应用型负载均衡

操作指南

腾讯云

版权所有：腾讯云计算（北京）有限责任公司

【版权声明】

©2013-2026 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【商标声明】

腾讯云

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【服务声明】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。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。

【联系我们】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或95716。

文档目录

版权所有：腾讯云计算（北京）有限责任公司

操作指南
ALB 实例
创建 ALB 实例

最近更新时间：2026-06-26 17:23:11

本文介绍如何创建 ALB 实例。

前提条件

应用型负载均衡（ALB）当前处于公测阶段，暂未全面开放。使用前需申请开通资格。请提交 工单申请 。

操作步骤

1. 登录 ALB 购买页 。
2. 按需选择以下 ALB 相关配置。

参数	说明
计费模式	支持按量计费。
地域	选择 ALB 实例所属地域，地域支持情况请以控制台页面为准。
网络类型	网络类型分为公网和内网两种类型。 • 公网：使用 ALB 分发来自公网的请求，公网 ALB 通过 EIP 提供公网服务，EIP 所产生费用详见 EIP 计费文档 。 • 内网：使用 ALB 分发来自腾讯云内网的请求。选择内网类型无需配置网络计费模式。
IP 版本	支持 IPv4。
所属网络	ALB 实例和目标组需位于同一个 VPC，ALB 实例创建后，VPC 无法修改。
可用区	多可用区部署：若所选地域支持多可用区，建议至少选择 2 个可用区并配置对应子网。 公网访问：当实例网络类型选择公网时，支持以下两种方式： • 绑定已有 EIP：为实例关联已有的弹性公网 IP。 • 自动分配公网 IP：系统将创建按使用流量计费的 EIP 并绑定至 ALB 实例。 EIP 一致性约束：同一个 ALB 实例下，不同可用区绑定的 EIP 类型需保持一致。
网络计费模式	支持按使用流量、共享带宽包。 • 按使用流量：按实际使用的流量计费，适合流量波动较大的场景。 • 共享带宽包：多个 EIP 共享带宽额度，适合多 IP 聚合计费，可降低公网费用。

	若在可用区处选择的 EIP 类型均为按使用流量计费，支持选择加入共享带宽包。
共享带宽包	若网络计费模式选择共享带宽包，则需要选择要加入的共享带宽包。
实例名	仅支持字母、数字、中文、句号、下划线、短划线，不能超过80个字符。
标签	选择标签键和标签值，也可选择添加标签，详情请参见 创建标签 。

3. 完成以上参数配置，勾选**服务协议与公测说明**后，单击**立即购买**创建实例。

删除 ALB 实例

最近更新时间：2026-06-26 17:23:31

本文介绍如何删除 ALB 实例。

前提条件

当您确认 ALB 实例已无流量，监控中出入带宽、出入流量已归0，不需要继续使用后，您可以删除 ALB 实例。

操作步骤

> **注意：**
> 删除后，ALB 实例本身、关联的监听器、转发规则等配置将被彻底清除且不可恢复，请谨慎操作。

1. 登录 应用型负载均衡控制台，在左侧导航栏选择**应用型负载均衡 ALB > 实例管理**。
2. 在实例列表上方选择地域，找到目标实例，单击最右侧操作栏下的**更多 > 删除**。
3. 弹出最终确认对话框，确认需要删除该实例后，单击**确定**即可删除。

版权所有：腾讯云计算（北京）有限责任公司

ALB 监听器
创建和管理监听器

最近更新时间：2026-07-02 15:10:23

监听器（Listener）用于检查客户端的连接请求，并根据您配置的转发策略将请求转发到后端目标组。您可以在应用型负载均衡（ALB）实例上添加 HTTP 或 HTTPS 监听器，对来自客户端的七层请求进行识别和转发。
HTTP/HTTPS 协议适用于需要对请求内容进行识别的应用，如 Web 应用、App 服务等。
本文介绍如何在 ALB 实例上创建、编辑、管理和删除监听器。

前提条件

- 已创建应用型负载均衡 ALB 实例。如未创建，请参见 创建实例 。
- 已创建目标组，并向目标组中添加了后端服务。如未创建，请参见 创建和管理目标组 。ALB 监听器通过转发到目标组的方式向后端服务转发请求，因此在创建监听器前请先准备好目标组。
- 如需创建 HTTPS 监听器，请确保已在证书管理中准备好服务器证书。可以选择 SSL 证书平台 中已有的证书，或新建上传证书。

创建监听器

步骤1：配置监听器

1. 登录 应用型负载均衡控制台 ，在左侧导航栏选择**应用型负载均衡 ALB > 实例管理**。
2. 在实例列表上方选择地域，单击目标实例的 ID/名称进入实例详情页。
3. 选择**监听器管理**页签，单击**创建监听器**。
4. 配置基础参数如下：

配置项	说明
监听协议	监听器使用的协议，支持 HTTP 和 HTTPS 两种。
名称	监听器的名称，不超过 255 个字符。留空时系统将自动生成名称。
监听器端口	监听器用来接收请求并向后端转发请求的端口，端口范围为 1 - 65535。同一个 ALB 实例内，相同协议的监听端口不可重复。
服务器证书	HTTPS 加密通信所使用的服务器证书，ALB 使用该证书与客户端完成 TLS 握手。可以选择 SSL 证书平台 中已有的证书，或新建上传证书。
TLS 安全策略	指定 HTTPS 监听器支持的 TLS 协议版本与加密套件，用于控制加密通信的安全等级。如需自定义协议版本与加密套件，请参见 TLS 安全策略 。

双向认证	默认关闭。开启后，除服务端验证外，客户端也需提供证书供 ALB 校验，实现双向身份验证，适用于对安全性要求较高的场景。开启后需配置 CA 证书。
HTTP 2.0	默认关闭。开启后监听器支持 HTTP/2 协议，可提升页面加载性能与多路复用能力。
标签	选择标签键和标签值，也可选择添加标签，详情请参见 创建标签 。

5. 配置高级选项：单击**隐藏高级选项 / 展开高级选项**，可进一步配置以下参数。如无特殊需求，可保持默认值。

配置项	说明
连接空闲超时（秒）	连接在空闲状态下保持的最长时间，超过该时间无新请求时连接将被断开。输入范围为 1 - 600 秒，默认值 15 秒。
请求超时（秒）	ALB 等待后端响应的最长时间，超过该时间未收到响应则视为请求超时。输入范围为 1 - 600 秒，默认值 60 秒。
Gzip 压缩	开启后 ALB 对响应内容进行 Gzip 压缩，可减少传输流量、提升响应速度。
附加 HTTP 头字段	配置 ALB 向后端服务转发请求时附加的 HTTP 头字段，用于向后端传递客户端及负载均衡的相关信息。详见下方附加 HTTP 头字段说明。

附加 HTTP 头字段说明

通过 X-Forwarded-For 头字段获取来访者客户端 IP，支持以下三种处理方式（**单选**）：

- 附加：ALB 将请求转发给后端服务之前，把客户端 IP 加入到最后一跳的 XFF 头字段中。
- 删除：ALB 将请求转发给后端服务之前删除 XFF 头，无论请求是否携带 XFF 头字段。
- 透传：ALB 保持 X-Forwarded-For 头不变，直接透传给后端服务，不做任何修改。

此外，可按需勾选以下头字段（**多选**）：

头字段	说明
X-Forwarded-Proto	通过该头字段获取负载均衡的监听协议（HTTP/HTTPS）。
X-Forwarded-Port	通过该头字段获取负载均衡实例的监听端口。
X-Forwarded-Host	通过该头字段获取访问负载均衡实例客户端的域名。
X-Forwarded-Client-srcport	通过该头字段获取访问负载均衡实例客户端的端口。

X-Forwarded-Clientcert-subjectdn	仅 HTTPS 监听器支持。通过该头字段获取访问负载均衡实例客户端证书的所有者信息（subject DN）。
X-Forwarded-Clientcert-issuerdn	仅 HTTPS 监听器支持。通过该头字段获取访问负载均衡实例客户端证书的颁发者信息（issuer DN）。
X-Forwarded-Clientcert-fingerprint	仅 HTTPS 监听器支持。通过该头字段获取访问负载均衡实例客户端证书的指纹取值。
X-Forwarded-Clientcert-clientverify	仅 HTTPS 监听器支持。通过该头字段获取访问负载均衡实例客户端证书的校验结果。

步骤2：配置转发动作

转发动作配置项如下，单击**确认**后完成配置监听器默认转发动作，系统将返回监听器管理列表页。

配置项	说明
转发动作类型	监听器命中后执行的动作。当前支持转发到目标组，即将请求转发至指定目标组内的后端服务。
选择目标组	选择请求需要转发到的目标组。如尚未创建目标组，可单击**新建后端目标组**前往创建。目标组的协议需与监听器协议匹配。
目标组详情	选定目标组后，自动展示该目标组的目标组 ID、目标组名称及关联的私有网络信息，便于确认所选目标组是否正确。

编辑监听器

1. 在监听器管理列表中，找到目标监听器。
2. 单击**监听器 ID**进入监听器详情。
3. 单击**编辑监听器**按钮，可对监听器属性、转发动作进行修改。
4. 修改完成后单击**完成**。

管理监听器

管理标签

1. 在监听器管理列表中，找到目标监听器。
2. 在目标监听器操作列单击**更多 > 编辑标签**。
3. 修改完成后单击**确定**。

管理证书（仅 HTTPS 监听器）

1. 在监听器管理列表中，找到目标监听器。

2. 单击**监听器 ID**进入监听器详情，切换至**证书管理**页签。
3. 在**证书管理**页签中支持**添加扩展证书**、**替换证书**、**删除证书**。

修改 SSL 解析方式或 TLS 安全策略（仅 HTTPS 监听器）

1. 在监听器管理列表中，找到目标监听器。
2. 单击**监听器 ID**进入监听器详情，在证书信息中单击 **SSL 解析方式**或 **TLS 安全策略**后的**编辑**图标完成修改。

删除监听器

1. 在监听器管理列表中，勾选一个或多个待删除的监听器，单击列表上方的**删除**；或在目标监听器操作列单击**更多 > 删除**。
2. 在**确认**弹窗中确认后完成删除。

> **注意：**
> 删除监听器后，该监听器上的转发规则将同步删除，相关业务流量将无法继续转发，请谨慎操作。

配置转发规则

最近更新时间：2026-07-02 15:38:31

转发规则（Rule）用于定义应用型负载均衡（ALB）监听器如何处理客户端请求。
创建监听器后，系统会自动为该监听器创建一条默认转发规则。您也可以在监听器下自定义添加多条转发规则，将不同特征的客户端请求按指定条件路由到不同的后端目标组，实现基于内容的精细化流量分发。
本文介绍如何创建、编辑、删除转发规则，以及如何调整规则优先级。

转发规则说明

每条转发规则由转发条件和转发动作两部分组成：转发条件用于匹配客户端请求，转发动作定义匹配成功后对请求执行的处理。

默认转发规则

创建监听器后，系统会自动为该监听器创建一条默认转发规则。该转发规则的转发动作为将请求转发至创建监听器时配置的目标组。

- 默认转发规则不支持删除，但支持修改其转发动作（如更换目标组）。
- 默认转发规则的优先级最低，且不支持调整优先级。
- 当客户端请求未匹配到任何自定义转发规则时，将按默认转发规则进行转发。

匹配原理

客户端请求会按照转发规则的优先级顺序逐条匹配，优先级数值越小，优先级越高。

- 请求匹配到某条转发规则时，ALB 立即按该规则执行转发动作，不再继续匹配后续规则；
- 请求未匹配到任何自定义转发规则时，则按默认转发规则转发。

> **说明：**
> ALB 严格按照优先级数值从小到大依次匹配转发规则，不会自动按域名或路径的精确程度排序。如需让更精确的规则优先匹配，请手动将其优先级数值调小（即排在更靠前的位置）。

前提条件

- 已创建应用型负载均衡 ALB 实例。如未创建，请参见 创建 ALB 实例 。
- 已为 ALB 实例创建 HTTP 或 HTTPS 监听器。如未创建，请参见 创建和管理监听器 。
- 已创建目标组，并向目标组中添加了后端服务。如未创建，请参见 创建和管理目标组 。转发动作需要将请求转发到目标组，因此请先准备好目标组。

创建转发规则

1. 登录 应用型负载均衡控制台 ，在左侧导航栏选择**应用型负载均衡 ALB > 实例管理**。

2. 在实例列表上方选择地域，单击目标实例的 ID/名称进入实例详情页。
3. 选择**监听器管理**页签，找到目标监听器，单击进入监听器详情页
4. 选**择转发规则**页签，单击**创建转发规则**。或在操作列单击**配置转发规则**。
5. 在**创建转发规则**页面，根据下表配置规则信息，配置完成后单击**确定**。

配置项	说明
规则名称	转发规则的名称，最多 255 个字符，支持数字、字母、中文、-、_、. 等字符。留空时系统将自动生成名称。
转发条件	设置请求的匹配条件，所有条件满足时，流量方可转发。 在下拉框中选择**条件类型**并填写对应的值。单击 + **增加条件** 可添加多个条件。 **说明：** 最多支持10个转发条件。 各条件类型说明，请参见 转发条件 。
转发动作	设置匹配成功后对请求执行的动作，满足任一动作时，流量即被转发。 在下拉框中选择**动作类型**并完成配置。单击 + **增加动作** 可添加多个动作。 **说明：** 最多支持5个转发动作。 各动作类型说明，请参见 转发动作 。
优先级	转发规则在监听器内生效的顺序。取值范围为 1 - 10000，数值越小优先级越高。 • 同一监听器内优先级不能重复。 • 单击**查看已有监听规则**可展开当前监听器下已存在规则的优先级与规则 ID，便于合理规划，避免冲突。

转发条件

一条转发规则中可以配置一个或多个转发条件（最多支持10个），条件之间遵循以下逻辑：

- 不同类型条件之间为“与（AND）”关系：例如同一规则同时配置了域名条件和路径条件，则请求必须同时满足域名和路径两个条件才会匹配该规则。
- 同一类型条件的多个值之间为“或（OR）”关系：例如域名条件添加了多个域名值，则请求域名匹配其中任意一个即满足该域名条件。

转发条件支持以下 7 种类型：

条件类型	说明

域名	根据客户端请求的域名进行匹配。 • 支持精准匹配及通配符、正则匹配（不区分大小写）。 • 支持 3-128 个字符。 • 支持 a-z、0-9、-、_、.、,等字符。
路径	根据客户端请求的 URL 路径进行匹配。 • 支持精准匹配及通配符、正则匹配（不区分大小写）、正则匹配（区分大小写）。 • 支持最多 128 个字符，精确匹配以/开头。
HTTP 标头	根据请求 HTTP 头字段的键值进行匹配。 在键字段输入标头名称、值字段输入标头内容，可添加一条或多条标头键值对。 • 键支持 1-40个字符，支持字母、数字和下划线，不支持 Host/Cookie。 • 值支持最多 128 个字符，不支持双引号，首尾不能是空格。
查询字符串	根据请求 URL 中的查询字符串（Query String）键值对进行匹配，可添加一个或多个键值对。 • 键支持 1-16 个字符，不支持大写字母、空格和#、[]、{}、\。 • 值支持最多 128 个字符，不支持空格、#、[]、{}、\、|、<>、&。
HTTP 请求方法	根据请求的 HTTP 方法进行匹配，支持 HEAD、GET、POST、OPTIONS、PUT、PATCH、DELETE。
Cookie	根据请求携带的 Cookie 键值对进行匹配，可添加一个或多个键值对。 • 键支持 1-64个字符，支持小写字母、数字、下划线和短划线。 • 值支持最多 128 个字符，不支持大写字母、控制字符、分号、逗号、空格、引号、反斜杠。
SourceIp	根据客户端来源 IP 地址或 IP 地址段进行匹配，可添加一个或多个 IP/IP 段。 • 支持 IPv4 CIDR 格式。 • 仅支持 /24 和 /32。 • 不支持0.0.0.0/x。

转发动作

转发条件满足时，按配置的一个或多个转发动作执行。转发动作支持以下 6 种类型：

动作类型	说明
转发至	将匹配的请求转发到指定的后端目标组。可选择一个或多个目标组，并通过设置各目标组的权重控制流量分发比例。
重定向至	将请求重定向到指定的协议、域名、端口、路径、查询、状态码。

返回固定响应	由 ALB 直接向客户端返回固定的响应状态码和响应正文，不再转发到后端。 • 响应状态码，取值范围： 200 - 299，400 - 599。 • 响应正文类型，支持 text/plain、text/css、text/html、application/javascript、application/json。仅支持 ASCII 字符，最大 1KB。
重写	在 ALB 内部将请求的域名、路径等修改为指定值后再转发至后端，客户端浏览器地址栏 URL 不变。
写入 Header	在请求中写入指定的 HTTP 头字段（键值对）。 • 键支持 1 - 40个字符，支持字母、数字、下划线和短划线。 • 值支持用户指定、引用、系统定义，最多 128 个字符，不支持双引号，首尾不能为空格。
删除 Header	删除请求中指定名称的 HTTP 头字段。支持 1 - 40 个字符，支持字母、数字、下划线和短划线。

编辑转发规则

1. 进入目标监听器详情页，选择**转发规则**页签。
2. 找到目标转发规则，在操作列单击**编辑**。
3. 在**编辑**页面修改**转发条件**、**转发动作**或**优先级**等信息，修改完成后单击**确定**，完成**编辑**。

> **注意：**
> - 编辑转发规则可能影响现有业务流量的转发路径，建议在业务低峰期进行操作，并在变更前充分评估和测试。
> - 默认转发规则仅支持修改转发动作（如更换目标组），不支持修改转发条件和优先级。

删除转发规则

1. 进入目标监听器详情页，选择**转发规则**页签。
2. 勾选要删除的一条或多条转发规则，单击列表上方的**删除**。
3. 在弹出的**确认**对话框中确认无误后，单击**确定**完成删除。

> **注意：**
> - 删除转发规则后，原本匹配该规则的请求将按其余规则或默认规则重新匹配转发，请确认不会影响线上业务后再操作。
> - 默认转发规则不支持删除。

版权所有：腾讯云计算（北京）有限责任公司

TLS 安全策略

最近更新时间：2026-06-26 17:42:50

为应用型负载均衡（ALB）实例配置 HTTPS 监听器时，TLS 安全策略决定了 ALB 与客户端进行 TLS 协商时所支持的 TLS 协议版本和加密算法套件（Cipher Suite）。在 TLS 握手过程中，客户端通过 Client Hello 发送其支持的协议版本和加密套件列表，ALB 根据所配置的 TLS 安全策略，从中选择双方都支持的协议版本与加密套件组合完成握手。

ALB 提供默认策略和自定义策略两类：

- 默认策略：系统预置的常用 TLS 安全策略，覆盖不同 TLS 版本与加密套件组合，可直接选择使用。
- 自定义策略：当默认策略无法满足特定的安全合规要求时，您可以自行创建自定义策略，灵活指定 TLS 版本和加密套件。

本文介绍如何在 ALB 控制台创建、编辑、删除自定义 TLS 安全策略。

前提条件

已创建应用型负载均衡 ALB 实例。如未创建，请参见 创建 ALB 实例 。

创建自定义策略

1. 登录 应用型负载均衡控制台 ，在左侧导航栏选择**应用型负载均衡 ALB > TLS 安全策略**。
2. 在页面上方选择地域，选择自定义策略页签，单击**新建自定义策略**。
3. 在右侧弹出的**新建自定义策略**面板中，根据下表配置策略信息，配置完成后单击**确定**。

配置项	说明
TLS 安全策略名称	自定义 TLS 安全策略的名称。 TLS 安全策略名称长度为 2 - 128 个字符，必须以字母或中文开头，可包含字母、数字、中文、句号、下划线、短划线。
最低的 TLS 版本	该策略支持的最低 TLS 协议版本。 • 在下拉框中选择，可选范围包括：TLS 1.2、TLS 1.1及以上、TLS 1.0及以上。 • 如业务无特殊兼容性要求，**建议**选择 TLS 1.2及以上版本以保障安全性。
启用 TLS 1.3版本	是否额外启用 TLS 1.3协议版本。 • 开启后，TLS 策略在已选最低 TLS 版本基础上，额外支持 TLS 1.3。 • 在业务兼容的前提下，**建议**启用 TLS 1.3协议版本，以提升通信的安全性与效率。
加密套件	选择 TLS 版本支持的加密算法套件。 面板分为左右两栏： • 左侧：可选加密套件列表。 • 右侧：当前已选中的套件。

标签	选择标签键和标签值，也可选择添加标签，详情请参见 创建标签 。

注意：
若同时启用了 TLS 1.2和 TLS 1.3版本，则必须至少选择一个支持 TLS 1.2的加密套件，否则无法创建。请确保为每个启用的 TLS 版本都选择了对应的加密套件。

编辑自定义策略

1. 进入 **TLS 安全策略**页面，选择**自定义策略**页签。
2. 找到目标策略，在操作列单击**编辑**。
3. 在编辑面板中修改 **TLS 版本、加密套件等信息**，修改完成后单击**确定**。

注意：
修改 TLS 安全策略会影响已关联监听器的 TLS 协商行为，建议在业务低峰期进行操作，并在变更后验证客户端访问是否正常。若出现异常，可立即改回原策略进行回滚。

删除自定义策略

1. 进入 **TLS 安全策略**页面，选择**自定义策略**页签。
2. 找到目标策略，在操作列单击**更多 > 删除**。
3. 在弹出的**确认**对话框中确认无误后，单击**确定**完成删除。

注意：
- 若自定义策略已被 HTTPS 监听器关联引用，需先修改该监听器的 TLS 安全策略（改为其他策略）或删除该监听器后，方可删除此自定义策略。
- 系统默认策略不支持编辑和删除。

ALB 目标组
创建和管理目标组

最近更新时间：2026-06-26 17:33:31

目标组（Target Group）是应用型负载均衡（ALB）用于管理一组后端服务的逻辑分组。ALB 的转发规则通过转发至目标组的方式将客户端请求分发到目标组内的后端服务。一个目标组可被多个监听器或转发规则关联引用，同一组后端服务因此可以在不同监听场景中复用。
本文介绍如何在 ALB 控制台创建目标组、配置目标组后端服务、编辑或删除目标组。

前提条件

- 已创建应用型负载均衡 ALB 实例。如未创建，请参见 创建实例 。
- 已在与目标组相同的私有网络（VPC）中创建了用于承载业务的后端服务（如云服务器 CVM、弹性网卡等），并完成业务应用的部署。

> ⚠ **注意：**
> - 目标组的后端协议、所属网络（VPC）等关键属性在创建后不支持修改。请提前根据业务规划确认配置。
> - 目标组内添加的后端服务须与目标组属于同一私有网络（VPC）。

创建目标组

1. 登录 应用型负载均衡控制台 ，在左侧导航栏选择**应用型负载均衡 ALB > 目标组管理**。
2. 在页面上方选择目标组所在的地域，单击**新建**，弹出**创建目标组**窗口。
3. 配置**基本信息**参数，配置完成后单击**下一步：健康检查**。

参数	说明
目标组类型	选择目标组承载的后端服务种类，当前支持实例。
目标组名称	目标组名称，便于识别和管理。支持最多60个字符。
私有网络	选择目标组所属的私有网络（VPC）。目标组内添加的后端服务须属于同一 VPC。如现有网络不合适，可单击**新建私有网络**前往创建。
后端协议	目标组与后端服务之间的通信协议，支持 HTTP、HTTPS、gRPC、gRPCs 四种。
均衡方式	目标组内多个后端服务之间的流量分配算法。支持**加权轮询**、**加权最小连接数**两种。详情请参见 均衡方式 。

会话保持	开启后，来自同一客户端的请求会被分配到同一台后端服务。详情请参见 会话保持 。
启用长连接	开启后，ALB 与后端服务连接数范围在请求[QPS, QPS*60]区间波动，具体数值取决于连接复用率。 若后端服务对连接数上限有限制，请谨慎操作。
标签	选择标签键和标签值，也可选择添加标签，详情请参见 创建标签 。

4. 配置**健康检查**参数，配置完成后单击**完成**。

   健康检查用于探测目标组内各后端服务的可用性，开启后将自动检查并移除异常的后端服务，从而保证业务高可用。健康检查配置请参见 配置健康检查 。

配置目标组后端服务

添加后端服务

创建目标组后，需向目标组中添加后端服务，业务流量才能真正转发到后端。

1. 在目标组管理列表中，单击**目标组名称**进入目标组详情页。
2. 选择**目标组内实例**页签，单击**添加**。
3. 在**添加后端服务**面板中，选择**后端服务**，单击**下一步：配置端口和权重**。
4. 为所选后端服务设置端口和权重，设置完成后单击**确定**。
   - 端口可配置范围：1 - 65535；
   - 权重越大转发的请求越多，默认为10，可配置范围为0 - 100。当权重设置为0，该服务器不会再接受新请求。

管理后端服务

进入目标组详情页的**目标组内实例**页签，可对后端服务进行以下管理操作：

- 添加：单击**添加**，按照上文中 添加后端服务 向目标组追加后端服务。
- 修改权重：在目标列表的**权重**列单击**编辑**图标，调整单个后端服务的权重。也可选中需要调整的后端服务，单击**修改权重**进行调整。
- 移出：在目标记录的操作列单击**移出**，将后端服务从目标组中移出。

> **说明：**
> 移出后端服务前，建议先将该后端服务的权重调整为 0，待其上的存量连接处理完毕后再移除，以避免业务中断。移出操作不会影响后端服务本身（不会销毁实例）。

编辑或删除目标组

- 编辑目标组：在目标组管理列表中，找到对应目标组，单击操作列的**编辑**。或单击目标组名称进入详情页，在**基本信息**页签中，单击对应属性后的**编辑**图标完成修改。

- 删除目标组：在目标组管理列表中，找到对应目标组，单击操作列的**删除**。在**确认**弹窗中单击**确认**后完成删除。

> **注意：**
> - 关联 ALB 实例的目标组不支持删除，请先在对应的监听器 > 转发规则中解除引用后再删除。
> - 删除目标组不会影响其中的后端服务本身，仅解除目标组对后端服务的逻辑分组。

均衡方式

最近更新时间：2026-06-26 17:24:29

均衡方式是应用型负载均衡（ALB）向目标组分配业务流量的算法，根据不同的均衡方式可以达到不同的均衡效果。

加权轮询算法

加权轮询算法（Weighted Round-Robin Scheduling）是以轮叫的方式、依次请求调度不同的服务器。加权轮询调度算法可以解决服务器间性能不一的情况，它用相应的权值表示服务器的处理性能，按权值的高低和轮询方式分配请求到各服务器。加权轮询算法根据新建连接数来调度，权重高值的服务器先收到连接，权重值越高被轮询到的次数（概率）也越高，相同权值的服务器处理相同数目的连接数。

- **优势**：简洁实用，无需记录当前所有连接的状态，是一种无状态调度。
- **劣势**：相对简单，在请求服务时间变化较大或每个请求消耗时间不一致的情况下，容易导致服务器间的负载不平衡。
- **适用场景**：当每个请求所占用的后端时间基本相同时，负载情况最好。常用于短连接服务，例如 HTTP 等。
- **用户推荐**：已知每个请求所占用后端时间基本相同、后端服务器处理的请求类型相同或者相似时，推荐您选择加权轮询的方式。请求时间相差较小时，也推荐您使用加权轮询的方式，因为该实现方式消耗小，无需遍历，效率较高。

加权最小连接数算法

在实际情况中，客户端的请求服务在服务器停留的时间会有较大的差异。随着工作时间的延伸，采用简单的轮询或随机均衡算法，每台服务器上的连接进程数目可能会有极大的不同，导致没有达到真正的负载均衡。最小连接调度是一种动态调度算法，与轮询调度算法相反，它通过服务器当前所活跃的连接数来估计服务器的负载情况。调度器需要记录各个服务器已建立连接的数目，当一个请求被调度到某台服务器时，其连接数加一；当连接中止或超时，其连接数减一。加权最小连接数算法（Weighted Least-Connection Scheduling）是在最小连接数调度算法的基础上，根据服务器的不同处理能力，给每个服务器分配不同的权值，使其能够接受相应权值数的服务请求，是在最小连接数调度算法的基础上的改进。

> **说明：**
> 假设各台后端服务器的权值依次为 wi，当前连接数依次为 ci，依次计算 ci/wi,值最小的后端服务器实例作为下一个分配的实例。如果存在 ci/wi 相同的后端服务器实例，再使用加权轮询的方式调度。

- **优势**：此算法适合长时处理的请求服务，如 FTP 等应用。
- **劣势**：由于接口限制，目前**最小连接数**和**会话保持**功能不能同时开启。
- **适用场景**：每个请求所占用的后端时间相差较大的场景。常用于长连接服务。
- **用户推荐**：如果用户需要处理不同的请求，且请求所占用后端时间相差较大，如3ms 和3s 等数量级差距，推荐使用加权最小连接数算法，实现负载均衡。

版权所有：腾讯云计算（北京）有限责任公司

会话保持

最近更新时间：2026-06-26 17:30:53

会话保持可使得来自同一 IP 的请求被转发到同一台后端服务上。默认情况下，负载均衡会将每个请求分别路由到不同后端服务上。但是，您可以使用会话保持功能使特定用户的请求被路由到同一台后端服务器上，这样可以使某些需要保持会话的应用程序（如购物车）合理地工作。

背景信息

七层协议（HTTP/HTTPS）支持基于 Cookie 插入的会话保持能力（由负载均衡器向客户端植入 Cookie），会话保持时间设置支持30 - 86400秒，会话保持与均衡方式相关：

均衡方式	特点	支持会话保持
加权轮询	根据后端服务器的权重分发请求	支持基于 Cookie 插入的会话保持
加权最小连接数	根据服务器负载和权重来综合调度	不支持会话保持

配置会话保持

1. 登录 应用型负载均衡控制台，在左侧导航栏选择**应用型负载均衡 ALB > 目标组管理**。
2. 在目标组管理列表中，找到对应目标组，单击操作列的**编辑**。
3. 在**基本信息**页签中，单击**会话保持**属性右侧的**编辑**图标。
4. 在弹窗中开启会话保持，并设置**保持时间**，单击**确定**即可。

版权所有：腾讯云计算（北京）有限责任公司

ALB 健康检查
配置健康检查

最近更新时间：2026-07-02 15:40:09

健康检查（Health Check）是应用型负载均衡（ALB）保障业务高可用的核心机制。开启健康检查后，ALB 会持续探测目标组内各后端服务的运行状况，自动将客户端请求只转发到健康检查正常的后端服务，并在某台后端服务异常时自动将其从流量分发中屏蔽，待其恢复后再自动重新纳入，从而避免单点故障影响整体业务。
本文介绍 ALB 健康检查的工作机制、各项参数含义，以及如何在控制台配置和修改健康检查。

前提条件

- 已创建应用型负载均衡 ALB 实例。如未创建，请参见 创建 ALB 实例 。
- 已创建目标组并向目标组中添加了后端服务。如未创建，请参见 创建和管理目标组 。

健康检查工作机制

ALB 以目标组为维度配置健康检查，每个目标组可独立开启并配置健康检查策略。其工作机制如下：

- 开启健康检查后，ALB 通过健康探测源 IP 周期性地向目标组内每个后端服务发起探测请求，根据响应结果判定后端服务的健康状态。
- 后端服务需连续多次（健康阈值）探测成功才会被判定为健康，以避免网络抖动造成的误判；连续多次（不健康阈值）探测失败则被判定为异常。
- 当某台后端服务被判定为异常时，ALB 自动将其从流量分发中剔除，新的请求会按调度策略分发到其他健康的后端服务；当其恢复正常后，ALB 自动将其重新纳入流量分发。
- 健康检查为短连接探测，每次探测完成后连接即关闭。
- 当目标组内所有后端服务均为异常时，ALB 仍会按调度策略尝试将请求转发至这些后端服务，以最大程度避免业务完全中断。

> **说明：**
> 权重为0的后端服务不接收新请求，也不参与健康检查。

健康检查方式

ALB 目标组支持以下健康检查方式。检查方式的可选项与目标组的后端协议联动：

- 后端协议选择 HTTP、HTTPS 时，检查方式支持 TCP、HTTP、HTTPS；
- 后端协议选择 gRPC 时，检查方式支持 TCP、gRPC；
- 后端协议选择 gRPCs 时，检查方式支持 TCP、gRPCs。

检查	检查原理	适用场景

方式		
TCP	通过向后端服务发起 TCP 三次握手（SYN 包）来探测端口是否存活，仅校验端口连通性，不感知应用层状态。	后端为通用 TCP 服务或仅需确认端口可达的场景。性能开销小，是默认的检查方式。
HTTP	通过向后端服务发送 HTTP 请求（HEAD 或 GET），根据返回的 HTTP 状态码判定后端应用是否正常。	后端为 HTTP 应用，需要从应用层确认服务可用性的场景。
HTTPS	通过向后端服务发送基于 TLS 加密的 HTTP 请求，根据返回的 HTTP 状态码判定后端应用是否正常。	后端为 HTTPS 应用，需要在加密链路上确认服务可用性的场景。
gRPC	通过向后端服务发送 gRPC 健康检查请求，根据返回的 gRPC 状态码判定后端应用是否正常。	后端为 gRPC 应用的场景。
gRPCs	通过向后端服务发送基于 TLS 加密的 gRPC 健康检查请求，根据返回的 gRPC 状态码判定后端应用是否正常。	后端为加密 gRPCs 应用的场景。

配置健康检查

健康检查可在 创建目标组 时配置，也可在目标组创建后进行编辑。本节以创建目标组流程中的健康检查步骤为例。

1. 登录 应用型负载均衡控制台 ，在左侧导航栏选择**应用型负载均衡 ALB > 目标组管理**。
2. 在页面上方选择地域后，单击**新建**，在创建目标组窗口中完成**基本信息**配置，单击**下一步：健康检查**。
3. 在健康检查步骤中，按下表配置健康检查参数，配置完成后单击**完成**。

基础参数

参数	说明
健康检查	是否开启健康检查。开启后 ALB 会自动检查并移除异常的后端服务。建议保持开启。
检查方式	选择健康检查使用的协议。可选项随目标组后端协议联动： 后端协议选择 HTTP、HTTPS 时，检查方式支持 TCP、HTTP、HTTPS； 后端协议选择 gRPC 时，检查方式支持 TCP、gRPC； 后端协议选择 gRPCs 时，检查方式支持 TCP、gRPCs。
检查端口	健康检查探测使用的端口。默认为后端服务的端口；若后端服务的健康检查端口与业务端口不同，可在此指定特定端口。 除特殊需要外建议留空。

HTTP/HTTPS 检查方式专属参数

当检查方式选择 HTTP 或 HTTPS 时，除基本参数外还需配置以下应用层参数：

参数	说明
检查域名	• 长度限制：1 - 80个字符。 • 默认为转发域名。 • 不支持正则表达式，当您的转发域名为通配域名时，需要指定某一固定域名（非正则）为健康检查域名。 • 支持的字符集为：a-z 0-9 . -。
检查路径	• 健康检查路径可设置为后端服务器根目录或指定的 URL： • 长度限制：1 - 200个字符。 • 默认为 /，且必须以 / 开头。 • 不支持正则表达式，建议指定某个固定 URL 路径（静态页面）进行健康检查。 • 支持的字符集为：a-z A-Z 0-9 . - _ / = ?。
HTTP 请求方式	健康检查选择 HTTP 请求方式，可选：GET 或 HEAD，默认为 HEAD。 • 若使用 HEAD 方法，服务器仅返回 HTTP 头部信息，可降低后端开销，提升请求效率，对应的后端服务需支持 HEAD。 • 若使用 GET 方法，则后端服务支持 GET 即可。
HTTP 状态码检测	判定后端服务存活的响应状态码集合，探测返回命中其中任一即视为存活。可勾选 http_1xx、http_2xx、http_3xx、http_4xx、http_5xx。当返回状态码命中所勾选的任一类别时，认为后端服务器存活。

gRPC/gRPCs 检查方式专属参数

当检查方式选择 gRPC 或 gRPCs 时，除基本参数外还需配置以下参数：

参数	说明
检查域名	• 长度限制：1 - 80个字符。 • 默认为转发域名。 • 不支持正则表达式，当您的转发域名为通配域名时，需要指定某一固定域名（非正则）为健康检查域名。 • 支持的字符集为：a-z 0-9 . -。
检查路径	健康检查选择 gRPC 协议时，路径默认为 `/TCLOUD.CLB/healthcheck`，同时可按 `/Package.Class/method` 格式进行自定义设置。
gRPC 状态码检测	判定后端服务存活的 gRPC 状态码，默认值为12，数值范围为0 - 99。 输入值可为单个数值、多个数值、范围以及相互组合，如20、20,25、0-99或12，25，30-40的组合。当 gRPC 返回状态码与设置的状态码匹配时，认为后端服务器存活。

高级选项

单击**显示高级选项**，可通过滑块调整健康检查的探测频率与判定阈值：

参数	说明	取值范围	默认值
响应超时	单次健康检查的最大等待时间。若后端服务在该时间内未正确响应，则判定本次探测失败。	2 – 60 秒	2 秒
检测间隔	相邻两次健康检查之间的时间间隔。间隔越短，异常发现越及时，但探测开销越大。	2 – 300 秒	5 秒
不健康阈值	连续探测失败达到该次数后，将后端服务判定为异常并剔除流量。	2 – 10 次	3 次
健康阈值	连续探测成功达到该次数后，将后端服务判定为健康并恢复流量。	2 – 10 次	3 次

说明：
响应超时时间应小于检测间隔。若后端服务启动较慢，可适当增大检测间隔或不健康阈值，避免服务启动期间被误判为异常；若网络延迟较高，可适当增大响应超时时间。

查看健康检查状态

为目标组开启健康检查并将其关联到监听器后，您可在**监听器管理**页面中查看各后端服务的健康检查状态。

1. 登录 应用型负载均衡控制台 ，进入目标实例，
2. 进入目标实例后，切换到**监听器管理**页面，在健康状态列中可见健康状态。后端服务的健康检查状态含义如下：

状态	说明
健康	后端服务连续通过健康检查（达到健康阈值），正常参与流量分发。
异常	后端服务连续未通过健康检查（达到不健康阈值），已被剔除出流量分发。
未知	绑定的目标组内无后端服务。
探测中	新绑定的后端服务器在检查间隔 × 健康阈值时间内的状态。
已关闭	关闭健康检查。CLB 向所有后端服务转发流量。

修改健康检查

目标组创建后，您可随时修改健康检查配置。

1. 在目标组管理列表中单击**目标组 ID** 进入详情页。

2. 在**基本信息**页的**健康检查**区域单击**编辑**。
3. 修改完成后，单击**保存**。

> **注意：**
> - 关闭健康检查后，ALB 将不再探测后端服务。一旦某台后端服务故障，流量无法自动切换至其他正常后端服务，可能导致业务受损，请谨慎操作。
> - 增大检测间隔会延长异常后端服务被发现的时间，请结合业务对故障切换时效的要求合理设置。