

云原生智能网关

常见问题

产品文档



【版权声明】

©2013–2026 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

【商标声明】

及其他腾讯云服务相关的商标均为腾讯集团下的相关公司主体所有。另外，本文档涉及的第三方主体的商标，依法由权利人所有。

【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

文档目录

常见问题

云原生网关常见问题

AI 网关常见问题

常见问题

云原生网关常见问题

最近更新时间：2026-09-22 18:50:07

本文汇总了云原生网关在使用过程中的常见问题及解答。

产品咨询类

云原生网关是什么？

云原生网关是腾讯云基于开源微服务网关（Kong）推出的一款高性能高可用的网关产品。作为云上微服务架构的流量入口，集成请求分发、API 管理、流量监控、访问限制等功能，提供：

- **流量网关**：南北向流量统一接入与智能调度。
- **微服务网关**：东西向流量治理与服务间通信。
- **安全网关**：全链路安全防护与认证鉴权。

云原生网关和 AI 网关有什么区别？

- **云原生网关**面向传统微服务与云原生流量治理场景，解决南北向、东西向流量的接入、路由、限流、熔断、安全防护等问题。
- **AI 网关**面向大模型与智能化场景，专注于解决企业接入、调度和管理多种 AI 模型时面临的协议复杂、治理困难、成本不可控及存量业务改造门槛高等核心问题。

两者同属云原生智能网关产品，可按业务需要独立选用。

云原生网关提供托管服务是独享实例吗？

云原生网关为每个用户提供物理隔离、资源独占的实例，保障性能与安全。

购买与计费类

如何购买云原生网关？

登录 [云原生网关](#) 控制台，在云原生网关相关页面进行实例新建与购买。云原生网关支持按节点规格和数量购买，也支持在购买后调整节点规格和数量。

云原生网关目前的新购限制是怎样的？

目前云原生网关仅对微服务场景支持开通白名单新购，其余场景不支持新接入。

云原生网关有哪些节点规格？

云原生网关提供 2 核 4G、4 核 8G、8 核 16G、16 核 32G 等多种节点规格，您可以根据业务需求选择合适的资源大小，并支持购买后进行节点规格变更和节点数量变更。

使用类

云原生网关创建实例时选择的 VPC 有什么作用？

云原生网关创建实例时选择的 VPC 主要用于与客户其他资源进行通信。网关会从该 VPC 子网中使用一个 IP 作为网关的内网入口。同时，当网关访问后端服务时，会根据网关实例节点的数量，从子网中使用相应数量的 IP 用于与 VPC 中其他资源进行通信。因此，请确保所选子网有足够的可用 IP。

云原生网关支持哪些服务来源？

云原生网关通过服务来源关联后端服务，支持的服务来源包括：

- K8s 服务（使用 TKE 或 EKS）。
- 注册中心服务（使用 TSF 或自建的 Nacos/Consul/Polaris）。
- 容器服务（TKE、EKS）。
- Serverless 云函数（SCF）。
- 腾讯云服务器（CVM）上的服务。
- 未依赖任何服务发现机制的域名、固定 IP、IP 列表。
- 私有域名解析（PrivateDNS）。

云原生网关支持哪些协议？

云原生网关支持 TCP、UDP、HTTP/HTTPS、WebSocket，以及 HTTP/HTTPS 转 gRPC、HTTP/HTTPS 转 Dubbo 等协议转换。

云原生网关支持哪些管理方式？

云原生网关支持腾讯云控制台、Ingress、云 API、原生控制台（Konga）、Terraform 等多种方式，您可按使用习惯进行选择。

云原生网关如何保障高可用？

接入点（CLB）、网关节点、网关配置均支持跨多可用区部署，关键链路均实现跨可用区容灾；并提供开源增强的限流、灰度、熔断、流量镜像等流量防护能力，支持服务无损上线，支持模拟故障进行混沌演练。

云原生网关如何进行安全防护？

云原生网关支持以下安全能力：

- 对接腾讯云 Web 应用防火墙（WAF），支持实例级、服务级、路由级防护。
- 对接腾讯云 DDoS 高防（高防 IP 或高防包）。
- 使用腾讯云证书管理平台（SSL）进行证书管理。
- 支持常用 API 认证鉴权（用户密码、Key 鉴权、JWT、OAuth2 等）和自定义鉴权。
- 支持访问 IP 黑白名单控制。

云原生网关支持哪些监控和日志能力？

支持在控制台查看请求数、时延、CPU、内存、带宽、连接数等监控指标并配置自定义告警，支持对接 Prometheus；支持在控制台查看网关请求日志和错误日志，支持日志投递到腾讯云日志服务（CLS）进行分析，也支持投递到 CKafka。

性能与规格类

云原生网关的推荐运行水位是多少？

- 推荐水位：CPU 使用率 30%，内存使用率 30%。
- 警戒水位：CPU 使用率 60%，内存使用率 60%。

建议在日常运行中将资源使用率控制在 30%–60% 之间，既能保持较高的资源利用率，又能应对请求突增。在警戒水位下，网关可获得完整的 SLA 保障；当超过警戒水位时，网关进入高负载状态，请求成功率和耗时可能受影响，建议配置告警。

不同节点规格的容量指标是怎样的（单节点警戒水位）？

节点规格	2核4G	4核8G	8核16G	16核32G
最大连接数	48000	96000	192000	384000
HTTP 每秒新建连接数	4000	8000	16000	32000
HTTPS 每秒新建连接数	1200	2400	4800	9600

增加网关容量有哪些方式？

增大节点规格或增加节点数量都可以提高网关容量。在节点规格不变的情况下，网关容量随节点数量近似线性增长。

弹性伸缩类

云原生网关弹性伸缩支持哪些类型？

弹性伸缩支持指标规则和定时规则两种类型的弹性伸缩策略（仅专业版支持）：

- 指标规则：适用于突发流量和典型周期性流量场景，由指标名称（CPU 使用率、内存使用率、QPS）、阈值描述弹性扩缩容的触发条件和节点数量变化。
- 定时规则：适用于资源使用率有周期性规律的场景，达到设定的周期时间点后自动调整实例节点数量。

弹性伸缩策略创建后为什么没有生效？

弹性伸缩策略需绑定集群分组后才能生效，支持一次绑定一个或多个集群分组。请在弹性伸缩策略列表页，单击目标策略右侧的“绑定分组”完成绑定。

同时配置指标伸缩与定时伸缩会如何生效？

当两种策略同时配置时，网关会结合两种策略内容进行节点数量调整。定时伸缩并不直接进行节点数调整，而是通过调整指标伸缩节点数范围的方式来进行伸缩判断与调整，最终节点数会根据调整后的指标伸缩节点数范围进行相应调整。

迁移类

如何从自建 Kong 网关迁移到云原生网关？

可利用 Kong 官方的 `deck` 工具完成配置迁移，主要步骤如下：

1. 使用 `deck dump` 导出自建 Kong 网关的配置（services、routes、plugins）。
2. 使用 `deck diff` 查看配置差异，确认无误后使用 `deck sync` 将配置导入到云原生网关 Kong。
3. 在云原生网关控制台查看配置是否导入成功。

如何从 Spring Cloud Gateway（SCG）迁移到云原生网关？

主要迁移步骤如下：

1. **明确服务来源：**如果当前服务来源为 K8s 服务、注册中心服务（TSF/Nacos/Consul/Polaris）或域名/固定 IP/IP 列表，可直接迁移；其他类型注册中心需先将服务注册到 Nacos/Consul/Polaris。
2. **迁移配置：**SCG 的 service/routes 对应云原生网关的服务/路由；Predicate 通过路由匹配条件或灰度规则实现；Filter 通过 `tse-proxy-rewrite` 插件、限流能力、自定义插件等实现。
3. **配置监控告警。**
4. **流量迁移：**按灰度方式逐步切换流量，先非核心业务小规模验证，再扩展到核心业务，最后切换域名解析进行全量迁移。

云原生网关相比自建网关有哪些优势？

对比项	云原生网关	自建网关
搭建及运维成本	资源全托管免运维，支持规格/节点数/计费模式变更、版本升级、弹性伸缩	需自行采购资源搭建，人工成本高
高可用	接入点、节点、配置均支持跨可用区容灾，提供完善流量防护	需自行探索开发高可用体系
易用性	支持控制台、Ingress、云 API、Konga 等多种方式	需自行搭建运维
安全性	对接 WAF、DDoS 高防、SSL 证书，支持认证鉴权与黑白名单	需自行对接相关云产品
可观测性	控制台查看监控指标并告警，支持 Prometheus、CLS	仅基础监控，需自行搭建
云产品集成	与 TSF、TKE、SCF、CVM、WAF、CLS、SSL 等深度集成	不支持

AI 网关常见问题

最近更新时间：2026-09-22 18:50:08

本文汇总了 AI 网关在使用过程中的常见问题及解答。

产品咨询类

AI 网关是什么？

AI 网关面向大模型与智能化场景，提供大模型网关（多模型统一接入与智能调度）、MCP 网关（传统应用 AI 化改造与协议转换）能力。

云原生网关和 AI 网关有什么区别？

- 云原生网关面向传统微服务与云原生流量治理场景，解决南北向、东西向流量的接入、路由、限流、熔断、安全防护等问题。
- AI 网关面向大模型与智能化场景，专注于解决企业接入、调度和管理多种 AI 模型时面临的协议复杂、治理困难、成本不可控及存量业务改造门槛高等核心问题，提供模型统一接入、智能路由、协议转换（MCP/OpenAI）、Token 级限流等能力。

两者同属云原生智能网关产品，可按业务需要独立选用。

AI 网关使用类

AI 网关能解决什么问题？

AI 网关专注于解决企业接入、调度和管理多种 AI 模型时面临的协议复杂、治理困难、成本不可控及存量业务改造门槛高等核心问题。它作为企业智能化架构的流量入口与治理中枢，通过统一的协议适配、智能的路由调度与全方位的可观测能力，帮助企业高效、安全、经济地集成与使用 AI 能力。

AI 网关支持哪些协议？

AI 网关 100% 兼容开源网关生态，并全面适配 AI 领域标准协议，支持 MCP、OpenAI、SSE 等，提供传统 RESTful、gRPC 等协议的转换，实现一套网关覆盖所有流量。

如何通过 AI 网关调用混元等大模型？

核心流程为六个步骤：

1. **创建模型密钥**：在网关中安全配置访问大模型所需的 API 密钥。
2. **创建模型服务**：添加模型供应商（如混元），并关联模型密钥，选择模型协议（如 OpenAI 兼容）。
3. **创建模型 API**：创建对外提供能力的 API，绑定模型服务，配置请求协议、路由和 Base Path。
4. **创建消费者**：创建调用方身份，并为其添加 API Key 凭证。
5. **创建消费者组并授权**：将消费者分组，并授予访问模型 API 的权限。

6. **获取访问地址并发起调用**：在控制台获取网关入口地址与请求路径，使用消费者凭证（API Key）发起调用。

调用示例：

```
curl -i -X POST <访问地址> \
  -H "Content-Type: application/json" \
  -H "Authorization: Bearer <API_KEY>" \
  -d '{
    "model": "<MODEL_NAME>",
    "messages": [
      {"role": "user", "content": "你好，请介绍一下你自己。"}
    ]
  }'
```

如何使用自定义模型供应商对接第三方模型平台

1. 在第三方模型平台创建 API Key 。
2. 在 AI 网关创建模型密钥，粘贴第三方 API Key。
3. 新建模型服务，选择"自定义供应商"→ 填入第三方 OpenAPI 兼容端点 → 选择默认模型。
4. 新建模型 API，指定 Base Path 与请求协议，绑定模型服务。
5. 在消费者组中授权该模型 API。
6. 使用消费者 API Key 通过网关公网入口发起调用。

AI 网关如何控制模型调用成本和保障稳定性？

AI 网关提供从消费者、API 到模型等多个维度的限流、熔断、降级能力：

- 支持 Token 级限流（TPM/TPH/TPD）、QPM 级限流、并发数限流。
- 支持多模型权重路由、模型名称路由等智能调度策略。
- 支持模型服务间 Fallback 配置，主模型服务异常时自动切换备用服务。
- 提供 API 调用延迟、Token 消耗、模型费用等多维度监控，助力运维与成本优化。

AI 网关如何实现多租户隔离与权限管理？

AI 网关通过消费者、消费者组的多级权限模型，为不同租户分配独立 API Key，实现 AI 能力在不同团队、项目间的安全隔离与便捷共享，支撑平台化运营。

什么是 MCP 网关？

MCP 网关面向传统应用 AI 化改造场景，通过协议转换引擎，将存量业务系统提供的标准 API 自动包装成 AI 应用可调用的标准化工具（MCP Tool），实现业务能力的"零代码改造"AI 化，使存量系统快速具备被 AI Agent 以 MCP 协议调用的能力。