安全合规审核，违规内容可被处置（披露、删除、阻止）。
静态数据保护	在征得用户同意的前提下，MaaS收集的用户个人数据中的敏感信息，如用户邮箱和手机号，MaaS在数据库中进行加密处理。其中，加密算法采用了国际通用的AES算法。
数据隔离机制	数据存储是按照租户隔离，租户之间互相看不到数据。可通过主账号给不同子账号授予不同业务空间权限，不同业务空间的数据互不影响，从而实现隔离效果。

用户需要遵守的安全准则

- 您应对您的内容（含提示词、模型输出）的安全合规负责，确保不包含色情、暴力、毒品等不良信息。
- 确保内容来源合法，不侵犯他人知识产权、商业秘密和个人信息。
- 若内容含个人信息，应进行匿名化处理或取得相应授权。
- 对模型输出及使用行为承担责任，商业化使用前应仔细评估。
- 若服务具有舆论属性或社会动员能力，应依法开展安全评估和算法备案。

常见问题

Q1：MaaS是否保存模型调用时产生的数据?

华为云严格保护数据隐私，**不会将您的数据用于模型训练、优化等**。推理交互数据仅做处理，**不进行存储，服务完成后立即删除**。根据相关法律法规要求，管理面操作记录（如API Key创建/删除/编辑）通过CTS云审计留存以支持审计追溯，调用统计指标（非内容本身）通过 CES留存以支持监控查询。

Q2：缓存命中是否有内容和隐私安全问题?

MaaS推理链路对用户输入数据不缓存、不存储、不留存，服务完成后立即删除，因此**不存在基于您输入的缓存命中导致内容或隐私泄露的风险**。若您在自有侧（如应用网关、CDN）部署缓存，应在缓存层对敏感内容设置不缓存或按权限隔离策略。

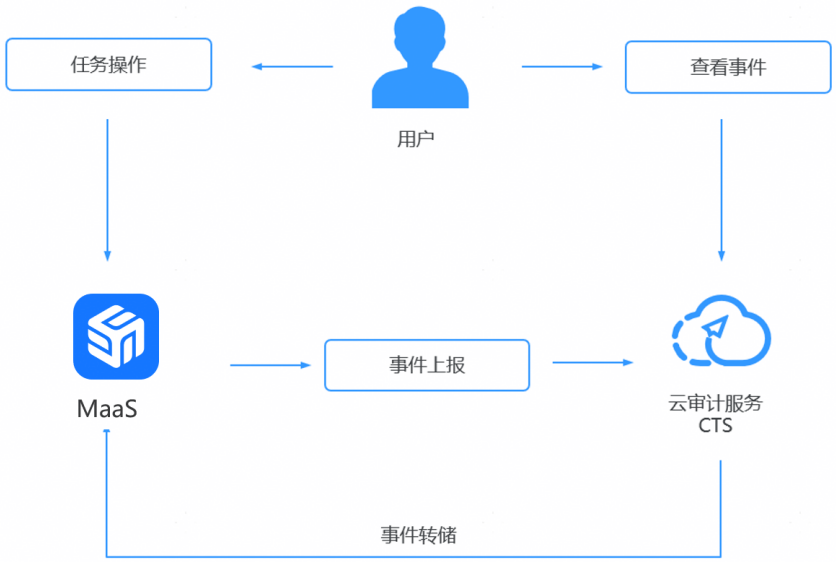
5 操作审计与模型监控

操作审计

云审计服务（Cloud Trace Service，CTS），是华为云安全解决方案中专业的日志审计服务，提供对各种云资源操作记录的收集、存储和查询功能，可用于支撑安全分析、合规审计、资源跟踪和问题定位等常见应用场景。

用户开通云审计服务并创建和配置追踪任务后，CTS可记录MaaS的管理事件和数据事件用于审计。事件审计流程如下：

图 5-1 云审计服务



CTS支持追踪的MaaS管理事件和数据事件列表，请参见**表5-1**。

表 5-1 MaaS 支持审计的关键操作列表

操作名称	资源类型	事件名称
创建API Key	api_key	CreateApiKey
删除API Key	api_key	DeleteApiKey
更新API Key	api_key	UpdateApiKey
签署服务声明	compliance	SignDisclaimer
获取监控信息	metrics	QueryServiceMetrics
创建推理服务	job	CreateService
删除推理服务	job	DeleteService
更新风控配置	job	UpdateModeration
开通模型服务	job	ActivateModelServices
更新推理服务配置	job	UpdateService
查询新版在线服务	job	ListUserInferV2ServicesByRegions
创建接入点	job	CreateCustomEndpoint
启停接入点	job	ControlCustomEndpoint
更新接入点	job	UpdateCustomEndpoint
删除接入点	job	DeleteCustomEndpoint

- CTS的详细介绍和开通配置方法，请参见**CTS入门指引**。
- 开启云审计服务后，系统会记录MaaS的相关操作，控制台默认保存最近7天的操作记录。查看MaaS操作事件记录请参考**查看审计日志**。

模型调用统计与监控

- **模型调用统计**

  统计预置服务、自定义接入点在指定时间段内的调用数据，如总调用Tokens数、输入Tokens数、输出Tokens数、端到端时延等信息，并以分钟为最小时间粒度展示数据趋势，帮助您了解服务的使用情况和性能变化，从而更有效地进行模型评估、问题定位、故障排除和性能优化。统计详情请前往**调用统计**页面查看。

- **云监控 CES**

  MaaS对接了云监控服务（CES），监控在线服务和对应模型负载，如RPM/TPM、Token用量、模型调用成功率/失败率等监控指标，且支持指标仪表盘可视化。监控指标详情请参见**监控指标说明**。

- **SMN 告警**

  MaaS对接了消息通知服务（SMN），支持对流控、异常请求（错误码等）配置短信、邮件、语音等告警方式。

- **运行日志记录**
  客户侧网关数据可记录服务调用、Token上下文等日志，按需存入OBS实现过程追溯。

6 认证证书

合规证书

华为云服务及平台通过了多项国内外权威机构（ISO/SOC/PCI等）的安全合规认证，用户可自行**申请下载**合规资质证书。

图 6-1 合规证书下载

资源中心

华为云还提供以下资源来帮助用户满足合规性要求，具体请查看**资源中心**。

图 6-2 资源中心

合规资质证书

华为云安全服务提供了网络安全专用产品安全检测证书、软件著作权等证书，供用户下载和参考。具体请查看**合规资质证书**。

图 6-3 网络安全专用产品安全检测证书&软件著作权证书